Linking the Genetic Signature of Asymbiotic Soil Diazotrophs with Nitrogen Fixation under Land-Use Change in the Amazon Rainforest

By

RACHEL ELIZABETH DANIELSON

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Soils and Biogeochemistry

In the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Jorge Rodrigues, Chair

_____
Brendan Bohannan

_____
William Horwath

_____
Douglas Nelson

Committee in Charge
2023

Acknowledgements

I would like to thank my graduate committee including Dr. Brendan Bohannan, Dr. William Horwath, Dr. Douglas Nelson, and Dr. Jorge Rodrigues. These members were critical in providing scientific guidance and feedback during the dissertation preparation process. I am extremely grateful to Dr. Jorge Rodrigues for serving as my graduate major professor- you helped me see this immense challenge through to the very end. Thank you to Dr. Amelie Gaudin for serving as my graduate advisor. Thank you to Dr. William Horwath, Dr. Douglas Nelson, Dr. Sanjai Parikh, and Dr. Kate Scow (chair) for serving on my qualifying exam committee. I am also grateful for the guidance imparted by Dr. Denneal McClung during my biotechnology internship, and the Innovation Institute for Food and Health as well as Stephanie Dorsey and Corey Jones of E^2JDJ for providing me with training in venture capital during the final leg of my degree.

I would like to thank Dr. Tad Doane for providing guidance on a myriad of soil physicochemical analytical techniques and laboratory protocols. Thank you to Christian Erickson for copious assistance with bioinformatics and computational processing, and to Jordan Sayre for providing assistance with microbial wet lab work optimization. Thanks to Christian, Dr. Jonathan Lin, Alex Gulachenski, and especially Jordan for building group morale and friendship inside and outside the lab. Carlotta Sainato and Corinne Butler were essential in completing laboratory work- and great gals!. Wagner Piccinini, Alexandre Pedrinho, and Dr. Jorge Rodrigues offered critical assistance in field work, including equipment transport, sample collection, and processing. Thank you to all extended Brazilian collaborators for your help and support in-country. All co-authors included throughout the chapters of this dissertation contributed to the development of this body of work. And finally, the most invaluable support was provided by my family including my mom, Laura, dad, Richard, brothers and cousins Reed, Ross, Alex, and Alaina, my partner Jesse, and friends Jordan, Zoe, Courtney and Eric, Bruce (dog), and of course my cat Pepper, who was integral. I could not have made it to this point without your support.

I am grateful for the financial support provided by the National Science Foundation – Dimensions of Biodiversity (DEB 14422214), the USDA NIFA Pre-doctoral Fellowship (Award No. 2019-67011-29554), the Henry A. Jastro Research Award, the UC Davis Soils and Biogeochemistry Graduate Group, and the UC Davis Innovation Institute for Food and Health.

Table of Contents

List of Tables and Figures

Abstract

Net primary productivity and carbon sequestration are dependent on an adequate source of nitrogen (N) in terrestrial ecosystems. Biological nitrogen fixation, the microbially-mediated conversion of atmospheric N_2 to biochemically reactive ammonia (NH_3) serves as an important source of new N to many natural ecosystems. Despite its crucial role in global biogeochemistry, controls over the rate of this microbially mediated process remain poorly understood. This is particularly true in the context of land use (LU) alteration, a major driver of climate change and ecological disruption. In the Amazon Basin, LU change from primary tropical forest to agricultural operations, particularly cattle pasture, has been prevalent and ongoing for decades. To better understand shifts in N inputs resulting from LU change, this dissertation investigates the activity and diversity of free-living soil diazotrophic communities across the LU dichotomy of forest and pasture in the Amazon Basin. First, to establish a broader perspective on the microbial response to LU change in the Amazon, we performed a comprehensive review of all studies investigating microbial communities or microbially mediated processes across variable LUs and successional states in the region, highlighting both generalized and repeated findings, as well as points of disagreement which may indicate regional heterogeneity or methodological biases. This meta-analysis revealed that although trends in the alpha-diversity of microbial communities vary somewhat across studies, spatial biotic homogenization in pastures compared to primary forests has been a near-universal finding among prokaryotes, fungi, and several other taxonomic or functional subgroups. Additionally, a large proportion of studies have identified soil Al^{3+} content and extractable acidity to be major factors shaping community structure, emphasizing the importance of tropical mineralogy. Significant shifts in microbially-driven biogeochemical cycling with LU change have been identified as well. Most studies have found that soil C stocks increase slightly with pasture conversion, simultaneous with elevated methane emissions and a decline in the community proportion and taxonomic richness of methanotrophs. However, the most drastic and consequential nutrient cycle shift has been to the N cycle. Studies have found that rates of net mineralization and nitrification decline sharply in maturing pastures, coincident with reductions in nitrous oxide emissions and inorganic N pools. Recent studies have indicated that free-living diazotrophs increase in abundance and diversity with pasture conversion, but measurements of asymbiotic nitrogen fixation (ANF) have not been made across LUs to bolster these community findings.

Addressing this knowledge gap in conjunction with community profiling was a primary focus of the original research presented here.

To fulfill this aim, we surveyed three primary forests and three pastures (converted in or around 1972) in the state of Rondônia, Brazil near the end of the wet season in April of 2017. At each site, we established a 100 m^2 quadrat and collected four replicate soil cores (0-10 cm in depth) from seven locations each. After homogenization, fresh soil was utilized for $^{15}N_2$ gas incubations to calculate soil N incorporation attributable to ANF. Additionally, preserved soil samples were used for nucleic acid extraction and analysis, quantifying a suite of physicochemical measurements, and profiling the bulk soil metabolome. Community nucleic acids were utilized for marker-gene targeted amplification to serve as a proxy of absolute abundance and functional community structure, as well as to obtain profiles of potential diazotrophs within the broader soil metagenome.

From this analysis, we concluded that soil ANF is indeed stimulated (47x increase) in active cattle pastures coincident with an augmentation in soil *nifH* copy number (18x increase; in line with previous observations), but that the two are not directly related within LUs. Using soil physicochemical parameters (including various pools of C and N, natural isotopic abundance, soil texture, pH, P, cation exchange capacity, and enzyme metallocluster constituents including Mo and V) for variable selection in multiple linear regression, we were unable to identify variables strongly associated with *nifH* community augmentation from forest to pasture. Substantially lower soil NO_3^- concentrations provided significant, but modest value in explaining the stimulation of ANF in mature, active cattle pastures compared to primary forests. Additionally, when forests were considered separately from pastures, *nifH* copy number and ANF rates (which were consistently near zero) were poorly explained by physicochemical parameters. Together these findings indicate that factors such as high inorganic N concentration as well as alternative, more productive N input pathways (i.e., canopy lichens, nodules, or the surface litter layer), which were not measured in this study, could largely suppress the activity of potential diazotrophic bacteria in primary forests. Within pasture soils, however, we found that *nifH* copy number was primarily associated with C pools. We found a modest, but significant association between pasture soil ANF rate and the ratio of low molecular weight extractable organic C to N, as well as the fraction of total N in the dissolved organic form, suggesting that this energy-intensive reaction is stimulated by a limited availability of N combined

with sufficient fuel in the form of low molecular weight C compounds. Additionally, using a multigene approach to profile soil diazotrophs within the broader microbial community revealed that the genetic potential for asymbiotic diazotrophy is one of the most (if not the most) enriched soil microbial functions accompanying pasture conversion. This realization speaks to the immense influence that large-scale LU change can have on microbial communities, and the strong pressure for N replenishment in grazed (but unfertilized) pastures of the Amazon Basin.

After observing that the absolute abundance of potential diazotrophs (based on DNA copy number) did not scale with ANF measurements in either LU, we aimed to further explore any meaningful relationships between diazotrophic communities and this crucial biogeochemical process they mediate. A previous analysis of the potential diazotroph community in forest and pasture soils of Rondônia found significant shifts in its structure. However, given the phylogenetic and trophic breadth of potential diazotrophs (i.e., microorganisms bearing nitrogenase-encoding genes, but not necessarily contributing to ANF), it is not clear whether these shifts have occurred independent of the larger soil prokaryotic community. By comparing potential diazotroph community structure (using DNA-based amplification of the nitrogenase marker gene, *nifH*) with that of all prokaryotes (by amplifying the 16S rRNA gene), we found that potential diazotroph community alpha-diversity was significantly higher in pasture soils, while the overall prokaryotic community did not reflect an increase. However, both profiles reflected a similar degree of community compositional dissimilarity with respect to LU change. Additionally, both communities exhibited significantly lower groupwise compositional dispersion in pasture compared to forest soils, agreeing with several earlier studies observing biotic homogenization of soil microbial communities with LU change. The compositional dissimilarities of both communities were associated with a similar subset of physicochemical conditions including clay content, pH, and total sulfur content, as well as the proportion of nitrogen in inorganic forms.

We further investigated how these potential diazotrophic communities relate to the subset of taxa that are both metabolically active, and actively transcribing the nitrogenase enzyme (via RNA-based sequencing of *nifH*), to determine if the latter community exhibits a community structural response to LU change differing from that of DNA-based communities, and whether active communities may better explain ANF activity. In contrast to our expectations, we found that active diazotrophs did not reflect an

increase in alpha-diversity with LU change and exhibited only a modest compositional response with no environmental correlates. Surprisingly, active, RNA-based communities across LUs were more similar to each other than they were to their DNA-based counterparts, showing an opposite trend in community dispersion with respect to LU. This increase in dispersion was related to ANF rates; we identified 17 of 882 taxa whose relative abundance scaled linearly and significantly (Pearson r = 0.88) with ANF. Of these, two OTUs, annotated as *Bradyrhizobium* and Enterobacteriaceae, reflected strong correlations with ANF on individual bases and were found to have the third and twentieth highest relative abundances among active pasture taxa, respectively. While paired potential and active diazotrophic communities (i.e., derived from the same sample) reflected a high degree of compositional dissimilarity from each other overall, by far the most drastic shift observed was the more than 100x enrichment of the photosynthetic cyanobacterial family Aphanizomenonaceae in active diazotroph profiles, irrespective of LU. Therefore, soil surface diazotrophs may also play an important role in providing N to pastures and forests alike, suggesting further work is needed to capture measurements of ANF activity under lighted conditions.

Chapter 1: Impacts of Land-Use Change on Soil Microbial Communities in the Amazon

Rachel E. Danielson[1], Jorge L. Mazza Rodrigues[1,5] *

[1] Department of Land, Air, and Water Resources, University of California – Davis, Davis, CA, USA

[5] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA USA.

ORCID #

Rachel Danielson: 0000-0001-7044-0747

Jorge L.M. Rodrigues: 0000-0002-6446-6462

***Corresponding author:** jmrodrigues@ucdavis.edu; University of California – Davis, 1110 Plant and Environmental Sciences Building, Davis CA 95616

Citation:

Danielson, R. E., & Mazza Rodrigues, J. L. (2022). Impacts of land-use change on soil microbial communities and their function in the Amazon Rainforest. In (Vol. 175, pp. 179-258). doi: 10.1016/bs.agron.2022.04.001

Abstract

The Amazon Rainforest is a global diversity hotspot that has experienced a significant level of deforestation over the past half century, primarily for the establishment of cattle pasture. Characterizing the impact of this large-scale ecosystem conversion on the composition and activity of the soil microbial community is crucial for understanding potentially consequential shifts in nutrient and greenhouse gas cycling, as well as adding to the body of knowledge concerning how tropical ecosystems respond to human disturbance. Research to date has shown that locally, communities of soil microorganisms tend to become more diverse upon conversion of forest to pasture. However, these communities undergo taxonomic homogenization at landscape-level spatial scales, mirroring the homogenization of plant communities across pastures. Microbial community structure is distinct between forest and pasture soil communities across several studies, and specific taxa, such as Firmicutes and Acidobacteria, show consistent association with pasture and forest soils, respectively. In addition, shifts in microbial community functions with pasture conversion have relevant impacts on both carbon and nitrogen cycling at the ecosystem scale: the abundance and diversity of methane-cycling prokaryotes shifts in conjunction with increased methane flux in pastures. Further, quantitation and community profiling of free-living nitrogen fixers has demonstrated that this functional group is favored in pastures and suggests that asymbiotic N_2 fixation may be a significantly augmented process. While human-driven deforestation is continuing, a large percentage of once-converted pastures are undergoing the process of secondary forest succession. Assessment of microbial communities in secondary forests compared to primary forests and pastures suggests convergence towards a recovery of functionality and community composition with reforestation.

Keywords: *tropical deforestation; microbial diversity; biotic homogenization; nutrient cycling; metagenomics; forest succession*

Introduction: The Amazon is an Ecosystem of Global Importance

The Amazon Basin contains the largest tropical forest on Earth, covering a cumulative area of 6.3 million square kilometers across eight different countries (plus the Territory of Guiana) in South America, although over two-thirds of its total area is contained within Brazil's borders (Butler (2020); (Hansen et al., 2020). The basin plays a major role in the planet's biosphere, supplying one fifth of the total freshwater flow to the oceans, controlling regional climate parameters, including temperature and precipitation, and regulating the exchange of atmospheric gases (Coe et al., 2017; Davidson et al., 2012). The Amazon Rainforest also contains a disproportionately high number of plants and animals in comparison to any other ecosystem on the planet (i.e., a hotspot), potentially housing 25% of all terrestrial species, including approximately 16,000 tree and 1,300 bird species (Barlow et al., 2018; Butler, 2020; Dion, 2010; Gaston, 2000; Oliviera and ter Steege, 2013).

Total rates of deforestation in the Amazon are difficult to assess due to differential coverage of satellite data across the region over the past several decades. Likely the most comprehensive data source available, Brazil's National Institute for Space Research recorded approximately 457,00 square kilometers of deforestation between 1988 and 2020 using satellite imagery (**Figure 1-1a**). Another 345,000 square kilometers is estimated to have been lost between 1970 and 1987, amounting to a cumulative projection of 18–20% of Brazil's historic coverage (INPE, 2020). During the first half of the 1970s, the widespread development of highways was likely an important catalyst for increased rates of deforestation. These highways afforded new access to the Amazon and created a feedback loop of speculative buying, land clearing, and subsequent inflation of value, driving further speculative buying (Fearnside, 1987). The progression of deforestation throughout the Amazon Rainforest Biome has left small, disconnected fragments of primary forest across a disturbed landscape (Lovejoy et al., 1986). Annual rates of loss peaked in the early 2000s before falling dramatically in 2005, though rates have been on the rise again since 2015, particularly within the Brazilian Legal Amazon (BLA; **Figure 1-1a**).

By far the most common cause of forest clearing in the Amazon Basin is for the establishment of cattle pasture, accounting for approximately 68% of land-use change-driven clearing between 2000 and 2013 in the BLA (Tyukavina et al., 2017). The establishment of small- and large-scale agriculture accounted for 13% and 11% of forest clearing over this same period, respectively (**Figure 1-1b**). The

remaining <8% of clearing resulted from activities including logging, construction, and mining (Tyukavina et al., 2017). The prevalence of pasture across the Amazon has heavily influenced the focus of investigations research into the environmental and ecological implications of this conversion type, and consequently, forest-to-pasture comparative studies in the BLA will be the main focus of this chapter.

Pasture conversion across the region typically begins with selective logging to remove valuable timber, followed by clear cutting and burning of the remaining vegetation. To establish a pasture, sites are aerially reseeded with fast-growing African perennial bunchgrass species (Mueller et al., 2014; Navarrete et al., 2015a). Dominant species include *Urochloa brizantha*, *Urochloa decumbens*, *Brachiaria humidicola*, and *Panicum maximum*, of which *U. brizantha* and *U. decumbens* are estimated to cover approximately 75% of total pasture area (Jank et al., 2014; Nogueira, 2012). The use of fertilizers on pasture is atypical throughout much of the Amazon due to cost limitations (Jank et al., 2014; Mueller et al., 2014); therefore, productivity of grasses may be very vulnerable to overgrazing. The carrying capacity of the pastures is typically considered to be 1.1 hectares per cattle head with improved cultivars of *U. brizantha*, but large variation in seasonality and rotational periods across managed ranches may not reflect the reality of this estimate (Jank et al., 2014; Pedreira et al., 2015).

The process of forest removal has a calculated net efflux of 325 Tg carbon (C) per year through biomass burning and degradation of remaining forest margins (Baccini, 2017). Data further shows a curtailing of the South American monsoonal circulation in response to deforestation (Boers et al., 2017). The reduction in condensational latent heat over once-forested areas results in lower inflow of atmospheric moisture from the Atlantic Ocean (Boers et al., 2017; Ciemer et al., 2020), resulting in conditions such as recurrent drought that accelerate tree mortality (Phillips et al., 2009). Consequently, future deforestation may approach a dangerous tipping point where forest-derived latent heat is insufficient to maintain the Atlantic moisture feedback (Boers et al., 2017), leading some to suggest the potential savannization of the entire ecosystem (Silvério et al., 2013). It is crucial, therefore, to understand current and future biological response to Amazon land-use change.

In terms of landscape-scale ecological response, perhaps the most conspicuous consequence of forest conversion in the Amazon is the loss of highly diverse and endemic floral and faunal communities, and replacement with a small number of forage grasses, crops, and cattle (Ferraz et al., 2003).

Unfortunately, some undesirable species have thrived; populations of *Anopheles darlingi,* a mosquito vector that transmits malaria, has gained habitat range due to the process of forest clearing removing physical barriers to their dispersal. This has ultimately increased the risk of malaria infection up to 300-fold across the region (Vittor et al., 2006).

A microcosm of the global biosphere with a somewhat ambiguous response to anthropogenic disturbance is the soil microbial community. Soil is known to harbor the highest known taxonomic and functional diversity across different environments, but its composition is largely undescribed (Bahram et al., 2021; Torsvik and Øvreås, 2002). The focus of this chapter will be to review what is currently known about the effects of human-induced land-use change in the Amazon Rainforest on soil microbial communities in terms of diversity and composition, as well as in relation to important microbially-mediated biogeochemical cycles. Also to be discussed are the specific factors of land-use conversion by fire and the recovery of secondary forests from abandoned pasture, which at present comprises a significant portion of the Amazon Rainforest. Current limitations in knowledge and considerations for interpreting data in light of regional differences, such as climate and endemic soil conditions, will be highlighted. In this chapter, we aim to convey the potential importance of expanding our understanding of soil microbial response to large-scale land-use change and will provide insight as to future directions for research efforts.

Why do Microbes Matter?

At first thought, considering the impact of land-use disturbance on soil microbial communities may seem somewhat abstruse, given that their species diversity and environmental activity is unseen to the human eye. Yet, closer examination reveals their relevance as major drivers of a myriad of crucial biogeochemical processes within the soil-plant-atmosphere continuum. Plants interact intimately with microbial community constituents, including archaea, bacteria, and fungi at-or-near their root surfaces (i.e., the rhizosphere). In many cases, plants rely on microbes for nutrient acquisition. Microorganisms collectively produce a complex suite of extracellular enzymes, aide in extension of soil exploration via hyphal networks, and can directly provide plants with nitrogen (N) through root nodule structures in exchange for carbon-rich root exudates (Brzostek et al., 2012; Desbrosses and Stougaard, 2011; Huang

et al., 2014; Linderman, 1991). These relationships are important considerations for understanding or predicting plant community succession or overall ecosystem function and sustainability following land-use change.

In addition to their direct relationships with plants, soil microbial communities mediate biogeochemical cycles relevant to greenhouse gas emissions and climate change. Heterotrophic soil microorganisms are primary agents of soil organic matter decomposition, typically accounting for the bulk of CO_2 released from soils (Yuste et al., 2011). Evidence suggests that substrate affinity competition between various microbial functional groups controls the rate of C mineralization (Fontaine et al., 2004; Fontaine and Barot, 2005). Therefore, shifts in organic matter composition or soil conditions, such as pH, temperature, moisture, and bulk density, with land-use change may impact decomposition rates and C use efficiency by altering the composition and interactions of the microbial community. This in turn may influence CO_2 emissions and long-term soil C storage (Öquist et al., 2016).

At present, our collective understanding of the relationship between the distribution of organisms within a given soil habitat and their functional importance is limited (Torsvik and Øvreås, 2002). In the context of large-scale anthropogenic land-use change, shifts in soil microbial processes and physiological plasticity could have key consequences in biogeochemical cycles (Mackelprang et al., 2011). Augmenting our understanding of these relationships has the potential to greatly enhance our ability to predict nutrient transformations and improve the representation of microbial processes in nutrient cycling models (Bradford et al., 2016; Treseder et al., 2011). Studies suggest that microbial community features, like composition and diversity, correspond to rates of activity at varying scales (Chen et al., 2019; Delgado-Baquerizo et al., 2016; Peter et al., 2011; Philippot et al., 2013; Strickland et al., 2009; van Elsas et al., 2012); however, more work is needed to understand the functional overlap across community members and how specific biogeochemical processes are impacted by changes in these communities.

The task of linking 1) the impact of land-use disturbance on the soil environment, 2) microbial community parameters including diversity, composition, and abundance, and 3) shifts in the soil functions they mediate, is immense. Attempting to resolve these complex abiotic and biotic relationships relies upon a variety of molecular methods typically targeting specific groups or whole soil communities, paired with activity measurements and relevant biogeochemical parameters. The next section will briefly review

the modern methodologies available to researchers in order to characterize soil microbial communities. Additionally, important terms used to define microbial community diversity will be explained. This will serve as a useful pretext in understanding the work being done by researchers to understand the response of soil microbial communities to land-use change throughout the Amazon Rainforest.

Characterizing Soil Microbial Communities: Current Methodologies and Metrics

Methods to Study Complex Communities

A straightforward, but fairly low-resolution method of gauging microbial community response to disturbance is through bulk biomass measurement, with the underlying assumption that a reduction in biomass corresponds to a negative impact, or vice versa. In soils, such measurements are made through chloroform fumigation and subsequent extraction and quantification of biomass-derived C and N (DeLuca et al., 2019). This can be put into the context of a community-wide activity response by measuring processes, such as respiration (either under ambient conditions or in response to a variety of substrates), in order to infer catabolic diversity or stress response based on C use efficiency (Degens and Harris, 1997; Wardle and Ghani, 1995). These methods do not, however, provide a dimension of genetic differentiation among community members and lack a detailed understanding of microbial diversity.

A comprehensive means of understanding the physiology and genetics of a microorganism is through culturing. When studying communities as complex as those found in soil, though, culturing is a rather impractical way of gauging diversity or microbial response to environmental change. Only a handful of microorganisms relative to global diversity have been successfully cultured (Handelsman, 2004; Overmann et al., 2017; Schloss and Handelsman, 2004), and even just a gram of soil is likely to contain billions of cells and thousands of distinct organisms (Trevors, 2010). Methods like phospholipid fatty-acid analysis (PLFA), terminal restriction fragment length polymorphism (T-RFLP), and denaturing gradient gel electrophoresis (DGGE) are not dependent on culturing and allow limited characterization of community profiles, but lack the sensitivity of molecular-based approaches and risk misinterpretation of community composition (Dickie et al., 2002; Nakatsu et al., 2000; Schoug et al., 2008).

The most common modern method employed to study microbial communities is the extraction of total soil community DNA and molecular amplification by polymerase chain reaction (PCR). A simple approach to measuring abundance of microbial groups is through quantitative PCR, which allows calculation of gene copy number using a fluorescently tagged DNA polymerase (Zemb et al., 2020). Profiling of communities based on composition can be achieved through high-throughput sequencing, which is commonly used to target the variable regions within the 16S ribosomal RNA (rRNA) gene or Internal Transcribed Spacer (ITS) region for prokaryotes and fungi, respectively (Preheim et al., 2013). While less commonly employed, functional genes may also be targeted. Sequence processing pipelines have become fairly streamlined and software platforms, such as QIIME2 (Bolyen et al., 2018) and DADA2 (Callahan et al., 2016), ultimately allow the translation of sequences to amplicon sequence variants (ASVs), the microbial approximation of a species (Callahan et al., 2016). The utilization of curated databases can also assign taxonomy to varying degrees of certainty (Nilsson et al., 2019; Yoon et al., 2017), affording more meaning to community profiles and allowing exploration of functional potential (Nguyen et al., 2016). However, due to our limited ability to culture and comprehensively study most soil organisms (Overmann et al., 2017), taxonomic assignment is fairly limited. Amplicon sequence data may also be used to calculate phylogenetic relatedness within and across samples.

Numerous studies over the last two decades have utilized these methods to describe microbial composition and diversity metrics in the context of comparing primary versus perturbed environments. Nonetheless, a more comprehensive community analysis can be achieved by sequencing *all* community DNA, rather than targeting genes of specific microbial groups, a strategy called metagenomics. In most cases, metagenomics avoids the issue of potential PCR bias incurred by amplicon-based studies, and also allows simultaneous analysis of taxonomic/phylogenetic diversity and function from a single dataset (Mendes et al., 2017; New and Brito, 2020). Because no specific gene is being targeted by amplification, comparative analyses can be semi-quantitative as well as compositionally descriptive.

Defining Soil Microbial Diversity

The magnitude of soil microbial diversity has been gradually realized over the past few decades (Gans et al., 2005; Torsvik et al., 1990). Still, complete characterization has remained elusive, despite the

ability to obtain hundreds of millions of DNA sequences from a single sample. Comparative studies among various ecosystems unequivocally show that microbial diversity in soil is the greatest of any environment on Earth (Locey and Lennon, 2016). There are an estimated 100–9,000 distinct prokaryotic taxa (bacteria and archaea; operationally referred to as ASVs) per cubic centimeter of soil, constituting approximately 4–20 billion cells. Fungi are relatively less diverse with approximately 200–235 taxa and 10,000 individuals per gram soil on average (Bardgett and van der Putten, 2014).

The diversity of soil microbial communities is comprised of 1) the total number of distinct taxa/ASVs (richness) and 2) how proportionally abundant (evenness) they are in the environment (Shade, 2017). The use of common diversity metrics borrowed from community ecology, such as alpha (within a sample) and beta (across samples of an environment) diversity (Whittaker, 1972), allows statistical testing of differences in community composition from the local to the landscape scale (Maron et al., 2011). These metrics provide a mechanism to account for spatial heterogeneity effects on community structure at different scales, and furthermore allows comparison across ecosystems and changing environmental conditions. Ultimately, characterization of microbial communities using these metrics is done in an attempt to extend theoretical frameworks used for macroscale populations down to the microscale (Shade, 2017).

Another particularly important aspect of microbial community diversity in soils is the relative abundance of identified species within diverse ecosystems (Gaston, 1994). Variation in the prominence of organisms over time or space may indicate a differential ecosystem function or suitability. Assessing relative changes in the abundance of all organisms across space or environment type provides a metric of community dissimilarity. The variation in community dissimilarity with geographical distance is known as the distance-decay relationship (Bell, 2010) and it is particularly informative when comparing and contrasting environments such as primary forests and pastures. Differences in the distance-decay relationship indicate that disturbance affects the turnover of species across space and provides information about differential taxa dispersal abilities. For example, taxa capable of a broad distribution across soil samples, known as generalists, may be governed by life-history strategies that allow increased dispersal, while specialist taxa are restricted to certain environmental conditions and have limited dispersal (Barberan et al., 2012; Sriswasdi et al., 2017).

Finally, any sound definition of soil microbial diversity requires the understanding of its functional diversity, defined as variation in traits between taxa (Escalas et al., 2019). Ecological traits, however, encompass a variety of ecosystem-relevant functions, such as biogeochemical processes (e.g., methanogenesis, denitrification, biological nitrogen fixation, C utilization, etc.) or cellular regulation processes like stress response. Each of these pathways and processes requires traits that vary in complexity and require intricate genetic machinery in order to be carried out (Martiny et al., 2015). Therefore, assessment of functional profiles, either through gene-targeted sequencing or metagenomics is essential for gaining a deeper understanding of biogeochemical shifts, as well as for gauging the potential for community resilience or multifunctionality under land-use change.

How Has Land-Use Change Impacted Microbial Communities in the Amazon?

The following sections will discuss what is known to-date concerning the impacts of land-use alteration on the diversity, community composition, and functional potential of soil microbiomes in the Amazon Rainforest. Studies discussed have focused on several aspects of the microbial community using a variety of analysis techniques across a limited number of established locations (**Table 1, Figure 1-2**). This creates continuity of datasets and ease of repeated sampling but imparts regional bias when assessing impacts of soil properties, climate, and land management, an important consideration when comparing results.

Shifts in Community Diversity

The first culture-independent study of soil microbial biodiversity in unaltered Amazon Rainforest was reported by Borneman and Triplett (1997) with sequencing of 98 bacterial 16S rRNA gene sequences from two different forest soils in the Brazilian State of Pará. The authors observed that all sequences were unique, and further estimates of species richness concluded that proper assessment of diversity would require sampling over 10,000 sequences per sample (Schloss and Handelsman, 2005). Contradicting the above results, subsequent investigation suggested that Peruvian Amazon soils contain the lowest diversity among soil samples collected across North and South America (Fierer and Jackson, 2006; Lauber et al., 2009). These previous studies, although important, were limited to a few samples.

Early analysis of the effects of deforestation on microorganisms showed that total microbial biomass increases with pasture conversion, and that microbial community structure under land-use change is significantly different from communities of primary forest (Cenciani et al., 2009). These differences were more pronounced in the dry season compared to the wet season. A study using a T-RFLP fingerprinting approach indicated that within-sample (i.e., local) biodiversity increased in converted land-use systems, both in pasture and agricultural plots (da C Jesus et al., 2009). Local-scale response to land-use change is therefore seemingly divergent between macro- and microbiota, and this result could be interpreted as implying that microbial diversity should be excluded when assessing biodiversity losses in tropical systems. However, this is only part of the story.

In 2009, our group created the Amazon Rainforest Microbial Observatory (ARMO) to expand upon the understanding of soil microbial response to forest-to-pasture conversion . Using a nested quadrat, we collected soil samples that spanned centimeter to kilometer intervals from primary forests and actively grazed pastures in the western Amazon, State of Rondônia, Brazil. The area was selected because it represents an extreme case of agricultural development, with higher total deforestation rates than any other state in BLA. The spatially-explicit sampling scheme allowed for the assessment of not only microbial richness as a measure of alpha diversity, but also community compositional variation and turnover as components of beta diversity. Contrary to the hypothesized trends, our research has shown that local taxonomic and phylogenetic richness increase with forest-to-pasture conversion. Spatially, forest communities turn over (i.e., become more dissimilar) much more rapidly with increasing distance between samples, while pasture communities are fairly homogenous, particularly phylogenetically (Rodrigues et al., 2013). This process of increasing the similarity of community members over space and/or time in the pasture was not the result of taxa invasion (e.g., microorganisms with increased dispersal abilities), but was mainly due to losses of endemic taxa from forest communities and increases in the range of existing taxa. This pattern of loss also drives overall distinct structural changes in the community. A temporal study in Ipiranga do Norte, Mato Grosso, also measured consistently higher alpha diversity, but found much greater seasonal variation in community diversity of pastures compared to forests, potentially indicating more severe seasonal stressors, such as higher exposure to rain and solar effects- in pasture soils (Mendes et al., 2015b). The finding of increased alpha diversity with pasture

conversion has been shown elsewhere as well (Mendes et al., 2015a; Navarrete et al., 2015a; Pedrinho et al., 2019) but in some cases overall diversity metrics and taxonomic richness show disparate trends (Pedrinho et al., 2019), indicating that community evenness may decline. Other studies have found that taxonomic richness does not change (Lammel et al., 2015b) or even declines (Melo et al., 2021).

This pattern of biotic spatial homogenization with forest conversion has been mirrored in other studies, including in total bacterial communities of converted pastures across cerrado and rainforest biomes (Lammel et al., 2015b), active pastures of Ipiranga do Norte, Mato Grosso (Mendes et al., 2015b), and converted no-till cropping systems in Querência municipality, Mato Grosso (Goss-Souza et al., 2019), as well as within communities of the phylum Acidobacteria in ARMO pastures (Navarrete et al., 2015b), and whole fungal communities in the Mutum-Paraná River Basin (Cerqueira et al., 2018). This trend has held across a range of pasture ages, soil types, climates, and locales (see **Table 1; Figure 1-2**), indicating that the conversion of forest to pasture, the common factor in each study, is likely the *driving* factor in spatial diversity shifts. The finding of spatial microbiotic homogenization along with monospecific transformation of the aboveground floral community is perhaps unsurprising. However, the functional implication of this spatially dependent shift in biodiversity is not clear; taxonomic and phylogenetic diversity as metrics offer at-best limited indication of shifts in relevant ecosystem functions including greenhouse gas and nutrient cycles.

Shifts in Community Composition

The shift in community composition across a land-use or disturbance gradient is a distinct measurement from diversity change, that is, communities may be just as diverse but experience significant shifts in the abundance of some taxa relative to others. Assessing consistency of trends in taxonomic shifts across studies (especially conducted at different locations) may identify taxa particularly responsive to forest-to-pasture conversion. **Figure 1-3** shows a taxonomic network of 16S rRNA gene-based community member (bacteria and archaea) relative abundances from soils sampled in three forests and three pastures in Agropecuaria Nova Vida, Rondônia, Brazil (ARMO; Rodrigues et al., 2013). Overall, the majority of taxonomic levels show no significant alteration in abundance across land-use types, but more groups are associated with forests compared to pastures, consistent with Rodrigues et al.

(2013) findings, which concluded that biotic homogenization in pastures is linked to the loss of endemic species from forests.

In the analysis presented in **Figure 1-3**, the largest taxonomic group significantly associated with pasture soils was the phylum Firmicutes, and specifically the order Bacillales and family Planococcaceae. Conversely, Thaumarchaeota, Acidobacteria, Verrucomicorbia, Gemmatimonadetes, Planctomycetes, and much of the Proteobacteria were found to be significantly associated with forest soils (**Figure 1-3**). An increase in the proportion of Firmicutes, a functionally-broad phylum, has likewise been shown in pastures of the same region sampled several years earlier, implying consistent long-term trends in the shifts of community composition (Rodrigues et al., 2013). A significant favoring of Firmicutes was also detected in pastures of the Tapajos National Forest (Pará, Brazil) relative to primary forests. While this was true year round, differences were particularly pronounced during the wet season (Pedrinho et al., 2019). Similarly, in active pastures of Ipiranga do Norte, Mato Grosso, Firmicutes increased in relative abundance by three to four-fold compared to primary forest (Mendes et al., 2015a; Mendes et al., 2015b). A T-RFLP-based survey of compositional changes in Benjamin-Constant (Amazonas State) reflected an opposite response, with Firmicutes decreasing significantly in pastures (da C Jesus et al., 2009). Multiple 16S-based studies have reached the same conclusion concerning a significantly decreased proportion of Acidobacteria, a phylum of Gram-negative, nonspore-forming bacteria typically favoring acidic environments (Dedysh and Damsté, 2018) within the prokaryotic community following pasture conversion (Khan et al., 2019; Navarrete et al., 2015b; Rodrigues et al., 2013). Subgroups 2 and 13 in particular showed a consistent trend, and reflected patterns of spatial biotic homogenization noted by whole community analysis (Navarrete et al., 2015b; Rodrigues et al., 2013). A survey of the response of Verrucomicrobia to land-use change in the same region using phylum-targeted sequencing generally showed greater diversity and relative abundance of subgroup 3 in pastures, but also found a significant favorability of the class Spartobacteria in forests, consistent with **Figure 1-3** (Ranjan et al., 2015; Rodrigues et al., 2013). While the comparison of relative abundance in **Figure 1-3** suggests favorability of Verrucomicrobia in forests, absolute quantification indicates proliferation with pasture conversion (Ranjan et al., 2015). Despite differences in sequencing techniques and pasture age, the agreement of findings from various studies conducted in the same geographic area builds confidence in its biological reality.

Using deeply sequenced metagenomic profiles, Kroeger et al. (2018) found significant abundance changes in 13 of 34 dominant phyla across forests and ca. 38-year-old pastures of the ARMO site. Gemmatimonadetes, Fusobacteria, Aquificae, Lentisphaerae, and Korarchaeota were among phyla identified as being significantly negatively impacted by forest-to-pasture conversion. Additionally, the abundance of Thaumarchaeota, an archaeal phylum containing many known ammonia-oxidizers, was extremely low in pasture communities, in agreement with **Figure 1-3** as well as Hamaoui et al. (2016) and Khan et al. (2019). Nitrospirae, a small bacterial phylum containing some lineages of nitrite-oxidizing bacteria, was also significantly negatively impacted by pasture conversion (Khan et al., 2019; Kroeger et al., 2018; Rodrigues et al., 2013). These results are not surprising since several members of the plant genera *Brachiaria* and *Urochloa*, common forage grass species in Amazonian pastures, are known to secrete a nitrification-inhibiting cyclic diterpene called brachialactone from roots. This compound works by blocking ammonia monooxygenase and hydroxylamine oxidoreductase enzymatic pathways, limiting NO_2^- /NO_3^- formation and energy production for these chemolithotrophs (Subbarao et al., 2009).

Across the studies discussed, differences in the responses of taxonomic abundance (sometimes reflecting opposite trends) with respect to land-use differences may be due to variability in locale, management, age since conversion, or operational variation such as different sequencing conditions, sample replication, and sequence processing. However inconsistent responses among studies may also be random, pointing to a poor understanding of microbial ecology. Further, explaining the compositional shifts observed is difficult. Thaumarchaeota and Nitrospirae are relatively narrow taxonomic groups with high intra-phylum functional similarity. Drawing inference regarding mechanistic controls over shifts in the relative abundance of Thaumarchaeota and Nitrospirae in response to land-use change is therefore somewhat straightforward. On the other hand, for much larger prokaryotic phyla, such as Proteobacteria, Acidobacteria, or Firmicutes, intra-phylum functional and physiological diversity is sufficiently high that few generalizations can be made, obscuring any deeper understanding as to why taxonomic groups may respond to land-use change in a particular way. Since most compositional analyses across land-use gradients in the Amazon have been done via 16S rRNA gene-based amplicon sequencing, inference is limited to relative abundance shifts, further conflating which taxonomic changes are true environmental responses.

Fungal Communities

Studies discussed thus far have focused primarily on assessing the response of prokaryotic (16S rRNA gene-based) or whole microbiotic (metagenome-based) communities to land-use change in the Amazon Rainforest. Although fungi are technically included as part of a whole community assessment, their proportional sequence abundance is typically dwarfed by bacterial-derived sequences. One meta-analysis indicated that bacterial rRNA sequences outweigh those of fungi by 20:1 on average and enzyme-encoding sequences derived from bacteria outweigh those of fungi by 163:1 on average, masking impacts related to important ecosystem functions that fungal communities perform (Bahram et al., 2021). Comparatively few studies have focused specifically on the response of fungal functional guilds or whole fungal communities to land-use change in the Amazon. Nonetheless, fungi are important components of forest and pasture ecosystems alike, with mycorrhizal fungi in particular serving as a symbiotic partner to over 90% of terrestrial plant species (Smith and Read, 2010).

Tropical ecosystems are typically dominated by arbuscular mycorrhizal fungi (AMF) (Schimann et al., 2017). Species richness and diversity of AMF do not appear to be significantly reduced by pasture conversion (Leal et al., 2013; Leal et al., 2009). This result is divergent from ecological theory, as well as a survey across forests of the western Amazon, which found a positive correlation between richness of forest plant communities and AMF from the order Glomerales (Peay et al., 2013; Wardle et al., 2004). This likely reflects an important distinction between how trees and forage grasses interact with AMF. Tropical host trees have been shown to make relatively small numbers of selective associations with AM species, and these associations change depending on the age of the tree (Husband et al., 2002). Conversely, forage grasses, like *Brachiaria*, associate with a wide range of AM species (Rodrigues and Dias-Filho, 1996; Teasdale et al., 2019). Despite similar decreases in plant species richness in alternate land-use types, including agroforestry and crop systems (in comparison to forest), AM fungal richness and diversity may actually increase in some cases (Sturmer and Siqueira, 2011).

There is also evidence that spore abundance increases with the conversion of primary forest to pasture (Leal et al., 2013; Leal et al., 2009; Sturmer and Siqueira, 2011), which is consistent with the observation that the forage grass *B. decumbens* induces high rates of sporulation (Carneiro et al., 1995).

This likely has important consequences for above and belowground biomass productivity (Cavagnaro et al., 2014), and should be studied further in the future. Despite overall compositional and spore count differences between systems, species of the genera *Glomus* and *Acaulospora* are abundant and cosmopolitan across land-use types (Leal et al., 2013; Sturmer and Siqueira, 2011).

Fungal communities as a whole play numerous roles in the soil: they may be community-regulating pathogens, plant symbionts, or important drivers of decomposition-related nutrient cycling (Maron et al., 2011; Martinez et al., 2009; Moore et al., 2015; Treseder and Lennon, 2015). Analysis of whole fungal community response to land-use conversion from forest to pasture reflects similar patterns observed for AMF: communities between land-use types are significantly distinct (Cerqueira et al., 2018; Mueller et al., 2014; Mueller et al., 2016). Similar conclusions were also drawn in comparing primary forest with converted monospecific- plantation fungal communities (Schimann et al., 2017). Plant community composition (but not richness) has been found in some cases to act as a significant driving factor in determining fungal composition (Mueller et al., 2014; Schimann et al., 2017). Conversion of forest to pasture appears to induce a decrease in the beta-diversity of the fungal community, indicating greater spatial homogenization, similar to the response of the prokaryotic community (Cerqueira et al., 2018; Rodrigues et al., 2013). Pastures also appear to favor colonization by generalist fungi, concomitant with decreased species richness and independent of factors such as pasture age (Mueller et al., 2016). This may be the result of greater niche competition by fungi able to tolerate extreme conditions.

Some disagreement in the assessment of fungal response to forest-to-pasture conversion has arisen across studies. While Mueller et al. (2016) and Fracetto et al. (2013) found a significant decrease in taxa richness with pasture conversion, Cerqueira et al. (2018) observed increased diversity overall. In addition, while Mueller et al. (2014) observed a significant reduction of the phylum Basidiomycota in pastures, Cerqueira et al. (2018) and (Fracetto et al., 2013) found increased representation of this phyla in pastures. Differences across studies may be explained by factors such as regionality or seasonality at the time of sampling. Owing to the different functions that soil fungi can provide to an ecosystem, it is imperative that we continue to expand our understanding of how soil fungal communities respond to land-use conversion in the Amazon Rainforest.

Genomic Features and Novel Organisms

Inclusion of all genetic material in the profiling of Amazon microbial communities allows evaluation of general shifts in taxonomic and functional composition, but also enables broad assessment of genomic features. Metagenomic analysis of soils from forests and pastures in the ARMO site showed clear genomic alterations caused by land-use change (**Figure 1-4**). DNA reads with low GC content (35–55%) appear significantly depleted in pasture soils. This is a nonspecific genomic feature, and the driving force or functional consequence of this pattern across land-use types is not readily apparent. However, bacterial GC content, particularly among Gram-negative bacteria, appears positively correlated with genome size (Li and Du, 2014). Some studies have further shown environmental selection mechanisms on GC content, and specific physiological features, such as aerobiosis, have also been associated with high levels of GC (Foerstner et al., 2005; Naya et al., 2002). Further analysis is needed to understand the significance, if any, of this large genomic shift.

Amplification of soil DNA and subsequent attempts to taxonomically annotate sequences typically result in high proportions of unidentified taxa (i.e., no culture match from publicly-deposited sequences; Bach et al., 2018). In an amplicon-based survey of Verrucomicrobia, nearly half of sequences were unidentified, with disproportionate representation in forest soils (Ranjan et al., 2015). While amplicon or unassembled metagenomic sequence data can provide detailed profiles of functional and taxonomic diversity within a community, sequence data lack information concerning the full genetic potential of community members, as well as the ability to describe the genetic potential of novel organisms. However, deep metagenomic sequencing data may be utilized to assemble complete or nearly-complete genomes from the pool of community sequence reads, known as Metagenome Assembled Genomes (MAGs). Employing this technique on soil communities of Amazon forests and pastures produced 28 MAGs, many of which were exclusively identified in pasture compared to forest samples; just a few were found in exclusively in forest samples (Kroeger et al., 2018). Some MAGs identified within the phylum Acidobacteria were placed in lineages containing no currently cultured organisms, and a phylum Melainabacteria MAG found only in pasture soils was placed within a new lineage in the candidate order Obscuribacter. Additional work exploring MAGs identified very small genomes from the Candidate Phyla Radiation (CPR) Patescibacteria in pasture soils, a vast and previously uncultured group (Nascimento

Lemos et al., 2020). The ability of MAGs to recover this unusually small genome calls into question current paradigms concerning the favorability of large genomes in the soil environment, which are based on 16S rRNA gene- or culture-based studies (Nascimento Lemos et al., 2020). Successful recovery of MAGs is computationally limiting, and likely for this reason has not been broadly applied to exploration of land-use change impact on soil microbial communities. However further integration of MAGs into microbial ecology studies could be essential in exploring the unculturable biosphere as well as examining the multifunctional genetic potential of microbial community members.

Soil Physicochemical Effects

The physical and chemical conditions of the soil environment have clear impacts on the composition, diversity, and function of microbial communities. Shifts in these soil-environmental conditions associated with land-use change are in large part what cumulatively shapes microbial community response. Some variables may be affected more strongly by land-use change (e.g., Diochon and Kellman, 2008) or may have a greater impact on microbial communities (Bending et al., 2002; Jones et al., 2019). However due to high spatial heterogeneity even within land-use types (Ritter et al., 2019), physicochemical conditions can also be a confounding factor that should be accounted for. Many soil chemical variables interact with each other: C and N content, for example, have highly constrained ratios in biological tissues (Cleveland and Liptzin, 2007). pH affects the availability of several other soil nutrients such as phosphorus (P; Penn and Camberato, 2019). Additionally, the chemical or structural composition of substrate pools may be an important consideration, even if absolute pool size is unaltered (Ng et al., 2014).

Tropical soils require unique regional considerations in determining the impacts of land-use change on soil physicochemical conditions, and subsequently soil microbial communities. The most common soil types found in the BLA are red-yellow podzols and latosols, equivalent to ultisols and oxisols in the United States-based soil taxonomy system (Moraes et al., 1995). These soils are moderately to highly weathered, predominated by secondary minerals, including kaolinite, as well as iron and aluminum oxides, with soil solution pH typically below 6.0 (Kitagawa and Möller, 1980). Aluminum (Al) saturation tends to be high, and cation exchange capacity (CEC) is low, typically less than 10 $cmol_c\ kg^{-1}$ soil

(Bernoux et al., 1998; Dalling et al., 2016). The general paradigm concerning nutrient availability in tropical systems, especially lowland forests of the Amazon Basin, is that most nutrients are immobilized in biomass rather than soil, and that P specifically is limiting to net primary production (Dalling et al., 2016; Vitousek, 1984). N is typically considered more abundant in tropical forests, particularly in comparison to temperate systems (Hedin et al., 2009).

Many studies considering the impact of land-use change on microbial communities in the Amazon measure soil environmental variables as covariates, and some consensus has been reached among studies as to important land-use change-related physicochemical factors shaping microbial community response. Across the studies discussed here, Al content is consistently identified as a significant correlate with microbial community metrics. In most cases, Al content varies significantly with overall community structure. This has been reported from several study regions, including Benjamin Constant, Amazonas (da C Jesus et al., 2009), Ariquemes, Rondônia (ARMO; Cenciani et al., 2009; Khan et al., 2019), Querência municipality, Mato Grosso (Goss-Souza et al., 2019), the Tapajós National Forest, Pará (Merloti et al., 2019; Pedrinho et al., 2019), and across primary forests, crop systems, and pastures ranging from 20–40 years since conversion (see **Table 1**). In one report, Al content was correlated with both community taxonomic and functional structure (Pedrinho et al., 2020). Several diversity metrics have also been shown to correlate negatively with Al content (Goss-Souza et al., 2019). Additionally, saturation index and total content of Al are among the most universally significant correlation factors with abundance of various taxonomic groups across forests and pastures (Mendes et al., 2015a; Navarrete et al., 2015a). While the precise importance of Al is unknown in the context of forest-to-pasture conversion, in tropical soils a large proportion of exchange sites are likely occupied by acidic cations (H^+ and Al^{3+}). Therefore Al concentration may be an indicator of CEC or soil fertility rather than a direct control over communities (Carvalho et al., 2009). This explanation is supported by a simultaneous correlation of community metrics with base saturation, total soil acidity, and CEC (Khan et al., 2019; Mendes et al., 2015a). Under this hypothesis, microbial communities may be shaped by plant community response to CEC. Alternatively, soil solution exchangeable Al has been shown to correlate negatively with microbial C use efficiency as pH declines below 5.5, and therefore may directly influence microbial community structure and diversity by selecting for organisms able to tolerate this exogenous stress through detoxification (Auger et al.,

2013; Jones et al., 2019). Primary forests in the Amazon Basin typically have a soil pH ranging from ~3.8 to ~5.3 (see references in Table 1, but e.g., Lammel et al., 2015b; Mendes et al., 2015a; Neill, 1995; Pedrinho et al., 2019). In land-use transition, slash-and-burn clearing methods typically lead to an increase of several tenths to over 1 pH unit (potentially crossing the aforementioned 5.5 transition point), with gradual acidification over several decades of use. This may explain community differences both across land-use types as well as in various regions of the basin.

In some reports, pH is also an important factor influencing the attributes of microbial communities such as compositional structure and phylum-level taxonomic abundance (da C Jesus et al., 2009; Lammel et al., 2015b; Mendes et al., 2015a). The community composition of AMF across a forest-to-pasture land-use gradient in the same region has also been shown to correlate significantly with pH (Leal et al., 2013). Studies across many other ecotypes have similarly shown pH to be an important determinant of fungal and bacterial communities, likely because it serves as a control over multiple soil conditions including nutrient availability and mobility, exoenzyme activity, and concentration of cellular stressors (Jones et al., 2019; Lauber et al., 2009; Puissant et al., 2019; Rousk et al., 2010). Mechanistically, moderate ash-induced increases in soil pH following slash-and-burn conversion of forest to pasture may be responsible for the frequently observed increase in local community diversity and richness (e.g., Rodrigues et al., 2013), since this theoretically allows for greater nutrient availability and release from growth constraints (i.e., stressors) imposed by the preexisting acidic conditions in primary Amazon Rainforest soils (de Souza Braz et al., 2013). Concentrations of nutrients and enzymatic co-factors including P, K, S, Ca, Mg, Fe, Mn, B, Cu, Zn, and Mo have pH-dependent availability, with the majority of these increasing as pH increases. Availability of these nutrients has also been shown to play a role in shaping microbial communities at the phylum level (Mendes et al., 2015a) and across functional groups (Pedrinho et al., 2020). Stress response has been identified as a significantly divergent functional characteristic across forest and pasture (Pedrinho et al., 2019), but a more in-depth investigation of stress tolerance genes related to pH-dependent cellular toxins, such as Al, may provide additional insights into the mechanisms driving functional assembly of a community across land-use types.

Several other factors including soil C and N content have also been identified as meaningful correlates across studies at the scale of whole communities, as well as prominent phyla (Cenciani et al.,

2009; Cerqueira et al., 2018; Navarrete et al., 2015a; Ranjan et al., 2015). This is perhaps unsurprising given the impact these nutrients have on microbial metabolism and the control they exert over biomass stoichiometry (Cleveland and Liptzin, 2007). Soil water holding capacity has additionally been shown to be significantly related to overall nutrient availability and average water content in the soil environment, therefore likely influencing community composition through differential preferences and tolerances for soil moisture conditions among taxa (Pedrinho et al., 2019; Zhao et al., 2016). The relationship between land-use disturbance, soil physicochemical conditions, and microbial communities is interactive. The response of the microbial community to land-use change is mediated by shifts in the physical and chemical conditions of their environment, which are in turn dependent on pre-existing edaphic and climatic conditions. Therefore, physical and chemical soil attributes are important considerations to understand mechanistic drivers of community change, as well as account for variation in response on the regional scale. While many studies have measured these variables and analyzed associations with microbial communities across land-use types, our body of knowledge would be greatly improved by hypothesis-driven and experimentally controlled studies in order to draw causal inference.

Beyond Taxonomy: Functional Diversity and Community Interaction

Along with obtaining an inventory of taxa altered by ecosystem disturbance, a central goal of studying the microbial communities of the Amazon Rainforest is to understand their role in mediating soil biological processes. The first comprehensive study of the functional gene diversity of soil microbial communities under land-use change in the Amazon was performed at the ARMO site. The study took advantage of the GeoChip 4.0, a microarray containing 83,992 probes targeting 410 gene families associated with the biogeochemical cycles of C, N, P, and S (Tu et al., 2014). This high-throughput microarray approach detected genes for 409 different families, underscoring the general richness of genes present in Amazon soils. However, reported losses of total gene richness with conversion of primary forest to young pasture (~6 years old) were significant—up to 31.8% (Paula et al., 2014). Genes related to C and N cycles, particularly to the processes of methane oxidation, nitrification, and denitrification, were significantly associated with forest sites while their abundances were reduced in pastures. Such a dramatic negative shift is compelling since local-scale taxonomic diversity increases

with pasture conversion (Rodrigues et al., 2013). In contrast, a metagenomic-based study in the Tapajós National Forest, Pará, Brazil, found that functional diversity increased significantly in conjunction with taxonomic diversity (Pedrinho et al., 2019).

Further, ARMO metagenomes have revealed that carbohydrate metabolism, sporulation, and cell signal regulation functional geness were significantly more common in pasture, while RNA metabolism and cofactors, vitamins, and pigments functional groups were more prevalent in forests (Kroeger et al., 2018). These functional profile shifts with land-use change are consistent with another study conducted in the Tapajós National Forest, Pará, Brazil, where the same trends hold true in microbial communities across both the wet and dry season (Pedrinho et al., 2019). The consistency of these studies is intriguing, but future work is needed to connect these broad community functional shifts to taxonomic representation in the context of land-use change, in order elucidate the mechanistic underpinnings that drive community response. Using metagenomic profiles, Pedrinho et al. (2020) assessed the ecological response of N-cycling community members to land-use change and determined structural alterations associate significantly with (perhaps unsurprisingly) nitrate (NO_3^-) and ammonium (NH_4^+) concentrations. Further, pasture communities contained a high proportion of specialists compared to primary forests.

An aspect of soil microbial ecology that is exceedingly difficult to study through experimentation is functional interactions among taxa. Instead, statistical methods are employed to infer interactions based on co-occurrence across samples, a potential indication of shared niche habitation between taxa. Forests and pastures of Rondônia have distinctly different co-occurrence networks that self-sort by land-use type (Khan et al., 2019). Additionally, nodes clustering taxa by the same functional potential suggests shifts in soil N cycling with land-use conversion (Khan et al., 2019), which is supported by a network analysis performed by Pedrinho et al. (2020). Analysis of rhizosphere versus bulk soil communities in soybean fields converted from primary forest suggest directional, niche-based assembly of communities near roots, compared to neutral (i.e., randomized) community assembly in bulk soil (Goss-Souza et al., 2020), but no such analysis has yet been done to compare forest and converted systems. Functional shifts in soil communities with land-use conversion are clear, but more process-based focus is needed to fully understand nutrient cycle shifts and explain variation across studies.

Microbial Impacts Associated with Carbon Cycling

The vast expanse of the Amazon rainforest makes it a critical global carbon storage hotspot. However, the initiation of forest-to-pasture conversion through biomass burning definitively leads to a net loss of C per area of former forest. While aboveground biomass accounts for an estimated 400 Mg C stored per hectare of primary forest across the basin (mainly in trees with >10 cm breast-height diameter), pastures store less than one-sixth of this, approximately 63 Mg C per hectare (Hughes et al., 2002; Nascimento and Laurance, 2002). Taking into account soil class-specific differences in organic matter content and distribution across the Amazon Basin, average C density within primaryforest soils has been estimated at 98–103 kg C per hectare—on a similar scale as the aforementioned aboveground stocks (Batjes, 1999; Moraes et al., 1995). Fifty-two percent of C stocks are estimated to be held in the top 30 cm of soil, which are most susceptible to disturbance with land-use conversion (Batjes and Dijkshoorn, 1999). However the net impact of land-use conversion on soil C is far less clear as compared to impacts on plant biomass-stored C. Various studies have found that soil C increases (de Moraes, 1996; Durrer et al., 2021; Neill et al., 1997a), decreases (Fearnside, 1997; Maia et al., 2010), or does not change appreciably (Durigan et al., 2017; Rittl et al., 2017) following land-use conversion. This suggests that the impact of land-use change on soil C storage is dependent either on pre-existing conditions, such as initial C stocks, soil texture, and nutrient status, or management factors such as the frequency of burns or grazing intensity. In addition, aspects of the microbial community, including net carbon use efficiency, microbial biomass, and genetic potential for degradations are likely relevant.

Respiration, Microbial Biomass, and C Degradation

Across pasture chronosequences in Rondônia, Brazil, soil C concentration consistently shows an increase with pasture age, and isotopic- $\delta^{13}C$ values become significantly less negative, indicating gradual turnover of forest-derived C and replacement with pasture-derived C (Durrer et al., 2021; Neill et al., 1996). Yet, the change in $\delta^{13}C$ value of microbial respiration greatly outpaces that of soil C stocks; in one study, for example, pasture-derived C of 3-year-old pastures constituted 17% of soil stocks, but 69% of microbial-respired C (Neill et al., 1996). These findings indicate that microbial activity in pastures is driven primarily by fresh inputs, including root exudates and root and shoot tissue. Another study found

accumulation of pasture-derived C to be highest in the particulate organic matter fraction (Lisboa et al., 2009), again indicating that elevated proportions of pasture-derived organic C are either respired or converted to biomass before accumulation in smaller fractions. The mechanisms driving retention of forest-derived C are unclear, but losses appear to be greatest from the silt-sized soil fraction, suggesting soil texture has an interactive role in C storage dynamics under land-use change.

Microbial activity has also been shown to depend heavily on total soil organic C content and pasture age, with metabolic quotients (respiration per unit biomass) highest in young, surface-soil pastures (1–2 years old, 0–2 cm in depth) which contain lower total (and presumably) pasture-derived C as well as lower microbial biomass (MB) C compared to primary forests or older pastures (Melo et al., 2012). Older pastures in this study (5–12 years) did not accumulate significant soil C or MB C, but metabolic quotients did return to reduced levels similar to forests, indicating shifts in C usage by microbial communities over time as soil pools change in their quality and composition. Other studies have found that MB C typically increases at depths to 20 cm with pasture conversion and variable pasture age (Cenciani et al., 2009; Cruz et al., 2019). Overall, however, concentrations of MB C across pastures may vary eight-fold based on season (dry versus wet) and soil type, accounting for differences across studies and potentially influencing how microbial communities respond to land-use change (Cerri et al., 2006).

Community profiles of catabolic metabolism using substrate-induced respiration across forests and pastures have confirmed differential activity under varying land-use types, with forest communities responding to malonic, malic, and succinic acid, and pasture communities responding to carboxylic and amino acids (Mazzetto et al., 2016). Correspondingly, shifts in the functional profiles related to carbohydrate metabolism derived from metagenomes (Kroeger et al., 2018) further indicate that microbial communities respond to differential organic-matter profiles across land-use types. Among annotated protein-encoding reads of metagenomes, a decrease in lignin-degradation genes, such as superoxide dismutase, was observed in pastures compared to forests (Kroeger et al., 2018). Further, the chemical-structural composition of soil organic matter (i.e., substrate) changes with forest-to-pasture conversion, resulting in an increased concentration of hemicellulose and a decreased concentration of lignin in pastures compared to primary forest, even if total pool size is unchanged (Lammel et al., 2015b). Additionally, pasture soils contain a significantly higher concentration of permanganate-oxidizable C,

which serves as an indicator of easily-catabolized C (Durrer et al., 2021). Several other metagenomic-based studies reporting bulk annotation of genes related to degradation of aromatic compounds did not find differences between land-use types (Mendes et al., 2015b; Pedrinho et al., 2019). Paula et al. (2014) found that abundance of C degradation-associated genes was negatively impacted in young (6-year-old) but not older (38-year-old) pasture compared to forests. The young pasture sampled in this study demonstrative significant overall decline in functional diversity, which may be indicative of an important time-dependent response. However, more intermediary-aged pastures between 6 and 38 years old are needed to test this theory. In future studies, shifts in gene expression rather than presence may prove more useful in characterizing soil microbial processes related to C degradation.

Overall, understanding how land-use change impacts C degradation and storage in Amazon soils is limited by a dearth of hypothesis-driven studies and a poor understanding of the relationship between community activity and genetic profiles related to C metabolism. Due to the relatively narrow scope of genes and microorganisms involved in its cycling, the impact of land-use change on methane flux in the Amazon has garnered far more attention and is better understood in terms of ecosystem-scale consequences and functional underpinnings.

Methane Flux

Methane (CH_4) is a climatically-relevant gas with a potency 34 times that of carbon dioxide over 100 years (Myhre et al., 2013). The impact of forest-to-pasture conversion on methane flux has shown consistent trends across the Amazon Basin. Process rate studies in Rondônia (Western Amazon) and Pará (Eastern Amazon) have shown steady, annual sink-to-source trends with conversion from forest (-470 mg CH_4 m^{-2}) to pasture (+270 mg CH_4 m^{-2}), but no clear trend in emissions has been found with respect to time since pasture conversion (Steudler et al., 1996). Another study across a conversion chronosequence in Paragominas (Eastern Amazon) found that pastures appear to return to a methane sink as they age (Verchot et al., 2000). A previous study has also indicated a consistent sink and source status across all seasons in forests and pastures, respectively, and that pastures act as a much stronger source in wet seasons (up to +614 mg CH_4 m^{-2} yr^{-1}, or +1682 µg CH_4 m^{-2} d^{-1}) compared to dry (Fernandes et al., 2002). Measurements conducted in the same region approximately two decades later confirmed

pastures to be a persistent source of methane averaging 3454 ± 9482 μg CH_4 m^{-2} d^{-1} (Meyer et al., 2020). However, Meyer et al. (2020) found that forests were also a weak source of methane (9.8 ± 120.5 μg CH_4 m^{-2} d^{-1}). Another investigation in Sinop, Mato Grosso (Southern Brazilian Amazon) found net daily efflux rates from a 25-year-old pasture of 1104 μg CH_4 m^{-2} d^{-1} and low daily uptake rates (-168 μg CH_4 m^{-2} d^{-1}) from primary forest (Lammel et al., 2015a).

An early study focusing on process rates and physicochemical shifts in soil profiles under land-use change concluded water-filled pore space to be the key factor in sink-to-source transition of converted pastures, with ~40% filled pores considered the tipping point (Steudler et al., 1996). Indeed, oxygen concentration (inversely related to water-filled pore space) has been demonstrated to be an important influence over production (Yang and Chang, 1998). The proposed mechanism of an altered methane cycle is that soil bulk density increases with pasture conversion due to cattle movement. This causes a decrease in soil porosity and fractures soil aggregates (Reiners et al., 1994), increasing the frequency of oxygen-devoid microsites that favor the anaerobic process of methane production (Fernandes et al., 2002). However, this physicochemical-focused understanding ignores the direct and indirect impacts of forest conversion on methane-cycling soil microbial community members. To fully understand the differences in net methane flux with land-use change, the impact on the abundance, structure, and activity of soil microbial communities must be considered.

A net flux of soil CH_4 from soils results from a balance between production by methanogenic archaea and consumption by methanotrophic bacteria (Conrad, 2007). Under anaerobic conditions, CH_4 is generated through the reduction of C1 carbon compounds such as CO_2 and methanol or disproportionation of compounds such as acetate (Schäfer, 2013). Methanotrophic bacteria of upland soils, meanwhile, mostly rely on aerobic conditions to oxidize methane to methanol using O_2 as an electron acceptor with either particulate- or soluble methane monooxygenases (Conrad, 2007; Guerrero-Cruz et al., 2021). Although thought to be important mainly in wetland and marine systems, canonical methanotrophic archaea also perform methanotrophy anerobically, though recent work suggests a greater importance of these organisms to upland soil CH4 cycling than previously thought (Guerrero-Cruz et al., 2021; Ho et al., 2019). Metagenomic sequencing of forest and pasture soil communities has indicated that these CH_4-consuming communities (methanotrophs) are strongly impacted by land-use

conversion, with decreased relative abundance and shifted taxonomic composition of methanotroph 16S rRNA genes (Meyer et al., 2017), which confirms a prior GeoChip-based assessment that found a significant association of the abundance of methane monooxygenase genes (denoting methanotrophs) with forests compared to pastures (Paula et al., 2014). In particular, a significant drop in the relative abundance of methanotrophs within the Alphaproteobacteria was detected (Meyer et al., 2017). Furthermore, annotation of community metagenomes of Rondônian soils (ARMO) revealed that all essential genes encoding methyl coenzyme M reductase (*mcr*) were significantly enriched in pasture soils, and particulate methane monooxyenase (*pmo*) genes were significantly enriched in forest soils (Kroeger et al., 2018). Using a 16S rRNA gene amplicon-based approach confirmed the alteration of functional community patterns across two locations in the Amazon in Rondônia and Pará, with significant increases in methanogen composition and relative abundance, as well as significant decreases in richness and relative abundance of methanotrophs in pastures (Meyer et al., 2020). Similarly, absolute gene quantification has shown a significant increase in *mcrA* (~2.5x) and a significant decrease in *pmoA* (~0.5x) gene copy numbers per gram soil in pastures compared to forests (Lammel et al., 2015a).

Taking microbial activity into account expands on these findings. In the 16S rRNA gene-based analysis discussed above, the flux of CH_4 across soil types appeared to be associated with richness and relative abundance of methanogenic communities after accounting for sample covariate structure (Meyer et al., 2020). **Figure 1-5**, reproduced with permission from Meyer et al. (2020), reflects a positive relationship between both community attributes and CH_4 flux ($R^2 = 0.42$ for both) when land-use types are considered together, but linear relationships appear particularly driven by pasture communities. In another study, an absolute quantitative analysis of CH_4 -cycling genes (*mcrA* and *pmoA*) did not show a significant direct explanatory relationship with flux rates, which did not vary between forest and pasture (Lammel et al., 2015a). This finding may speak to the spatial and temporal heterogeneity of the process. While 526 taxa were highly associated with methane flux in pasture soils, just 41 taxa were associated with flux in primary forest. Moreover, few of these taxa were known methane-cyclers, indicating a wide range of organisms associated or co-correlated with the process, but not directly mediating it (Meyer et al., 2020). Use of an isotopic tracer (^{13}C) to enrich and identify microorganisms involved in methane cycling under a given set of conditions has also indicated an increase in the abundance and diversity of

active methanogens in pastures (Kroeger et al., 2021). These results highlight the importance of integrating process-based measurements with microbial community profiles to understand their role in ecosystem function. It should be noted that while the ecology of CH_4 cycling across this large-scale land-use gradient represents a complex and important area of study in soil microbial ecology, the contribution of upland soils only amounts to approximately 5% of the overall CH_4 emissions across the Amazon Basin, with biomass burning and cattle accounting for the major sources (Steudler et al., 1996).

Microbial Impacts Associated with Nitrogen Cycling

Nitrogen, an essential nutrient for both plants and soil microorganisms, is tightly linked in its cycling with C through both soil organic matter and biomass. Soil N comprises both organic and inorganic nutrient pools and is utilized for biomass assimilation as well as dissimilatory energetic reactions (Pajares and Bohannan, 2016). A thorough review of N cycling in tropical soils with discussion pertaining to climatic and pedological factors unique to this region is provided in Pajares and Bohannan (2016), although the focus of the review remains on forest ecosystems generally rather than on land-use conversion gradients. Unlike temperate soils which are geologically younger, highly weathered tropical soils are high in N, but limited in P and base cations (Hedin et al., 2009). However, disturbance related to pasture conversion may alter this pattern. The reality that a high proportion of pastures are eventually abandoned (Chazdon et al., 2009) presents a pressing need to better understand how microbial communities mediate N cycling across land-use change gradients. This is especially relevant since N limitation is a suspected contributor to productivity decline. Similar to C pools, total N has been reported to increase (Navarrete et al., 2015b; Neill et al., 1997b), decrease (Lammel et al., 2015a; Neill et al., 1997b; Pedrinho et al., 2020), or remain stable (Durigan et al., 2017; Melo et al., 2012; Neill et al., 1997a) with forest-to-pasture conversion. Regionality, intensity of conversion practices, and pasture age/management could account for many of these differences; however, significant shifts have been reported in both transformation rates and pool sizes of N in pastures (Neill, 1999).

Several aspects of land-use conversion to pastures are plausible contributors to the above shifts, which are conceptualized in **Figure 1-6**. Common forage species used throughout pastures in the Amazon are known to exude a nitrification inhibiting compound called brachialactone as a strategy for

scavenging N (Egenolf et al., 2020; Subbarao et al., 2015). It is not clear how soil N availability scales with brachialactone production (e.g., whether exudation is obligate or facultative), but its presence in the rhizosphere is likely to impact community structure and activity of N-cycling soil microbial communities. The effect of cattle grazing in pastures is additionally a relevant consideration. Cattle grazing exports nutrients from the ecosystem through the animal itself, but also through displacement and inefficient nutrient recovery following excretion; this may contribute to N loss through NH_3 volatilization, leaching, or erosion (Dias-Filho et al., 2001). This could cumulatively have a substantial impact on available N supply with a net export being estimated up to 20 kg N ha^{-1} yr^{-1} (Dias-Filho et al., 2001). Limited reporting on the relationship between grazing intensity/rotational practices and soil C and N dynamics in tropical systems suggests that soil C and N cycling and storage may be unaffected by light grazing, but depressed under medium or heavy stocking rates (Cantarutti et al., 2002; Silva et al., 2008). Since soil microbial functional groups mediate key steps in the soil N cycle, such as nitrogen fixation, mineralization, nitrification, and denitrification, their response to land-use alteration in the Amazon deserves thorough examination to better understand shifts in the N cycle across the region.

Nitrogen Fixation

Nitrogen fixation, the conversion of atmospherically-derived N_2 to a biologically reactive form (NH_3) is an imperative function for providing new N to terrestrial ecosystems, especially in the early stages of succession. Despite the relative abundance of N in tropical forest soils, both symbiotic and associative/free-living N_2 fixation (SNF and ANF, asymbiotic nitrogen fixation) in soil, leaf litter, and the tree canopy are thought to be important processes in response to considerable loss through NO_3^- leaching and denitrification, as well as in order to maintain extracellular phosphatase activity required to combat P limitation (Hedin et al., 2009; Pajares and Bohannan, 2016). A key difference in potential diazotrophy with forest-to-pasture conversion is that forage grasses sown in pastures are not leguminous, and therefore do not engage in symbiotic nodulation or SNF. This is likely to have effects on community composition, regardless of N_2 fixation rates. Based on the potential loss of pathways associated with landscape conversion to grazed pastures, it is reasonable to postulate that associative and free-living diazotrophs may have an elevated role in N-cycling (via ANF) following land-use change.

SNF- diazotrophs have the benefit of O_2-depleted conditions (O_2 reversibly inactivates the nitrogenase enzyme) and a direct supply of C and energy from root nodules, in exchange for some of the N_2 they fix. ANF diazotrophs, on the other hand, are subject to more extreme environmental conditions and must fund the steep energetic cost of ANF themselves. It is then probable that their activity is almost entirely limited to the rhizosphere, where plant root deposits may serve as an easily-utilized C source, stimulating microbial activity (Figure 1-6; Bürgmann et al., 2005). The impact of grazing on plant root C allocation appears highly dependent on forage species, grazing intensity, and soil nutrients (Dawson et al., 2000; Hamilton III and Frank, 2001). However, based on limited data, low cattle stocking rates may stimulate belowground C allocation (Dias-Filho et al., 2001; Durigan et al., 2017; Trumbore et al., 1995). Due to high energetic costs, ANF may also be heavily regulated by NH_3 concentration (Peters et al., 2013), bringing into question the potential impact of nitrification inhibition by brachialactone. Depending on the efficiency with which forage species scavenge available NH_3, environmental concentrations may be high enough to inhibit ANF. On the other hand, ANF also requires a supply of N (Peters et al., 2013), so increased NH_3 concentration may stimulate activity in conjunction with C-rich root exudates.

Lima et al. (2009) performed a valuable survey of viability, diversity, and efficiency of symbiotic (nodulating) diazotrophs across primary forests and converted landscapes including pasture. Using a promiscuous legume, nodulation rates were highest in agroforestry soils and lowest in primary forests; however, the most efficient strains (performing SNF in the presence of high N content) were identified from forest soils, whereas pastures were typified by mostly inefficient strains. Given the lack of leguminous plant species in pastures, their continued presence suggests alternative ecological roles and a strong potential for root colonization in a pasture abandonment/secondary forest succession scenario. A free-living, diazotroph-targeted culture-based experiment found that pasture soils yielded the highest cell densities when compared to primary forest, agricultural, and agroforestry systems (Silva et al., 2011). Subsequent protein profiling identified *Burkholderia* and *Bacillus* among pasture-derived isolates. Nitrogenase activity was variable across strains within and among land-use types, but the highest performing strain was isolated from pasture soils. Due to the inherent limitations of assessing environmental biodiversity using culture-based methods, conclusions from these studies should be taken with care but serve as a baseline for comparison with molecular-based methods.

Assessment of diazotrophs at the ARMO sites was conducted by sequencing gene clones, as well as directly amplifying the *nifH* gene, which encodes an essential subunit of the nitrogenase enzyme (Mirza et al., 2020; Mirza et al., 2014). Quantitative analysis first indicated a significant increase in *nifH* gene abundance per gram of pasture soil compared to that in forest soil (Mirza et al., 2014), and community structure varied significantly by land-use type, both taxonomically and phylogenetically. Analysis of community turnover revealed the same distance-decay pattern observed for total prokaryotic communities (Rodrigues et al., 2013), indicating spatial biotic homogenization in pastures compared to forests. However, this result was more apparent taxonomically than phylogenetically. Similar to the local diversity increase observed in total prokaryotic communities, taxonomic and phylogenetic richness increased significantly in pasture soil (Mirza et al., 2020). Compositional comparison of sequenced clones have revealed that Deltaproteobacteria, uncultured Spirochaetes, Verrucomicrobia, and Archaea arre favored in pastures while Cyanobacteria, several Firmicutes, and an uncultured Archaea were found exclusively in forests. While the finding of strong compositional differences is interesting, the inability to annotate below the phylum-level in this study limits interpretation (Mirza et al., 2014). Co-occurrence networks built from Rondônian forest and pasture microbiomes have further suggested an important role of diazotrophic taxa in pasture communities (Khan et al., 2019). A survey conducted in the Tapajós National Forest in Pará compared total N fixation-associated gene sequence counts from metagenomes across forests and pastures and found a significant association with pasture soils across both the wet and dry season, which is in agreement with findings from Rondônia (Pedrinho et al., 2020). This was also bolstered by *nifH* gene amplification via qPCR, which reflected significant increases in pasture. Gene abundance has also been measured at significantly higher concentrations in corn/soy cropping systems two years following forest conversion, compared to primary forests (Merloti et al., 2019).

However, quantification of *nifH* genes by Lammel et al. (2015b) and Paula et al. (2014) via GeoChip profiling conversely inferred significantly larger diazotrophic communities in primary forest compared to pasture. The cause for disagreement between groups of studies is not readily apparent but may be due to regional heterogeneity, the impact of grazing intensity on nutrient status and belowground C allocation by forage grass roots, or operational factors such as taxonomic coverage of the primers selected. Furthermore, diazotrophy is phylogenetically widespread across prokaryotes (Gaby and

Buckley, 2014), and presence of taxa capable of ANF does not imply activity. The future inclusion of rate measurements to support molecular data is imperative in order to more conclusively understand diazotrophic response to land-use change.

Nitrification, Denitrification, and N Oxide Flux

Nitrification, a multistep reaction whereby NH_4^+ is converted to NO_3^- (**Figure 1-6**), is crucial in the overall N cycle of terrestrial ecosystems. This process impacts N oxide emissions, leaching potential, and preferential nutrient availability to plants and microbes (Pajares and Bohannan, 2016). Nitrification is mediated by chemolithoautrophic ammonia oxidizers of both the archaeal (AOA) and bacterial (AOB) domains, nitrite-oxidizing bacateria, and heterotrophic bacteria and fungi that can utilize organic N as a substrate (Zhu et al., 2014). A study of N cycling rates across forest and pasture chronosequences in Rondônia, Brazil, measured a decline in NO_3^- pool sizes with forest-to-pasture conversion, which typically coincided with consistently reduced net nitrification rates. Gross rates of nitrification were subsequently shown to decline with pasture conversion (Neill et al., 1997b; Neill, 1999). Furthermore, these patterns appear stable across both dry and wet seasons.

More than a decade later, a comprehensive survey of ammonia oxidizers with pasture conversion at ARMO sites in Rondônia revealed a significant decrease (more than 10-fold) in Thaumarchaeal *amoA* gene copy numbers, as discussed previously (Hamaoui et al., 2016). This trend was present irrespective of pasture age since conversion and the results are consistent with earlier findings at ARMO sites using GeoChip 4.0 (Paula et al., 2014). Intriguingly, AOB were not detectable in this study, even with the use of several primer sets. This may be due to the low pH in these environments or stronger inhibition of AOB communities by brachialactone production (Hamaoui et al., 2016; Hatzenpichler, 2012; Subbarao et al., 2015). Overall, the community structure of Thaumarchaeal ammonia oxidizers was significantly dissimilar between land-use types, and compositional analysis revealed the most drastic shift to be an almost complete removal of taxa from the genus *Nitrosotalea* in pastures. While the driving mechanism of *Nitrosotalea* loss is unknown, it may have important implications for depressed rates of nitrification (Shen et al., 2013). In contrast to the above study, a significant increase in absolute abundance of AOA was observed in 25-year-old pastures near Sinop, Mato Grosso (Lammel et al., 2015a). However, in Lammel

et al. (2015a), AOB were detected in pasture soils, but their abundance was approximately half that measured in forest soils. In addition, NO_3^- pool size was significantly lower in pastures compared to forests, consistent with nitrification inhibition by brachialactone and/or overall N limitation. It is possible that differential abundance trends in Hamaoui et al. (2016) and Lammel et al. (2015a) are due to variable selection pressures across AOA and AOB community members with pasture conversion, but the latter study did not measure compositional differences. Earlier analysis of total archaeal *amoA* gene diversity from primary forest and pasture in Benjamin Constant found slight reductions in diversity with pasture conversion, but increases in richness (Navarrete et al., 2011). In addition, no reduction in NO_3^- pool size was observed. Forage species present in these pastures were not specified, so it is uncertain whether nitrification inhibition occurs in these soils. If pastures are overgrazed, depression in labile subsurface C availability may also stimulate nitrification by decreasing N immobilization in microbial biomass (Neill, 1999). Further, the potential role of heterotrophic nitrification in NO_3^- production is unknown.

Nitrous oxide (N_2O) is a potent greenhouse gas with 300 times the warming potential of carbon dioxide over a 100-year period (Forster et al., 2007; UNEP, 2013). A meta-analysis of N_2O flux studies throughout the Amazon has revealed a consistent pattern: young pasture soils (<10 years) may increase slightly in annual emissions compared to forests (median 2.52 kg N ha^{-1} versus 2.42 kg N ha^{-1}), but older pastures show drastically reduced emissions (median 0.9 kg N ha^{-1}) that in some cases act as a slight sink (Meurer et al., 2016). Further, the productivity status of pastures appears to influence N_2O flux, with degraded pastures emitting less annually than active pastures, likely as a result of ecosystem N depletion (Verchot et al., 1999).

N_2O is yielded through the reduction of NO_3^- or NO_2^- (**Figure 1-6**), which may be performed as an incomplete anaerobic respiration reaction by heterotrophic denitrifiers or autotrophic nitrifiers (Bateman and Baggs, 2005; Patureau et al., 2000; Wrage et al., 2001). A gene-flux paired study in Mato Grosso found that nitrite reductase (*nirK*) and nitrous oxide reductase (*nosZ*) absolute gene abundances decreased in a 25-year-old pasture compared to a forest. An alternate nitrite reductase gene, *nirS*, was also measured in this study, and conversely showed a two-fold increase in pasture. While there is no clear understanding of divergent trends in *nirK*/*nirS* abundance given they perform the same function, *nosZ* and *nirK* counts may agree due to genomic co-occurrence (Jones et al., 2008). N_2O flux

measurements agreed with the meta-analysis discussed above (Meurer et al., 2016) both in trend and rate magnitude; emissions were roughly eight-fold higher in the forest than pasture. Further, the *nosZ* (clade I) gene count was a decent predictor of N_2O flux ($R= 0.61$, $p < 0.003$). Unsurprisingly, soil water content and NO_3^- concentration were also important physicochemical predictors. Using GeoChip profiling, (Paula et al., 2014) found no significant differentiation in *nirK* or *nosZ* gene abundance in pastures compared to forests, but did identify an overall association of denitrification genes in forest soils. Additionally, co-occurrence networks by (Khan et al., 2019) suggested a greater importance of denitrification in Rondônian forests compared to pastures.

In a quantitative analysis of N-cycling communities, Pedrinho et al. (2020) conversely identified significant increases in absolute abundance of *nirK* (~three-fold) and *nosZ* (~two to three-fold) genes in pastures compared to forests, with consistent wet- and dry-season trends. Metagenomic read annotation reflected a significant increase in total denitrification-related genes in pastures during the dry season, but no difference during the wet season. The pasture surveyed in this study was approximately 13 years old, so although N_2O flux was not measured, trends from the meta-analysis suggest that emissions should be reduced compared to forests (Meurer et al., 2016). The genetic potential to contribute to denitrification is phylogenetically widespread (Chen et al., 2014; Shoun et al., 1992; Wei et al., 2015), so it may be that increases in *nirK* and *nosZ* genes are co-occurring with microbes that have an increased abundance in pastures due to an alternate set of conditions.

Nitric oxide (NO) is another biotically and abiotically produced gas that impacts ozone concentration in the atmosphere. It may be produced as a 'leaky pipe' intermediate of nitrification, denitrification, or physicochemical processes (Pilegaard, 2013). Data on NO flux in the Amazon are limited, but one study found that similar to N_2O, annual NO production was considerably higher in primary forest (1.5 kg N ha^{-1} yr^{-1}) compared to pasture (0.6 kg N ha^{-1} yr^{-1}) (Verchot et al., 1999). *nirK/S* genes catalyze the production of NO, but it is unclear how strong of a predictor their abundance may be for emission rates, and to the authors' knowledge, no studies related to land-use change in the Amazon have investigated this. Further molecular work paired with flux measurements is required to understand NO flux through the soil and atmosphere in relation to land-use change. Additionally, while Pedrinho et al. (2020) reported significant compositional shifts in N-cycling taxa with pasture conversion, no specific analyses

have been conducted to explore diversity or compositional changes in communities capable of contributing to denitrification processes. This may be a valuable line of inquiry for future studies.

Conversion by Fire

The cumulative impact of land-use change on the activity, abundance, and community composition of soil microbes is likely impacted by the process of land-use conversion itself, most commonly through slash-and-burn clearing. This is important for consideration in the Amazon, where slash-and-burn practices occur on a range of scales, from large fires for pasture establishment (the primary land-use conversion type addressed in this chapter), to smaller-scale subsistence farming practices called ‘shifting cultivation’. Forest burn in the BLA corresponds with total rates of deforestation, which has likely been intensified by climate change-related warming and drought over the past decades, releasing approximately 1600 kg CO_2 per ton dry biomass burned on average (Barkhordarian et al., 2017; Cochrane and Laurance, 2008; Silva et al., 2021; van Marle et al., 2017). Fire alters above and belowground plant communities, and to varying degrees may impact soil textural and structural properties, biogeochemical pools, nutrient ratios, soil organic matter quality, and live microbial abundance, depending on heat intensity, preexisting organic matter content, or soil texture and bulk density (Butler et al., 2017; Cochrane and Laurance, 2008; Mataix-Solera et al., 2009). Within one day to several months, direct sterilization may reduce microbial biomass C significantly. Previous studies have indicated a biomass reduction of 64% and 74% for soil depths to 5 or 10 cm, respectively, following slash-and-burn clearing (Luizao, 1992; Prieto-Fernández et al., 1998). Overall, bacteria have typically been found to be more resistant to the effects of fire than fungi in both temperate and tropical forests. Fungal sensitivity may be attributed to their broad hyphal networks (Aguilar-Fernandez et al., 2009; Barraclough and Olsson, 2018; Rashid et al., 1997); however, studies have found that microbial biomass has been negatively impacted for several years post-burn (Prieto-Fernández et al., 1998). Also impacted post-burn are biomass C (Luizao, 1992), spore diversity, and viability (Aguilar-Fernandez et al., 2009), which eventually return to pre-burn values or increase, suggesting the potential for rapid rebound of populations.

Aside from direct heat effects, communities are likely to be indirectly impacted by the availability of easily metabolized C and N compounds. An initial flush of surface layer (0–10 cm) extractable C and N compounds immediately following forest burn commonly coincides with microbial biomass losses through heat sterilization; in the medium-term (2 months to several years following burn), surface soil concentrations of organic matter are typically similar to or lower-than unburned forest stands (Navarrete et al., 2015a; Neill, 1999; Prieto-Fernández et al., 1998). Additionally, fire is likely to lower the ratio of available C:N (Bomfim et al., 2020; Prieto-Fernández et al., 1998) given the volatility of C compounds. This shift in C and N availability may impact the composition and diversity of recovering microbial communities, particularly if burns are repeated (Zarin et al., 2005). Organic matter in soil may also be changed to a 'pyromorphic' humus, negatively impacting its susceptibility to microbial degradation and water-holding properties (Gonzalez-Perez et al., 2004). Microbial communities of tropical forest soils may also be impacted by post-fire P (a limiting nutrient in tropical forests) availability, which has a high volatilization temperature (Butler et al., 2017), meaning it should be enriched in post-fire ash relative to pre-fire bulk surface soil. The availability of other crucial cations should also be enhanced given that ash increases soil pH (Neill, 1999; Ribeiro Filho et al., 2015). In tropical systems, a large proportion of ecosystem nutrients (C, N, P, etc.) are stored in biomass rather than soil pools (Wan et al., 2002). While fire may initially release these nutrients, the replacement of biomass-dense forest stands with perennial bunchgrass means that a smaller quantity of nutrients can be immobilized as plant biomass per unit area, and negatively charged inorganic molecules are particularly at risk of leaching from soil over time, especially under heavy rainfall. Specific impacts of these post-burn nutrient shifts on soil microbial communities are largely unknown.

Surprisingly, limited work has been done exploring the short-, medium-, and long-term impacts of controlled or uncontrolled forest burns on soil microbial community structure, either in tropical systems or otherwise. There appears to be a consensus across several studies that in the short- and medium-term, microbial diversity increases or stays the same, and that community composition shifts in association with changes in pool size of available soil nutrients, particularly C and/or N (Fontúrbel et al., 2012; Lucas-Borja et al., 2019; Navarrete et al., 2015a; Prendergast-Miller et al., 2017). In particular, Navarrete et al. (2015a) found that two-to-four months after slash-and-burn deforestation, prokaryotic alpha diversity

(measured via Simpson index) increased significantly (~70%) at two of three surveyed sites in Mato Grosso, Brazil. Additionally, for copiotrophic taxa, such as Actinomycetales, an increase in richness and relative abundance following burn appeared to be related to N availability. The study also found a decrease in Planctomycetes, Chlamydiae, and Verrucomicrobia following slash-and burn. The latter finding is interesting given that Verrucomicrobia have been shown to increase in diversity once pastures are established (Ranjan et al., 2015), suggesting the ecological resilience of this group. Indeed, Verrucomicrobia appear to be adapted to a low concentrations of substrate, as would be the case post-burn (Noll et al., 2005). Shifts in favored taxa have been observed elsewhere as well, such as increased abundance of Firmicutes, some of which are capable of endospore formation, a robust survival tactic under unfavorable conditions (Ferrenberg et al., 2013; Lucas-Borja et al., 2019; Prendergast-Miller et al., 2017). This is consistent with many studies previously discussed, which found an increase in Firmicutes with pasture conversion (e.g.,Rodrigues et al., 2013). In the northern Amazon region, a 16S rRNA gene-based community composition study following the fire preparation of a pasture site reflected lower species diversity and richness in comparison to surveyed primary forests (Melo et al., 2021); however, the functional consequence of this is unknown.

Shifts in microbial function as a result of slash-and-burn in the Amazon have been indicated in a few studies. Just months after fire conversion, genes related to protein metabolism decreased up to 30% in burned areas, while genes related to DNA metabolism increased in pasture plots in Mato Grosso (Navarrete et al., 2015a). The latter finding may serve as an indication of survival and maintenance of genetic material during periods of unfavorable conditions. In the Amazonia-Cerrado transition zone, a study on the effects of fire on seasonally-flooded forest soils indicated that rates of a key soil microbial process, ANF, were on average 24% lower in burned compared to unburned surface soils (0–10 cm) (Bomfim et al., 2020). However, variable frequency of burn does not appear to add significant effect to this difference, and rates below 10 cm are unaffected by burning of any frequency. The mechanism of control over decreased activity is the shift in ratio of C to other nutrients such as N and P (Bomfim et al., 2020). This is consistent with previous work, which determined the quantity and characteristics of soil C post-burn to be significant determinants of ecosystem function (Gonzalez-Perez et al., 2004; Prieto-Fernández et al., 1998). The ANF study also found that while activity rates scaled linearly with nutrient

ratios in unburned forests, relationships were highly nonlinear in burned forests (Bomfim et al., 2020), potentially indicating an additional unmeasured impact of forest fires on diazotrophic soil communities, ultimately affecting activity rates.

Limited knowledge of the role of fire on microbial communities in the Amazon Rainforest and along its transitional zones presents a serious gap in knowledge. As previously mentioned, fire in the Amazon has been increasing in recent decades and will likely continue to do so for several reasons (Cochrane and Laurance, 2008). First, forest-clearing activities, including pasture conversion and shifting cultivation practices, present new vulnerabilities for unplanned fires at forest margins (Barlow et al., 2016; Cochrane, 2001; Cochrane and Schulze, 1999). Indeed, a recent analysis has spatially linked the outbreak of accidental forest fires to ongoing deforestation (MAAP, 2019). Second, forest clearing has a positive feedback effect on climate change-related shifts in the hydrologic cycle through the release of C to the atmosphere as well as a direct alteration of local and regional scale air moisture circulation patterns (Betts et al., 2009). Therefore, understanding soil microbial community response to Amazonian forest fires will be crucial in further elucidating the explanatory factors of long-term compositional and metabolic shifts of soil microbial communities in response to abrupt disturbance and long-term land-use change.

Secondary Forest Recovery

This chapter has focused primarily on the conversion of primary forest to cattle pasture in the Amazon, as this has been the most widespread cause of land-use change throughout the region. However, it is not uncommon for pastures to be abandoned within 5–15 years of establishment due to the loss of forage grass productivity (Asner et al., 2004). In the years following abandonment, secondary forests begin to form. At present, it is estimated that approximately 30 to 50% of previously converted pastureland is in some stage of succession (Chazdon et al., 2009; Pacheco, 2012), a reality that raises important questions as to whether microbial communities of these secondary forests rebound to a similar state of taxonomic and functional diversity as primary forests. Furthermore, it is unknown what implications this has for the biogeochemical cycles that microbial communities in secondary forests mediate. Some studies have compared communities of primary forest and pasture to that of secondary forests, and many conclude that microbial diversity indices of secondary forest are more similar to primary

forests than active pastures. Evidence related to functional recovery is more mixed, with results varying by the specific function considered. **Table 2** summarizes these findings and includes relevant metadata from studies discussed throughout this section. **Figure 1-7** correspondingly provides the geographic locations of these studies.

Microbial composition and abundance in succession

Mycorrhizal populations are presumably essential in forest succession. They are crucial in successful seedling recruitment by providing vulnerable roots with nutrients, water, and pathogen protection (Igwe and Vannette, 2019; Nara, 2006; Ueki et al., 2018; van der Heijden et al., 2016; Van Der Heijden, 2004). In a study of AM fungal diversity on the Eastern Amazonia margin, species richness and spore abundance were indistinguishable between young, degraded-secondary forests and mature rainforests during the rainy season. Intriguingly, spore abundance and diversity were actually significantly higher in young secondary forests during the dry season (Reyes et al., 2019). Species composition was similarly more affected by seasonality than forest type or age, indicating that AMF are resilient members of the soil microbial community despite high levels of historic disturbance. A survey of forests in succession across several Brazilian biomes has demonstrated that plant species involved in early succession engage in dense rates of AM colonization, likely investing in fungal recruitment to maximize capacity for nutrient acquisition (Zangaro et al., 2012); this finding lends mechanistic support to the notion that AM fungal communities recover quickly with reforestation. Total fungal communities of secondary forests are also more similar to those of primary forests than pastures, but this similarity appears dependent on geographic distance from forests, potentially indicating that primary forests can act as species reservoirs for the recolonization of secondary forests in succession (Mueller et al., 2016). This result also suggests that considering inter-sample distance at scales as large as several kilometers may be an important predictor of fungal composition.

Similarly, prokaryotic communities appear to respond to reforestation over time. Comparison of several secondary forests of varying age has shown greater similarity in community structure and composition between primary and older secondary forests (~5–30 years) compared to younger secondary forests (<5 years), indicating a trajectory to recovery (da C Jesus et al., 2009). This pattern has also been

reflected in a narrower community scope: a survey of Verrucomicrobia using phylum-targeted amplification and sequencing revealed greater similarity between secondary and primary forest community structures compared to pasture, although diversity of taxa within the phylum was higher in secondary than primary forests (Ranjan et al., 2015). Additionally, targeting archaeal communities using PCR-DGGE has indicated greater similarity of primary forest communities with those of secondary forests, compared to pastures and cropping systems (Navarrete et al., 2011). A metagenomic-based analysis found greater taxonomic similarity between primary and secondary forest compared to pasture, but also discovered that the functional diversity of a ~14 year old secondary forest, similar to pastures, was elevated in comparison to primary forest (Pedrinho et al., 2019). Functional profiling using the GeoChip 4.0 showed greater similarity of overall functional gene richness and diversity between primary and secondary forests compared to pastures, but substantial differences were found in specific gene composition (Paula et al., 2014). The results of these studies may indicate that despite vegetational succession, legacy effects of pasture on soil conditions, such as nutrient status, may present a lag in the recovery of soil functions compared to taxonomic representation.

Microbial function in succession

Studies comparing microbial-mediated C degradation across primary forest, pasture, and secondary forest have been fairly limited and have mixed results. One study across a land-use chronosequence at ARMO, State of Rondônia, Brazil, analyzed C dynamics in 13 to18-year-old secondary forests that were abandoned after 7–10 years of pasture use. The study found that the organic C profiles (quantity and chemical composition) of secondary forest soil closely resemble that of primary forests, with concomitant increases in C-cycling enzyme (ß- glucosidase) activity (Durrer et al., 2021); this pattern is perhaps intuitive because the plant community and detrital substrate of secondary forest is much more similar to original forests than pastures filled with exotic grasses. The interrelationship of CH_4-C emissions and methane-cycling microbial functional groups with forest recovery is of distinct interest. In a recent study of methane-cycling communities in Rondônia, the richness and relative abundance of methanogens in secondary forest showed a significant decline relative to pasture, while the richness and relative abundance of methanotrophs showed a significant increase (Meyer et al., 2020). Active

methanotrophic communities of secondary forests are also more structurally comparable to primary forests than pastures (Kroeger et al., 2021). These findings are supported by flux measurements that identified secondary forest as a weak sink (-10.2 $\pm$ 35.7 μg CH_4 m^{-2} d^{-1}) compared to pasture, which typically acts as a source (Meyer et al., 2020). A similar conclusion was reached in a timeseries land-use gradient study, namely, secondary forests (as well as degraded pastures) were found to be a year-round weak net sink of methane (Verchot et al., 2000). However, in the latter study, seasonal patterns of methane flux differed substantially across all land-use types, indicating some persisting differences in methane-cycling microbial communities, and potentially soil physicochemical conditions.

As discussed previously, pastures may become N limited with age, particularly if land management is unsustainable. It stands to reason that following pasture abandonment, recuperation of N is likely to be an important aspect of reforestation, particularly in the early stages (Davidson et al., 2007). Elucidating shifts in N-cycling microbial community members in secondary forests, compared to primary forests and pastures, is likely to improve our understanding of how N flux rates and pool sizes change with forest succession. Evidence from multiple studies would suggest that, with secondary forest succession, the abundance of diazotrophs decreases in comparison to pasture, contracting toward primary forest levels (Mirza et al., 2014; Pedrinho et al., 2020). Community taxonomic and phylogenetic diversity at the local scale comparably decreases, but turnover at the landscape scale increases in secondary forests (Mirza et al., 2020; Silva et al., 2011), indicating a reversal of the biotic homogenization effect of pasture establishment identified by Rodrigues et al. (2013). Additionally, compositional structure of diazotrophs is highly similar between primary and secondary forests compared to distinct pasture community structure (Mirza et al., 2020). A study that used GeoChip profiling similarly found that the trend of nitrogen fixer abundance of secondary forests became more similar to that of primary forests (Paula et al., 2014). Interestingly, study of SNF potential suggests this may be an important process as forests recover (Lima et al., 2009). A claybox mesocosm experiment in abandoned pastures using a ^{15}N tracer revealed that early secondary forest legume species may obtain 75% of biomass N content from SNF-diazotrophs (Davidson et al., 2000). Survey of legume density in forests of central Amazonia has shown that in secondary forests, depression of leaf litter $\delta^{15}N$ is positively correlated with higher legume density compared to primary forests (Gehring et al., 2005). This suggests higher rates of SNF in secondary

forests than in primary forests. The study further concluded that this trend is consistent in successional forests spanning multiple decades in age. Taken together, these findings indicate a potentially divergent role of associated and free-living versus symbiotic N_2 fixers in recovering secondary forests, warranting further study.

Ammonia oxidizers of the Thaumarchaeota, an archaeal phylum, reflect one of the most dramatic drop-offs of any group in converted pasture soils. An ARMO-based analysis demonstrated community revival with secondary forest succession: the absolute abundance of an *amoA* marker gene in secondary forest soil was akin to that of primary forests and approximately an order of magnitude higher than in pastures. While distinct, ammonia-oxidizing archaeal community structure of secondary forests showed greater similarity to primary forest than pastures (Hamaoui et al., 2016). Although gross and net rates of N mineralization and nitrification were not directly measured in this study, concentrations of NH_4^+ (~5.5 µg-N g^{-1} soil) and NO_3^- (~3.7 µg-N g^{-1} soil) were nearly identical in primary and secondary forests, compared to much higher NH_4^+ (~14 µg-N g^{-1} soil) and virtually no NO_3^- present in pastures. These results suggest recovery of this microbial functional group is concomitant with restoration of the environmental processes they mediate. The results of this study were supported by previous GeoChip profiling, which showed that nitrification genes were significantly associated with primary and secondary forest compared to pasture (Paula et al., 2014). In contrast, a PCR clone-based study in Benjamin Constant found considerable differentiation in archaeal *amoA*-based community structure between primary and secondary forest types, with primary communities more similar to those of pastures (Navarrete et al., 2011). Additionally, richness and diversity metrics of *amoA* communities were appreciably lower in secondary forests compared to primary forests or pastures. This variability across studies may be explained in part by large differences in trends of inorganic N pool sizes in opposing land-use types. While NH_4^+ concentrations did increase in pastures relative to primary forest, concentrations of NH_4^+ in secondary forests were significantly lower than both. Further, no difference was observed in NO_3^- concentrations across land-use types. Of course, factors like secondary forest age and N limitation status at the time of abandonment certainly play a role in the trajectory of N-cycling as forests regrow.

A comprehensive analysis of all N-cycling microbes as interpreted from metagenomic data suggests that taxonomic and functional community structure shifts with pasture abandonment but remains

distinct from primary forests (Pedrinho et al., 2020). Patterns of diversity in the N-cycling community showed inconsistent trends: taxonomically, primary and secondary forests were similar compared to elevated diversity in pastures, but in terms of functional diversity (and abundance of reads relative to whole metagenomes), secondary forests were more similar to pastures. It appears that, while secondary forest N-cycling communities do trend toward recovery (i.e., a primary forest-like state), initial persistent N limitation remains, potentially for several decades. This is reflected in metrics of foliar N, litterfall mass to N ratio, NO_3^- concentration, and N_2O production across primary forests, active/degraded pastures, and secondary forests (Davidson et al., 2007; Verchot et al., 1999). N_2O emissions from secondary forest in particular appear to have intermediary rates between primary forests and pastures (Verchot et al., 1999). Overall, the recovery of N-cycling microbial groups is likely dependent on the degree of N limitation imposed by agricultural use, and secondary forest age is likely an important consideration when assessing recovery status related to N cycling.

More broadly, land-use history and intensity may impact many long-term aspects of reforestation such as rate of regrowth, forest density, and soil C stocks (Uhl, 1988; Zarin et al., 2005). Studies investigating the response of soil microbial communities to pasture abandonment and reforestation in the Amazon are sparse and often lack a process-based measurement. Future research should focus on assessing the role of land-use history, including native forest conditions, land management decisions, and conversion/abandonment timelines (Fearnside and Guimarães, 1996; Zarin et al., 2005), on shaping the composition and function of the soil microbial community in secondary forests.

Conclusions

Considering the future of the Amazon

Concerns over not only biodiversity loss, but also contributions to climate change-related warming with deforestation in the Amazon Rainforest has helped increase public awareness of its global importance (Cerri et al., 2018). With climate change, it has become clear that more research is needed to understand how the Amazon Rainforest cycles and stores carbon. The complex role microbes play in this ecological function is still poorly understood and should be an essential focus of future microbial research

in order to increase their representation in ecosystem models. Despite concerns, Amazon Rainforest losses have been increasing over the past ~5-6 years (Figure 1-1a; INPE, 2020). While the drivers of Amazon Rainforest losses are not yet apparent, clearing of primary forest appears to be disproportionately more likely in small forest fragments compared to larger forest fragments (Hansen et al., 2020). Conservation efforts should therefore be focused on these larger contiguous fragments of forest. Ongoing fragmentation also indicates that further work is needed to understand the impact of edge-effects on soil microbial community diversity and function, since at present, this is a severely understudied aspect of microbial ecology. Recent analysis of forest loss in the Brazilian Amazon concluded that secondary rather than primary forests are now the dominant type being cleared, accounting for 72% of total deforestation as of 2014 (Wang et al., 2020), with over 90% of re-cleared land again converted to cattle pasture (Tyukavina et al., 2017). It is imperative that future studies—especially those focused on young pastures—take into account and assess the effect of complex land-use histories that may include multiple forest/pasture cycles. Not only will this likely begin to explain variation in the data, but it also presents a unique opportunity to determine the extent to which soil microbiomes recover their community assembly when the land use shifts. Additionally, a recent effort towards grazing intensification in order to slow new deforestation (Barbosa et al., 2015) brings into question the response of soil microbes, both compositionally and functionally, to variable grazing pressures.

The Complexity of Microbial Communities in Response to Land-Use Change

Attention to microbiomes of tropical forests, particularly the Amazon, has only gathered momentum in the past 15 years, fueled by our concerns over continued forest loss and degradation, and by a poor understanding of the role microbial diversity and function play in these vital ecosystems. A complex interaction of factors shapes soil microbial community response to land-use change in the Amazon, as has been discussed throughout this chapter. **Figure 1-8** attempts to conceptualize these interactions but is by no means exhaustive. Land-use change factors (such as initial slash-and-burn clearing, and changes in aboveground vegetation with pasture reseeding), and land management (including grazing practices, controlled burns, and abandonment)- impact the microbiome directly and indirectly through complex feedbacks on soil physicochemical conditions. These changing conditions

influence several aspects of the soil microbial community including abundance, compositional structure, and taxonomic, phylogenetic, and functional diversity.

This chapter has discussed in detail what is currently known of microbial community response to land-use change in the Amazon Rainforest and has put this in the context of shifts in important biogeochemical cycles such as C and N. However, there are still vast gaps in the understanding of regional variability in microbial community response, how environmental and management conditions shape these responses, and how shifts in community composition and structure relate to their environmental function and the biogeochemical cycles they mediate. Henceforth, employing computational approaches to better utilize the vast quantities of data generated through techniques, such as metagenomics, will hopefully impart a greater understanding of the ecological role microbial communities assume in tropical soils.

Acknowledgment of support

The Soil Ecogenomics Laboratory received financial support from the National Science Foundation – Dimensions of Biodiversity (DEB 14422214) and the U.S. Department of Agriculture – NIFA (Award # 2009-35319-05186). This work was supported the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. Support is also acknowledged through the USDA NIFA Pre-doctoral Fellowship to RD (Award No. 2019-67011-29554).

References

Aguilar-Fernandez, M., Jaramillo, V. J., Varela-Fregoso, L., and Gavito, M. E. (2009). Short-term consequences of slash-and-burn practices on the arbuscular mycorrhizal fungi of a tropical dry forest. *Mycorrhiza* **19**, 179-186.

Asner, G. P., Townsend, A. R., Bustamante, M. M., Nardoto, G. B., and Olander, L. P. (2004). Pasture degradation in the central Amazon: linking changes in carbon and nutrient cycling with remote sensing. *Global Change Biology* **10**, 844-862.

Auger, C., Han, S., Appanna, V. P., Thomas, S. C., Ulibarri, G., and Appanna, V. D. (2013). Metabolic reengineering invoked by microbial systems to decontaminate aluminum: implications for bioremediation technologies. *Biotechnology advances* **31**, 266-273.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D. and Houghton, R.A. (2017). Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science* **358**, 230-234.

Bach, E. M., Williams, R. J., Hargreaves, S. K., Yang, F., and Hofmockel, K. S. (2018). Greatest soil microbial diversity found in micro-habitats. *Soil biology and Biochemistry* **118**, 217-226.

Bahram, M., Netherway, T., Frioux, C., Ferretti, P., Coelho, L. P., Geisen, S., Bork, P., and Hildebrand, F. (2021). Metagenomic assessment of the global diversity and distribution of bacteria and fungi. *Environ Microbiol* **23**, 316-326.

Barberan, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**, 343-51.

Barbosa, F. A., Soares Filho, B. S., Merry, F. D., Azevedo, H. d. O., Costa, W. L. S., Coe, M. T., BATISTA, E. d. S., Maciel, T., Sheepers, L., and OLIVEIRA, A. d. (2015). Cenários para a pecuária de corte amazônica. *Universidade Federal de Minas Gerais: Belo Horizonte, Brazil*, 154.

Bardgett, R. D., and van der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505-11.

Barkhordarian, A., von Storch, H., Zorita, E., Loikith, P. C., and Mechoso, C. R. (2017). Observed warming over northern South America has an anthropogenic origin. *Climate Dynamics* **51**, 1901-1914.

Barlow, J., Franca, F., Gardner, T. A., Hicks, C. C., Lennox, G. D., Berenguer, E., Castello, L., Economo, E. P., Ferreira, J., Guenard, B., Gontijo Leal, C., Isaac, V., Lees, A. C., Parr, C. L., Wilson, S. K., Young, P. J., and Graham, N. A. J. (2018). The future of hyperdiverse tropical ecosystems. *Nature* **559**, 517-526.

Barlow, J., Lennox, G. D., Ferreira, J., Berenguer, E., Lees, A. C., Mac Nally, R., Thomson, J. R., Ferraz, S. F., Louzada, J., Oliveira, V. H., Parry, L., Solar, R. R., Vieira, I. C., Aragao, L. E., Begotti, R. A., Braga, R. F., Cardoso, T. M., de Oliveira, R. C., Jr., Souza, C. M., Jr., Moura, N. G., Nunes, S. S., Siqueira, J. V., Pardini, R., Silveira, J. M., Vaz-de-Mello, F. Z., Veiga, R. C., Venturieri, A., and Gardner, T. A. (2016). Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **535**, 144-7.

Barraclough, A., and Olsson, P. (2018). Slash-and-Burn Practices Decrease Arbuscular Mycorrhizal Fungi Abundance in Soil and the Roots of Didierea madagascariensis in the Dry Tropical Forest of Madagascar. *Fire* **1**.

Bateman, E. J., and Baggs, E. M. (2005). Contributions of nitrification and denitrification to N2O emissions from soils at different water-filled pore space. *Biology and Fertility of Soils* **41**, 379-388.

Batjes, N., and Dijkshoorn, J. (1999). Carbon and nitrogen stocks in the soils of the Amazon Region. *Geoderma* **89**, 273-286.

Batjes, N. H. a. D., J.A., 1999 (1999). Carbon and nitrogen stocks in the soils of the Amazon Region. *Geoderma* **89**, 273-286.

Bell, T. (2010). Experimental tests of the bacterial distance-decay relationship. *ISME J* **4**, 1357-65.

Bending, G. D., Turner, M. K., and Jones, J. E. (2002). Interactions between crop residue and soil organic matter quality and the functional diversity of soil microbial communities. *Soil Biology and Biochemistry* **34**, 1073-1082.

Bernoux, M., Cerri, C., Arrouays, D., Jolivet, C., and Volkoff, B. (1998). Bulk densities of Brazilian Amazon soils related to other soil properties. *Soil Science Society of America Journal* **62**, 743-749.

Betts, A. K., Fisch, G., Von Randow, C., Silva Dias, M. A. F., Cohen, J. C. P., Da Silva, R., and Fitzjarrald, D. R. (2009). The Amazonian boundary layer and mesoscale circulations. Amazonia and global change. *Geophysical monograph* **186**, 163-181.

Boers, N., Marwan, N., Barbosa, H. M., and Kurths, J. (2017). A deforestation-induced tipping point for the South American monsoon system. *Sci Rep* **7**, 41489.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E., Da Silva, R., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B., Kang, K. B., Keefe, C. R., Keim, P., Kelley, S. T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M. G. I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J., Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Robeson, I. I. M. S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ul-Hasan, S., van der Hooft, J. J. J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., et al. (2018). QIIME 2: Reproducible, interactive, scalable,
and extensible microbiome data science. *PeerJ*.

Bomfim, B., Silva, L. C. R., Marimon-Júnior, B. H., Marimon, B. S., Doane, T. A., and Horwath, W. R. (2020). Fire Affects Asymbiotic Nitrogen Fixation in Southern Amazon Forests. *Journal of Geophysical Research: Biogeosciences* **125**.

Borneman, J., and Triplett, E. W. (1997). Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied and environmental microbiology* **63**, 2647-2653.

Bradford, M. A., Berg, B., Maynard, D. S., Wieder, W. R., Wood, S. A., and Cornwell, W. (2016). Understanding the dominant controls on litter decomposition. *Journal of Ecology* **104**, 229-238.

Brzostek, E. R., Greco, A., Drake, J. E., and Finzi, A. C. (2012). Root carbon inputs to the rhizosphere stimulate extracellular enzyme activity and increase nitrogen availability in temperate forest soils. *Biogeochemistry* **115**, 65-76.

Bürgmann, H., Meier, S., Bunge, M., Widmer, F., and Zeyer, J. (2005). Effects of model root exudates on structure and activity of a soil diazotroph community. *Environmental Microbiology* **7**, 1711-1724.

Butler, O. M., Lewis, T., and Chen, C. (2017). Fire alters soil labile stoichiometry and litter nutrients in Australian eucalypt forests. *International Journal of Wildland Fire* **26**.

Butler, R. A. (2020). The Amazon Rainforest: The World's Largest Rainforest. Vol. 2020.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-3.

Cantarutti, R., Tarré, R., Macedo, R., Cadisch, G., de Rezende, C., Pereira, J., Braga, J., Gomide, J., Ferreira, E., and Alves, B. (2002). The effect of grazing intensity and the presence of a forage legume on nitrogen dynamics in Brachiaria pastures in the Atlantic forest region of the south of Bahia, Brazil. *Nutrient Cycling in Agroecosystems* **64**, 257-271.

Carneiro, M., Siqueira, J., Vale, F., and Curi, N. (1995). Nutritional limitation and the effect of soil pre-cultivation with Brachiaria decumbens and inoculation with Glomus etunicatum on the growth of tree seedlings in degraded soil. *Science and Practice* **19**, 281-288.

Carvalho, J. L. N., Carlos Eduardo Pelegrino, C., Feigl, B. J., Píccolo, M. d. C., Godinho, V. d. P., Herpin, U., and Cerri, C. C. (2009). Conversion of Cerrado into agricultural land in the south-western Amazon: Carbon stocks and soil fertility. *Scientia Agricola* **66**, 233-241.

Cavagnaro, R. A., Oyarzabal, M., Oesterheld, M., and Grimoldi, A. A. (2014). Screening of biomass production of cultivated forage grasses in response to mycorrhizal symbiosis under nutritional deficit conditions. *Grassland Science* **60**, 178-184.

Cenciani, K., Lambais, M. R., Cerri, C. C., Azevedo, L. C. B. d., and Feigl, B. J. (2009). Bacteria diversity and microbial biomass in forest, pasture and fallow soils in the southwestern Amazon basin. *Revista Brasileira de Ciência do Solo* **33**, 907-916.

Cerqueira, A. E. S., Silva, T. H., Nunes, A. C. S., Nunes, D. D., Lobato, L. C., Veloso, T. G. R., De Paula, S. O., Kasuya, M. C. M., and Silva, C. C. (2018). Amazon basin pasture soils reveal susceptibility to phytopathogens and lower fungal community dissimilarity than forest. *Applied Soil Ecology* **131**, 1-11.

Cerri, C., Piccolo, M., Feigl, B., Paustian, K., Cerri, C., Victoria, R., and Melillo, J. (2006). Interrelationships among soil total C and N, microbial biomass, trace gas fluxes, and internal N-cycling in soils under pasture of the Amazon region. *Journal of sustainable agriculture* **27**, 45-69.

Cerri, C. E. P., Cerri, C. C., Maia, S. M. F., Cherubin, M. R., Feigl, B. J., and Lal, R. (2018). Reducing Amazon deforestation through agricultural intensification in the Cerrado for advancing food security and mitigating climate change. *Sustainability* **10**, 989.

Chazdon, R. L., Peres, C. A., Dent, D., Sheil, D., Lugo, A. E., Lamb, D., Stork, N. E., and Miller, S. E. (2009). The potential for species conservation in tropical secondary forests. *Conservation biology* **23**, 1406-1417.

Chen, Q.-L., An, X.-L., Zheng, B.-X., Gillings, M., Peñuelas, J., Cui, L., Su, J.-Q., and Zhu, Y.-G. (2019). Loss of soil microbial diversity exacerbates spread of antibiotic resistance. *Soil Ecology Letters* **1**, 3-13.

Chen, Y., Zhou, W., Li, Y., Zhang, J., Zeng, G., Huang, A., and Huang, J. (2014). Nitrite reductase genes as functional markers to investigate diversity of denitrifying bacteria during agricultural waste composting. *Appl Microbiol Biotechnol* **98**, 4233-43.

Ciemer, C., Rehm, L., Kurths, J., Donner, R. V., Winkelmann, R., and Boers, N. (2020). An early-warning indicator for Amazon droughts exclusively based on tropical Atlantic sea surface temperatures. *Environmental Research Letters* **15**.

Cleveland, C. C., and Liptzin, D. (2007). C: N: P stoichiometry in soil: is there a “Redfield ratio” for the microbial biomass? *Biogeochemistry* **85**, 235-252.

Cochrane, M. A. (2001). Synergistic interactions between habitat fragmentation and fire in evergreen tropical forests. *Conservation Biology* **15**, 1515-1521.

Cochrane, M. A., and Laurance, W. F. (2008). Synergisms among fire, land use, and climate change in the Amazon. *Fire Ecology and Management* **37**, 522-527.

Cochrane, M. A., and Schulze, M. D. (1999). Fire as a Recurrent Event in Tropical Forests of the Eastern Amazon: Effects on Forest Structure, Biomass, and Species Composition. *Biotropica* **31**, 2-16.

Coe, M. T., Brando, P. M., Deegan, L. A., Macedo, M. N., Neill, C., and Silvério, D. V. (2017). The forests of the Amazon and cerrado moderate regional climate and are the key to the future. *Tropical Conservation Science* **10**.

Conrad, R. (2007). Microbial Ecology of Methanogens and Methanotrophs. *In* "Advances in Agronomy", Vol. 96, pp. 1-63. Academic Press.

Cruz, L. G., Bastidas, A. T. C., Suárez, L. R., and Salazar, J. C. S. (2019). Microbial Properties of Soil in Different Coverages in the Colombian Amazon. *Floresta e Ambiente* **26**.

da C Jesus, E., Marsh, T. L., Tiedje, J. M., and de, S. M. F. M. (2009). Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J* **3**, 1004-11.

Dalling, J. W., Heineman, K., Lopez, O. R., Wright, S. J., and Turner, B. L. (2016). Nutrient availability in tropical rain forests: the paradigm of phosphorus limitation. *In* "Tropical tree physiology", pp. 261-273. Springer.

Davidson, E. A., de Araujo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., MM, C. B., Coe, M. T., DeFries, R. S., Keller, M., Longo, M., Munger, J. W., Schroeder, W., Soares-Filho, B. S., Souza, C. M., Jr., and Wofsy, S. C. (2012). The Amazon basin in transition. *Nature* **481**, 321-8.

Davidson, E. A., de Carvalho, C. J., Figueira, A. M., Ishida, F. Y., Ometto, J. P., Nardoto, G. B., Saba, R. T., Hayashi, S. N., Leal, E. C., Vieira, I. C., and Martinelli, L. A. (2007). Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature* **447**, 995-8.

Davidson, E. A., Michael Keller, Heather E. Erickson, Louis V. Verchot, a., ", E. V., .", and Bioscience 50, n.-. (2000). Testing a conceptual model of soil emissions of nitrous and nitric oxides: using two functions based on soil nitrogen availability and soil water content, the hole-in-the-pipe model characterizes a large fraction of the observed variation of nitric oxide and nitrous oxide emissions from soils. *Bioscience* **50**, 667-680.

Dawson, L., Grayston, S., and Paterson, E. (2000). Effects of grazing on the roots and rhizosphere of grasses. *Grassland ecophysiology and grazing ecology*, 61-84.

de Moraes, J. F., Volkoff, B.C.C.C., Cerri, C.C. and Bernoux, M., (1996). Soil properties under Amazon forest and changes due to pasture installation in Rondônia, Brazil. *Geoderma* **70**, 63-81.
de Souza Braz, A. M., Fernandes, A. R., and Alleoni, L. R. F. (2013). Soil Attributes after the Conversion from Forest to Pasture in Amazon. *Land Degradation & Development* **24**, 33-38.
Dedysh, S. N., and Damsté, J. S. S. (2018). Acidobacteria. *eLS*, 1-10.
Degens, B., and Harris, J. (1997). Development of a physiological approach to measuring the catabolic diversity of soil microbial communities. *Soil Biology and Biochemistry* **29**, 1309-1320.
Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., Berdugo, M., Campbell, C. D., and Singh, B. K. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat Commun* **7**, 10541.
DeLuca, T. H., Pingree, M. R. A., and Gao, S. (2019). Assessing soil biological health in forest soils. *In* "Global Change and Forest Soils", pp. 397-426.
Desbrosses, G. J., and Stougaard, J. (2011). Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host Microbe* **10**, 348-58.
Dias-Filho, M., Davidson, E. A., and Carvalho, C. (2001). The Biogeochemistry of the Amazon Basin. Oxford University Press New York.
Dickie, I. A., Xu, B., and Koide, R. T. (2002). Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* **156**, 527-535.
Diochon, A., and Kellman, L. (2008). Natural abundance measurements of13C indicate increased deep soil carbon mineralization after forest disturbance. *Geophysical Research Letters* **35**.
Dion, P. (2010). Towards a New Purpose for Traditional and Other Forms of Soil Knowledge. *In* "Soil Biology and Agriculture in the Tropics" (P. Dion, ed.), pp. 299-316. Springer Berlin Heidelberg, Berlin, Heidelberg.
Durigan, M., Cherubin, M., de Camargo, P., Ferreira, J., Berenguer, E., Gardner, T., Barlow, J., Dias, C., Signor, D., Junior, R., and Cerri, C. (2017). Soil organic matter responses to anthropogenic forest disturbance and land Use change in the eastern Brazilian Amazon. *Sustainability* **9**.
Durrer, A., Margenot, A. J., Silva, L. C. R., Bohannan, B. J. M., Nusslein, K., van Haren, J., Andreote, F. D., Parikh, S. J., and Rodrigues, J. L. M. (2021). Beyond total carbon: conversion of amazon forest to pasture alters indicators of soil C cycling. *Biogeochemistry* **152**, 179-194.
Egenolf, K., Conrad, J., Schone, J., Braunberger, C., Beifuss, U., Walker, F., Nunez, J., Arango, J., Karwat, H., Cadisch, G., Neumann, G., and Rasche, F. (2020). Brachialactone isomers and derivatives of Brachiaria humidicola reveal contrasting nitrification inhibiting activity. *Plant Physiol Biochem* **154**, 491-497.
Escalas, A., Hale, L., Voordeckers, J. W., Yang, Y., Firestone, M. K., Alvarez-Cohen, L., and Zhou, J. (2019). Microbial functional diversity: From concepts to applications. *Ecol Evol* **9**, 12000-12016.
Fearnside, P. M. (1987). Causes of Deforestation in the Brazilian Amazon. *In* "The Geophysiology of Amazonia: Vegetation and CLimate Interactions", pp. 37-61.
Fearnside, P. M. (1997). Greenhouse gases from deforestation in Brazilian Amazonia: net committed emissions. *Climatic Change* **35**, 321-360.
Fearnside, P. M., and Guimarães, W. M. (1996). Carbon uptake by secondary forests in Brazilian Amazonia. *Forest ecology and management* **80**, 35-46.
Fernandes, S. A. P., Bernoux, M., Cerri, C. C., Feigl, B. J., and Piccolo, M. C. (2002). Seasonal variation of soil chemical properties and CO2 and CH4 fluxes in unfertilized and P-fertilized pastures in an Ultisol of the Brazilian Amazon. *Geoderma* **10**, 227-241.
Ferraz, G., Russell, G. J., Stouffer, P. C., Bierregaard, R. O., Jr., Pimm, S. L., and Lovejoy, T. E. (2003). Rates of species loss from Amazonian forest fragments. *Proc Natl Acad Sci U S A* **100**, 14069-73.
Ferrenberg, S., O'Neill, S. P., Knelman, J. E., Todd, B., Duggan, S., Bradley, D., Robinson, T., Schmidt, S. K., Townsend, A. R., Williams, M. W., Cleveland, C. C., Melbourne, B. A., Jiang, L., and Nemergut, D. R. (2013). Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J* **7**, 1102-11.
Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**, 626-631.
Foerstner, K. U., von Mering, C., Hooper, S. D., and Bork, P. (2005). Environments shape the nucleotide composition of genomes. *EMBO Rep* **6**, 1208-13.
Fontaine, S., Bardoux, G., Abbadie, L., and Mariotti, A. (2004). Carbon input to soil may decrease soil carbon content. *Ecology Letters* **7**, 314-320.

Fontaine, S., and Barot, S. (2005). Size and functional diversity of microbe populations control plant persistence and long-term soil carbon accumulation. *Ecology Letters* **8**, 1075-1087.
Fontúrbel, M. T., Barreiro, A., Vega, J. A., Martín, A., Jiménez, E., Carballas, T., Fernández, C., and Díaz-Raviña, M. (2012). Effects of an experimental fire and post-fire stabilization treatments on soil microbial communities. *Geoderma* **191**, 51-60.
Forster, P., Ramaswamy, V., A., P., , Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., and Nganga, J. (2007). Changes in atmospheric constituents and in radiative forcing. *In* "Climate Change: The Physical Science Basis.".
Fracetto, G. G., Azevedo, L. C., Fracetto, F. J., Andreote, F. D., Lambais, M. R., and Pfenning, L. H. (2013). Impact of Amazon land use on the community of soil fungi. *Scientia Agricola* **70**, 59-67.
Gaby, J. C., and Buckley, D. H. (2014). A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database* **2014**.
Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387-1390.
Gaston, K. J. (1994). What is rarity? *In* "Rarity", pp. 1-21. Springer.
Gaston, K. J. (2000). Global patterns in biodiversity. *Nature* **405**, 220-227.
Gehring, C., Vlek, P. L. G., de Souza, L. A. G., and Denich, M. (2005). Biological nitrogen fixation in secondary regrowth and mature rainforest of central Amazonia. *Agriculture, Ecosystems & Environment* **111**, 237-252.
Gonzalez-Perez, J. A., Gonzalez-Vila, F. J., Almendros, G., and Knicker, H. (2004). The effect of fire on soil organic matter--a review. *Environ Int* **30**, 855-70.
Goss-Souza, D., Mendes, L. W., Borges, C. D., Rodrigues, J. L. M., and Tsai, S. M. (2019). Amazon forest-to-agriculture conversion alters rhizosphere microbiome composition while functions are kept. *FEMS Microbiol Ecol* **95**.
Goss-Souza, D., Mendes, L. W., Rodrigues, J. L. M., and Tsai, S. M. (2020). Ecological Processes Shaping Bulk Soil and Rhizosphere Microbiome Assembly in a Long-Term Amazon Forest-to-Agriculture Conversion. *Microb Ecol* **79**, 110-122.
Guerrero-Cruz, S., Vaksmaa, A., Horn, M. A., Niemann, H., Pijuan, M., and Ho, A. (2021). Methanotrophs: discoveries, environmental relevance, and a perspective on current and future applications. *Front Microbiol* **12**, 678057.
Hamaoui, G. S., Rodrigues, J. L. M., Bohannan, B. J. M., Tiedje, J. M., and Nüsslein, K. (2016). Land-use change drives abundance and community structure alterations of thaumarchaeal ammonia oxidizers in tropical rainforest soils in Rondônia, Brazil. *Applied Soil Ecology* **107**, 48-56.
Hamilton III, E. W., and Frank, D. A. (2001). Can plants stimulate soil microbes and their own nutrient supply? Evidence from a grazing tolerant grass. *Ecology* **82**, 2397-2402.
Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews* **68**, 669-685.
Hansen, M. C., Wang, L., Song, X. P., Tyukavina, A., Turubanova, S., Potapov, P. V., and Stehman, S. V. (2020). The fate of tropical forest fragments. *Science advances* **6**.
Hatzenpichler, R. (2012). Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Applied and environmental microbiology* **78**, 7501-7510.
Hedin, L. O., Brookshire, E. J., Menge, D. N., and Barron, A. R. (2009). The nitrogen paradox in tropical forest ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **40**, 613-635.
Ho, A., Lee, H. J., Reumer, M., Meima-Franke, M., Raaijmakers, C., Zweers, H., de Boer, W., Van der Putten, W. H., and Bodelier, P. L. E. (2019). Unexpected role of canonical aerobic methanotrophs in upland agricultural soils. *Soil Biology and Biochemistry* **131**, 1-8.
Huang, X.-F., Chaparro, J. M., Reardon, K. F., Zhang, R., Shen, Q., and Vivanco, J. M. (2014). Rhizosphere interactions: root exudates, microbes, and microbial communities. *Botany* **92**, 267-275.
Hughes, R. F., Kauffman, J. B., and Cummings, D. L. (2002). Dynamics of aboveground and soil carbon and nitrogen stocks and cycling of available nitrogen along a land-use gradient in Rondônia, Brazil. *Ecosystems* **5**, 244-259.
Husband, R., Herre, E. A., Turner, S. L., Gallery, R., and Young, J. P. W. (2002). Molecular diversity of arbuscular mycorrhizal fungi and patterns of host association over time and space in a tropical forest. *Molecular Ecology* **11**, 2669-2678.
Igwe, A. N., and Vannette, R. L. (2019). Bacterial communities differ between plant species and soil type, and differentially influence seedling establishment on serpentine soils. *Plant and Soil* **441**, 423-437.

INPE (2020). Monitoring Program of the Amazon and other biomes. Deforestation- Legal Amazon. (E. O. G. C. National Institute for Space Research, ed.).
Jank, L., Barrios, S. C., do Valle, C. B., Simeão, R. M., and Alves, G. F. (2014). The value of improved pastures to Brazilian beef production. *Crop and Pasture Science* **65**.
Jones, C. M., Stres, B., Rosenquist, M., and 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.
Jones, D. L., Cooledge, E. C., Hoyle, F. C., Griffiths, R. I., and Murphy, D. V. (2019). pH and exchangeable aluminum are major regulators of microbial energy flow and carbon use efficiency in soil microbial communities. *Soil Biology and Biochemistry* **138**.
Khan, M. A. W., Bohannan, B. J. M., Nusslein, K., Tiedje, J. M., Tringe, S. G., Parlade, E., Barberan, A., and Rodrigues, J. L. M. (2019). Deforestation impacts network co-occurrence patterns of microbial communities in Amazon soils. *FEMS Microbiol Ecol* **95**.
Kitagawa, Y., and Möller, M. R. F. (1980). Kaolin minerals in the Amazon soils. *Soil Science and Plant Nutrition* **26**, 255-269.
Kroeger, M. E., Delmont, T. O., Eren, A. M., Meyer, K. M., Guo, J., Khan, K., Rodrigues, J. L. M., Bohannan, B. J. M., Tringe, S. G., Borges, C. D., Tiedje, J. M., Tsai, S. M., and Nusslein, K. (2018). New Biological Insights Into How Deforestation in Amazonia Affects Soil Microbial Communities Using Metagenomics and Metagenome-Assembled Genomes. *Front Microbiol* **9**, 1635.
Kroeger, M. E., Meredith, L. K., Meyer, K. M., Webster, K. D., de Camargo, P. B., de Souza, L. F., Tsai, S. M., van Haren, J., Saleska, S., Bohannan, B. J. M., Rodrigues, J. L. M., Berenguer, E., Barlow, J., and Nusslein, K. (2021). Rainforest-to-pasture conversion stimulates soil methanogenesis across the Brazilian Amazon. *ISME J* **15**, 658-672.
Lammel, D. R., Feigl, B. J., Cerri, C. C., and Nusslein, K. (2015a). Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. *Front Microbiol* **6**, 1057.
Lammel, D. R., Nüsslein, K., Tsai, S. M., and Cerri, C. C. (2015b). Land use, soil and litter chemistry drive bacterial community structures in samples of the rainforest and Cerrado (Brazilian Savannah) biomes in Southern Amazonia. *European Journal of Soil Biology* **66**, 32-39.
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**, 5111-20.
Leal, P. L., Siqueira, J. O., and Stürmer, S. L. (2013). Switch of tropical Amazon forest to pasture affects taxonomic composition but not species abundance and diversity of arbuscular mycorrhizal fungal community. *Applied Soil Ecology* **71**, 72-80.
Leal, P. L., Stürmer, S. L., and Siqueira, J. O. (2009). Occurrence and diversity of arbuscular mycorrhizal fungi in trap cultures from soils under different land use systems in the Amazon, Brazil. *Brazilian Journal of Microbiology* **40**, 111-121.
Li, X. Q., and Du, D. (2014). Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS One* **9**, e88339.
Lima, A. S., Nóbrega, R. S. A., Barberi, A., da Silva, K., Ferreira, D. F., and Moreira, F. M. d. S. (2009). Nitrogen-fixing bacteria communities occurring in soils under different uses in the Western Amazon Region as indicated by nodulation of siratro (Macroptilium atropurpureum). *Plant and Soil* **319**, 127-145.
Linderman, R. G. (1991). Mycorrhizal interactions in the rhizosphere. In The rhizosphere and plant growth pp. 343-348. Springer.
Lisboa, C. C., Conant, R. T., Haddix, M. L., Cerri, C. E. P., and Cerri, C. C. (2009). Soil Carbon Turnover Measurement by Physical Fractionation at a Forest-to-Pasture Chronosequence in the Brazilian Amazon. *Ecosystems* **12**, 1212-1221.
Locey, K. J., and Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* **113**, 5970-5.
Lovejoy, T. E., Bierregaard Jr, R. O., Rylands, A. B., Malcolm, J. R., Quintela, C. E., Harper, L. H., Brown Jr, K. S., Powell, A. H., Powell, G. V. N., Schubart, H. O. R., and Hays, M. B. (1986). Edge and other effects of isolation on Amazon forest fragments.
Lucas-Borja, M. E., Miralles, I., Ortega, R., Plaza-Alvarez, P. A., Gonzalez-Romero, J., Sagra, J., Soriano-Rodriguez, M., Certini, G., Moya, D., and Heras, J. (2019). Immediate fire-induced changes

in soil microbial community composition in an outdoor experimental controlled system. *Sci Total Environ* **696**, 134033.
Luizao, R. C. C., Bonde, T.A., Rosswall, T. (1992). Seasonal variation of soil microbial biomass- the effects of clearfelling a tropical rainforest and establishment of pasture in the central Amazon. *Soil Biology and Biochemistry* **24**, 805-813.
MAAP (2019). "MAAP synthesis: 2019 Amazon deforestation trends and hotspots."
Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., Rubin, E. M., and Jansson, J. K. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**, 368-71.
Maia, S. M. F., Ogle, S. M., Cerri, C. E. P., and Cerri, C. C. (2010). Soil organic carbon stock change due to land use activity along the agricultural frontier of the southwestern Amazon, Brazil, between 1970 and 2002. *Global Change Biology* **16**, 2775-2788.
Maron, J. L., Marler, M., Klironomos, J. N., and Cleveland, C. C. (2011). Soil fungal pathogens and the relationship between plant diversity and productivity. *Ecol Lett* **14**, 36-41.
Martinez, D., Challacombe, J., Morgenstern, I., Hibbett, D., Schmoll, M., Kubicek, C. P., Ferreira, P., Ruiz-Duenas, F. J., Martinez, A. T., Kersten, P., Hammel, K. E., Vanden Wymelenberg, A., Gaskell, J., Lindquist, E., Sabat, G., Splinter BonDurant, S., Larrondo, L. F., Canessa, P., Vicuna, R., Yadav, J., Doddapaneni, H., Subramanian, V., Pisabarro, A. G., Lavín, J. L., Oguiza, J. A., Master, E., Henrissat, B., Coutinho, P. M., Harris, P., Magnuson, J. K., Baker, S. E., Bruno, K., Kenealy, W., Hoegger, P. J., Kües, U., Ramaiya, P., Lucas, S., Salamov, A., Shapiro, H., Tu, H., Chee, C. L., Misra, M., Xie, G., Teter, S., Yaver, D., James, T., Mokrejs, M., Pospisek, M., Grigoriev, I. V., Brettin, T., Rokhsar, D., Berka, R., and Cullen, D. (2009). Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences* **106**, 1954-1959.
Martiny, J. B., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science* **350**, aac9323.
Mataix-Solera, J., Guerrero, C., García-Orenes, F., Bárcenas, G. M., Torres, M. P., and Barcenas, M. (2009). Forest fire effects on soil microbiology. . *In* "Fire effects on soils and restoration strategies," Vol. 5, pp. 133-175, Boca Raton, FL.
Mazzetto, A. M., Feigl, B. J., Cerri, C. E., and Cerri, C. C. (2016). Comparing how land use change impacts soil microbial catabolic respiration in Southwestern Amazon. *Braz J Microbiol* **47**, 63-72.
Melo, V. F., Barros, L. S., Silva, M. C. S., Veloso, T. G. R., Senwo, Z. N., Matos, K. S., and Nunes, T. K. O. (2021). Soil bacterial diversities and response to deforestation, land use and burning in North Amazon, Brazil. *Applied Soil Ecology* **158**.
Melo, V. S., Desjardins, T., Silva Jr, M. L., Santos, E. R., Sarrazin, M., and Santos, M. M. L. S. (2012). Consequences of forest conversion to pasture and fallow on soil microbial biomass and activity in the eastern Amazon. *Soil Use and Management* **28**, 530-535.
Mendes, L. W., Braga, L. P. P., Navarrete, A. A., Souza, D. G., Silva, G. G. Z., and Tsai, S. M. (2017). Using Metagenomics to Connect Microbial Community Biodiversity and Functions. *Curr Issues Mol Biol* **24**, 103-118.
Mendes, L. W., de Lima Brossi, M. J., Kuramae, E. E., and Tsai, S. M. (2015a). Land-use system shapes soil bacterial communities in Southeastern Amazon region. *Applied Soil Ecology* **95**, 151-160.
Mendes, L. W., Tsai, S. M., Navarrete, A. A., de Hollander, M., van Veen, J. A., and Kuramae, E. E. (2015b). Soil-borne microbiome: linking diversity to function. *Microb Ecol* **70**, 255-65.
Merloti, L. F., Mendes, L. W., Pedrinho, A., de Souza, L. F., Ferrari, B. M., and Tsai, S. M. (2019). Forest-to-agriculture conversion in Amazon drives soil microbial communities and N-cycle. *Soil Biology and Biochemistry* **137**.
Meurer, K. H., Franko, U., Stange, C. F., Dalla Rosa, J., Madari, B. E., and Jungkunst, H. F. (2016). Direct nitrous oxide (N2O) fluxes from soils under different land use in Brazil—a critical review. *Environmental Research Letters* **11**, 023001.
Meyer, K. M., Klein, A. M., Rodrigues, J. L., Nusslein, K., Tringe, S. G., Mirza, B. S., Tiedje, J. M., and Bohannan, B. J. (2017). Conversion of Amazon rainforest to agriculture alters community traits of methane-cycling organisms. *Mol Ecol* **26**, 1547-1556.
Meyer, K. M., Morris, A. H., Webster, K., Klein, A. M., Kroeger, M. E., Meredith, L. K., Braendholt, A., Nakamura, F., Venturini, A., Fonseca de Souza, L., Shek, K. L., Danielson, R., van Haren, J., Barbosa de Camargo, P., Tsai, S. M., Dini-Andreote, F., de Mauro, J. M. S., Barlow, J., Berenguer,

E., Nusslein, K., Saleska, S., Rodrigues, J. L. M., and Bohannan, B. J. M. (2020). Belowground changes to community structure alter methane-cycling dynamics in Amazonia. *Environ Int* **145**, 106131.
Mirza, B. S., McGlinn, D. J., Bohannan, B. J., Nüsslein, K., Tiedje, J. M., and Rodrigues, J. L., 2020. . 86(10). (2020). Diazotrophs show signs of restoration in Amazon rain forest soils with ecosystem rehabilitation. *Applied and environmental microbiology,* **86**.
Mirza, B. S., Potisap, C., Nusslein, K., Bohannan, B. J., and Rodrigues, J. L. (2014). Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. *Appl Environ Microbiol* **80**, 281-8.
Moore, J. A. M., Jiang, J., Post, W. M., and Classen, A. T. (2015). Decomposition by ectomycorrhizal fungi alters soil carbon storage in a simulation model. *Ecosphere* **6**.
Moraes, J. L., Cerri, C. C., Melillo, J. M., Kicklighter, D., Neill, C., Skole, D. L., and Steudler, P. A. (1995). Soil carbon stocks of the Brazilian Amazon basin. *Soil Science Society of America Journal* **59**, 244-247.
Mueller, R. C., Paula, F. S., Mirza, B. S., Rodrigues, J. L., Nusslein, K., and Bohannan, B. J. (2014). Links between plant and fungal communities across a deforestation chronosequence in the Amazon rainforest. *ISME J* **8**, 1548-50.
Mueller, R. C., Rodrigues, J. L. M., Nüsslein, K., Bohannan, B. J. M., and Treseder, K. (2016). Land use change in the Amazon rain forest favours generalist fungi. *Functional Ecology* **30**, 1845-1853.
Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J. F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J. H., Zhang, K., Zhang, H., and Zhou, C. (2013). Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations. *Atmospheric Chemistry and Physics* **13**, 1853-1877.
Nakatsu, C. H., Torsvik, V., and Øvreås, L. (2000). Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. *Soil Science Society of America Journal* **64**, 1382-1388.
Nara, K. (2006). Ectomycorrhizal networks and seedling establishment during early primary succession. *New Phytol* **169**, 169-78.
Nascimento, H. E., and Laurance, W. F. (2002). Total aboveground biomass in central Amazonian rainforests: a landscape-scale study. *Forest ecology and management* **168**, 311-321.
Nascimento Lemos, L., Manoharan, L., William Mendes, L., Monteiro Venturini, A., Satler Pylro, V., and Tsai, S. M. (2020). Metagenome assembled-genomes reveal similar functional profiles of CPR/Patescibacteria phyla in soils. *Environ Microbiol Rep* **12**, 651-655.
Navarrete, A. A., Taketani, R. G., Mendes, L. W., Cannavan, F. d. S., Moreira, F. M. d. S., and Tsai, S. M. (2011). Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. *Revista Brasileira de Ciência do Solo* **35**, 1527-1540.
Navarrete, A. A., Tsai, S. M., Mendes, L. W., Faust, K., de Hollander, M., Cassman, N. A., Raes, J., van Veen, J. A., and Kuramae, E. E. (2015a). Soil microbiome responses to the short-term effects of Amazonian deforestation. *Mol Ecol* **24**, 2433-48.
Navarrete, A. A., Venturini, A. M., Meyer, K. M., Klein, A. M., Tiedje, J. M., Bohannan, B. J., Nusslein, K., Tsai, S. M., and Rodrigues, J. L. (2015b). Differential Response of Acidobacteria Subgroups to Forest-to-Pasture Conversion and Their Biogeographic Patterns in the Western Brazilian Amazon. *Front Microbiol* **6**, 1443.
Naya, H., Romero, H., Zavala, A., Alvarez, B., and Musto, H. (2002). Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *J Mol Evol* **55**, 260-4.
Neill, C., Fry, B., Melillo, J. M., Steudler, P. A., Moraes, J. F., and Cerri, C. C. (1996). Forest-and pasture-derived carbon contributions to carbon stocks and microbial respiration of tropical pasture soils. *Oecologia* **107**, 113-119.
Neill, C., Melillo, J. M., Steudler, P. A., Cerri, C. C., de Moraes, J. F., Piccolo, M. C., and Brito, M. (1997a). Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern Brazilian Amazon. *Ecological Applications* **7**, 1216-1225.
Neill, C., Piccolo, M. C., Cerri, C. C., Steudler, P. A., Melillo, J. M., and Brito, M. (1997b). Net nitrogen mineralization and net nitrification rates in soils following deforestation for pasture across the southwestern Brazilian Amazon Basin landscape. *Oecologia* **110**, 243-252.

Neill, C., Piccolo, M.C., Melillo, J.M., Steudler, P.A. and Cerri, C.C., (1999). Nitrogen dynamics in Amazon forest and pasture soils measured by 15N pool dilution. *Soil Biology and Biochemistry,* **31**, 567-572.
Neill, C., Piccolo, M.C., Steudler, P.A., Melillo, J.M., Feigl, B.J. and Cerri, C.C., 1 (1995). Nitrogen dynamics in soils of forests and active pastures in the western Brazilian Amazon Basin. *Soil Biology and Biochemistry* **27**, 1167-1175.
New, F. N., and Brito, I. L. (2020). What Is Metagenomics Teaching Us, and What Is Missed? *Annual Review of Microbiology* **74**, 117-135.
Ng, E. L., Patti, A. F., Rose, M. T., Schefe, C. R., Wilkinson, K., Smernik, R. J., and Cavagnaro, T. R. (2014). Does the chemical nature of soil carbon drive the structure and functioning of soil microbial communities? *Soil Biology and Biochemistry* **70**, 54-61.
Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., and Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* **20**, 241-248.
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., and Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* **47**, D259-D264.
Nogueira, M. (2012). Rally da Pecuária faz diagnóstico de nossas pastagens. *Balde Branco*, 43-44.
Noll, M., Matthies, D., Frenzel, P., Derakshani, M., and Liesack, W. (2005). Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. *Environmental Microbiology* **7**, 382-395.
Oliviera, S. M. D., and ter Steege, H. (2013). Floristic overview of the epiphytic bryophytes of terra firme forests across the Amazon basin. *Acta Botanica Brasilica* **27**, 347-363.
Öquist, M. G., Erhagen, B., Haei, M., Sparrman, T., Ilstedt, U., Schleucher, J., and Nilsson, M. B. (2016). The effect of temperature and substrate quality on the carbon use efficiency of saprotrophic decomposition. *Plant and Soil* **414**, 113-125.
Overmann, J., Abt, B., and Sikorski, J. (2017). Present and Future of Culturing Bacteria. *Annual review of microbiology* **71**, 711-730.
Pacheco, P. (2012). Actor and frontier types in the Brazilian Amazon: Assessing interactions and outcomes associated with frontier expansion. *Geoforum* **43**, 864-874.
Pajares, S., and Bohannan, B. J. (2016). Ecology of Nitrogen Fixing, Nitrifying, and Denitrifying Microorganisms in Tropical Forest Soils. *Front Microbiol* **7**, 1045.
Patureau, D., Zumstein, E., Delgenès, J.-P., and Moletta, R. (2000). Aerobic denitrifiers isolated from diverse natural and managed ecosystems. *Microbial ecology* **39**, 145-152.
Paula, F. S., Rodrigues, J. L., Zhou, J., Wu, L., Mueller, R. C., Mirza, B. S., Bohannan, B. J., Nusslein, K., Deng, Y., Tiedje, J. M., and Pellizari, V. H. (2014). Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Mol Ecol* **23**, 2988-99.
Peay, K. G., Baraloto, C., and Fine, P. V. (2013). Strong coupling of plant and fungal community structure across western Amazonian rainforests. *ISME J* **7**, 1852-61.
Pedreira, C. G., Silva, L. S., and Alonso, M. P. (2015). Use of grazed pastures in the brazilian livestock industry: a brief overview. *Forages Warm Clim* **7**, 7-18.
Pedrinho, A., Mendes, L. W., Merloti, L. F., Andreote, F. D., and Tsai, S. M. (2020). The natural recovery of soil microbial community and nitrogen functions after pasture abandonment in the Amazon region. *FEMS Microbiol Ecol* **96**.
Pedrinho, A., Mendes, L. W., Merloti, L. F., da Fonseca, M. C., Cannavan, F. S., and Tsai, S. M. (2019). Forest-to-pasture conversion and recovery based on assessment of microbial communities in Eastern Amazon rainforest. *FEMS Microbiol Ecol* **95**.
Penn, C., and Camberato, J. (2019). A Critical Review on Soil Chemical Processes that Control How Soil pH Affects Phosphorus Availability to Plants. *Agriculture* **9**.
Peter, H., Beier, S., Bertilsson, S., Lindstrom, E. S., Langenheder, S., and Tranvik, L. J. (2011). Function-specific response to depletion of microbial diversity. *ISME J* **5**, 351-61.
Peters, J. W., Boyd, E. S., D'Adamo, S., Mulder, D. W., Therien, J., and Posewitz, M. C. (2013). Hydrogenases, Nitrogenases, Anoxia, and H2 Production in Water-Oxidizing Phototrophs. *In* "Algae for Biofuels and Energy" (M. A. Borowitzka and N. R. Moheimani, eds.), pp. 37-75. Springer Netherlands, Dordrecht.

Philippot, L., Spor, A., Henault, C., Bru, D., Bizouard, F., Jones, C. M., Sarr, A., and Maron, P. A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. *ISME J* **7**, 1609-19.

Phillips, O. L., Aragão, L. E., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., Malhi, Y., Monteagudo, A., Peacock, J., and Quesada, C. A. (2009). Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344-1347.

Pilegaard, K. (2013). Processes regulating nitric oxide emissions from soils. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130126.

Preheim, S. P., Perrotta, A. R., Friedman, J., Smilie, C., Brito, I., Smith, M. B., and Alm, E. (2013). Computational methods for high-throughput comparative analyses of natural microbial communities. *Methods Enzymol* **531**, 353-70.

Prendergast-Miller, M. T., de Menezes, A. B., Macdonald, L. M., Toscas, P., Bissett, A., Baker, G., Farrell, M., Richardson, A. E., Wark, T., and Thrall, P. H. (2017). Wildfire impact: Natural experiment reveals differential short-term changes in soil microbial communities. *Soil Biology and Biochemistry* **109**, 1-13.

Prieto-Fernández, A., Acea, M. J., and Carballas, T. (1998). Soil microbial and extractable C and N after wildfire. *Biology and Fertility of Soils* **27**, 132-142.

Puissant, J., Jones, B., Goodall, T., Mang, D., Blaud, A., Gweon, H. S., Malik, A., Jones, D. L., Clark, I. M., Hirsch, P. R., and Griffiths, R. (2019). The pH optimum of soil exoenzymes adapt to long term changes in soil pH. *Soil Biology and Biochemistry* **138**.

Ranjan, K., Paula, F. S., Mueller, R. C., Jesus Eda, C., Cenciani, K., Bohannan, B. J., Nusslein, K., and Rodrigues, J. L. (2015). Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. *Front Microbiol* **6**, 779.

Rashid, A., Ahmed, T., Ayub, N., and Khan, A. G. (1997). Effect of forest fire on number, viability and post-fire re-establishment of arbuscular mycorrhizae. *Mycorrhiza,* **7**, 217-220.

Reiners, W. A., Bouwman, A., Parsons, W., and Keller, M. (1994). Tropical rain forest conversion to pasture: changes in vegetation and soil properties. *Ecological Applications* **4**, 363-377.

Reyes, H. A., Ferreira, P. F. A., Silva, L. C., da Costa, M. G., Nobre, C. P., and Gehring, C. (2019). Arbuscular mycorrhizal fungi along secondary forest succession at the eastern periphery of Amazonia: Seasonal variability and impacts of soil fertility. *Applied Soil Ecology* **136**, 1-10.

Ribeiro Filho, A. A., Adams, C., Manfredini, S., Aguilar, R., and Neves, W. A. (2015). Dynamics of soil chemical properties in shifting cultivation systems in the tropics: a meta-analysis. *Soil Use and Management* **31**, 474-482.

Ritter, C. D., Faurby, S., Bennett, D. J., Naka, L. N., Ter Steege, H., Zizka, A., Haenel, Q., Nilsson, R. H., and Antonelli, A. (2019). The pitfalls of biodiversity proxies: Differences in richness patterns of birds, trees and understudied diversity across Amazonia. *Sci Rep* **9**, 19205.

Rittl, T. F., Oliveira, D., and Cerri, C. E. P. (2017). Soil carbon stock changes under different land uses in the Amazon. *Geoderma Regional* **10**, 138-143.

Rodrigues, J. L., Pellizari, V. H., Mueller, R., Baek, K., Jesus Eda, C., Paula, F. S., Mirza, B., Hamaoui, G. S., Jr., Tsai, S. M., Feigl, B., Tiedje, J. M., Bohannan, B. J., and Nusslein, K. (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc Natl Acad Sci U S A* **110**, 988-93.

Rodrigues, K. F., and Dias-Filho, M. B. (1996). Fungal endophytes in the tropical grasses Brachiaria brizantha cv. marandu and B. humidicola. *Embrapa Amazônia Oriental-Artigo em periódico indexado (ALICE*.

Rousk, J., Baath, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., and Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**, 1340-51.

Schäfer, G. (2013). Membrane-Associated Energy Transduction in Bacteria and Archaea. *In* "Encyclopedia of Biological Chemistry (Second Edition)" (W. J. Lennarz and M. D. Lane, eds.), pp. 28-35. Academic Press, Waltham.

Schimann, H., Bach, C., Lengelle, J., Louisanna, E., Barantal, S., Murat, C., and Buee, M. (2017). Diversity and Structure of Fungal Communities in Neotropical Rainforest Soils: The Effect of Host Recurrence. *Microb Ecol* **73**, 310-320.

Schloss, P. D., and Handelsman, J. (2004). Status of the microbial census. *Microbiology and molecular biology reviews* **68**, 686-691.

Schloss, P. D., and Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**, 1501-6.
Schoug, Å., Fischer, J., Heipieper, H. J., Schnürer, J., and Håkansson, S. (2008). Impact of fermentation pH and temperature on freeze-drying survival and membrane lipid composition of Lactobacillus coryniformis Si3. *Journal of Industrial Microbiology and Biotechnology* **35**, 175-181.
Shade, A. (2017). Diversity is the question, not the answer. *ISME J* **11**, 1-6.
Shen, J.-P., Cao, P., Hu, H.-W., and He, J.-Z. (2013). Differential response of archaeal groups to land use change in an acidic red soil. *Science of the total environment* **461**, 742-749.
Shoun, H., Kim, D.-H., Uchiyama, H., and Sugiyama, J. (1992). Denitrification by fungi. *FEMS Microbiology Letters* **94**, 277-281.
Silva, C. A., Santilli, G., Sano, E. E., and Laneve, G. (2021). Fire Occurrences and Greenhouse Gas Emissions from Deforestation in the Brazilian Amazon. *Remote Sensing* **13**.
Silva, K. d., Nóbrega, R. S. A., Lima, A. S., Barberi, A., and Moreira, F. M. d. S. (2011). Density and diversity of diazotrophic bacteria isolated from Amazonian soils using N-free semi-solid media. *Scientia Agricola* **68**, 518-525.
Silva, L. C. R., Sternberg, L., Haridasan, M., Hoffmann, W. A., Miralles-Wilhelm, F., and Franco, A. C. (2008). Expansion of gallery forests into central Brazilian savannas. *Global Change Biology* **14**, 2108-2118.
Silvério, D. V., Brando, P. M., Balch, J. K., Putz, F. E., Nepstad, D. C., Oliveira-Santos, C., and Bustamante, M. M. (2013). Testing the Amazon savannization hypothesis: fire effects on invasion of a neotropical forest by native cerrado and exotic pasture grasses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120427.
Smith, S. E., and Read, D. J. (2010). "Mycorrhizal Symbiosis," Elsevier Science.
Sriswasdi, S., Yang, C. C., and Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nat Commun* **8**, 1162.
Steudler, P. A., Melillo, J. M., Feigl, B. J., Neill, C., Piccolo, M. C., and Cerri, C. C. (1996). Consequence of forest-to-pasture conversion on CH4fluxes in the Brazilian Amazon Basin. *Journal of Geophysical Research: Atmospheres* **101**, 18547-18554.
Strickland, M. S., Lauber, C., Fierer, N., and Bradford, M. A. (2009). Testing the functional significance of microbial community composition. *Ecology* **90**, 441-451.
Sturmer, S. L., and Siqueira, J. O. (2011). Species richness and spore abundance of arbuscular mycorrhizal fungi across distinct land uses in western Brazilian Amazon. *Mycorrhiza* **21**, 255-67.
Subbarao, G. V., Nakahara, K., Hurtado, M. P., Ono, H., Moreta, D. E., Salcedo, A. F., Yoshihashi, A. T., Ishikawa, T., Ishitani, M., Ohnishi-Kameyama, M., Yoshida, M., Rondon, M., Rao, I. M., Lascano, C. E., Berry, W. L., and Ito, O. (2009). Evidence for biological nitrification inhibition in Brachiaria pastures. *Proc Natl Acad Sci U S A* **106**, 17302-7.
Subbarao, G. V., Yoshihashi, T., Worthington, M., Nakahara, K., Ando, Y., Sahrawat, K. L., Rao, I. M., Lata, J. C., Kishii, M., and Braun, H. J. (2015). Suppression of soil nitrification by plants. *Plant Sci* **233**, 155-164.
Teasdale, S. E., Caradus, J. R., and Johnson, L. J. (2019). Fungal endophyte diversity from tropical forage grass Brachiaria. *Plant Ecology & Diversity* **11**, 611-624.
Torsvik, V., Goksøyr, J., and Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Applied and environmental microbiology* **56**, 782-787.
Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current opinion in microbiology*, 240-245.
Treseder, K. K., Balser, T. C., Bradford, M. A., Brodie, E. L., Dubinsky, E. A., Eviner, V. T., Hofmockel, K. S., Lennon, J. T., Levine, U. Y., MacGregor, B. J., Pett-Ridge, J., and Waldrop, M. P. (2011). Integrating microbial ecology into ecosystem models: challenges and priorities. *Biogeochemistry* **109**, 7-18.
Treseder, K. K., and Lennon, J. T. (2015). Fungal Traits That Drive Ecosystem Dynamics on Land. *Microbiology and Molecular Biology Reviews* **79**, 243-262.
Trevors, J. T. (2010). One gram of soil: a microbial biochemical gene library. *Antonie Van Leeuwenhoek* **97**, 99-106.
Trumbore, S., Davidson, E., Camargo, P., Nepstad, D., and Martinelli, L. (1995). Belowground cycling of carbon in forests and pastures of Eastern Amazonia. *Global Biogeochem Cycles* **9**, 515-528.

Tu, Q., Yu, H., He, Z., Deng, Y., Wu, L., Van Nostrand, J. D., Zhou, A., Voordeckers, J., Lee, Y. J., and Qin, Y. (2014). GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Molecular ecology resources* **14**, 914-928.

Tyukavina, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Smith-Rodriguez, K., Okpa, C., and Aguilar, R. (2017). Types and rates of forest disturbance in Brazilian Legal Amazon, 2000–2013. *Science Advances* **3**.

Ueki, A., Kaku, N., and Ueki, K. (2018). Role of anaerobic bacteria in biological soil disinfestation for elimination of soil-borne plant pathogens in agriculture. *Appl Microbiol Biotechnol* **102**, 6309-6318.

Uhl, C., Buschbacher, R. and Serrao, E.A.S., (1988). Abandoned Pastures in Eastern Amazonia. I. Patterns of Plant Succession. *The Journal of Ecology,* **76**, 663-681.

UNEP (2013). "Drawing down N2O to protect climate and ozone layer.." United Nations Environment Programme (UNEP), Nairobi, Kenya

van der Heijden, M. G., de Bruin, S., Luckerhoff, L., van Logtestijn, R. S., and Schlaeppi, K. (2016). A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. *ISME J* **10**, 389-99.

Van Der Heijden, M. G. A. (2004). Arbuscular mycorrhizal fungi as support systems for seedling establishment in grassland. *Ecology Letters* **7**, 293-303.

van Elsas, J. D., Chiurazzi, M., Mallon, C. A., Elhottova, D., Kristufek, V., and Salles, J. F. (2012). Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proc Natl Acad Sci U S A* **109**, 1159-64.

van Marle, M. J., Field, R. D., van der Werf, G. R., Estrada de Wagt, I. A., Houghton, R. A., Rizzo, L. V., Artaxo, P., and Tsigaridis, K. (2017). Fire and deforestation dynamics in Amazonia (1973-2014). *Global Biogeochem Cycles* **31**, 24-38.

Verchot, L. V., Davidson, E. A., Cattânio, H., Ackerman, I. L., Erickson, H. E., and Keller, M. (1999). Land use change and biogeochemical controls of nitrogen oxide emissions from soils in eastern Amazonia. *Global Biogeochemical Cycles* **13**, 31-46.

Verchot, L. V., Davidson, E. A., Cattânio, J. H., and Ackerman, I. L. (2000). Land-Use Change and Biogeochemical Controls of Methane Fluxes in Soils of Eastern Amazonia. *Ecosystems* **3**, 41-56.

Vitousek, P. M. (1984). Litterfall, nutrient cycling, and nutrient limitation in tropical forests. *Ecology* **65**, 285-298.

Vittor, A. Y., Gilman, R. H., Tielsch, J., Glass, G., Shields, T., Lozano, W. S., Pinedo-Cancino, V., and Patz, J. A. (2006). The effect of deforestation on the human-biting rate of Anopheles darlingi, the primary vector of falciparum malaria in the Peruvian Amazon. *The American journal of tropical medicine and hygiene* **74**, 3-11.

Wan, S., Hui, D., and Luo, Y. (2002). Fire effects on nitrogen pools and dynamics in terrestrial ecosystems: a meta-analysis *Ecological Applications,* **11**, 1349-1365.

Wang, Y., Ziv, G., Adami, M., Almeida, C. A. d., Antunes, J. F. G., Coutinho, A. C., Esquerdo, J. C. D. M., Gomes, A. R., and Galbraith, D. (2020). Upturn in secondary forest clearing buffers primary forest loss in the Brazilian Amazon. *Nature Sustainability* **3**, 290-295.

Wardle, D., and Ghani (1995). A critique of the microbial metabolic quotient (qCO2) as a bioindicator of disturbance and ecosystem development. *Soil Biology and Biochemistry* **27**, 1601-1610.

Wardle, D. A., Bardgett, R. D., Klironomos, J. N., Setälä, H., Van Der Putten, W. H., and Wall, D. H. (2004). Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629-1633.

Wei, W., Isobe, K., Shiratori, Y., Nishizawa, T., Ohte, N., Ise, Y., Otsuka, S., and Senoo, K. (2015). Development of PCR primers targeting fungal nirK to study fungal denitrification in the environment. *Soil Biology and Biochemistry* **81**, 282-286.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* **21**, 213-251.

Wrage, N., Velthof, G., Van Beusichem, M., and Oenema, O. (2001). Role of nitrifier denitrification in the production of nitrous oxide. *Soil biology and Biochemistry* **33**, 1723-1732.

Yang, S. S., and Chang, H. L. (1998). Effect of environmental conditions on methane production and emission from paddy soil. *Agriculture, ecosystems & environment* **69**, 69-80.

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., and Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* **67**, 1613-1617.

Yuste, J. C., PeÑUelas, J., Estiarte, M., Garcia-Mas, J., Mattana, S., Ogaya, R., Pujol, M., and Sardans, J. (2011). Drought-resistant fungi control soil organic matter decomposition and its response to temperature. *Global Change Biology* **17**, 1475-1486.

Zangaro, W., Alves, R. A., Lescano, L. E., Ansanelo, A. P., and Nogueira, M. A. (2012). Investment in fine roots and arbuscular mycorrhizal fungi decrease during succession in three Brazilian ecosystems. *Biotropica* **44**, 141-150.

Zarin, D. J., Davidson, E. A., Brondizio, E., Vieira, I. C., Sá, T., Feldpausch, T., Schuur, E. A., Mesquita, R., Moran, E., and Delamonica, P. (2005). Legacy of fire slows carbon accumulation in Amazonian forest regrowth. *Frontiers in Ecology and the Environment* **3**, 365-369.

Zemb, O., Achard, C. S., Hamelin, J., De Almeida, M. L., Gabinaud, B., Cauquil, L., Verschuren, L. M. G., and Godon, J. J. (2020). Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *Microbiologyopen* **9**, e977.

Zhao, C., Miao, Y., Yu, C., Zhu, L., Wang, F., Jiang, L., Hui, D., and Wan, S. (2016). Soil microbial community composition and respiration along an experimental precipitation gradient in a semiarid steppe. *Scientific reports* **6**, 1-9.

Zhu, T., Meng, T., Zhang, J., Zhong, W., Müller, C., and Cai, Z. (2014). Fungi-dominant heterotrophic nitrification in a subtropical forest soil of China. *Journal of Soils and Sediments* **15**, 705-709.

Table 1-1: Summary of findings across microbial studies investigating the impact of land use conversion from primary forest to pasture in the Amazon on microbial community diversity, composition, function, and activity. Tables include relevant metadata, including the land use effect investigated, parameters measured, type of land use conversion compared to primary forest (primarily pastures, but some cropping systems), age of converted systems, climate information including Köppen's class designation, annual temperature, and annual rainfall, soil type, location of study, major findings, and the article referenced. See footnote for list of abbreviations.

LUS Effect Measured	Parameters Measured	Land Types	Yr since conver-sion	Köppens Class, Annual Temp, Precip	Soil Type	Location(s)	Finding (*= statistically signif)	Citation
Soil - Phys/Chem prop controls over bact div/comm str	-T-RLFP cloning/seq of bact comm. -Phys/Chem prop	-Active past -Crop -Agrofor	past: 23-25yr	Af 25.7° C 2562mm	Incept	Benjamin Constant, Amazonas State, BR	*Comms shaped by soil attributes, including base saturation, Al^{3+}, pH *past and crop comms more div than Agrofor	1) Da C Jesus et al. 2009
MB and bact div	-PCR-DGGE of bact comms -MB C -Phys/Chem prop	-Active past -Fallow "capoeira" past	Active: 20yr Fallow: 15yr	Af 25.7° C 2562mm	Kand	Ariquemes, Rondônia, BR (ARMO)	*MB C and N higher in past than pfor and fallow across seasons *Comm diff across LUSs more pronounced in dry vs wet season *DNA profiles related to Al content, and MB C:Total C ratio	2) Cenciani et al. 2009
Whole comm tax and phylo div, spatial comm turnover	-Pyroseq of 16S rRNA -Phys/Chem prop	-Active past	22yr	Af 25.7° C 2562mm	Kand	Ariquemes, Rondônia, BR (ARMO)	*pfor and past comms are distinct by comp str *Within-sample tax and phylo div inc in past, but more homogenous across space *Acidobacteria dec and Firmicutes inc. in past	3) Rodrigues et al. 2013
Comp and relative abund of the phylum Verruco-microbia	-Verruco-microbia - specific 16S gene comp -Phys/Chem prop	-Active past	17yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Verrucomicrobia div higher in past, esp. subgroup 3 (tax, phylo) *Comp shifts, esp an inc. in subgroup 3 and dec. in Spartobacteria *absolute abund. assoc w/C content	4) Ranjan et al. 2015
Comp and relative abund of the Acidobacteria phylum	-16S rRNA comm comp -Phys/Chem prop	-Active past	22yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Acidobacteria are nearly 2x as abund. in pfor as past *Comm. comp. of most abund. subgroups homogenized across past soils	5) Navarrete et al. 2015

Abund and comm str of Thaumarchaeal ammonia oxidizers	-qPCR of *amoA* gene -Pyroseq Rhaumarchaeal *amoA*, 16S	-Active past	13yr 102yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Thaumarchaea over 10x lower in past compared to pfor. (per ng DNA), but not correlated w/ Phys/Chem prop **amoA*-based comms less div. **Nitrosotalea* (cluster 4) disappear in past	6) Hamaoui et al. 2016
Div and comm str of archaeal domain	-PCR DGGE of 16S rRNA gene -clone library of *amoA* gene -Phys/Chem prop	-Crop -Active past	Crops; 2yr; past: 38yr; sugarcane previously	Af 25.7°C 2562mm	Incept	Benjamin Constant, Solimões River, Amazonas state, BR	*Distinct archaeal comms across LUS *Div. of ammonia oxidizing archaea lower in converted LUS compared to pfor.	7) Navarrete et al. 2011
Comm resp to LUS change and repeated burning	-16S rRNA seq -Phys/Chem prop	-Aband past -Active past -'Area in preparation' -Subsist-ence ag	NP	Af 20-32°C 1700-2000	Typic Hapludalf Hapludult	Tepequém settlement, Amajari, Roraima, BR	-Species rich dec following pfor. burn (conversion to past) -Conventional plantations and intensively managed agriculture have lowest species rich and div.	8) Melo et al. 2021
Comp and func div as related to soil phys/chem prop	-16S rRNA pyroseq, qpCR -MG -Phys/Chem prop	-Defor sites (slash and burn)	2-4mo.	Am 28° C 2000mm	Ox	Porto dos Gaúchos & Ipiranga do Norte, Mato Grosso, BR	*Dec in SOM, inc in pH, base saturation *Inc in alpha div, Actinomycetales (related to N) *Dec in Planctomycetes (related to Al), Verrucomicobia, Chlamydiae *Inc in DNA repair/ protein maintenance genes	9) Navarrete et al. 2015a
Rhiz microbe tax and func comp along a LUS chrono-sequence	-MG -Greenhouse experiment to simulate rhiz -Phys/Chem prop	-Crop (no-till)	1, 10, 20yr	Am 27° C 1400mm	NP	Alto Xingú, Querência municipality, Mato Grosso, BR (12◦22'S; 52◦15'W)	*Homogenization in cropping systems- bulk soil and rhiz *func shifts occur in bulk soil related to virulence, K metab, disease defense *soybean rhiz is stable over time *Al and cation negatively impact tax div metrics	10) Goss-Souza et al. 2019

Whole comm div and func resp to soil conditions under LUS change	-MG -Phys/Chem prop	-Active past	20yr.	Am 26° C 2150mm	Ox	Tapajós National Forest, Pará, BR (2°51'23.9' S, 54◦57'28.4'W)	*Comms have distinct tax between LUS, assoc. w/ soil water holding capacity, but no diff. in rich *past soil comms are more func. div and assoc. w/Al concentration	11) Pedrinho et al. 2019
Linking func and tax div	-MG -Phys/Chem prop	-Defor -Crop (soy) -past	NP	Am 28° C 2000mm	Ox	Ipiranga do Norte, Mato Grosso, BR	*Alpha div incs, beta div decs in past *Seasonal trends in div more dramatic in defor sites *Tax and func struct diff between defor and pfor sites *Proteobacteria dec, Chloroflexi, Firmicutes inc in defor sites *Virulence and disease resp, aromatic metab genes higher in for, protein metab higher in defor.	12) Mendes et al. 2015(b)
Catabolic respiration	-Substrate-induced respiration profiles	-Degraded past -Improved past -Crop (no-till) -Crop (till)	Various	Various	Various	Rondônia, Mato Grosso	*Divergent profiles of catabolic resp *pasts respond to carboxylic and amino acids, pfors/cerrado respond to malonic, malic and succinic acid *soil type and climate are less important than LUS	13) Mazzetto et al. 2016
Func gene comp shifts with LUS change	-GeoChip4.0 gene probe -Phys/Chem prop	-Active past	6yr 38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Func. gene rich, and number of gene families dec. in young past (32% lower), but begins to recover in older past (16% lower). Func. gene redundancy higher in past *Denitrification, methane monooxygenase linked to pfor sites *pfor. soil comms have more genes assoc with C fixation and degradation, CH_4 oxidation, N cycling (ammonification, annamox, assimilatory N reduction, nitrification, etc.) *Only pullalanase and isopullulanase were assoc with past sites	14) Paula et al. 2014
Div of whole fungal comm	-DGGE of 18S rRNA gene	-Active past -Agrofor -Crop	NP	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	*Shift in dominant fungal groups between past and other LUS: Basidiomycota in past, Zygomycota in pfor and crop	15) Fracetto et al. 2013

AM fungi abund	-Spore count (trap cultures)	-Active past -Agrofor -Crop	NP	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	*trap culture spore counts higher in past than pfor	16) Leal et al. 2009
AM fungi div	-Spore count -Spore tax -Plant species	-Active past -Agrofor -Crop	41yr (sugarcane prior)	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	*Inc in spore count in past *Inc in spore div/count in crop/ agrofor	17) Sturmer and Siqueira 2011
AM fungi div	-Spore count -Spore tax -Phys/Chem prop -SOM	-Active past	41yr (sugarcane prior)	Af 25.7° C 2562mm	Incept	Benjanin Constant, Alto Solimões, Amazon State, BR	*Inc spore count in past *Comm comp diff	18) Leal et al. 2013
Fungal/ plant comm relationship	-Fungal rDNA comp -Plant *trnL* comp	-Active past	38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′5″S and 62°49′ 27″ W)	*Comp diff with LUS *Basidiomycota dec in past	19) Mueller et al. 2014
Fungal comm/ distribution patterns across LUS types	-Fungal rDNA comp, richness	-Active past	6yr 38yr 99yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′5″S and 62°49′ 27″ W)	-Rich decreases in past *Comm comp shifts *Generalist fungi in pasts, regardless of age *Geographical dist from pfor is strong predictor	20) Mueller et al. 2016
Fungal/plant comm div interaction	-Fungal rDNA comp in litter and soil	-Plantations	Plantation seedlings ~3yr	Af 25.7° C 3041mm	Acrisol	Paracou, French Guinana (5° 18′ N, 52° 53′ W)	-No variation in richs or evenness across plantations or pfor. *High spatial heterogeneity	21) Schimann et al. 2017

Fungal comm/ phytopathogen prevalence	-Fungal rDNA comp, div	-Active past	NP	Aw 20° C 1600-1900mm	NP	Mutum-Paraná River, Rondônia, BR	*Comp diff with LUS *Div incs in past *Comms more homogenous across past *potential phytopathogens inc in past	22) Cerqueira et al. 2018
Assembly processes in rhiz comms following conversion	-MG -Greenhouse rhiz -Phys/Chem prop -Enzyme activity	-Crop (no-till)	1, 10, 20yr	Am 27° C 1400mm	NP	Alto Xingú, Querência municipality, Mato Grosso, BR (12◦22'S; 52◦15'W)	*Comm. assembly in soybean bulk soil fit neutral assembly model *Comm assembly in rhiz fit niche-based model, leading to a point of permanent distribution state	23) Goss-Souza et al. 2020
Variation in microbial genomes	-Metagenome-assembled genomes (MAGs)	-Active past	38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Carbohydrate metab, cell signaling, dormancy genes inc in pasts. RNA metab and cofactor inc in pfor *Methanogensis genes inc, methanotrophy genes dec *Thaumarchaeota disappear in past, tax profiles vary *past comms more tax homogenous *28 MAGs recovered. Several lineages only found in past *Some MAGs from lineages containing no cultured organisms	24) Kroeger et al. 2018
Bact comm str in relation to soil and litter chemistry	-Bact T-RFLP -qPCR *nosZ, mcrA, pmoA,* 16S -Phys/chem prop	-Active past -Crop (soy)	20+yr	Aw 2000 mm	Red Ox	-Sinop, Mato Grosso, BR	*Inc'd pH, nutrient status, OM lability *Comm str shifts in LUSs -Bact richness does not decrease	25) Lammel et al. 2015a
Relation of soil Phys/Chem props and bact comm metrics	-Phys/Chem prop -qPCR and T-RFLP of 16S rRNA gene -MG	-Ag field -Defor -Active past	5yr <1yr >10yr	Am 28°C 2000mm	Ox	Ipiranga do Norte, Mato Grosso, BR	*Acidobacteria and Chlamydiae more abund in pfor soil *Firmicutes more abund in past *Nitrospira, Deinococcus-Thermus in crop systems *Actinobacteria in defor sites *Soil chem properties shape bact comms across LUS: pH, C, N, NO_3^-, and K *Al, base saturation index, Mg, Ca are correlated with many phyla	26) Mendes et al. 2015(a)

Co-occurrence patterns of prokaryotic comms	-16S rRNA seq	-Active past	38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*co-occurrence networks between pfor and past are distinct. *Modules of larger networks reflect potential shifts in N cycling in past *Props including temp, C/N, and H^{+}+Al^{3+} impact comm comp and network str	27) Khan et al. 2019
C cycling activity and SOC characterization along a LUS chrono-sequence	- ß- glucosidase enzyme activity - Fourier-transform spectroscopy -POXC	-Active past	100, 39, 24, 7yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*ß- glucosidase activity incs in past soils in absolute terms, but decs when normalized by C content *SOC incs with past age, POXC fraction decs with past age	28) Durrer et al. 2021
Microbial activity resp, C and N cycling	-MB, MR -Phys/Chem prop	-Plantation, -silvo-past -Active past	NR	Af 24-25° C 2500-4000mm	NR	SW of Caquetá Department, Columbia	*MB C inc in silvopastoral and past systems compared to pfor and plantation *SOC content is a controlling factor of microbial activity	29) Cruz et al. 2019
Methano-genic and -trophic comm div and life strategy	-MG	-Active past	38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Methanotroph taxa and *pmoA* genes lower in past vs. pfor., particularly alpha-Proteobacteria **mcRA* higher in past *ruderal life history (disturbance-specialist) favored in pasts.	30) Meyer et al. 2017
Relationship between CH_4 production and methanogen/troph comms	-CH_4 flux -amplicon seq, *pmoA* and *mcrA*	-Active past	NP	-2200mm (Rond.) -2000mm (Tapa.)	-Kand (r/y pod-lat; Rond.) -Ults, Oxs, Incepts (Tapa.)	-Ariquemes, Rondônia -Tapajos National Forest, Pará	* CH_4 flux higher in past, particularly Rondônia *Methanotrophic rich/relative abund lower, methanogen richness/ relative abund higher in past *Methangen rich and rel. abund correlate to flux rate *526 past taxa highly assoc w/flux rate compared to 41 in pfor.	31) Meyer et al. 2020

Active methanotrophic comms	-SIP amplicon seq and MG (methanogens/trophs)	-Active past	NP	NP	NP	-Ariquemes, Rondônia -Tapajos National Forest, Pará	*Actively-fixing methanogens (particularly acetoclastic types) inc in abund and div in past *Many active taxa id'd in past compared to for	32) Kroeger et al. 2021
Func potential for C& N cycling	-qPCR of func marker genes: *amoA*, *nirK*, *nirS*, *norB*, *nosZ*, *nifH*, *mcrA*, *pmoA*, 16S, 18SrRNA -Fluxes of NO_3-, N_2O, CO_2, CH_4	-Crop (soy) -Active past	-Crop: 2yr, 25yr -past: 25yr	Am 24.1° C 2171mm	Red Ox with clay texture	-Sinop, Mato Grosso, BR	*Compared to pfor., *nifH, amoA* (bacteria), *mcrA, pmoA, nirK, nosZ* genes dec in past, crop (2yr and 25yr) **amoA* (archaea), *nirS* inc in past, crop (2yr and 25yr) *past is CH_4 source, pfor. and crop are sink or neutral *pfor. is N_2O source, past and crop are neutral	33) Lammel et al. 2015b
Microbial-mediated C and N cycling across past	-MB C -Soil C,N mineralization/ nitrification -CH_4, N_2O, CO_2 flux	-Degraded past (no pfor. comparison)	-19yr	Am 25. 6°C 2200mm	Ox, Ult	-Fazenda Nova Vida, Rondônia, BR	*MB C varied 8x across pasts. based on soil type *past are either source or sink of CH_4, typically a source of N_2O, and CO_2 efflux loosely related to MB C *High rates of N immobilization for MB growth	34) Cerri et al. 2006
Changes in soil MB and activity	-MB C -respiration	-Fallow past -Active past	-Fallow: 9yr -past: 2, 6, 11yr	Awi 26°C 2082mm	Dystrophic Ox	Itupiranga , Pará, BR	*Young past contain less C and MB C than pfor and fallow past *Metab quotient highest in young surface pasts. but decreases in older pasts.	35) Melo et al. 2012
Impact on N cycling on bact comms	-16S rRNA amplicon seq - qPCR (*nifH*, *amo*A, *amo*B, *nir*K, *nos*Z, 16S) -Phys/Chem prop	-Crop (no-till; corn, soy)	2yr 8yr 20yr	Am 26°C 2150mm	Typic Haplustox (Ox)	Tapajós National Forest and Belterra municipality, Pará, BR	*Inc. bact div, dec archaeal div w/conversion *Inc N fixation in young ag soils *Greater potential for N_2O reductase ($N_2O \rightarrow N_2$) in crop soils over pfor *Greater potential for nitrite reductase ($NO_2^- \rightarrow NO$) in pfors over crop soils *Ca, Al, NH_4^+ and total N correlated to comm str	36) Merloti et al. 2019

N-cycling comm str, comp and func.	-MG -qPCR of N-cycling genes (*nifH*, *nirK*, *nosZ*) -Phys/Chem prop	-Active past	21yr	Am 26° C 2150mm	Ox	Belterra municipality, State of Pará, BR	*Al saturation and NO_3^- corr w/ tax and func comm str related to N cycling *pasts more func and taxon rich in N cyclers *Enriched taxa include *Anaeromyxobacter, Bacillus, Geobacter, Sorangium, Koribacter, Streptosporangium, and Conexibacter. Mycobacterium decrease* **nifH*, *nirK*, *nosZ* gene counts inc *denitrification gene trends differ by season *ammonia assimilation and nitrosative stress higher in past *nitric oxide synthse lower in past *past soils contain more specialists (unique)	37) Pedrinho et al. 2020
Isolation of free-living diazs to determine div	-Isolation of culturable free-living diaz + protein profiling -nitrogenase activity	-Agrofor -Crop -Active past	NP	Af 25.7° C 2562mm	NP	Benjamin Constant, Alto Solimões, Amazon State, BR	*Cell densities were highest in past-soil derived cultures for two of three medias tested *Id'd isolates from past included *Burkholderia* and *Bacillus* *nitrogenase activity varied, but highest performer id'd from past soil only	38) Silva et al. 2011
Viability, div, and efficiency of symbiotic (nodulating) diaz	-Nodulation density and efficiency of promiscuous legume	-Agrofor -Crop -Active past	38yr	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	*Highest nodulation number from agrofor and crop systems, followed by past, and pfor with lowest number *LUS contain similar number of 'efficient' nodule-forming strains, but crop and past contain highest number of 'inefficient; strains. pfor. contained the most 'high efficiency' strains	39) Lima et al. 2009
Diaz comm div	-*nifH* clone seq -qPCR of *nifH* gene	-Active past	5yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	-Rich and div were not diff *past comm str diff taxally and phylo *More gene copies in past than pfor *Diaz Firmicutes enriched in pfor, Spirochaetes, delta-proteobacteria, Verrucomicrobia and uncultured favored in past	40) Mirza et al. 2014

Distance-decay relationship of diaz comms	-*nifH* gene seq -Phys/Chem prop	-Active past	38yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia, BR (ARMO)	*Local (alpha) div incs, but comm turnover (beta) decs *LUS was a stronger determinant of comm str than geographic dist or Phys/chem prop *Pfor is particularly dissimilar w/ dist phylo	41) Mirza et al. 2020
Impact of fire frequency	-N_2 fixation rates -Soil Phys/Chem prop	-Pfor stands of variable fire frequency	NA	Aw 26° C 2000mm	NP	Araguaia State Park, Mato Grosso, BR	*Fire history decs N fixation rate in pfor areas, on average 24% -Frequency of fire has no signif effect *Positive linear relationship between N fixation rate and C:N ratio and P content in unburned For - Relationship between these factors was nonlinear in burned For.	42) Bomfim et al. 2020

Abbreviations: **Aband:** Abandoned.**Abund:** abundance/ abundant**. Agrofor:** Agroforestry**. Al:** aluminum**. AM:** Arbuscular mycorrhizal. ***amoA*:** ammonia monooxygenase gene. **ARMO:** Amazon Rainforest Microbial Observatory**. Assoc:** associated**. Bact:** bacteria. **BR:** Brazil. **C:** carbon**. CH_4:** methane**. Comm:** community. **Comp:** composition. **Crop:** Agricultural cropping system. **Dec:** decrease. **Defor:** Deforested. **Diaz:** diazotrophic. **Diff:** difference. **Dist:** distance. **Div:** diversity/diverse**. Esp:** especially. **Func:** functional/function. **Inc:** increase**. Incept:** Inceptisol. **K:** potassium. **Kand:** Kandiudult. **LUS:** LUS systems**.** ***mcrA*:** methyl coenzyme M reductase gene**. MB:** microbial biomass. **Metab:** metabolism/metabolic.**MG:** Metagonomic profile (DNA-based). **MR:** microbial respiration. **N:** nitrogen**. NA:** Not applicable. ***nifH*:** nitrogenase reductase**.** ***nirK*:** nitrite reductase gene**.** ***nosZ*:** nitrous oxide reductase gene**. NP:** Not provided.**Oxisol:** Oxisol. **Past:** pasture**. PCR DGGE:** polymerase chain reaction denaturation gradient gel electrophoresis**. Pfor:** primary forest. **Phys/Chem prop:** Physicochemical properties**.** ***pmoA*:** particulate methane monooxygenase**. POXC:** permanganate oxidizable carbon**. Phylo:** phylogenomic/phylogenetic. **Precip:** precipitation. **Pyroseq:** pyrosequencing**. qPCR:** quantitative Polymerase Chain Reaction. **Resp:** response.**Rhiz:** rhizosphere. **Rich:** richness. **r/y pod-lat:** red-yellow podzolic latosol. **Sfor:** secondary forest**. Seq:** sequencing**. Signif:** significant. **SOC:** soil organic carbon**. SIP:** Stable Isotope Probe**. SOM:** soil organic matter**. Str:** structure/structural. **Tax:** tax/tax. **Temp:** Temperature**. T-RLFP:** terminal restriction length fragment polymorphism**.** ***trnL*:** chloroplast intron gene. **Ult:** ultisol.**Yr:** year/s. **16SrRNA:** prokaryotic ribosomal RNA gene (DNA-based)

Table 1-2: Summary of findings across microbial studies investigating secondary forest recovery following primary forest to pasture conversion and abandonment in the Amazon. Studies investigate impacts on microbial community diversity, composition, function, and activity. Tables include relevant metadata, including the land use effect investigated, parameters measured, age of secondary forest, climate information including Köppen's class designation, annual temperature, and annual rainfall, soil type, location of study, major findings, and the article referenced. All studies referenced in Table 1-2 are also referenced in Table 1-1. See footnote for list of abbreviations.

LUS Effect Measured	Parameters Measured	SForest Age	Köppens Class Annual Temp, Precip	Soil Type	Location (s)	Findings (*=significant)	Citation
Div of bact and comm structural controls (soil - Phys/Chem prop.)	-T-RLFP cloning/seq of bact comm -Phys/Chem prop.	<5 yr 5-30yr	Af 25.7° C 2562mm	Incept	Benjamin Constant, Amazonas State, Brail	*Sfor become more similar (str and comp) to pfor w/succession age compared to other LUS	1) da C Jesus et al. 2009
Comp and relative abund of the phylum Verrucomicrobia	-Verrucomicrobia -specific 16S gene comp -Phys/Chem prop.	10yr	Aw 25.5° C 2200mm	Kand	Fazenda Nova Vida (10◦10′5″S and 62◦ 49′ 27″ W)	*Verrucomicrobia div higher in sfor (tax, but not phylo) *Absolute abund assoc w/C content *Comp str more similar to pfor than past	2) Ranjan et al 2015
Div and comm str of archaeal domain	-PCR DGGE of 16S rRNA gene -clone library of *amoA* gene -Phys/Chem prop.	2yr 16yr	Af 25.7°C 2562mm	Incept	Benjamin Constant, near Solimões River, Amazonas state, BR	*Archaeal comms more similar in pfor and sfor than past or crop *Div of NH_3- oxidizing archaea lower in sfor vs pfor, past, or crop systems	3) Navarrete et al. 2011
Func gene comp shifts with LUS change	-GeoChip4.0 gene probe -Phys/Chem prop.	13yr	Af 25.5° C 2200mm	Kand	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	-Suggests incomplete funct recovery in sfor *High spatial turnover of gene profiles	4) Paula et al. 2014
Abund and comm str of thaumarchaeal ammonia oxidizers	-qPCR of *amoA* gene -Pyroseq thaumarchaeal *amoA*, 16S	NR	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Thaumarchaeal abund similar across pfor and sfor -Comm str is more similar to pfor than past	5) Hamaoui et al. 2016

Whole comm div and func resp to soil conditions under land use change	-MG -Phys/Chem prop.	13-15 yrs	Am 26° C 2150mm	Ox	Tapajós National Forest, Pará, BR - 2◦51'23.9'S, 54◦57'28.4'W	*Tax div of sfors is similar to that of pfors rather than active past *Func div was higher than pfor, and more similar to pasts	6) Pedrinho et al. 2019
Div of whole fungal comm	-DGGE of 18S rRNA gene	NR	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	-Sfor comms more similar to pfor and agrofor sites than past	7) Fracetto et al. 2013
AM fungi div	-Spore count -Spore taxonomy -Plant species	>20 yr (old) <20 yr (young)	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State	*Inc'd spore div/count for both young and old sfor *Comm comp differs between pfor and sfor	8) Sturmer and Siqueira 2011
AM fungi div/ seasonality and soil prop relationship	-Spore count, -Species comp/div -Glomalin content	3-4yr (young) 6-8yr (mid) >120yr (mature)	Aw 2370mm	Typic Haplaustox	Alcântara county, Amazonia periphery, BR (2° 23′ 51″ S, 44° 24′ 16″ W)	*_Glomus_ and _Acaulospora_ dominate samples regardless of season -No difference in glomalin content (high % of SOC) *Seasonality more signif than successional age *Spore density higher in dry young-mid sfor compared to mature pfor	9) Reyes et al. 2019
Fungal/ plant comm relationship	-Fungal rDNA comp -Plant _trnL_ comp	12yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	ARMO, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Comp differs across for types *Basidiomycota dec in sfor	10) Mueller et al. 2014
Fungal comm/ distribution patterns across LUSs	-Fungal rDNA comp, rich	12yr 17yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Variable rich and comp divergence compared to pfor -No clear trends in generalist vs specialist fungi	11) Mueller et al. 2016

C cycling activity and SOC characterization along LUS chrono- sequence	- ß- glucosidase enzyme activity - Fourier-transform spectroscopy -POXC	12yr, 18yr	Af 25.5° C 2200mm	Kand	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Characterization of SOC based on organic derivative and absolute (POXC) is very similar to pfor, but POXC as fraction of total C is similar to past * ß- glucosidase activity is similarly high to pfor	12) Durrer et al. 2021
Relationship between CH_4 production and methanogen/troph comms	-CH_4 flux -amplicon seq, *pmoA* and *mcrA*	NR	-2200mm (Rond.) -2000mm (Tapa.)	-Kand (r/y pod-lat; Rond.) -Ultisols, Oxs, Incepts (Tapa.)	-Ariquemes, Rondônia -Tapajos National Forest, Pará	*CH_4 flux rates similar to pfor *Methanogen/troph rich and relative abund similar to pfor	13) Meyer et al. 2020
Active methanotrophic comms	-SIP amplicon seq (methanotrophs and methanogens) -MG	NR	NR	NR	-Ariquemes, Rondônia -Tapajos National Forest, Pará	*Active comms more similar to pfor than past *methanotrophs more active than methanogens	14) Kroeger et al. 2021
N-cycling comm comp and function	-MG -qPCR of N-cycling genes (*nifH*, *nirK*, *nosZ*) - Soil Phys/Chem	17yr	Am 26° C 2150mm	Ox	Belterra municipality, State of Pará, BR	*Comm tax and func str of N-cycling comms are distinct between pfor and sfor, -Div and rich are similar **nirK* and *nosZ* gene counts are higher in sfor	15) Pedrinho et al. 2020
Viability, div, and efficiency of symbiotic (nodulating) diaz	-Nodulation density and efficiency of promiscuous legume	<5yr (young) >5yr (old)	Af 25.7° C 2562mm	Incept	Benjamin Constant, Alto Solimões, Amazon State, BR	*Nodulation number slightly higher in sfor compared to pfor -Similar strain efficiency	16) Lima et al. 2009
Isolation of free-living diaz to determine div	-Isolation of culturable free-living diaz + protein profiling -nitrogenase activity	NR	Af 25.7° C 2562mm	NR	Benjamin Constant, Alto Solimões, Amazon State, BR	-Yielded a similarly low number of isolates as pfor, with moderate potential activity	17) Silva et al. 2011
Potential of symbiotic N_2 fixation in recovering forests	- $\delta^{15}N$ levels in forest leaf litter -Survey of legume species	2-25yr	Am 26.4° C 3000mm	Kaolinitic Ox (yellow latosol)	70km east of Manaus, central Amazonia, BR	*High density of N_2-fixing legumes assoc with low $\delta^{15}N$ in leaf litter in sfor, not pfor *No diff in trend over 25 yr of sfor succession	18) Gehring et al. 2005

Potential of N_2 fixation in degraded pasture	-^{15}N tracer analysis between forage grass and early successional species (legume and non-legume)	Aband pasture	Aw 26.5° C 1800mm	Kaolinitic yellow latosols	Fazenda Vitória, Paragominas, Pará, BR	*Successional legume species may obtain 75% of N from symbiotic fixation, 27% of total across species -atmospheric N not detected in forage grass	19) Davidson et al. 2018
Diaz comm div	-*nifH* clone seq -qPCR of *nifH* gene	10yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Tax but not phylo distinct from pfor *More gene copies than pfor, but less than past *Firmicutes, alpha, and beta proteobacteria enriched in sfor	20) Mirza et al. 2014
Distance-decay relationship of diaz comms	-*nifH* gene seq -soil Phys/Chem prop	12-17yr	Aw 25.5° C 2200mm	Kand (r/y pod-lat)	Ariquemes, Rondônia (ARMO; 10°10′18.71″S, 62°47′15.67″W)	*Tax rich and phylo div are similar to pfor, lower than past *Beta div (distance-decay) is higher than past, and similar to pfor	21) Mirza et al. 2020

Abbreviations: **Aband:** Abandoned.**Abund:** abundance/ abundant**. Agrofor:** Agroforestry**. Al:** aluminum**. AM:** Arbuscular mycorrhizal. ***amoA*:** ammonia monooxygenase gene. **ARMO:** Amazon Rainforest Microbial Observatory**. Assoc:** associated**. Bact:** bacteria. **BR:** Brazil. **C:** carbon**. CH_4:** methane**. Comm:** community. **Comp:** composition. **Crop:** Agricultural cropping system. **Dec:** decrease. **Defor:** Deforested. **Diaz:** diazotrophic. **Diff:** difference. **Dist:** distance. **Div:** diversity/diverse**. Esp:** especially. **Func:** functional/function. **Inc:** increase**. Incept:** Inceptisol. **K:** potassium. **Kand:** Kandiudult. **LUS:** LUS systems**.** ***mcrA*:** methyl coenzyme M reductase gene**. MB:** microbial biomass. **Metab:** metabolism/metabolic.**MG:** Metagonomic profile (DNA-based). **MR:** microbial respiration. **N:** nitrogen**. NA:** Not applicable. ***nifH*:** nitrogenase reductase**. *nirK*:** nitrite reductase gene**. *nosZ*:** nitrous oxide reductase gene**. NP:** Not provided.**Oxisol:** Oxisol. **Past:** pasture**. PCR DGGE:** polymerase chain reaction denaturation gradient gel electrophoresis**. Pfor:** primary forest. **Phys/Chem prop:** Physicochemical properties**. *pmoA*:** particulate methane monooxygenase**. POXC:** permanganate oxidizable carbon**. Phylo:** phylogenomic/phylogenetic. **Precip:** precipitation. **Pyroseq:** pyrosequencing**. qPCR:** quantitative Polymerase Chain Reaction. **Resp:** response.**Rhiz:** rhizosphere. **Rich:** richness. **r/y pod-lat:** red-yellow podzolic latosol. **Sfor:** secondary forest**. Seq:** sequencing**. Signif:** significant. **SOC:** soil organic carbon**. SIP:** Stable Isotope Probe**. SOM:** soil organic matter**. Str:** structure/structural. **Tax:** tax/tax. **Temp:** Temperature**. T-RLFP:** terminal restriction length fragment polymorphism**. *trnL*:** chloroplast intron gene. **Ult:** ultisol.**Yr:** year/s. **16SrRNA:** prokaryotic ribosomal RNA gene (DNA-based)

Figure 1-1: Patterns of deforestation in the Brazilian Legal Amazon from 1988-2020. **(A)** Annual rates of deforestation in square kilometers (dark gray bars, left axis) overlaid with cumulative forest loss (red line, right axis). Data obtained from Brazil's National Institute for Space Research (INPE) DETER satellite imagery **(B)** Causes of deforestation, shown as relative percent of total, normalized to exclude general 'fire' category using Tyukavina et al. (2017) data. Large-scale agriculture includes agro-industrial crop production such as soy as well as plantations, whereas small-scale agriculture refers to both subsistence farming and small commercial operations. The 'other' category may include mining, construction of roads and urban areas, dams, and more. Photo used for 'pasture' segment courtesy Jorge Rodrigues. Logging, small- and large-scale agriculture photos obtained from Unsplash.com.

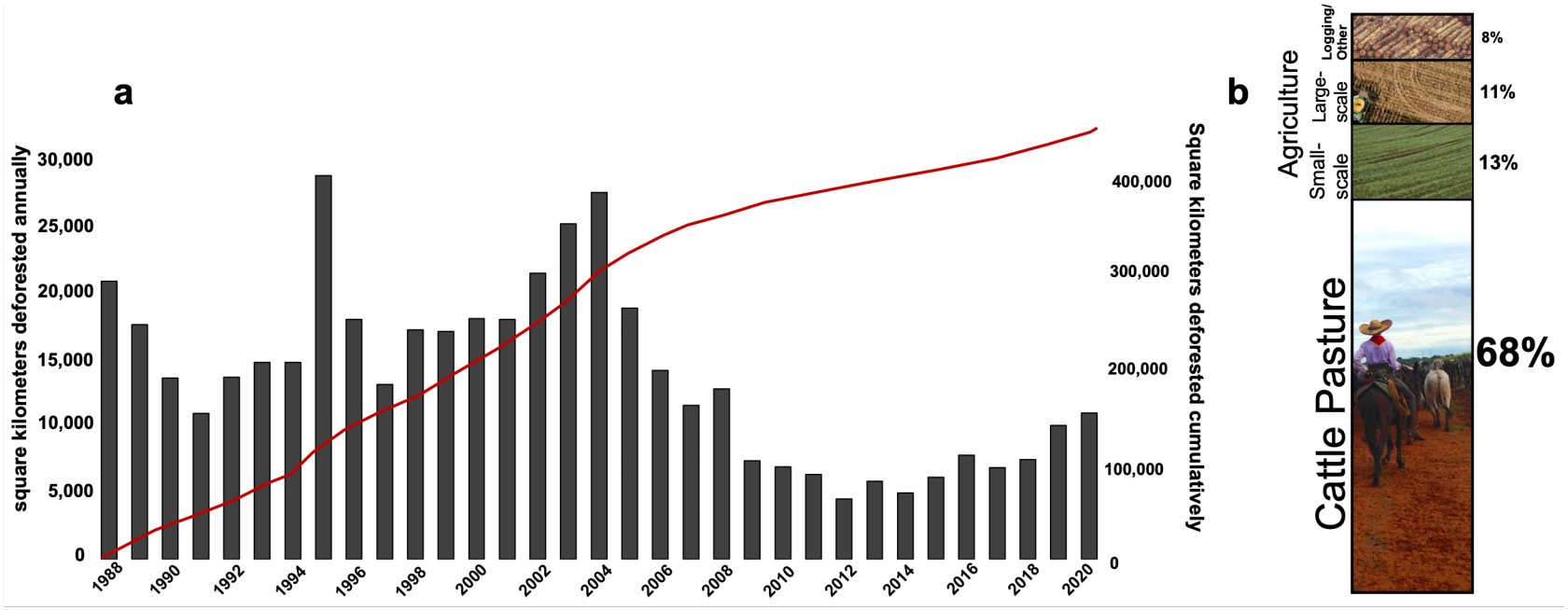


Figure 1-2: Geographical distribution of studies presented in Table 1-1. Labels correspond to number in 'Citation' column.

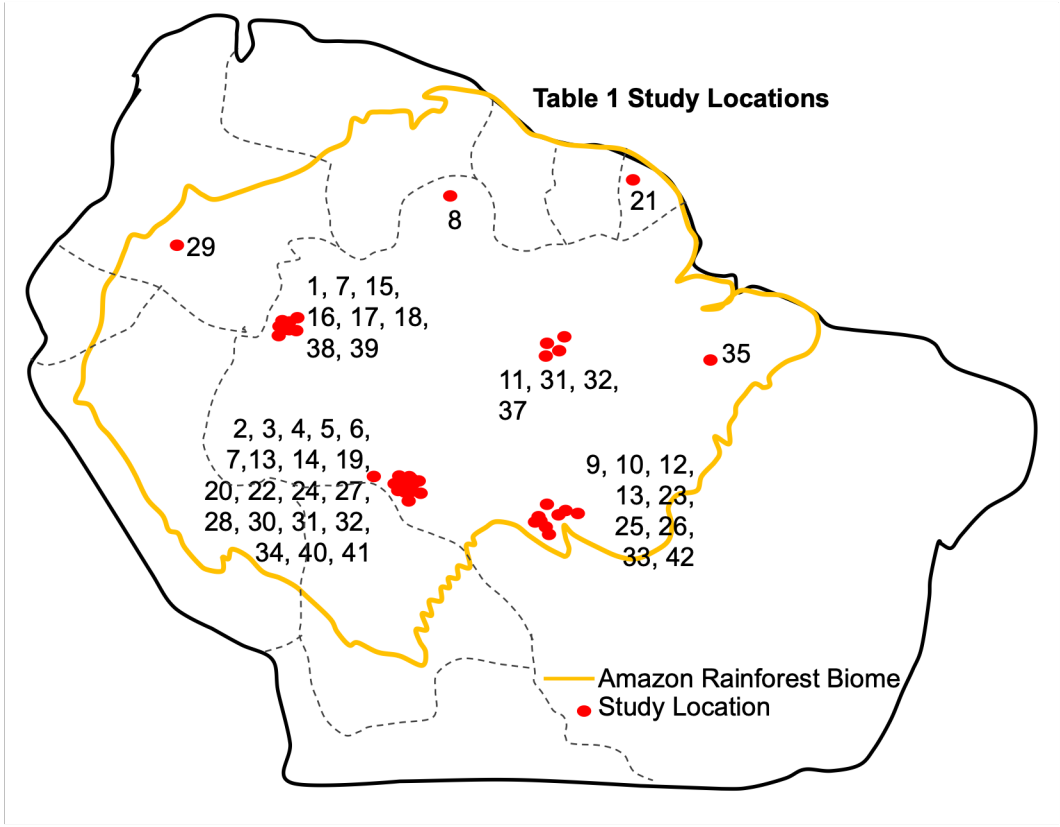


Figure 1-3: Hierarchical taxonomic network of prokaryotes (bacteria and archaea) identified in forest and pastures in Rondônia, Brazil based on 16S rRNA sequencing. Three forests and pastures each were sampled at 0-10cm in depth (n=42). Taxonomy included in analysis were subset based on presence in ≥ 50% of samples, and overall abundance ≥ 0.01% of sequences. Text and node size scale with total abundance across samples, and edge length is arbitrary. At each taxonomic level, a two-sample Bayesian test was used to determine whether abundance differed significantly (p-value adjusted) between forest and pasture.

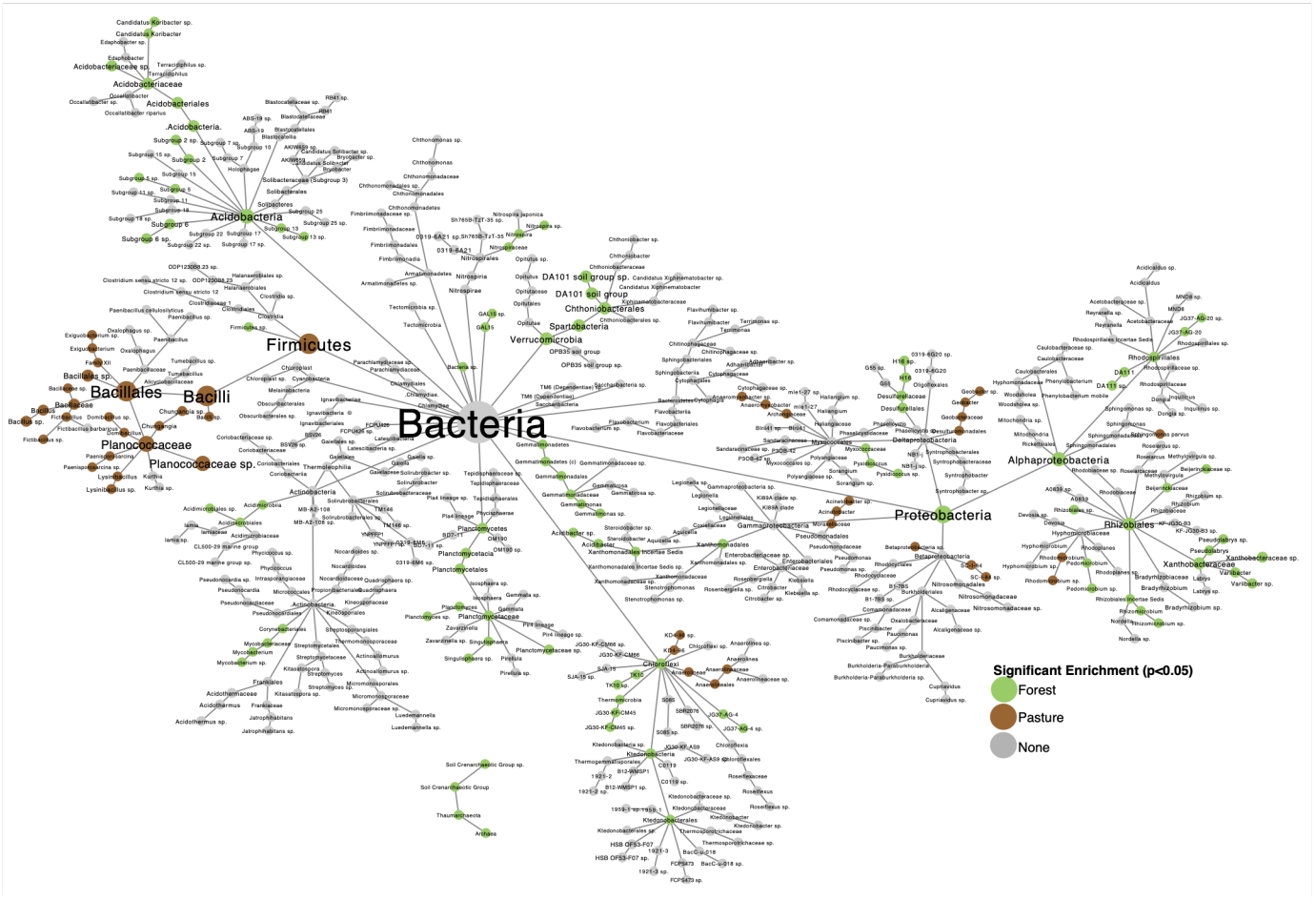


Figure 1-4: Coverage of reads at varying GC (guanine and cytosine) content within metagenomes of forest (n=5) and pasture (n=5) of Rondônia, Brazil. Blue dots represent individual reads passing quality filter check. Pasture samples show drastic loss in GC content between 30% and 50%, reflecting strong taxonomic changes. (From Rodrigues, unpublished data).

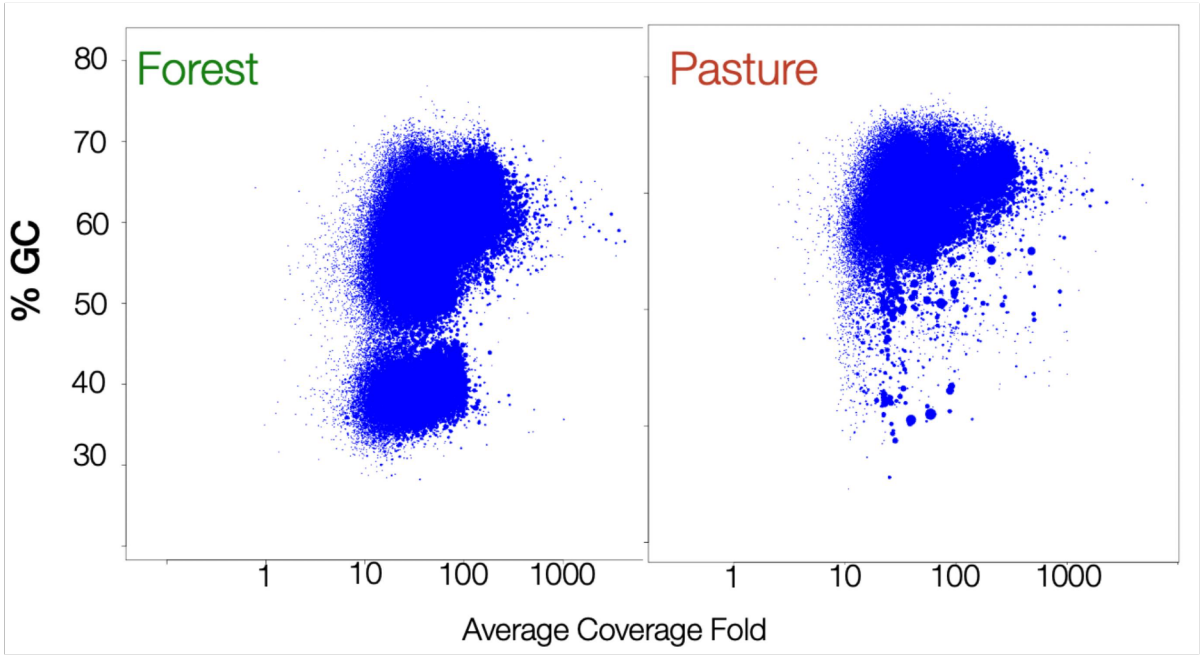


Figure 1-5: Reproduced with permission from Meyer et al. (2020) and Creative Commons License. Significant relationships between **(A)** methanogen richness and **(B)** methanogen relative abundance and log-transformed methane flux rate, taken from forest, pasture, and secondary forest in two Amazon states: Rondônia and Pará, Brazil. The R^2 values shown represent the proportion of variance in logarithmic methane flux explained by methanogen richness or relative abundance using a linear model approach. Dotted lines represent a flux rate of 0 ug CH_4 m^{-2} d^{-1} since a minimum value is added to data (+162).

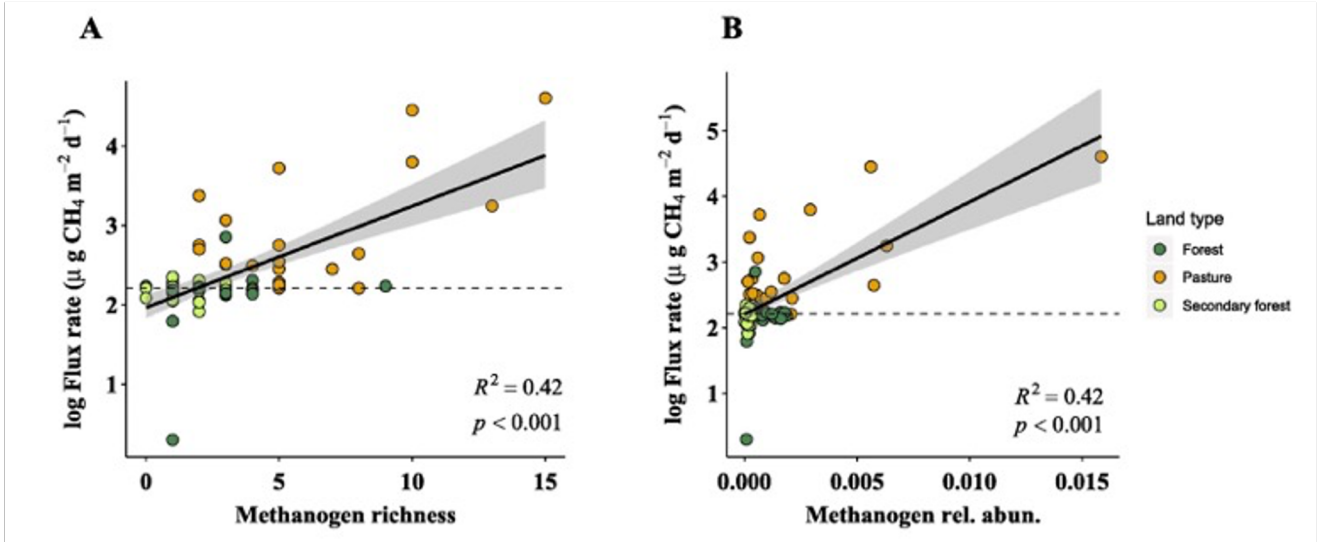


Figure 1-6: Conceptual diagram depicting crucial shifts in the nitrogen (N) cycle with conversion of primary tropical forest to pasture in the Amazon. Evidence from studies compiled so far suggest N is lost from the system upon conversion, and is further lost through cattle grazing, by export of biomass, and potentially inefficient reuptake of animal waste. Additionally, forage grasses such as those within the genera *Brachiaria* and *Urochloa* release a nitrification inhibitor called brachialactone from roots which inactivates enzymes catalyzing the oxidation of ammonia (NH_4^+) to nitrate (NO_3^-). This shifts the inorganic N pool from NO_3^- - to- NH_4^+- dominated, impacting availability of N for plants and microbes (both for biomass assimilation and redox reactions). Grazing impacts the physical soil environment by increasing bulk density and subsequently decreasing soil aeration and porosity. The shift in plant community has an inconsistent effect on total soil organic matter (SOM) storage across the Amazon. In the decades following conversion, SOM gradually shifts from forest-to-pasture derived, but the fraction actively metabolized by soil microbes is heavily pasture-derived even in very young pastures. pH increases by ash fertilization following forest conversion, potentially releasing microbial and plant communities from nutrient limitations including phosphorus (P). These factors are likely contributors to an observed increase in asymbiotic nitrogen fixers in the rhizosphere, in response to greater N limitation and favorable conditions. Decreased nitrification may also negatively impact nitrous oxide (N_2O) flux via denitrification, though results are mixed on the impacts of land use change on the abundance and diversity of denitrifying microbes.

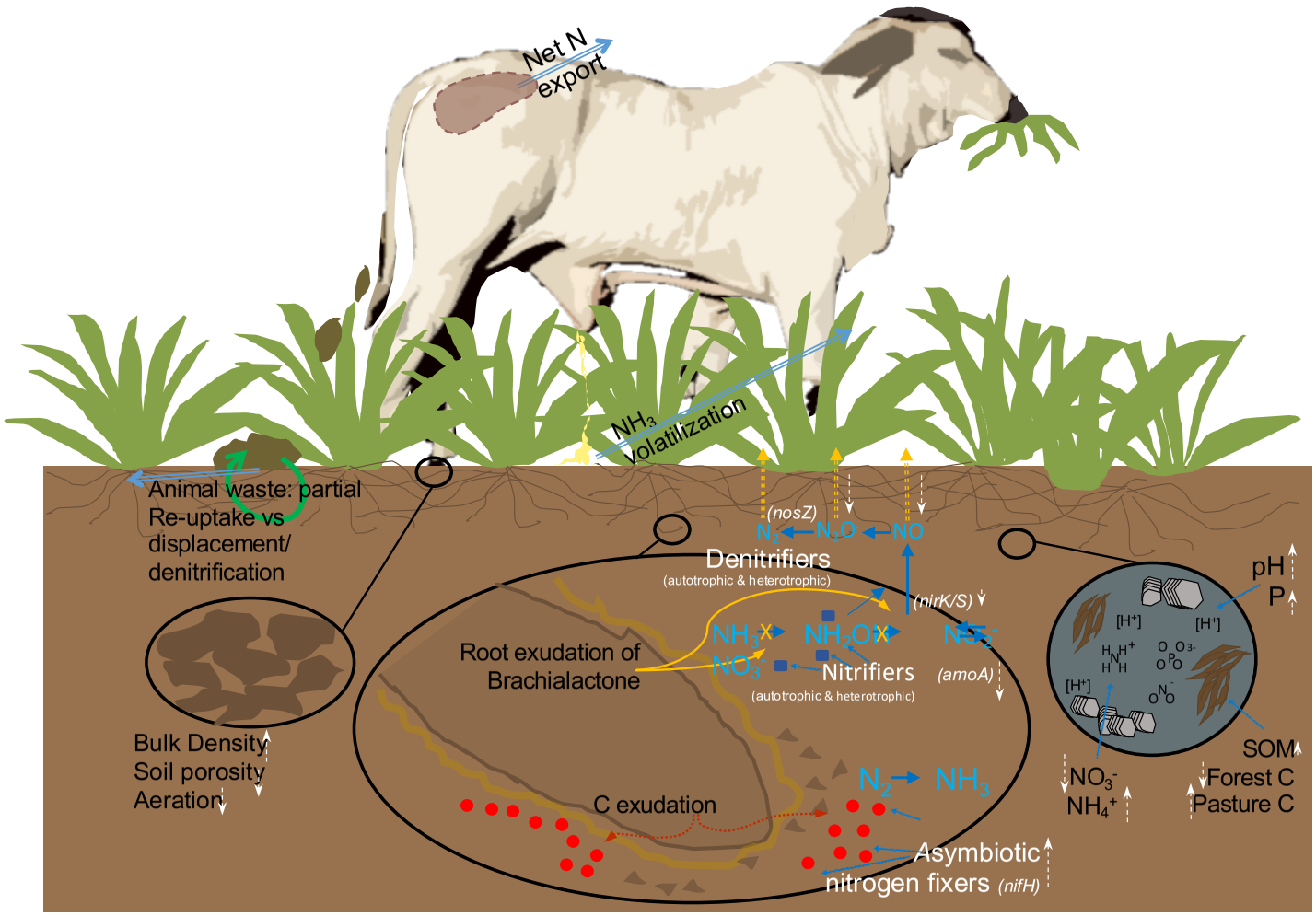

Net N export
NH_3 volatilization
Animal waste: partial Re-uptake vs displacement/ denitrification
(nosZ)
N_2 ← N_2O ← NO
Denitrifiers
(autotrophic & heterotrophic)
(nirK/S)
Root exudation of Brachialactone
NH_2OX
NO_3
Nitrifiers
(autotrophic & heterotrophic)
(amoA)
pH
P
Bulk Density
Soil porosity
Aeration
N_2 → NH_3
C exudation
Asymbiotic nitrogen fixers (nifH)
NO_3^-
NH_4^+
SOM
Forest C
Pasture C

Figure 1-7: Geographical distribution of studies presented in Table 1-2. Labels correspond to number in 'Citation' column.

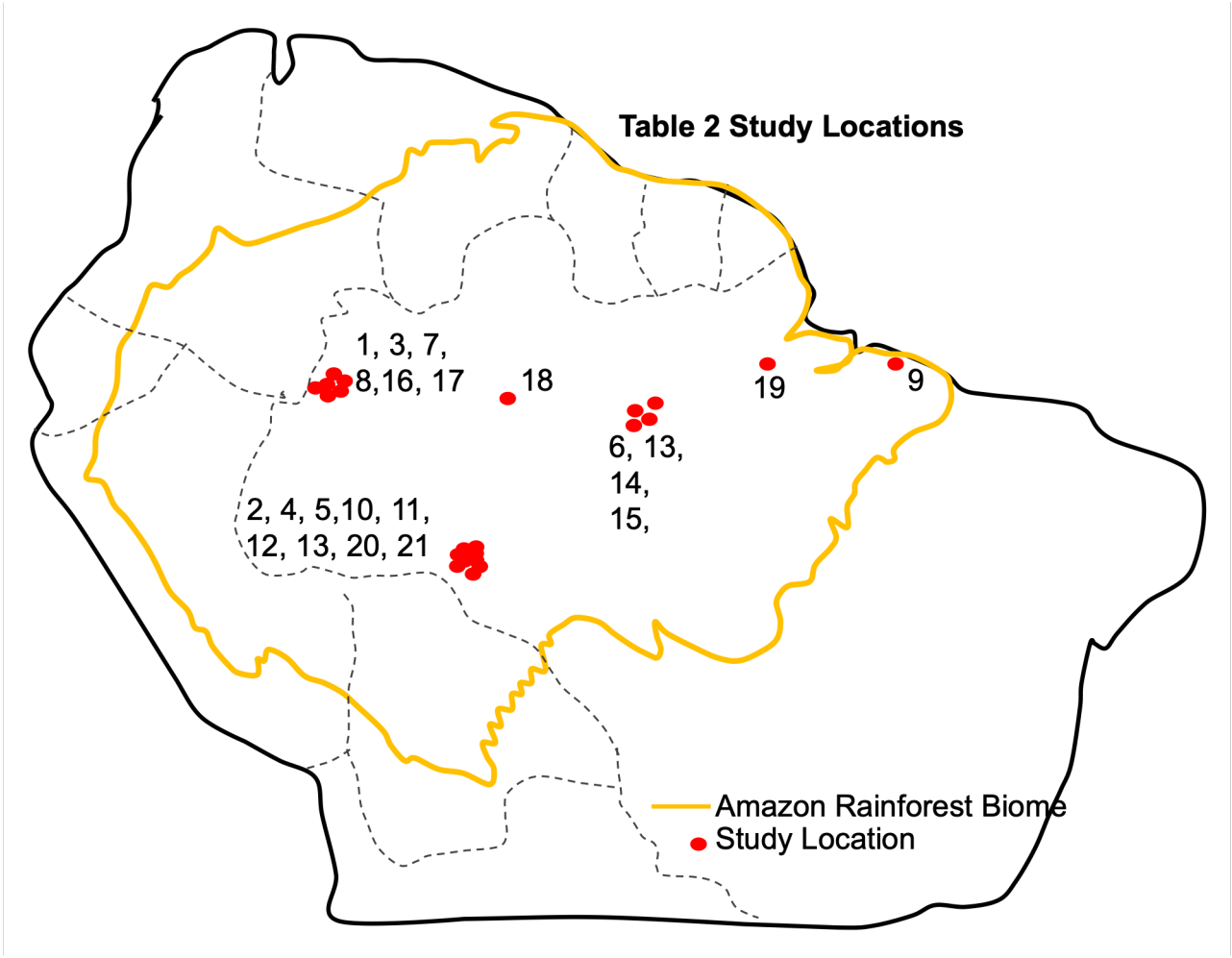
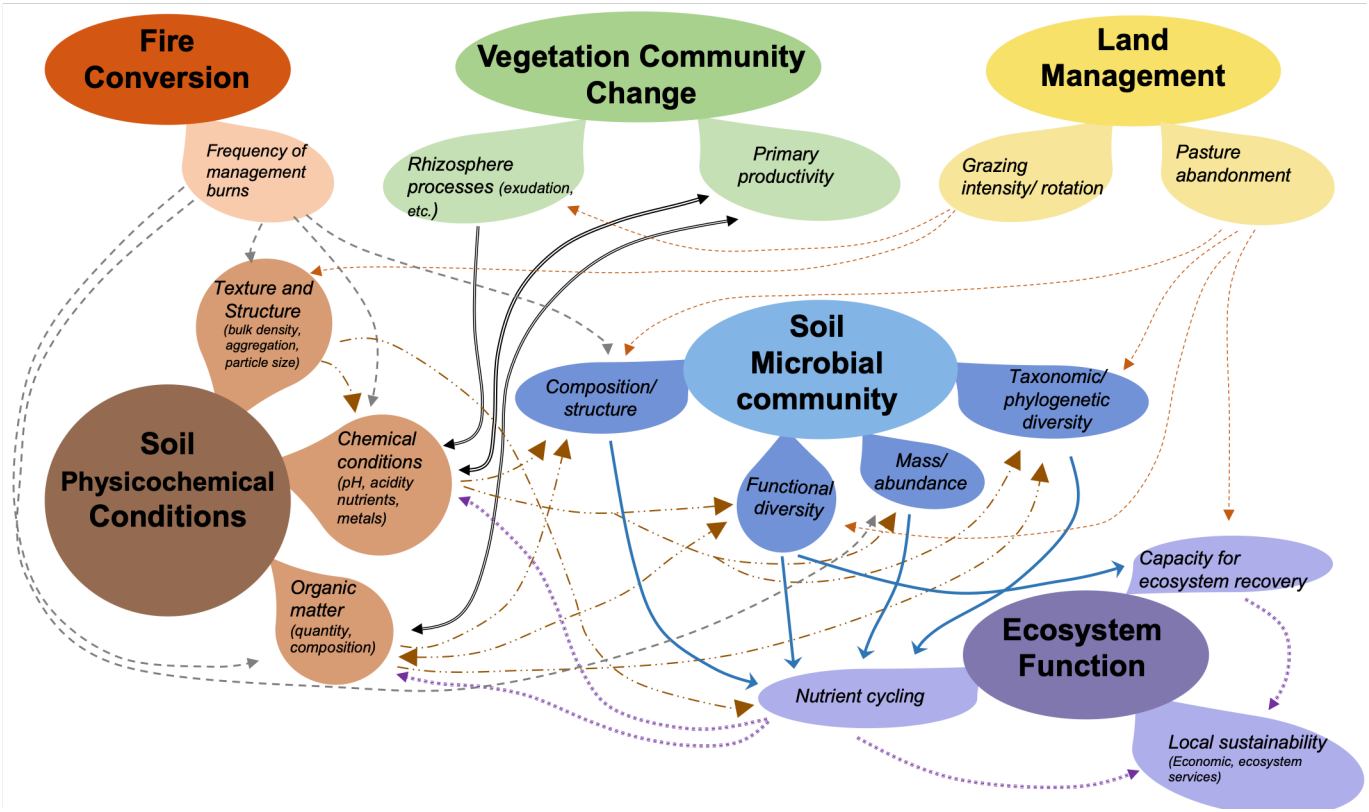


Figure 1-8: Conceptual Diagram of interactive forest-to-pasture land use disturbance in the Amazon. Changes in the plant community, use of fire for conversion and maintenance of pasture, and management vs. abandonment/reforestation influence a complex array of interactions between the soil environment, soil microbial communities, and ecosystem function. Interactions are direct, indirect, and may be two-way. Arrows included are based on causative or associative studies discussed throughout the chapter, but the diagram is not an exhaustive depiction of all ecological interactions. Arrow style is determined by the causative factor: Fire conversion- gray dashed; Vegetation community change- black double line; Land management- orange dots; Soil physicochemical conditions- brown dash/dot; Soil microbial community- blue solid; Ecosystem function- purple square dot.



Chapter 2: Asymbiotic diazotrophs provide nitrogen to soils following forest-to-pasture conversion in the Brazilian Amazon

Rachel E. Danielson[1], Jorge L. Mazza Rodrigues[1,2] *

[1] Department of Land, Air, and Water Resources, University of California – Davis, Davis, CA, USA

[2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA USA.

Running Title: Asymbiotic nitrogen fixation in Brazilian cattle pastures

ORCID #

Rachel Danielson: 0000-0001-7044-0747

Jorge L.M. Rodrigues: 0000-0002-6446-6462

***Corresponding author:** jmrodrigues@ucdavis.edu; University of California – Davis, 1110 Plant and Environmental Sciences Building, Davis CA 95616

Abstract

Asymbiotic diazotrophs replenish soil nitrogen (N) to natural terrestrial ecosystems through the fixation of atmospheric dinitrogen in the absence of obligate plant symbiosis. However, the relative importance of asymbiotic nitrogen fixation (ANF), as well as its environmental controls, are poorly understood in most ecosystems. This is especially true of the Amazon Basin, where more than 250,000 km^2 of tropical forest have been converted to cattle pasture. ANF has rarely been directly measured in primary forests in this region and has never been directly measured in active pastures, despite studies showing major shifts to the N cycle with land use (LU) conversion. We aimed to close this considerable knowledge gap by assessing how ANF activity rates, along with potential diazotrophic community abundance and soil physicochemical conditions, change across this LU dichotomy. Using $^{15}N_2$ -based ANF assays applied to surface soils (0-10 cm) of each LU, we revealed that significant augmentation of the potential soil diazotroph community abundance was concomitant with elevated rates of ANF in pastures, whereas measured rates were consistently at or near zero in forests. The abundance of *nifH* genes, however, did not help to explain process rates, and the two did not exhibit similar relationships with any physicochemical parameters. Increased ANF with LU change was modestly associated with a drastic decline in nitrate concentration. When pasture soils were considered alone, ANF rates were significantly associated with the ratio of soil-solution extractable organic carbon to nitrogen content, and to a lesser degree, with the ratio of total N to soil- solution extractable organic N. Since these factors were not related to forest ANF rates, it is likely that the entire range of N availability across surveyed forest soils was sufficiently replete so that low molecular weight N and C sources have no stimulatory effects on ANF. Additionally, metagenomic profiles revealed that the genetic potential for ANF is one of the most drastically augmented microbial functions across the entire soil microbial community. Protein subunit and cofactor biosynthesis genes were strongly correlated with other pasture-enriched genes, including those aiding in disaccharide C acquisition and dissimilatory reduction of inorganic N and sulfur compounds. The totality of this research has provided new insights into an important aspect of N cycling shifts associated with ongoing LU change in the Amazon Rainforest.

Keywords: Amazon Basin agroecosystems, Land-use change, Nitrogen cycling, Asymbiotic nitrogen fixation, Free-living diazotroph, Microbial ecology

Abbreviations: **AIC**: Akaike information criterion; **ANF:** asymbiotic nitrogen fixation; $\mathbf{C_{df}}$: forest-derived soil C; **C:** carbon; **DIG:** diazotrophic indicator gene; **DRV**: diazotrophic response variables (potential community size and ANF); **FC**: $\log_2$ fold-change; **GTA:** gene-targeted alignment; **LMW**: low molecular weight; **LU:** land-use; **mPSe**: mean Pearson and Spearman estimate; **mSe:** mean Spearman estimate; **N:** nitrogen; **P:** phosphorus; **RA**: read-assembly; **S**: sulfur; **UDP-NAG**: Uridine diphosphate N-acetylglucosamine

Introduction

Biological nitrogen fixation (BNF), the microbiological reduction of inert atmospheric N_2 to reactive ammonia (NH_3), is crucial in fueling primary productivity across Earth's ecosystems due to the relative scarcity of lithospheric N (Cleveland et al., 1999; Houlton et al., 2008). This functional trait, known as diazotrophy, is phylogenetically widespread amongst prokaryotes (referred to as diazotrophs), and is performed in terrestrial systems under a myriad of life strategies (Gaby and Buckley, 2012; Kaschuk and Hungria, 2017). The most well-known of these is the symbiosis occurring between select bacterial and plant species within special root structures called nodules. However, BNF is also performed by cyanobacterial diazotrophs in symbiosis with fungi (i.e., lichens), by endophytes in shoot or root tissue, and independently on soil surface detritus, the rhizosphere, and within bulk soil (Hedin et al., 2009; James et al., 2000). These 'independent' diazotrophs, often collectively referred to as free-living or asymbiotic, face the greatest barriers to activity, given the irreversible inactivation of the reaction-catalyzing nitrogenase enzyme in response to oxygen (O_2) exposure, and the steep energetic cost associated with breaking dinitrogen's (N_2) triple bond- at least 16 mol ATP per mol N_2 (Smercina et al., 2019; Soumare et al., 2020). Constrained modeling of N cycling has suggested its contribution to ecosystem-scale N inputs may outpace that of symbiotic BNF (Cleveland et al., 2009). Yet, this group of diazotrophs remains poorly understood, both in terms of genetic diversity, and response to environmental conditions.

In tropical systems such as the Amazon Rainforest, the role of BNF (collectively including symbiotic and asymbiotic) has been postulated as potentially significant in order to balance high rates of nitrate (NO_3^-) leaching (Hedin et al., 2009). However, extensive deforestation for the establishment of cattle pasture throughout the region has claimed upwards of 250,000 km^2 of historic forest area (Tyukavina et al., 2017). Such an immense land-use (LU) transformation has spurred investigation into its impact on nutrient cycling as well microbial communities that mediate these cycles (see Danielson and Mazza Rodrigues, 2022). While depth-integrated C stocks tend to increase in active pasture soils- potentially accounting for 50% of the aboveground C lost by forest conversion- N stocks do not appear to accumulate to the same degree (Neill et al., 1997a). Further, several studies have measured a multi-

magnitude decline in gross mineralization and nitrification rates (Neill et al., 1997b; Piccolo et al., 1994), as well as decreased emissions of N_2O and NO in pastures more than five years old (Neill et al., 2005). These observations collectively indicate that N cycling in pastures becomes decoupled from that of C, cycling more tightly within the ecosystem (i.e., minimizing loss pathways). Forests are generally converted to pasture through prescribed burning, resulting in ecosystem-scale N loss through volatilization and debris-leaching (Sharma et al., 2021). However, pasture management also plays a sustained role in N displacement, presumably intensifying with increased stocking rate (Dias-Filho et al., 2001). Due to economic constraints, it is uncommon for ranchers to apply fertilizer in order to combat potential N limitations on forage crops (Jank et al., 2014). The loss pathways imposed by pasture conversion as well as the removal of potentially N-fixing plant species such as legumes and canopy lichens raises the question of whether asymbiotic N_2 fixation (ANF) plays a role in microbial replenishment of soil N in grazed pastures.

Although ANF is a crucial segment of terrestrial N cycling, few studies have investigated process rates across this forest-pasture dichotomy in the Amazon. Surprisingly, just a handful of studies have utilized proxy measurements including ^{15}N dilution, acetylene reduction, and model-based mass balance to approximate or constrain rates in primary forest soils (e.g., Cleveland et al., 2009; Moreira et al., 2021; Salati et al., 1982; Wong et al., 2021; Wong et al., 2019). Other studies conducted throughout the Amazon region have assessed the impact of LU change on ANF potential, but indirectly through diazotroph marker gene abundance, with one study concluding a decreased abundance while another two demonstrating that active pastures foster a higher absolute quantity of diazotrophs (Lammel et al., 2015; Mirza et al., 2014; Pedrinho et al., 2020). However, given the reaction's energy demand, O_2 sensitivity, and widespread genetic incidence (Robson and Postgate, 1980; Smercina et al., 2019; Yan et al., 2022), quantification of a single marker gene cannot be relied upon to explicitly indicate an elevated influx of N via ANF in pasture soils, and to our knowledge, ANF rate has only been investigated in an abandoned and degraded tract using a ^{15}N dilution approach (Davidson et al., 2018). We aimed to address this crucial knowledge gap by simultaneously measuring two diazotrophic response variables (DRVs) across the LU dichotomy- potential diazotrophic community size, and ANF process rates, testing our hypothesis that augmentation of the soil diazotrophic community scales with increased ANF process

rates in active pastures compared to primary forests. To address this question, we employed high-sensitivity $^{15}N_2$ gas-based isotopic labeling rather than proxy measurements. Next, we profiled soil physicochemistry (including pools of C, N, macro- and micronutrients, pH, and soil texture) as well as untargeted soil metabolomics across the LU dichotomy, aiming to determine whether these environmental variables are associated with shifts in ANF rate and community size. In doing so, we hypothesized that augmentation of DRVs is related to greater availability of low molecular weight (LMW) C sources as well as decreased availability of inorganic N sources. Finally, we utilized a novel two-pronged analysis of the soil metagenome, aiming to characterize diazotrophs within their broader microbial communities, to quantify a suite of diazotrophic genes, assess how shifts in the functional potential for diazotrophy compares to total community functional potential, and identify potential biogeochemical interactions with diazotrophy through identification of correlated gene co-occurrence.

Materials and Methods

Site description and sample collection

Soil samples were collected in late March and early April of 2017 (end of the wet season) from sites at or near Agropecuaria Nova Vida, approximately 35 km southeast of Ariquemes, Rondônia, Brazil (10° 9’ 52.2” S, 62° 49’ 42.6” W). The Köppen’s classification for this region is Am (tropical monsoon), with an annual average precipitation of 2200- 2500 mm (Alvares et al., 2013). Soils have previously been classified as red-yellow podzolic latosols in this region (Neill et al., 1997a), with soils in this study mainly falling into sandy loam, loam, and clay loam textural classes. We sampled from three tracts of primary forest and pasture each. Pasture sites were selected to control for time since conversion at approximately 45 years. At each site, seven samples were collected within a 100 m^2 quadrat, at each corner point as well as three additional samples along the north-facing transect at 10 cm, 1 m and 10 m from the origin point. At each sampling point, any surface debris was removed, and four, 2.5 cm diameter cores were extracted with a sterilized corer from 0 – 10 cm and kept cool (~4-5° C) until homogenization and processing. A subsample was frozen for inorganic N measurements. Volumetric soil water content was

measured ten times at 5 and 10 cm depth and averaged, and three additional core replicates were extracted for field estimation of gravimetric water content.

Isotopic labeling experiment

To establish rate assays, soils were homogenized, sieved to 2 mm, and cleared of root fragments or plant debris. To establish ANF activity rate, 5 g dry weight-adjusted soil was added to 15 mL glass vials fitted with rubber septa and crimped to make them airtight. For each of the 42 samples, five replicate incubations were prepared, totaling 210 vials. Isotopic enrichment was performed by using a three-way syringe valve which allows accurate pulling of the desired volume of headspace and immediately replacing it with labeled gas mixtures enriched to 52 $\pm$ 0.01% of total headspace (normalized based on soil volume, gravimetric water content, and later adjusted for particle density). Gas mixtures added were 80% $^{14}N_2$ (control, n=2) or $^{15}N_2$ (enriched, n=3, Cambridge Isotope Laboratories, Cambridge, MA, USA) and 20% O_2 (Ultrapure; Airgas, Sacramento, CA, USA). Samples were kept in the dark at 30 °C for 14 days (temperature was chosen based on mean ambient temperature across sampling dates), then frozen to -80 °C until downstream processing. Enriched and natural abundance experimental samples as well as field-preserved 'baseline' soils were measured for total isotopic enrichment. ANF rate calculations were performed as previously described in Bomfim et al. (2018).

Soil DNA extraction and molecular analyses

Soil DNA was extracted with the DNeasy PowerLyzer Powersoil Kit (Qiagen Inc., Germantown, MD, USA) in triplicate from each sampling point using 0.3 g wet-weight soil per extraction. Quantitative polymerase chain reaction (qPCR) targeting an essential nitrogenase subunit-encoding gene (*nifH*) was used as a proxy of diazotroph abundance. The *nifH* gene was targeted using the PolF/PolR (Poly et al., 2001) primer set, since a previous assessment has shown this primer pair produces the most consistent quantification across species and variable PCR conditions (Gaby and Buckley, 2017). Extracted total soil DNA was also used for PCR-free metagenomic sequencing at the Department of Energy - Joint Genome Institute (Berkeley, CA). For metagenomic sequencing, a starting quantity of1.82 ng sample DNA was

sheared to 471bp then size-selected, end-repaired and ligated with Illumina adaptors. Samples were sequenced on a NovoSeq S4 2x151bp paired end with XP V1.5 reagent kit.

Soil physicochemical and metabolite profiling

The soil chemical factors measured in this study were meant to capture indicators of plant and microbial nutrient availability and were used to determine environmental factors important to diazotrophic abundance and activity. Physical metrics included soil texture, while soil chemical and nutrient status metrics included pH, CEC, and soil acidity, as well as total C (TC), N (TN), P, K, Ca, and Mg. The reactive N pool measurements included extractable organic N (EON), NH_4^+, and NO_3^-, and in combination, total extractable N (TE; Hood-Nowotny et al., 2010; Qiu et al., 1987). Proxies of 'microbial-accessible' C included loss-on-ignition organic matter (OM; Schulte and Hoskins, 1995), permanganate oxidizable C (POXC; Culman et al., 2012), and potassium sulfate-extractable C (EOC; Boyle et al., 2008). Soil $\delta^{15}N$ natural abundance and $\delta^{13}C$ were measured via dry combustion mass spectrometry at the UC Davis Stable Isotope Facility. Values were used as indicators of overall N cycling, and to estimate the derivation of soil C from forest or pasture sources, respectively. The use of untargeted soil metabolomics was used as an additional proxy for low molecular weight C compounds (Swenson et al., 2015). This approach has the added benefit of indicating the availability of organic N and P. Briefly, 1g field-preserved soil was submitted to the West Coast Metabolomics Center at UC Davis, and samples were processed using gas chromatography- time of flight mass spectrometry. Emissions peaks were used to identify approximately half of the metabolites found across samples. The remaining metabolites are identified by BinBase number, which can be queried at binvestigate.fiehnlab.ucdavis.edu/#/ for further information. Finally, additional metal pools including Fe, Mo, and V were quantified using ICPMS (**Supplementary [Supp.] Methods**), as these elements are vital to the nitrogenase enzyme.

Metagenomic sequence processing

Metagenomic data was processed using two custom pipelines in order to address different and specific questions by targeting diazotrophic genes of interest (DIGs). First, a read assembly (RA) approach was used. The goal of this approach was to achieve a robust diazotroph profile within the

context of the whole soil community metagenome. A second gene-targeted alignment (GTA) based approach was devised to compensate for some limitations of the above-described RA approach. Owing to the complexity of soil communities, the average assembly length in this study was approximately 540 bp, meaning longer genes were less likely to achieve annotation using an RA approach. In fact, only 19 of the 32 DIGs were identified using this method, and those not identified tended to be longer genes. Additionally, the RA method does not provide any taxonomic information associated with gene counts. Although this annotation pipeline was a custom development for this study, a similar approach has previously been described and applied with success (Brown et al., 2019).

Statistical analysis

Both parametric and non-parametric approaches were used to address the effects of LU change on diazotrophic response variables (DRVs; ANF rate and qPCR *nifH* copy number). A Kruskall-Wallis (KW) rank-sum test paired with Dunn's *post-hoc* analysis (non-parametric approach) was first used on untransformed DRVs to determine significant differences and groupings within and across LU types, applying a Benjamini-Hochberg p-value (pval) correction. When parametric methods were used (with the assumption of a linear response), DRVs were first natural-log [ln()] transformed. To robustly account for spatial heterogeneity of sampling locations before testing for LU significance, we employed variance component analysis (varComp package; Qu and Qu, 2017) and linear mixed-model regression using geospatial data as a random effect in a Matérn function (spaMM package; Rousset, 2017). Environmental pools including C, N, P, Mo, V, Fe, pH, textural class, and metabolites were tested across LUs using non-parametric KW tests as described above. To determine if physicochemical and metabolite profiles differed by LU, distance-based redundancy analysis was performed with log-transformed data. PERMANOVAs tested for the explanatory value of LUs, and a constrained correspondence permutation tested for groupwise (beta-) dispersion. Metabolites were screened individually for association with DRVs with correlation tests (Pearson's r parametric and Spearman's ρ non-parametric) for each LU separately.

To better understand associations between DRVs and soil physicochemical conditions, we applied interaction-based Lasso variable selection with multiple linear regression on natural log-transformed DRVs. Potential explanatory variables included the physicochemical parameters we

measured (see **Supp. Methods**). To account for the spatially heterogeneous sampling scheme, average values across the 0 m and 10 cm sampling points were calculated for each variable at each site, reducing the total sample size from 42 to 36. Correlation plots between DRVs, and physicochemistry were inspected, using the full data set as well as LU-specific data, to detect meaningful relationships visually, and to check collinearity amongst variables. Models were first built using the full dataset for each DRV based on Lasso selection using 'glinternet' in R (Lim and Hastie, 2015). Candidate models were manually refined based on variable significance and compared using ANOVAs, visual fit, and prioritizing reduced variable inclusion. Based on poor unifying fit (especially forest samples), we also built models separately for each LU (on non-transformed values) using AIC-based ordinary least squares model selection (due to small sample size). LU-specific variable selection was performed two ways: with physicochemical variables only, and with the inclusion of the most highly correlated metabolites (positive and negative, based on mean Pearson-Spearman estimates [mPSe]) for each response -LU combination. Best-fit models were again selected using the procedure detailed above, and physicochemical-only and metabolite-included models were compared using the Akaike information criterion (AIC).

Quasi-likelihood negative binomial generalized log-linear models were fit to metagenomic count data for both RA and GTA approaches. Genes with total gene counts within each LU of <10 were excluded from testing. Data were normalized for total library size, and negative binomial and Bayes tagwise dispersion (edgeR package; Robinson et al., 2010) before testing for significant $\log_2$ fold-changes (FC). To identify significant correlations among nitrogenase-encoding genes and the metagenome as a whole stemming from the RA approach, Spearman correlation analysis was run between 13 *nif* genes and all other metagenome genes. The *nif* genes were selected based on consistent FC response, and their essential role in BNF (core subunits and cofactors) to minimize mis-annotation bias. Genome-wide associations were considered meaningful if correlation estimates >|0.72| for at least 7 of the 13 *nif* genes, with p-values (p-vals) < 0.05 after Bonferroni correction. Using data from the GTA approach, read count abundance was tested by taxonomic annotation (aggregated at the class level) between LUs using the edgeR approach above. *nifH* reads were used for count comparison, but only after filtering read count taxa for additional annotation of *nifD* and *nifK*.

Supplemental information and data availability statement

For more detail on ANF experimental methods and rate calculations, quantitative PCR (qPCR) conditions, soil physicochemical and metabolomic protocols, specification of physicochemical parameters considered for full profiling and Lasso variable selection, metagenomic bioinformatic pipelines, the list of DIGs targeted, and spatial statistical analysis, see **Supp. Methods**. Raw and processed metagenomic sequences along with additional information regarding sample processing can be accessed at genome.jgi.doe.gov using GOLD Project IDs Gp0507946- Gp0507987.

Results

DRVs across forest and pasture

Bulk soil $^{15}N_2$-enriched incubations reflected a significant increase in ANF process rates between forest and pasture. Within forest samples, the mean rate was close to zero, at 0.031 $\pm$ 0.45 ng N g soil^{-1} day^{-1} (based on a 14-day incubation period). Pasture soils exhibited considerably higher rates (47.1x), with a mean value of 1.46 $\pm$ 0.34 ng N g soil^{-1} day^{-1}. A KW rank-sum test indicated significant differences in measured rates across the six sites (χ^2 = 22; p-val = 5.6*10^{-4}). *Post-hoc* analysis showed grouping by LU- except for sites F2 and P1 (**Figure 2-1a**), and pre-and post-incubation controls showed minimal experimental drift (**Supp. Figure 2-1a**). After controlling for geographic spatial distance of sample points, LU type was still a significant predictor of rates, based on variance component analysis (F-value [F-val] = 35; p-value [p-val] = 6.4*10^{-7}; **Table 2-1**) and spatial mixed-linear modeling (likelihood ratio χ^2 =10.1; p-val = 1.2*10^{-3}). As expected, the abundance of the *nifH* marker gene was significantly higher in pastures (1.2*10^9 $\pm$ 1.5*10^8) compared to forest (6.7*10^7 $\pm$ 1.5*10^7 copies g^{-1} soil), amounting to an 18-fold difference (χ^2 = 31.4; p-val = 1.6*10^{-5}; **Figure 2-1b**). This was also true on a per ng DNA basis (**Supp. Figure 2-1b**). Variance components analysis revealed that spatial distance (both within and across sites) accounted for 19% of the explanatory power of LU change on *nifH* copy numbers, but that LU was still a significant factor (F-val = 19; p-val = 1.4*10^{-4}; **Table 2-1**). Spatial mixed modeling supported this (likelihood ratio χ^2 =14.9; p-val = 1.1*10^{-4}). While both DRVs coincidently increased in pastures compared

to forest, correlation analysis revealed no relationship between the two for either LU (mPSe = 0.05 for forests and mPSe =0.03 for pastures; **Figure 2-1c**).

Soil chemistry and metabolomic profiles

Of the 18 soil physicochemical parameters (C, and N pools, P, metal cofactors, pH, and texture) we hypothesized to play an important role in diazotrophy, all varied significantly by site (p-val < $1*10^{-2}$) but five did so in an ecologically meaningful way (i.e., grouped within or nearly within LUs in *post-hoc* analysis; **Table 2-2**), and four of these reflected declines in pasture. These included the percent of soil C derived from forest (C_{df}; $\chi^2 = 33.2$; p-val = $3.4*10^{-6}$), the natural abundance of ^{15}N ($\chi^2 = 34.8$; p-val = $2.3*10^{-5}$), soil solution NO_3^- concentration ($\chi^2 = 34.6$; p-value = $2.3*10^{-5}$), and surprisingly, total Mo (KW χ^2 = 26.4; p-val = $2*10^{-4}$). Meanwhile, NH_4^+ concentration increased significantly (KW $\chi^2 = 28.9$; p-val = 10^{-4}). pH also tended to increase with LU change, but Dunn's grouping by LU was not as clear as the above parameters ($\chi^2 = 30.5$; p-val = $6*10^{-5}$). The remaining pools we measured, including total C, POXC, EOC, loss-on-ignition OM, total N, EON, P, V, Fe, as well as clay, silt and sand content, all varied significantly by site (KW $\chi^2 \geq 13.5$; p-val < 10^{-3}), but none grouped by LU in *post-hoc* testing. Overall soil physicochemical profiles, however, varied significantly by LU (F-statistic [F-stat] = 31; p-val = 10^{-3}; **Figure 2-2a; Supp. Table 2-1**).

A total of 180 unique metabolites were identified from bulk soil. Of these, 91 were known compounds, and just 11 out of 180 grouped statistically by LU (eight of which were unidentified; **Supp. Table 2-1**). The three identified compounds included phytol and nonadecanoic acid (lower in pastures), as well as 3-(4-hydroxyphenyl) propionic acid (higher in pastures). Overall metabolome profiles varied by LUs, and more so when unidentified compounds were included (F-stat = 11.2; p-val = 10^{-3}; **Figure 2-2b**). Profiles including just identifiable compounds exhibited more modest, but still significant differences across LUs (F-stat = 20; p-val = $1*10^{-3}$; **Figure 2-2c**). When known compounds were correlated against the two DRVs, we found drastic differences in the metabolites which correlated significantly within forest and pasture samples, and no overlap with those compounds found to differ across LUs. For example, for forest ANF rates, the most positively correlated metabolite, succinic acid (mPSe= 0.47) was one of the metabolites most negatively correlated with pasture ANF rates (mPSe = -0.53) as well as pasture *nifH*

copy numbers (mPSe = -0.46; **Supp. Table 2-2**). We also did not observe any discernible trends in positively vs. negatively correlated compounds when grouped by molecular class (i.e., the structure of potential C and N sources) within and across LUs.

Diazotroph-environmental associations: multiple linear regression

For ln(*nifH* copy number), the best-fit full model (adjusted [adj.] R^2 = 0.86; F-stat = 73; p-val= $2.0*10^{-14}$) included a large LU effect (pos. eff. [positive effect]; F-val = 206; p-val= $1.8*10^{-15}$), a small positive effect (pos. eff.) of clay content (F-val = 8.4; p-val= $6.6*10^{-3}$) as well as an interaction between LU and POXC (LU:POXC; pos. eff.; F-val = 4.7; p-val= $3.6*10^{-2}$; (**Table 2-3**). The model fit was extremely poor for forest samples (**Figure 2-3; Supp. Figure 2-2**). For the ln(ANF rate) full model (adj. R^2 = 0.56; F-stat = 15.9; p-val= $1.6*10^{-6}$), significant best-fit variables again included a significant LU term (negative [neg.] eff.; F-val= 36.6; p-val = $9.4*10^{-7}$), an interaction between LU and the ratio of EOC to EON (pos. eff.; F-val = 9.6; p-val = $4*10^{-3}$), and natural log-transformed NO_3^- (neg. eff; F-val = 6.0; p-val = .03). Similar to the full *nifH* copies model, the ANF model fit was incongruous between forest and pasture samples, with forest again exhibiting trivial fit.

Based on variable selection and goodness of fit for unifying models, we further considered regression on forest and pasture samples separately for each response variable. Crucially, we found that the best models predicting forest *nifH* copies and ANF rate were non-significant (adj. R^2= 0.1 and 0.08, respectively; **Supp. Table 2-2**), while pasture-only models for both *nifH* copies (adj. R^2= 0.66; F-stat= 11.9; p-val= $4*10^{-4}$) and ANF (adj. R^2= 0.26; F-stat= 4.0; p-val= .04; **Table 2-4a, Figure 2-4a**) were significant – albeit modestly significant for ANF. *nifH* copies within pasture soils were predicted by C_{df} (pos. effect; F-val= 16.3; p-val= $1.2*10^{-3}$), POXC (pos. effect; F-val= 15.5; p-val= $1.5*10^{-3}$), and total Mo (F-val= 3.8; p-val= 0.07), which surprisingly had a negative effect estimate. The pasture *nifH* copy model fit considerably better than the pasture ANF rate model (**Supp. Figure 2-3**), which included as predictive variables the ratio of EOC to EON (pos. effect; F-val= 5.4; p-val= 0.03) as well as the ratio of TN to EON (pos. effect; F-val= 2.5; p-val= 0.057).

To determine the potential utility of metabolite data in predicting DRVs, the most positively and negatively correlated metabolites for each DRV -LU combination were included in variable selection

procedures, improving the best-fit forest *nifH* copy model (adj. R^2= 0.81; F-stat = 24.9; p-val = $4*10^{-7}$): Combined clay and silt content (pos. effect; F-val= .12.1; pval= $1*10^{-3}$), as well as TC (neg. effect; F-val= 4.3; p-val= 0.05) together contributed a minor amount of explanatory value compared to the metabolite β-sitosterol (pos. effect; F-val= 58.5; p-val= $2.9*10^{-6}$), a ring carbon (**Table 2-4b; Figure 2-4b**). In contrast, inclusion of top-correlating metabolites did not add significant explanatory value to the best-fit pasture *nifH* copies model. The fit and significance of the pasture ANF model were improved substantially (AIC = 57 vs. 62.8 as compared with physicochemical parameters alone; adj. R^2= 0.70; F-stat= 14.2; p-val= $1.6*10^{-4}$) by including the effect by erythrose, a positively correlated monosaccharide (F-val= 25.2; p-val= $1.8*10^{-4}$) and uridine diphosphate-N-acetylglucosamine (UDP-NAG), a negatively correlated nucleotide sugar (F-val= 6.9; p-val=0.02), along with the physicochemical ratio TN/EON (F-val= 10.4 p-val= $6*10^{-3;}$ **Supp. Figure 2-3**). Including metabolites in variable selection still failed to produce a significant forest ANF model.

Quantifying diazotrophs in the soil metagenome: RA and GTA methods

On average, sample libraries produced 49 $\pm$ 2 M quality-filtered reads each, of which 15 $\pm$ 0.5% were annotated by the KEGG database, and a total of 5,152 genes were identified using the RA approach. Of these, 8.5% were enriched in forests, 12.7% were enriched in pastures, 54.3% showed no difference, and 24.5% fell below the abundance threshold for statistical testing. Of targeted DIGs, 19 out of the 32 were identified, averaging 24.8 $\pm$ 1.23 counts per million (CPM). Of these, 13 were highly enriched in pastures. When these genes were ranked by log_2fold change (FC), all but *nifJ* (FC = 2.3 p-val = $2.9*10^{-13}$) fell in the top 60 (i.e., top 7%), and *nifE* (FC = 9.8; p-val = $1.3*10^{-7}$), *nifD* (FC = 9.3; p-val = $1.3*10^{-7}$), and nif*N* (FC = 9.1; p-val = $1.1*10^{-6}$) fell in the top 10 (**Figure 2-5a**). The regulatory and transferase genes *nifA* and *nifS* did not shift in abundance. Notably, *nod* genes (indicating symbiotic diazotrophs) also showed no difference across LUs, and genes encoding alternative nitrogenases (*anf*, *vnf*) were not identified. Increased genetic potential for ANF in pastures was coincident (but not necessarily correlated) with shifts in other N cycling genes, including a decrease in ammonia monooxygenase (*amoA*) and nitrate reductase (*napA*), as well as increased hydroxylamine reductase (*hcp*) and nitrite reductase (nirS/*nirK*) protein subunits (**Supp. Figure 2-4**).

Quantifying genes using the GTA method identified all 32 DIGs queried and produced largely consistent results in terms of the directionality of significant FCs (**Figure 2-5b; Supp. Table 2-4**), with a few exceptions. First, *nifV*, did not reflect pasture enrichment, in contrast to the RA method. Additionally, GTA quantification indicated an increase in *nodC* (FC = 3.65; p-val =$1.3*10^{-2}$), a plant-targeting nodulation signal, and that two *vnf* (i.e., vanadium-based metallocluster subunit) genes, *vnfD* (FC = 3.8; p-val = $1.6*10^{-2}$) and *vnfH* (FC = 2.4; p-val = $1.6*10^{-2}$) increased in pastures, but counts (.03 $\pm$.008 CPM) were much lower as compared to *nif* (9.1 $\pm$.9 CPM) subunits. Additionally, using taxonomic annotations of *nifH* reads (*nifHDK* positive taxa only), we found 22 of 25 (88%) identified taxonomic classes increased significantly in pasture, with the greatest FC observed in Desulfobacteria (FC = 5.6; p-val = $2*10^{-8}$; **Supp. Figure 2-5**).

Soil diazotrophs and co-occurring functions

Across metagenomes, 78 genes were found to be significantly correlated with *nif* genes in terms of read counts (**Supp. Table 2-5**), all of which exhibited positive relationships, and were largely driven by pasture metagenomes. Several of these genes or gene sets were of notable function (**Figure 2-6**). For example, maltose 6' phosphate phosphatase (*mapP*) was the metagenome-gene most highly correlated with *nif* genes (driven by pasture metagenomes), with a median Spearman estimate (mSe) coefficient of 0.90 across the 13 *nif* protein subunit and cofactor genes tested. The *mapP* gene was also found to exhibit the greatest FC in pasture soils of all 5,152 genes identified (FC = 10.3; p-val =$9.5*10^{-15}$). NAD+ N_2 reductase ribosyltransferase (*draT*), involved in post-transcriptional regulation of the nitrogenase enzyme, was also highly correlated (mSe = 0.88) and in the top 10 most enriched pasture genes (FC = 9.36; p-val = $7.9*10^{-13}$). Additional highly correlated genes associated with dissimilatory environmental nutrient cycling included nitrite reductase (*nirF*, mSe = .83; *nirS*, mSe = 0.78; *nirC*, mSe =0.75), anaerobic sulfite reductase (*asrB*, mSe = 0.76; *dsrC*, mSe = 0.74; *hydD*, mSe = 0.83), and anaerobic CO dehydrogenase (*cooF*; mSe =0 .83). An additional 12 highly-correlated metagenome genes (such as *cheD*; mSe = 0.83) were associated with chemotaxis and flagellation. Three genes encoding choline transport (CTR; mSe = 0.77), and two N, N'-diacetylchitobiose transport permease proteins (*dasA* and

dasB; mSe = 0.74 and 0.72, respectively) were the only correlating genes not found in diazotroph genomes which are currently available in the KEGG database (**Supp. Table 2-5**).

Discussion

Nitrogen inputs to pasture soils through ANF and relationship with community size

The primary aim of this study was to establish whether soil ANF rates increase in pastures of Rondônia, Brazil, based on previous observations of drastic increases in the potential diazotroph community size (Mirza et al., 2014). Isotopically labeled soil incubations revealed near-zero rates (0.031 $\pm$ 0.45 ng N g soil^{-1} day^{-1}) of ANF in forest soils compared to an average of 1.46 $\pm$ 0.34 ng N g soil^{-1} day^{-1} in pastures (**Figure 2-1a**). Tropical forest soils are considered replete with N, and the role of ANF in sustaining these high inorganic N concentrations, given considerable NO_3^- leaching, has been an important line of inquiry (Hedin et al., 2009). Utilizing a constrained mass-balance approach, Cleveland et al. (2009) estimated total BNF inputs range from 1-8 kg N ha^{-1} yr^{-1} in Rondônian primary forests as a combination of tree canopy, litter layer, root nodule, and surface soil inputs. Although derived from one timepoint, our assay measurements, in conjunction with gene abundances and read counts, suggest surface soil ANF contributes minimally or infrequently to forest N inputs. Indeed, a recent study in the state of Pará (Moreira et al., 2021) found that tree canopy nitrogen fixation (via acetylene reduction) was more than 10-fold that of soil (0.3 vs. 0.03 kg N ha^{-1} yr^{-1}). Another study in the Southeastern Amazon also measured extremely low activity, particularly in the dry season (Wong et al., 2019). In contrast, an LU-change study in Brazil's Atlantic forest reported appreciable ANF rates in forest soils (median 1.25 ng N g soil^{-1} h^{-1}), with no overall difference between forest and pasture (Bomfim et al., 2018). Regional variation (e.g., soil morphology) between the two basins may explain this discrepancy.

The 47x increase in ANF and 18x increase in *nifH* copy number we observed in active cattle pastures of Rondônia indicate a major shift in the pathways of N inputs with LU change. Assuming a soil bulk density of 1.31 g cm^{-3} (Neill et al., 1995), annual N inputs via ANF may average 0.69 kg N ha^{-1}, with a maximum estimate around 3.3 kg N ha^{-1} (based on the maximum rate). We used a 14-day incubation

period (Buckley et al., 2007) to ensure sufficient labeling, but it is plausible that the majority of activity occurred in as little as the first 24 h, implying our average rate corresponds to a potential of 9.75 kg N ha^{-1} yr^{-1}. Despite this uncertainty, our range of annualized values generally agrees with grassland rates reported elsewhere. For example, a meta-analysis by Reed et al. (2011) estimated annual average inputs of 4.7 kg N ha^{-1} across temperate grasslands. Keuter et al. (2014; via acetylene reduction), using a temporally robust sampling scheme, calculated a range of 1.7-5.7 kg N ha^{-1} yr^{-1}, depending on the effects of fertilization, seasonality, and mowing frequency (analogous to pasture grazing) in a grassland of Lower Saxony, Germany. This latter study, along with others (e.g., Moreira et al., 2021; Smercina et al., 2021), highlights the significance of seasonality in dictating ANF rates due to fluctuations in moisture, temperature, and productivity (Smercina et al., 2019). Temperature and day length have minimal seasonality in Rondônia (~-10°S), but precipitation varies considerably; therefore, to robustly quantify N inputs from ANF, more temporal sampling, particularly during the peak wet and dry seasons, is needed.

While we have demonstrated the occurrence of ANF in lightly stocked active cattle pastures approximately 45 years in age, the reality of the region is that a high proportion of pastures are abandoned on the order of years rather than decades, due to major declines in grass productivity as a result of N limitation (Davidson et al., 2012). In contrast to our findings within an active pasture, a clay-box mesocosm experiment in Pará, Brazil estimated ANF- N inputs in a degraded pasture compared to an adjacent secondary forest (Davidson et al., 2018), and concluded that little, if any, ANF was occurring in the abandoned pasture, an observation potentially rooted in grazing pressure. In support of this, a study of N status in a mixed clover - *U. brizantha* pasture under variable stocking rates identified a threshold whereby legume inputs could not sustain forage productivity (Cantarutti et al., 2002), perhaps related to disproportionate increases in animal excretion- related N losses with increased stocking density (Boddey et al., 2004). On the other hand, the mowing study discussed above (Keuter et al., 2014) indicated that in the absence of fertilizer, low-intensity vegetation removal resulted in the highest annualized ANF estimates, possibly driving a flush of below-ground C allocation (Holland et al., 1996). As such, a balance between ANF inhibition by overgrazing, and stimulation by low levels of herbivory may influence grazing sustainability in the absence of fertilizer inputs.

LU shift in ANF linked to nitrate, but DRVs are poorly explained in forests

Another aim of this study was to identify credible causative associations between soil physicochemistry and the drastic shifts observed in DRVs across the LU dichotomy. We initially expected that changes to C, N, and micronutrient pools would help explain these observations. However, just a few measurements reflected significant LU alteration, including declines in C_{df}, $\delta^{15}N$, total inorganic N, and its partitioning (from NO_3^- to NH_4^+- dominated), but not TN or EON (**Table 2-2**). Through variable selection and model comparison, we found the strongest explanatory variable for both DRVs to be a categorical LU effect (0.81 and 0.43 of the explanatory power for ln(*nifH* copies) and ln(ANF), respectively; **Table 2-3**) which is largely indicative of complex physicochemical interactions that we did not identify. For LU-wise increases in ln(*nifH* copies), the only non-interactive physicochemical variable selected was clay, which offered extremely low explanatory power (0.02). We also found that $\ln(NO_3^-)$ was selected as a negative predictor (though again, with small unique explanatory power: 0.06) of the increase in ln(ANF rate) with LU change, suggesting that indeed, drastic reduction in the availability of inorganic N (the bulk of which was NO_3^- in forests) stimulates ANF activity in these pastures (Piccolo et al., 1994). The fact that increased NH_4^+ content did not add a significant negative effect likely relates to active N_2 fixing diazotrophs contributing to NH_4^+ pools as well as drawing from them (Chapin et al., 2002). Additionally, the forage species predominantly planted throughout Amazonian pastures (e.g. *Urochloa*), which are adapted to low-N soils, exude biological nitrification inhibitors from their roots (Subbarao et al., 2009), effectively preserving a large proportion of total inorganic N as NH_4^+ in pastures. In turn, these forage grasses may scavenge NH_4^+ more efficiently than soil microbial communities.

In graphing the two DRV regressions, we found forests were poorly predicted (**Figure 2-3a, b),** with trivial model fit (**Supp. Figure 2-2**). Attempting to identify variables explaining DRVs in forest soils only did not produce significant models (gray shading, **Figure 2-4a, Supp. Table 2-3**), with a parsimonious explanation being that we simply did not measure the variables driving forest DRVs (consistent with the large LU effect). Yet, additional factors may be in play. Potential diazotrophs assume many trophic strategies (Le Boulch et al., 2019), so community size in forests may not be driven by conditions favoring diazotrophs specifically. Also, forest ANF rates remained close to zero, hence activity

could be frequently suppressed by a large reactive N pool (Neill et al., 1997b), in turn fed by more efficient avenues of BNF such as canopy lichens (Hedin et al., 2009; Moreira et al., 2021).

Pasture DRVs are associated with physical, chemical, and isotopic properties

Potential diazotroph community abundance: First, we found that POXC commanded substantial explanatory effect (.29) over *nifH* copies in pastures (**Table 2-4a**). POXC is generally considered a proxy for C that can be readily utilized by microbes (Huang et al., 2021), consistent with our initial expectation that LMW C is key in fueling ANF. It was somewhat curious that this association was not observed in forest soils as well, independent of whether *nifH*-positive taxa perform ANF. On closer inspection, however, the lack of relationship appeared to be driven by three points strongly disagreeing with an overall positive trend.

Additionally, while the full *nifH* copy model fit most optimally with clay content, the pasture-only model selected for C_{df}, with a somewhat unexpected positive effect. We also found that these two parameters were moderately correlated (mPSe = 0.66), so they may synergistically influence *nifH* communities. An increase in clay content is positively associated with the volume of hypoxic or anoxic microsites (Keiluweit et al., 2018), and since most diazotrophs require extremely low exogenous O_2 for BNF (Khadem et al., 2010), inhabiting frequently O_2- depleted textural sites would be beneficial, especially for taxa genetically equipped for anaerobic respiration (e.g., sulfite/sulfate reducers; Le Boulch et al., 2019). The positive association between pasture *nifH* copies and C_{df} may be driven by the fact that endurance of C_{df} mainly occurs in the clay-sized fraction after LU change, since physical isolation and anaerobic microsites presumably provide some buffer against SOM turnover (Keiluweit and Fendorf, 2016). This seems contradictory to the equivalent explanatory effect of POXC (0.3 vs 0.29), but could be because clay also stores the greatest share of pasture-derived C (Desjardins et al., 2004), and association with clay does not explicitly imply recalcitrance (Kleber et al., 2011). Another interpretation is that both C indicators benefit populations in different soil compartments. We did not measure *nifH* abundance in rhizosphere vs. bulk soil, but the dense, shallow root system of *Urochloa* (Guenni et al., 2004) suggests sampling from 0-10 cm could capture both. Therefore, a positive association with POXC

may benefit rhizosphere diazotrophs, while C_{df} could represent greater access to C amongst clay-associated communities.

Soil Mo concentration also added a modest (0.07) explanatory effect to the prediction of pasture *nifH* copies. It is a key micronutrient due to its role in metallocluster formation (Reed et al., 2011), but it is disputed whether it limits diazotrophs in the Amazon (Moreira et al., 2021; Wong et al., 2021). We found that Mo largely decreased in pastures (**Table 2-2**), and its effect on *nifH* copies was actually *negative*. The total pool we measured may underrepresent the bioavailable fraction given a slight rise in pasture pH, but the explicitly negative effect we observed is unclear, potentially pointing to unknown collinear parameters.

Asymbiotic nitrogen fixation: Pasture ANF was best explained by a model with just two physicochemical parameters. First, the EOC:EON ratio served as a positive predictor, agreeing with our hypothesis regarding drivers of activity rates (**Table 2-4a**). It is noteworthy that EOC, presumably representing a fraction (42% ± 20%) of POXC based on extraction protocols, is only a relevant predictor of ANF rate when ratioed against EON. While EON does not explicitly proxy potentially mineralizable N in a monoculture, its concentration likely scales with inorganic and LMW N sources for microbial assimilation. To boot, diazotrophs prioritize the uptake of these N forms before investing in the nitrogenase enzyme (Norman and Friesen, 2017). The inverse impact of the proportion of total N in the dissolved organic form suggests that pools of potentially mineralizable, as well as directly utilizable reactive N stimulate the enzymatic investment in this energy-intensive reaction. Additionally, the importance of the soil solution EOC to EON ratio points to an incentive to execute ANF by means of ample energetic resources and limited alternative N. We found one corroborating marine-based study observing the same association (Hanson and Gundersen, 1977), and curiously found that Mirza et al. (2014) identified this ratio to be highly associated with *nifH* abundance. Overall, however, the explanatory value of these variables was modest. The large proportion of variation left unexplained is perhaps attributable to microsite conditions during the assay incubation period, such as oxygen concentration and distribution of dissolved nutrients.

Potential for the predictive power of metabolites

The role of plant metabolites in moderating microbial community structure and activity is increasingly being investigated (e.g., Guyonnet et al., 2017; Jacoby et al., 2020). Here, we asked whether these LMW compounds could serve as indicators of DRVs across LUs. Although metabolomic profiles changed drastically across LUs (**Figure 2-2a; Supp Table 2-1**), they did not predict DRVs (not shown). By directly correlating individual (known) metabolites with DRVs, we found a limited number of candidates for model variable selection (top correlates in **Supp Table 2-2**). In some cases, such as for forest *nifH* copies and pasture ANF, the inclusion of highly correlated metabolites did improve model R^2 (**Table 2-4b, Figure 2-4b**). The model of pasture ANF in particular was substantially improved by inclusion of erythrose, a LMW (120 g mol^{-1}) monosaccharide which could plausibly serve as an easily-utilized source of C, as well as UDP-NAG (651 g mol^{-1}), a potential source of organic N which could be prioritized over atmospherically- reduced N (Norman and Friesen, 2017). However, we cannot speak to the temporal stability of these associations. At least one study has utilized soil metabolomics for identifying selective pathways of community structure (RoyChowdhury et al., 2022), but based on our original aim, we concluded its application is hindered by a high proportion of unidentified compounds, and likely requires dense temporal sampling (Withers et al., 2020).

Diazotrophic genes are more abundant in the soil metagenome

Our final major aim was to robustly profile potential diazotrophs within the broader soil community across the LU dichotomy. Querying a full suite of DIGs across metagenomes using both RA and GTA analyses bolstered our conclusions regarding consistent and significant enrichment of diazotrophic soil microbes, agreeing with the qPCR results discussed above, as well as those presented previously (Mirza et al., 2014). RA analysis contextualizes this enrichment against all other metagenome genes, revealing that nitrogenase metallocluster subunits and cofactor biosynthesis-encoding genes (n=13), most of which are essential for nitrogen fixation (Dos Santos et al., 2012), fall in the top 7% of differentially-enriched pasture genes based on FC ranking (**Figure 2-5a, 2-6**). The inclusion of DIGs such as those encoding *nod* factors provided additional insight regarding shifts in diazotrophic life strategies with LU conversion. We expected that nodulation-initiation genes would decline in pastures, given the loss of symbiotic plant

hosts, yet *nod* genes were found in low abundance (i.e., no LU effect) in both annotation pipelines, except for *nodC* (encidong a plant targeting signal) in GTA analysis (**Figure 2-5b; Supp. Table 2-4**). Previous survey of viable symbiotic diazotrophs across the LU dichotomy demonstrated pasture isolates readily colonize host plants (Lima et al., 2009), suggesting they sustain an alternative ecological niche under LU change.

Multi-gene correlation analysis presents a unique method of identifying genes and microbial functions which co-occur significantly with DIGs, either within diazotrophic genomes themselves, or in closely associated organisms. The metagenome-gene with the highest median correlation (and largest overall FC change in pasture) was *mapP* (**Figure 2-6**), which encodes the novel intracellular enzyme maltose 6'-P phosphatase. This enzyme plays a role in maltose (and some other disaccharide) metabolism, cleaving alpha linkages into two glucose molecules as part of ATP production (Mokhtari et al., 2013). Therefore, it may play an important role in supporting the energy demands of biogeochemical processes such as ANF in pastures (as it was not identified in forests; Mokhtari et al., 2013). Another strongly correlated gene was *draT*, encoding a ribosyltransferase that reversibly regulates nitrogenase post-translationally (Huergo et al., 2012). Its general prevalence amongst diazotrophic genomes is poorly understood (Huergo et al., 2012), but we identified it to be present in 16% of those diazotrophic genomes available in the KEGG database. With the sixth highest ranking by FC, *draT* enrichment and correlation with *nif* genes adds further support to the magnitude of diazotrophic enrichment in pastures. It may also indicate a specific favoring of diazotrophic taxa known to employ this regulatory system, such as *Azospirillium, Rhodopseudomonas, Geobacter, and Methylobacter (Huergo et al., 2012)*. Genes associated with flagellation and chemotaxis (e.g., *cheD*) comprised an additional 15% of the 78 total nitrogenase-correlating genes (**Supp. Table 2-5**). The ability to translocate in response to exogenous cues such as O_2 or nutrient concentration is crucial to diazotrophs of all life strategies (Mandimba et al., 1986; Scharf et al., 2016), and while nearly 60% of diazotrophic genomes in the KEGG database bear the *cheD* gene, just 23% of all organisms with the *cheD* gene are also diazotrophic (Aramaki et al., 2020). This could suggest co-occurrence of nitrogen fixation and chemotaxis in the rhizosphere, enabling movement towards energy-rich exudates (Scharf et al., 2016). Multiple genes encoding dissimilatory nitrite (*nirS,C,F*) and sulfite reductase (*asrB, dsrC*) were also correlated. These genes are moderately

common among diazotrophic genomes (20-45%), and strong co-occurrence of these gene sets suggests a link between ANF, denitrification, and/or sulfur cycling through anaerobic energy conservation (Lukat et al., 2008; Shirodkar et al., 2011).

Conclusions

Our finding that active cattle pastures in Rondônia, Brazil exhibit significant increases in soil asymbiotic N_2 fixation, coupled with the insight that this biogeochemical pathway is one of the most, if not *the* most enriched microbial functions with LU change, speaks to an immense pressure for N replenishment under cattle grazing without fertilizer inputs. Yet, the rate of ANF within LUs was completely independent of potential diazotroph community abundance, likely related to the variation in trophic strategies and nano-scale soil conditions driving the metabolic activity of diazotrophs. While a reduction in the available pool of NO_3^- across LUs accounted for minor predictive power over ANF rate changes with pasture conversion, our general inability to identify variables 'bridging the gap' in N_2 activity across these LU types likely indicates unaccounted-for measurements like fluxes in N and LMW C cycles, complex physicochemical interactions, or 'tipping points' in nutrient concentration whereby biogeochemical cycling shifts dramatically. *Within* pastures however, we found that ANF rate, was positively associated with an increasing ratio of microbially-available C to N in soil solution. This is worthy of further investigation, given that a sizeable proportion of pastures in the Amazon quickly lose productivity, and may be abandoned for further deforestation, in many cases likely a result of N limitation. Gaining further insight to the interaction between forage grass rhizosphere activity, land management practices like grazing pressure, and ANF may aid in a better understanding how natural N fertilization may be stimulated in these agricultural systems.

Funding

This work was supported by the National Science Foundation – Dimensions of Biodiversity (DEB-14422214) grant, a USDA National Institute of Food and Agriculture predoctoral fellowship [project no. CA-D-LAW-2527-CG/accession no. 1019332], and The Joint Genome Institute (proposal: 504104) for sequencing services. Metagenomic work was conducted by the U.S. Department of Energy Joint Genome

Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, and is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Acknowledgments

The authors thank the owners and staff of Agropecuaria Nova Vida for logistic support and permission to work on their property. We are grateful to Alexandre Pedrinho and Wagner Piccinini for their assistance with the fieldwork.

References

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. d. M., and Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* **22**, 711-728.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252.

Boddey, R. M., Macedo, R., Tarré, R. M., Ferreira, E., de Oliveira, O. C., de P. Rezende, C., Cantarutti, R. B., Pereira, J. M., Alves, B. J. R., and Urquiaga, S. (2004). Nitrogen cycling in Brachiaria pastures: the key to understanding the process of pasture decline. *Agriculture, Ecosystems & Environment* **103**, 389-403.

Bomfim, B., Silva, L. C. R., Doane, T. A., and Horwath, W. R. (2018). Interactive effects of land-use change and topography on asymbiotic nitrogen fixation in the Brazilian Atlantic Forest. *Biogeochemistry* **142**, 137-153.

Boyle, S. A., Yarwood, R. R., Bottomley, P. J., and Myrold, D. D. (2008). Bacterial and fungal contributions to soil nitrogen cycling under Douglas fir and red alder at two sites in Oregon. *Soil Biology and Biochemistry* **40**, 443-451.

Brown, S. M., Chen, H., Hao, Y., Laungani, B. P., Ali, T. A., Dong, C., Lijeron, C., Kim, B., Wultsch, C., Pei, Z., and Krampis, K. (2019). MGS-Fast: Metagenomic shotgun data fast annotation using microbial gene catalogs. *Gigascience* **8**.

Buckley, D. H., Huangyutitham, V., Hsu, S.-F., and Nelson, T. A. (2007). Stable isotope probing with 15N2 reveals novel noncultivated diazotrophs in soil. *Applied and Environmental Microbiology* **73**, 3196-3204.

Cantarutti, R., Tarré, R., Macedo, R., Cadisch, G., de Rezende, C., Pereira, J., Braga, J., Gomide, J., Ferreira, E., and Alves, B. (2002). The effect of grazing intensity and the presence of a forage legume on nitrogen dynamics in Brachiaria pastures in the Atlantic forest region of the south of Bahia, Brazil. *Nutrient Cycling in Agroecosystems* **64**, 257-271.

Chapin, F. S., Matson, P. A., Mooney, H. A., and Vitousek, P. M. (2002). Principles of terrestrial ecosystem ecology.

Cleveland, C. C., Houlton, B. Z., Neill, C., Reed, S. C., Townsend, A. R., and Wang, Y. (2009). Using indirect methods to constrain symbiotic nitrogen fixation rates: a case study from an Amazonian rain forest. *Biogeochemistry* **99**, 1-13.

Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., Perakis, S. S., Latty, E. F., Von Fischer, J. C., Elseroad, A., and Wasson, M. F. (1999). Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. *Global Biogeochemical Cycles* **13**, 623-645.

Culman, S. W., Snapp, S. S., Freeman, M. A., Schipanski, M. E., Beniston, J., Lal, R., Drinkwater, L. E., Franzluebbers, A. J., Glover, J. D., and Grandy, A. S. (2012). Permanganate oxidizable carbon reflects a processed soil fraction that is sensitive to management. *Soil Science Society of America Journal* **76**, 494-504.

Danielson, R. E., and Mazza Rodrigues, J. L. (2022). Impacts of land-use change on soil microbial communities and their function in the Amazon Rainforest. Vol. 175, pp. 179-258.

Davidson, E. A., de Araujo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., MM, C. B., Coe, M. T., DeFries, R. S., Keller, M., Longo, M., Munger, J. W., Schroeder, W., Soares-Filho, B. S., Souza, C. M., Jr., and Wofsy, S. C. (2012). The Amazon basin in transition. *Nature* **481**, 321-8.

Davidson, E. A., Markewitz, D., de O. Figueiredo, R., and de Camargo, P. B. (2018). Nitrogen Fixation Inputs in Pasture and Early Successional Forest in the Brazilian Amazon Region: Evidence From a Claybox Mesocosm Study. *Journal of Geophysical Research: Biogeosciences* **123**, 712-721.

Desjardins, T., Barros, E., Sarrazin, M., Girardin, C., and Mariotti, A. (2004). Effects of forest conversion to pasture on soil carbon content and dynamics in Brazilian Amazonia. *Agriculture, Ecosystems & Environment* **103**, 365-373.

Dias-Filho, M. B., Davidson, E. A., and Carvalho, C. (2001). Linking Biogeochemical Cycles to Cattle Pasture Management and Sustainability. *The biogeochemistry of the Amazon Basin.*, 84-105.

Dos Santos, P., Fang, Z., Mason, S., Setubal, J., and Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**.
Gaby, J. C., and Buckley, D. H. (2012). A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. *PLoS One* **7**, e42149.
Gaby, J. C., and Buckley, D. H. (2017). The use of degenerate primers in qPCR analysis of functional genes can cause dramatic quantification bias as revealed by investigation of *nifH* primer performance. *Microb Ecol* **74**, 701-708.
Guenni, O., Baruch, Z., and Marín, D. (2004). Responses to drought of five Brachiaria species. II. Water relations and leaf gas exchange. *Plant and soil* **258**, 249-260.
Guyonnet, J. P., Vautrin, F., Meiffren, G., Labois, C., Cantarel, A. A. M., Michalet, S., Comte, G., and Haichar, F. E. Z. (2017). The effects of plant nutritional strategy on soil microbial denitrification activity through rhizosphere primary metabolites. *FEMS Microbiol Ecol* **93**.
Hanson, R. B., and Gundersen, K. (1977). Relationship between nitrogen fixation (acetylene reduction) and the C: N ratio in a polluted coral reef ecosystem, Kaneohe Bay, Hawaii. *Estuarine and coastal marine science* **5**, 437-444.
Hedin, L. O., Brookshire, E. J., Menge, D. N., and Barron, A. R. (2009). The nitrogen paradox in tropical forest ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **40**, 613-635.
Holland, J. N., Cheng, W., and Crossley, D. (1996). Herbivore-induced changes in plant carbon allocation: assessment of below-ground C fluxes using carbon-14. *Oecologia* **107**, 87-94.
Hood-Nowotny, R., Umana, N. H.-N., Inselbacher, E., Oswald-Lachouani, P., and Wanek, W. (2010). Alternative methods for measuring inorganic, organic, and total dissolved nitrogen in soil. *Soil Science Society of America Journal* **74**, 1018-1027.
Houlton, B. Z., Wang, Y. P., Vitousek, P. M., and Field, C. B. (2008). A unifying framework for dinitrogen fixation in the terrestrial biosphere. *Nature* **454**, 327-30.
Huang, J., Rinnan, Å., Bruun, T. B., Engedal, T., and Bruun, S. (2021). Identifying the fingerprint of permanganate oxidizable carbon as a measure of labile soil organic carbon using Fourier transform mid-infrared photoacoustic spectroscopy. *European Journal of Soil Science* **72**, 1831-1841.
Huergo, L. F., Pedrosa, F. O., Muller-Santos, M., Chubatsu, L. S., Monteiro, R. A., Merrick, M., and Souza, E. M. (2012). PII signal transduction proteins: pivotal players in post-translational control of nitrogenase activity. *Microbiology (Reading)* **158**, 176-190.
Jacoby, R. P., Chen, L., Schwier, M., Koprivova, A., and Kopriva, S. (2020). Recent advances in the role of plant metabolites in shaping the root microbiome. *F1000Res* **9**.
James, E. K., Gyaneshwar, P., Barraquio, W. L., Mathan, N., and Ladha, J. K. (2000). Endophytic diazotrophs associated with rice. *The quest for nitrogen fixation in rice*, 119-140.
Jank, L., Barrios, S. C., do Valle, C. B., Simeão, R. M., and Alves, G. F. (2014). The value of improved pastures to Brazilian beef production. *Crop and Pasture Science* **65**.
Kaschuk, G., and Hungria, M. (2017). Diversity and importance of diazotrophic bacteria to agricultural sustainability in the tropics. *In* "Diversity and Benefits of Microorganisms from the Tropics", pp. 269-292. Springer.
Keiluweit, M., and Fendorf, S. (2016). Texture-dependent anaerobic microsites constrain soil carbon oxidation rates. *In* "EGU General Assembly Conference Abstracts".
Keiluweit, M., Gee, K., Denney, A., and Fendorf, S. (2018). Anoxic microsites in upland soils dominantly controlled by clay content. *Soil Biology and Biochemistry* **118**, 42-50.
Keuter, A., Veldkamp, E., and Corre, M. D. (2014). Asymbiotic biological nitrogen fixation in a temperate grassland as affected by management practices. *Soil Biology and Biochemistry* **70**, 38-46.
Khadem, A. F., Pol, A., Jetten, M. S. M., and Op den Camp, H. J. M. (2010). Nitrogen fixation by the verrucomicrobial methanotroph 'Methylacidiphilum fumariolicum' SolV. *Microbiology (Reading)* **156**, 1052-1059.
Kleber, M., Nico, P. S., Plante, A., Filley, T., Kramer, M., Swanston, C., and Sollins, P. (2011). Old and stable soil organic matter is not necessarily chemically recalcitrant: implications for modeling concepts and temperature sensitivity. *Global Change Biology* **17**, 1097-1107.
Lammel, D. R., Feigl, B. J., Cerri, C. C., and Nusslein, K. (2015). Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. *Front Microbiol* **6**, 1057.

Le Boulch, M., Dehais, P., Combes, S., and Pascal, G. (2019). The MACADAM database: a MetAboliC pAthways DAtabase for Microbial taxonomic groups for mining potential metabolic capacities of archaeal and bacterial taxonomic groups. *Database (Oxford)* **2019**.
Lim, M., and Hastie, T. (2015). Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics* **24**, 627-654.
Lima, A. S., Nóbrega, R. S. A., Barberi, A., da Silva, K., Ferreira, D. F., and Moreira, F. M. d. S. (2009). Nitrogen-fixing bacteria communities occurring in soils under different uses in the Western Amazon Region as indicated by nodulation of siratro (Macroptilium atropurpureum). *Plant and Soil* **319**, 127-145.
Lukat, P., Rudolf, M., Stach, P., Messerschmidt, A., Kroneck, P. M., Simon, J., and Einsle, O. (2008). Binding and reduction of sulfite by cytochrome c ntrite reductase. *Biochemistry* **47**, 2080-2086.
Mandimba, G., Heulin, T., Bally, R., Guckert, A., and Balandreau, J. (1986). Chemotaxis of free-living nitrogen-fixing bacteria towards maize mucilage. *Plant and Soil* **90**, 129-139.
Mirza, B. S., Potisap, C., Nusslein, K., Bohannan, B. J., and Rodrigues, J. L. (2014). Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. *Appl Environ Microbiol* **80**, 281-8.
Mokhtari, A., Blancato, V. S., Repizo, G. D., Henry, C., Pikis, A., Bourand, A., de Fatima Alvarez, M., Immel, S., Mechakra-Maza, A., Hartke, A., Thompson, J., Magni, C., and Deutscher, J. (2013). Enterococcus faecalis utilizes maltose by connecting two incompatible metabolic routes via a novel maltose 6'-phosphate phosphatase (MapP). *Mol Microbiol* **88**, 234-53.
Moreira, J. C. F., Brum, M., de Almeida, L. C., Barrera-Berdugo, S., de Souza, A. A., de Camargo, P. B., Oliveira, R. S., Alves, L. F., Rosado, B. H. P., and Lambais, M. R. (2021). Asymbiotic nitrogen fixation in the phyllosphere of the Amazon forest: Changing nitrogen cycle paradigms. *Sci Total Environ* **773**, 145066.
Neill, C., Melillo, J. M., Steudler, P. A., Cerri, C. C., de Moraes, J. F., Piccolo, M. C., and Brito, M. (1997a). Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern Brazilian Amazon. *Ecological Applications* **7**, 1216-1225.
Neill, C., Piccolo, M. C., Cerri, C. C., Steudler, P. A., Melillo, J. M., and Brito, M. (1997b). Net nitrogen mineralization and net nitrification rates in soils following deforestation for pasture across the southwestern Brazilian Amazon Basin landscape. *Oecologia* **110**, 243-252.
Neill, C., Piccolo, M. C., Steudler, P. A., Melillo, J. M., Feigl, B. J., and Cerri, C. C. (1995). Nitrogen dynamics in soils of forests and active pastures in the western Brazilian Amazon Basin. *Soil Biology and Biochemistry* **27**, 1167-1175.
Neill, C., Steudler, P. A., Garcia-Montiel, D. C., Melillo, J. M., Feigl, B. J., Piccolo, M. C., and Cerri, C. C. (2005). Rates and controls of nitrous oxide and nitric oxide emissions following conversion of forest to pasture in Rondônia. *Nutrient Cycling in Agroecosystems* **71**, 1-15.
Norman, J. S., and Friesen, M. L. (2017). Complex N acquisition by soil diazotrophs: how the ability to release exoenzymes affects N fixation by terrestrial free-living diazotrophs. *ISME J* **11**, 315-326.
Pedrinho, A., Mendes, L. W., Merloti, L. F., Andreote, F. D., and Tsai, S. M. (2020). The natural recovery of soil microbial community and nitrogen functions after pasture abandonment in the Amazon region. *FEMS Microbiol Ecol* **96**.
Piccolo, M. C., Neill, C., and Cerri, C. C. (1994). Net nitrogen mineralization and net nitrification along a tropical forest-to-pasture chronosequence. *Plant and soil* **162**, 61-70.
Poly, F., Monrozier, L. J., and Bally, R. (2001). Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil. *Research in microbiology* **152**, 95-103.
Qiu, X.-c., Liu, G.-P., and Zhu, Y.-Q. (1987). Determination of water-soluble ammonium ion in soil by spectrophotometry. *Analyst* **112**, 909-911.
Qu, L., and Qu, M. L. (2017). Package 'varComp'.
Reed, S. C., Cleveland, C. C., and Townsend, A. R. (2011). Functional ecology of free-living nitrogen fixation: A contemporary perspective. *Annual Review of Ecology, Evolution, and Systematics* **42**, 489-512.
Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
Robson, R. L., and Postgate, J. R. (1980). Oxygen and hydrogen in biological nitrogen fixation. *Annual Reviews in Microbiology* **34**, 183-207.

Rousset, F. (2017). An introduction to the spaMM package for mixed models.
RoyChowdhury, T., Bramer, L. M., Brown, J., Kim, Y. M., Zink, E., Metz, T. O., McCue, L. A., Diefenderfer, H. L., and Bailey, V. (2022). Soil metabolomics predict microbial taxa as biomarkers of moisture status in soils from a tidal wetland. *Microorganisms* **10**.
Salati, E., Sylvester-Bradley, R., and Victoria, R. L. (1982). Regional gains and losses of nitrogen in the Amazon basin. *In* "Nitrogen Cycling in Ecosystems of Latin America and the Caribbean", pp. 367-376. Springer.
Scharf, B. E., Hynes, M. F., and Alexandre, G. M. (2016). Chemotaxis signaling systems in model beneficial plant-bacteria associations. *Plant Mol Biol* **90**, 549-59.
Schulte, E., and Hoskins, B. (1995). Recommended soil organic matter tests. *Recommended soil testing procedures for the north eastern USA. northeastern regional publication* **493**, 52-60.
Sharma, M., Kaushal, R., Kaushik, P., and Ramakrishna, S. (2021). Carbon farming: prospects and challenges. *Sustainability* **13**.
Shirodkar, S., Reed, S., Romine, M., and Saffarini, D. (2011). The octahaem SirA catalyses dissimilatory sulfite reduction in Shewanella oneidensis MR-1. *Environ Microbiol* **13**, 108-115.
Smercina, D. N., Evans, S. E., Friesen, M. L., and Tiemann, L. K. (2019). To fix or not to fix: controls on free-living nitrogen fixation in the rhizosphere. *Appl Environ Microbiol* **85**.
Smercina, D. N., Evans, S. E., Friesen, M. L., and Tiemann, L. K. (2021). Temporal dynamics of free-living nitrogen fixation in the switchgrass rhizosphere. *GCB Bioenergy* **13**, 1814-1830.
Soumare, A., Diedhiou, A. G., Thuita, M., Hafidi, M., Ouhdouch, Y., Gopalakrishnan, S., and Kouisni, L. (2020). Exploiting biological nitrogen fixation: A route towards a sustainable agriculture. *Plants (Basel)* **9**.
Subbarao, G. V., Nakahara, K., Hurtado, M. P., Ono, H., Moreta, D. E., Salcedo, A. F., Yoshihashi, A. T., Ishikawa, T., Ishitani, M., Ohnishi-Kameyama, M., Yoshida, M., Rondon, M., Rao, I. M., Lascano, C. E., Berry, W. L., and Ito, O. (2009). Evidence for biological nitrification inhibition in Brachiaria pastures. *Proc Natl Acad Sci U S A* **106**, 17302-7.
Swenson, T. L., Jenkins, S., Bowen, B. P., and Northen, T. R. (2015). Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biology and Biochemistry* **80**, 189-198.
Tyukavina, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Smith-Rodriguez, K., Okpa, C., and Aguilar, R. (2017). Types and rates of forest disturbance in Brazilian Legal Amazon, 2000–2013. *Science Advances* **3**.
Withers, E., Hill, P. W., Chadwick, D. R., and Jones, D. L. (2020). Use of untargeted metabolomics for assessing soil quality and microbial function. *Soil Biology and Biochemistry* **143**.
Wong, M. Y., Neill, C., Marino, R., Silverio, D., and Howarth, R. W. (2021). Molybdenum, phosphorus, and pH do not constrain nitrogen fixation in a tropical forest in the southeastern Amazon. *Ecology* **102**, e03211.
Wong, M. Y., Neill, C., Marino, R., Silvério, D. V., Brando, P. M., and Howarth, R. W. (2019). Biological nitrogen fixation does not replace nitrogen losses after forest fires in the Southeastern Amazon. *Ecosystems* **23**, 1037-1055.
Yan, J., Han, X., Lu, X., Chen, X., and Zou, W. (2022). Land use indirectly affects the cycling of multiple nutrients by altering the diazotrophic community in black soil. *J Sci Food Agric* **102**, 3788-3795.

Tables and Figures

Table 2-1: Spatial heterogeneity between replicate LU sites accounted for using two modeling methods **(a)** Variance component modeling parses the variance from spatial distance **(b)** Spatial mixed modeling using a Matérn function on geospatial coordinates. Data normality was controlled for through natural-log transformation of DRVs

a.	**Variance component effects**			
		Fixed Effect Est. [CI]	**F-val (p-val)**	**Var. from spatial distance interaction [error]**
ln(ANF rate)	*Intercept*	-0.55 [±0.264]	18.73 (9.73E-05)	
	LU	1.07 [±0.37]	34.85 (6.42E-07)	0.00 [.34]
ln(*nifH* copies)	*Intercept*	18.11 [±0.56]	4359 (4.9E-38)	
	LU	2.2 [±1.02]	18.97 (1.35E-04)	0.19 [0.49]
b.	**Spatial mixed model with Matérn function partitioning**			
		Est. Effect (t-value)	**Likelihood Ratio χ^2 (p-val)**	**Bootstrap χ^2 (p-value)**
ln(ANF rate)	*Intercept*	-0.52 (-3.85)		
	LU	1.07 (5.3)	10.06 ($1.15*10^{-3}$)	5.05 (0.024)
ln(*nifH* copies)	*Intercept*	17.76 (95.4)		
	LU	2.997 (11.4)	14.88 ($1.14*10^{-4}$)	3.63 (0.05)

Table 2-2: Soil parameters expected to impact diazotrophic activity rates. Values represent mean ± standard error for each site. Kruskall-Wallis rank sum tests performed across sites, with test statistics reported as χ^2. P-values are adjusted for multiple comparisons using a Benjamini- Hochberg correction. Letters next to each mean indicate significant site grouping. Variables in bold are those which grouped by LU. * indicates variables with overall significant difference between forest and pasture (ignoring site effect).

		Forest			Pasture				
		F1	**F2**	**F3**	**P1**	**P2**	**P3**	**χ^2**	**p-value**
Carbon	TC	43.1±10.7 a	22.9±1.4 ab	17.3±1.8 b	18.4±0.7 b	17.6±1.2 b	34.2±1.9 a	21.10	$9.2*10^{-4}$
	EOC	0.52±0.07 ab	0.59±0.05 a	0.32±0.03 c	0.31±0.02 c	0.36±0.02 bc	0.53±0.05 a	22.92	$5.1*10^{-4}$
	POXC	2.16±0.57 ab	0.9±0.06 a	0.89±0.18 a	0.8±0.05 a	0.87±0.1 a	2.16±0.1 ab	16.13	$6.5*10^{-3}$
	OM	6.7±1.34 a	4.1±0.19 ab	2.8±0.29 c	3.1±0.19 bc	2.6±0.15 c	6.2±0.24 a	26.68	$1.9*10^{-4}$
	C_{df}	94.7±1.14 **a**	96.1±0.73 **a**	93.3±0.9 **a**	15.47±3.17 **b**	17.8±1.8 **b**	25.16±2.05 **b**	33.18	$3.5*10^{-6}$
Nitrogen	TN	3.5±0.71 a	1.9±0.08 ab	1.6±0.12 bc	1.5±0.07 bc	1.4±0.09 c	2.9±0.19 a	24.33	$3.2*10^{-4}$
	EON	0.21±0.01 a	0.2±0.01 a	0.14±0.02 b	0.18±0.01 ab	0.19±0.01 ab	0.3±0.03 c	20.77	$9.7*10^{-4}$
	*NH_4^+	5.5±1.0 ab	1.4±0.8 a	1.6±1.1 a	10.5±1.6 bc	8.3±2.6 bc	11.6±1.1 c	28.87	$1.1*10^{-4}$
	NO_3^-	106±6.6 **a**	55±4.1 **a**	54±7.4 **a**	2.57±1.8 **b**	2.54±0.97 **b**	0.74±0.36 **b**	34.62	$2.3*10^{-5}$
	^{15}N	8.9±0.24 **a**	10.6±0.3 **a**	9.8±0.2 **a**	7.2±0.2 **b**	6.8±0.1 **b**	7.1±0.3 **b**	34.79	$2.3*10^{-5}$
Cofactors	P	5.1±0.4 ab	6.9±0.34 a	3.6±0.3 bc	4.4±1.15 bc	6.3±1.85 ab	2.6±0.43 c	22.32	$1.6*10^{-6}$
	*Mo	2.35±0.35 a	1.1±0.03 ab	1.09±0.17 a	0.51±0.12 bc	0.64±0.22 bc	0.36±0.06 c	26.39	$1.9*10^{-4}$
	V	88.7±17.7 a	68.6±2.9 a	39.9±5.6 b	54.6±10.2 ab	56.9±23.6 b	236.1±22.4 c	24.70	$3.2*10^{-4}$
	Fe	44.8±5.3 ab	41.8±1.5 ab	30.1±2.5 ac	25.6±3.2 c	28.2±9.2 c	69.6±5.6 b	24.21	$3.2*10^{-4}$
Phys-chem	*pH	5.5±0.1 ab	4.7±0.04 a	5.7±0.1 bc	5.8±0.11 bc	6.1±0.06 cd	6.3±0.2 d	30.53	$5.9*10^{-5}$
	Clay	17.6±1.8 ab	35.9±1.2 c	21.9±1.2ac	13.77±1.3 b	12.1±0.9 b	20.6±1.82 a	29.99	$1.5*10^{-5}$
	Silt	28.33±1.5 a	28.6±1.1 a	31±1.09 ab	28.04±1.52 a	28.81±0.87 a	35.7±1.37 b	13.48	$1.9*10^{-2}$
	Sand	54.1±2.6 ab	35.4±1.9 c	47.0 ±1.6 a	58.2 ±2.7 b	59.1±1.5 b	43.7±3.1 ac	29.07	$2.2*10^{-5}$

Abbreviations: TC (Total Carbon): mg C g^{-1} soil; EOC (Extractable Organic Carbon): mg C g^{-1} soil; POXC (Permanganate-oxidizable Carbon): mg C g^{-1} soil; OM (Loss-on-ignition Organic Matter): percent mass; C_{df} plant (forest-derived soil carbon):percent mass soil C; TN (Total Nitrogen): mg N g^{-1} soil; EON (Extractable Organic Nitrogen): mg N g^{-1} soil; NH_4^+ (Ammonium): µg N g^{-1} soil; NO_3^- (Nitrate): µg N g^{-1} soil; Mo (Molybdenum): µg Mo g^{-1} soil; V (Vanadium): µg V g^{-1} soil; Fe (Iron): mg Fe g^{-1} soil; P (Mehlich-3 extractable orthophosphate P): µg P g^{-1} soil.

Table 2-3: Results of ANOVAs based on best unifying models (across LUs) for *nifH* copy number and ANF rate. Variables were selected using interaction lasso variable selection and manual model refining. Model residual values formatted as *Sum of Squares; mean Squares*. *indicates adjusted R^2

ln(*nifH* copies) Full	**Variable Stats**					**Model Stats**			
	Estimate	**Std. Err.**	**SS**	**Fval**	**pval**	**F-Stat**	**pval**	**RMSE**	**R^2**
(Intercept)	17.0	0.43				72.9	$2.1*10^{-14}$	0.64	.87
LU	2.7	044	82.9	205.6	$1.8*10^{-15}$		***		*.86
Clay	$3.4*10^{-2}$	$1.6*10^{-2}$	3.39	8.4	$6.6*10^{-3}$				
LU:POXC	$5.3*10^{-4}$	$2.4*10^{-4}$	1.93	4.7	$3.6*10^{-2}$				
Residual			*12.9; 0.4*						

ln(ANF) Full	**Variable Stats**					**Model Stats**			
	Estimate	**Std. Err.**	**SS**	**Fval**	**pval**	**F-Stat**	**pval**	**RMSE**	**R^2**
(Intercept)	-.03	0.34				15.9	$1.6*10^{-6}$	0.5	.6
LU	-.8	0.57	9.32	36.6	$9.4*10^{-7}$		***		*.56
LU:(EOC/EON)	0.71	0.29	1.6	9.6	0.02				
ln(NO_3^-)	-0.11	0.08	1.27	6.0	0.03				
Residual			*8.1; 0.25*						

Abbreviations: ln(ANF) = natural-log transformed asymbiotic nitrogen fixation rate (ng N g^{-1} soil day^{-1}); Clay (% composition); EOC= Extractable organic carbon (µg g^{-1} soil); EON = Extractable organic nitrogen (µg g^{-1} soil); F-stat = F statistic; Fval = F-value; LU = Land use; ln(NO_3^-) = natural-log transformed nitrate (µg g^{-1} soil); POXC = Permanganate oxidizable C (µg g^{-1} soil); Pval = p-value; RMSE = Root mean squared error; TN = Total nitrogen (µg g^{-1} soil).

Table 2-4: Results of ANOVAs based on best models built separately with forest and pasture samples for *nifH* copy number and ANF rate. Variables selected from **(a)** physicochemical parameters, or **(b)** with the most highly correlated metabolites for each response variable- LU combination. *Non-significant models not included*.

a	**Physicochemical parameters only**								
	Variable Stats					**Model Stats**			
***nifH* Past.**		**Estimate**	**Std. Err.**	**SS**	**F-val**	**p-val**	**F-Stat**	**p-val**	**RMSE**	***R^2**
	(Intercept)	$2.7*10^8$	$3.3*10^8$				11.9	$4*10^{-4}$	$4*10^8$	.71
	C_{df}	$2.3*10^7$	$1.9E*10^7$	$2.8*10^{18}$	16.3	$1.2*10^{-3}$		***		*.66
	POXC	$6.2*10^5$	$2.0E*10^5$	$2.7*10^{18}$	15.5	$1.5*10^{-3}$				
	Mo	$-5.5*10^8$	$2.6E*10^8$	$6.7*10^{17}$	3.8	0.07				
	Residual			*$2.48*10^{18}$; $1.77*10^{17}$*						
ANF Past.		**Estimate**	**Std. Err.**	**SS**	**F-val**	**p-val**	**F-Stat**	**p-val**	**RMSE**	***R^2**
	(Intercept)	-2.79	1.78				3.96	0.42	1.4	.35
	EOC/EON	1.54	0.80	10.89	5.42	0.03		*		*.26
	TN/EON	.23	0.15	4.88	2.5	0.057				
	Residual			*22.18; 1.48*						

b	**Physicochemical and Metabolites**								
	Variable Stats					**Model Stats**			
***nifH* For.**		**Estimate**	**Std. Err.**	**SS**	**F-val**	**p-val**	**F-Stat**	**p-val**	**RMSE**	***R^2**
	(Intercept)	$-1.4*10^8$	$6*10^7$				24.9	$7*10^{-7}$	$3.7*10^7$	.84
	Clay+Silt	$2.6*10^6$	$9.5*10^5$	$1.2*10^{16}$	12.1	$1*10^{-3}$		***		*.81
	TC	$-1.3*10^3$	$6.2*10^2$	$4.2*10^{15}$	4.3	0.05				
	β-sitosterol	$8.6*10^8$	$1*10^3$	$5.8*10^{16}$	58.5	$2.9*10^{-6}$				
	Residual			*$1.34*10^{16}$; $9.92*10^{14}$*						
ANF Past.		**Estimate**	**Std. Err.**	**SS**	**F-val**	**p-val**	**F-Stat**	**p-val**	**RMSE**	***R^2**
	(Intercept)	-1.2	1.06				14.2	$1.6*10^{-4}$	0.9	.75
	TN/EON	-0.25	$9.2*10^{-2}$	8.48	10.43	$6.1*10^{-3}$		***		*.70
	erythrose	$1.8*10^{-3}$	$4.4*10^{-4}$	20.5	25.3	$1.8*10^{-4}$				
	UDP-NAG	$-8.2*10^{-4}$	$3.2*10^{-4}$	5.6	6.88	0.02				
	Residual			*11.4; 0.81*						

Abbreviations: ANF= asymbiotic nitrogen fixation rate (ng N g^{-1} soil day^{-1}); β- sitosterol (peak value);Cdf (% C forest-derived); Clay/Silt (% composition); EOC= Extractable organic C (µg g^{-1} soil); EON = Extractable organic N (µg g^{-1} soil); erythrose (peak values); For. = Forest; Fval = F-value; F-stat = F statistic; Metabolites (peak values); Mo= Molybdenum (µg g^{-1} soil); Past. = Pasture; POXC = permanganate oxidizable C (µg g^{-1} soil); Pval = p-value; RMSE = Root mean squared error; Std. Err. = Standard error; SS = Sum of squares; TN/EON = Total: extractable organic N (organic + inorganic; µg g^{-1} soil); UDP-NAG = uridine diphosphate-N-acetylglucosamine (peak values)

Figure 2-1: Asymbiotic nitrogen fixation rate (ANF) and *nifH* gene abundance**.** Horizontal lines represent mean measurements of each LU. Letters denote grouping by *post-hoc* Dunn's tests, based on nonparametric Kruskall-Wallis rank-sum tests (top right). **(a)** ANF rates calculated from $^{15}N_2$ labeled soil incubations. **(b)** Abundance of *nifH* marker genes derived from qPCR, (copies per grams soil^{-1}). Baseline $^{15}N_2$ and copies calculated per ng^{-1} DNA in Supp. Figure 2-1. **(c)** Correlation between *nifH* copy number and ANF rate for forest (green) and pasture (brown), with dotted lines representing linear regression smoothing.

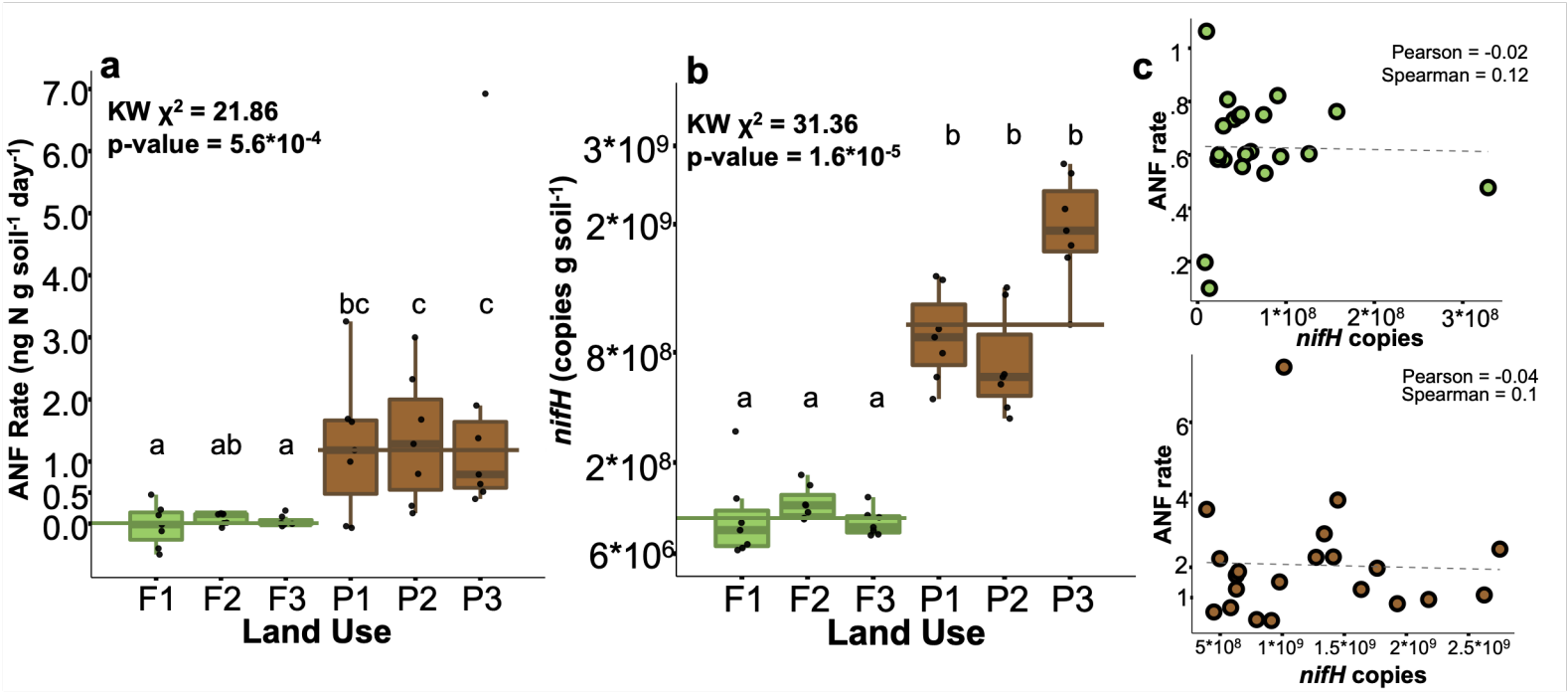
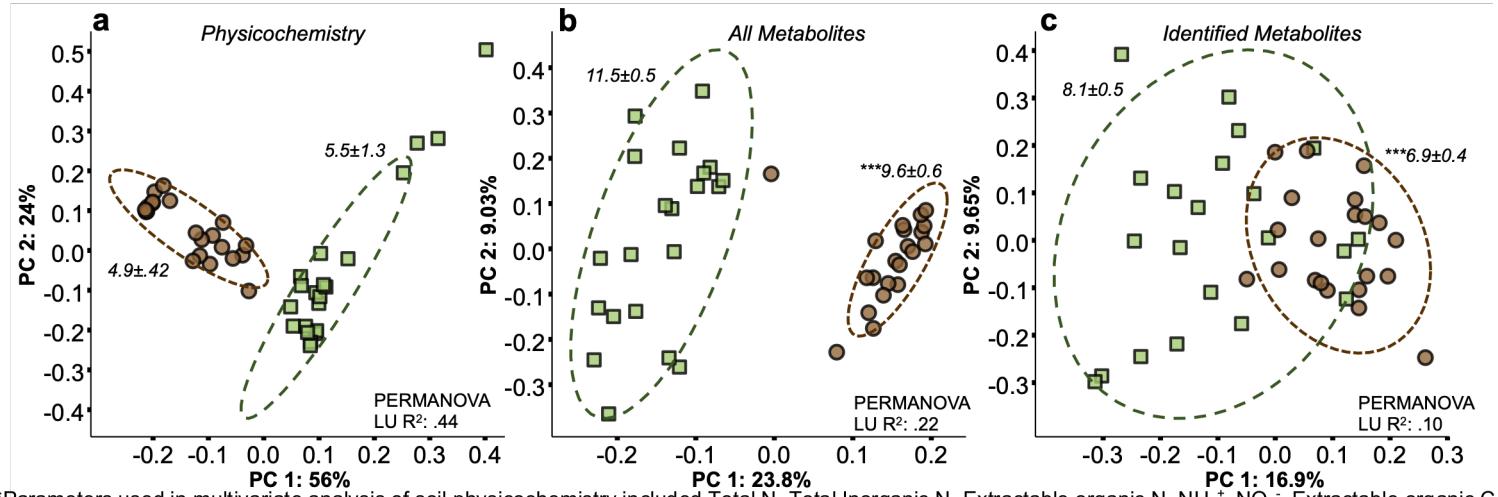
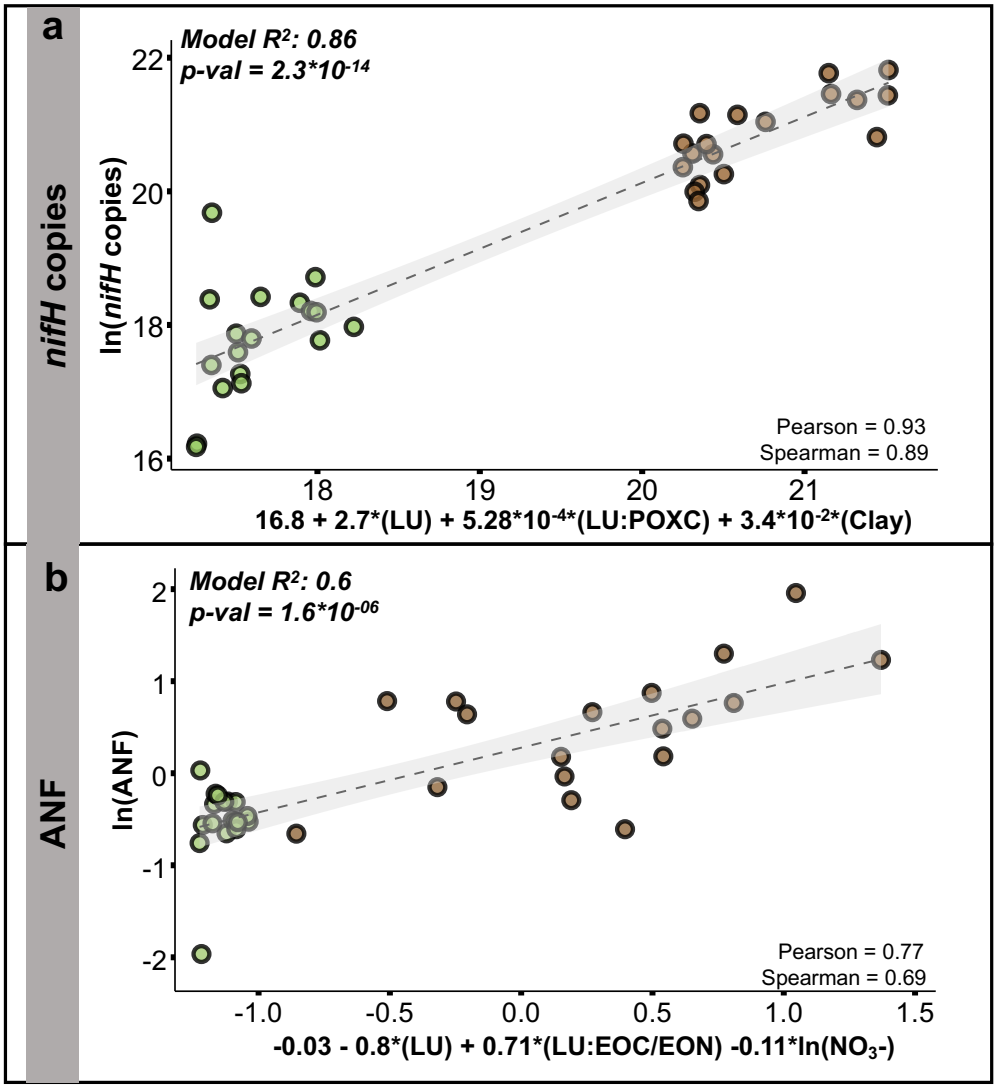


Figure 2-2: Unconstrained redundancy analysis plots for soil profiles**,** including **(a)** soil physicochemical profiles (including variables in Table 2-2), **(b)** all metabolites, including identified and unidentified compounds (n=180), and **(c)** identified metabolites only (n=91). Forest samples are represented by green squares and pasture samples are represented by brown circles. Ellipses represent 90% confidence intervals for land use grouping. Corresponding statistics are presented in Supp. Table 2-1. *** on italicized mean distance indicates significant difference.



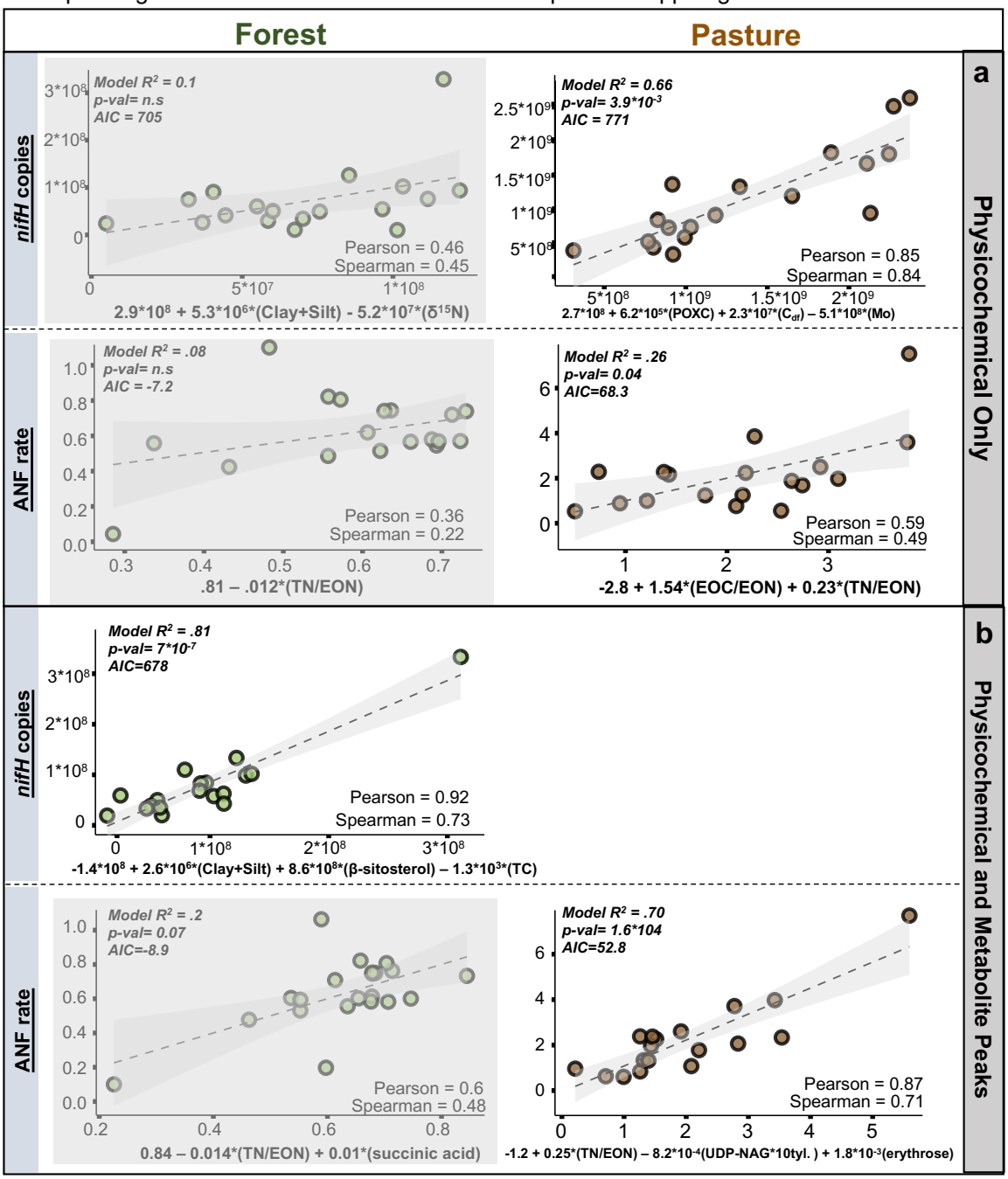
*Parameters used in multivariate analysis of soil physicochemistry included Total N, Total Inorganic N, Extractable organic N, NH_4^+, NO_3^-, Extractable organic C, POXC, organic matter (percentage), percent C derived from forest, K, Mg, Ca, Mo, V, Fe, S, P, extractable acidity, CEC, pH, Clay+Silt content, $\delta^{15}N$, Z, and Cu.

Figure 2-3: Best unifying models (across LUs) for **(a)** ln(*nifH* copy number) and **(b)** ln(ANF)**.** Dashed lines linear regression smoothing function, with gray shading representing a 90% confidence interval. Formula on x-axis represents the best fit equation. Correlation between prediction and response are in the lower right, with goodness of model fit in upper left. Corresponding statistics in Table 2-3. Goodness of fit plots are shown in Supp. Figure 2-2.



Abbreviations: ln(ANF) = natural-log transformed asymbiotic nitrogen fixation rate (ng N g^{-1} soil day^{-1}); Clay (% composition); EOC= Extractable organic carbon (µg g^{-1} soil); EON = Extractable organic nitrogen (µg g^{-1} soil); LU = Land use; ln(NO_3^-) = natural logarithm-transformed nitrate (µg g^{-1} soil); POXC = Permanganate oxidizable C (µg g^{-1} soil); p-val = p-value; TN = Total nitrogen (µg g^{-1} soil).

Figure 2-4: Best models built separately for forests and pastures. Best variables were selected with **(a)** physicochemical parameters only, and with **(b)** the highest correlated metabolites for each response variable- LU combination (bottom). Figures shaded in gray were non-significant. The missing *nifH* pasture plot indicates that inclusion of metabolites in variable selection did not improve model fit. Corresponding statistics in Table 2-4. Goodness of fit plots in Supp. Figure 2-3.



Abbreviations: AIC = Akaike Information Criterion; ANF = asymbiotic nitrogen fixation rate (ng N g^{-1} soil day^{-1}); β- sitosterol (peak value); C_{df} (% C forest-derived); Clay+Silt (% composition); EOC= Extractable organic carbon (µg g^{-1} soil); EON = Extractable organic nitrogen (µg g^{-1} soil); erythrose (peak value); Mo= Molybdenum (µg g^{-1} soil); POXC = Permanganate oxidizable carbon (µg g^{-1} soil); p-val = p-value; succinic acid (peak value); TC = Total carbon (µg g^{-1} soil); TE = Total extractable (organic + inorganic N; µg g^{-1} soil); TN = Total nitrogen (µg g^{-1} soil); UDP-NAG = uridine diphosphate N-acetylglucosamine (peak value); $\delta^{15}N$= delta value of ^{15}N (‰).

Figure 2-5: Volcano plots representing DIG read count log_2 Fold-change plotted against adjusted p-values. Data was derived from two analytical approaches including **(a)** read-assembly and **(b)** gene-targeted annotation (GTA). GTA statistics in Supp. Table 2-4. Gray dots denote non-significant (ns) changes. CPM = counts per million reads. Symbols in **(b)** represent gene functional groups. *Note:* the y-axis in **(b)** is square-root transformed

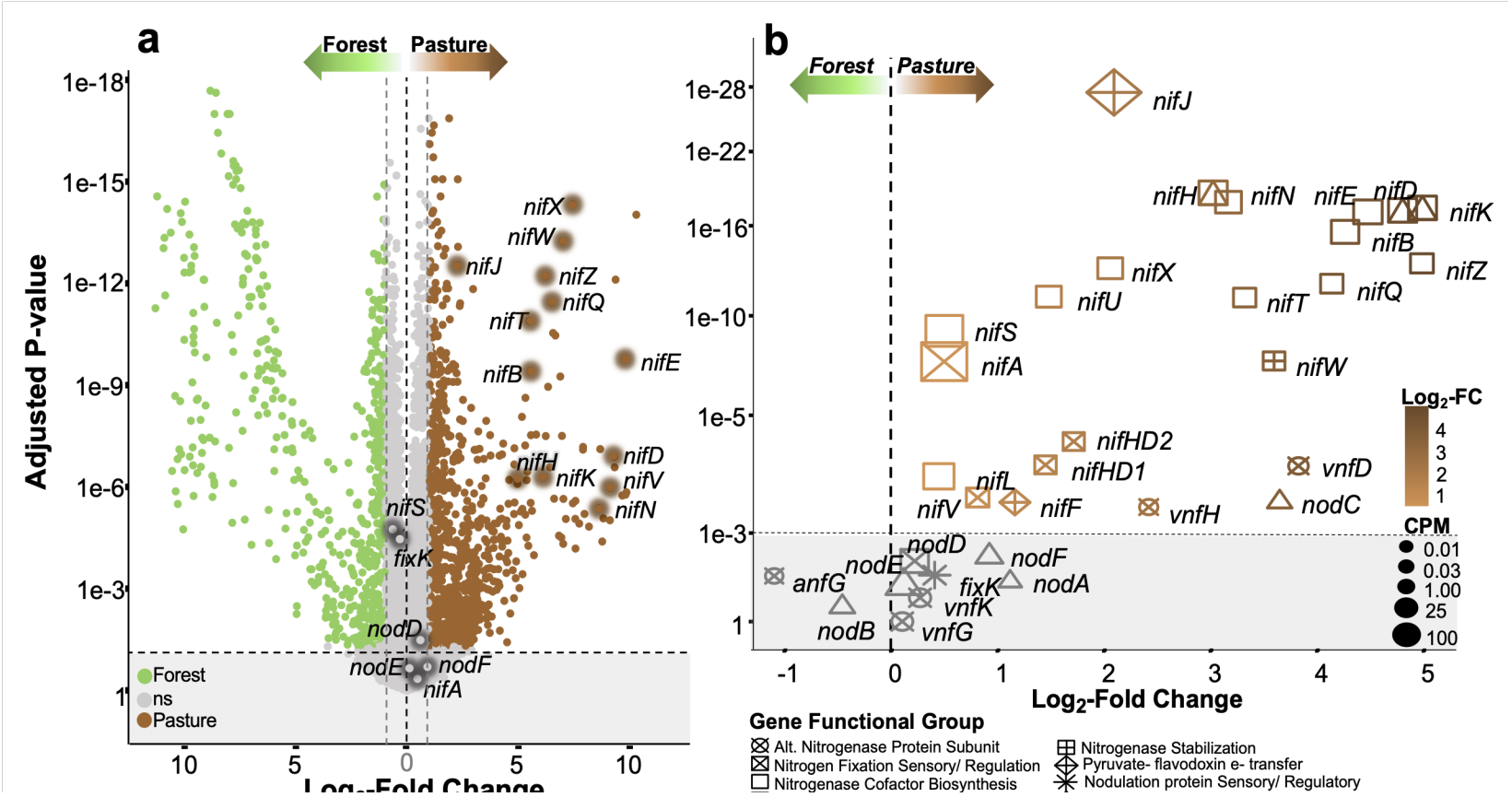
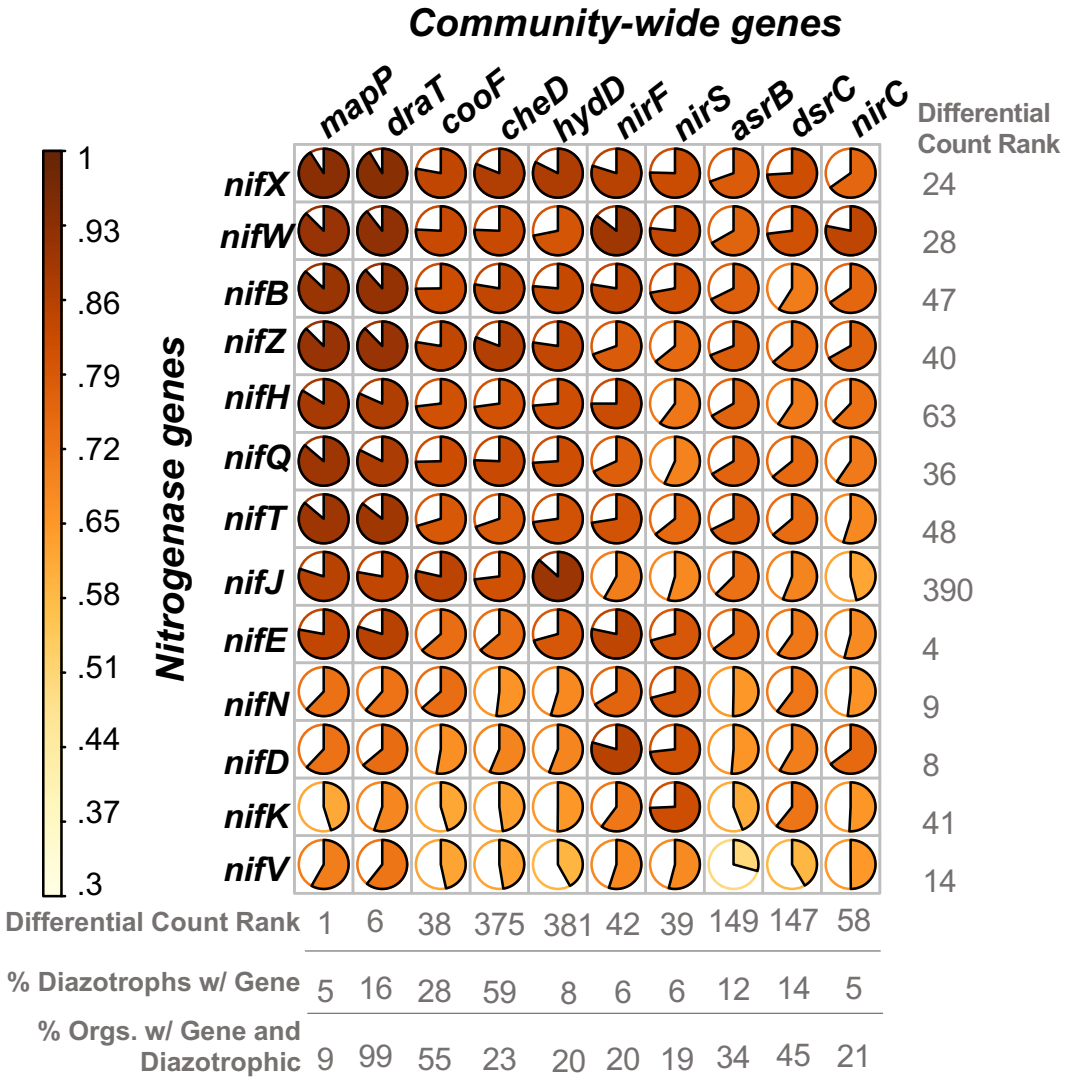


Figure 2-6: Using the RA approach, *nif* genes (rows) were correlated against community-wide genes, identifying 78 significant correlations. A subset of functionally or biogeochemically- relevant genes are shown here (columns). Results for all genes available in Supp. Table 2-5. Any Spearman correlation coefficients below 0.59 are considered non-significant based on a Bonferroni multiple comparisons adjustment. Differential count rank represents log_2-fold enrichment level in pastures. Community-wide genes were cross-referenced with nitrogenase genes in KEGG database to calculate the percent overlap (see values in gray)



Supplementary Methods

Field-estimation of gravimetric water content

To calculate dry-weight adjusted soil mass, three replicate cylindrical sections of the preserved field cores (wet) approximately 3 cm in length were cut and weighed, and gravimetric water content was estimated by multiplying the inverse of the core section bulk density by field-measured volumetric water content and averaged across replicates. Although this provides a rough approximation, a sufficient level of precision was achieved: dry mass per incubation averaged 4.98 $\pm$ 0.21 g (mean $\pm$ standard deviation).

Isotopic measurements

Soils were allowed to thaw and to air dry under ambient laboratory conditions, then any remaining rock fragments were removed before grinding soils to a fine powder using 1/4 and 1/8 -inch stainless steel agitator balls loaded onto a shaking apparatus running at 300 rpm for 30 minutes. Tin-encapsulated samples were analyzed for isotopic enrichment using an Elementary Vario EL Cube (Elementar Analysensysteme GmbH, Hanau, Germany) equipped with a PDZ Europa 20-20 isotope ratio mass spectrometer (Serco Ltd., Cheshire, UK) at the UC Davis Stable Isotope Facility (Davis, CA, USA). Both ^{15}N and ^{13}C data were obtained from these measurements.

Nitrogen fixation rate calculations

$$Headspace\ Volume = Vial\ volume - \frac{(Wet\ soil * (1 - grav.water\ percent)}{soil\ particle\ density} - \frac{Water\ mass}{water\ density}$$

$$Gas\ enrichment\ \% = \frac{(\%\ 15N\ in\ added\ gas * mL\ exchanged) + (\%\ 15N\ Natural\ Abund * mL\ ambient)}{Headspace\ volume}$$

$$\%\ excess\ increase = \ mean\ 15N\ incubation\ atom\ \% - mean\ 14N\ control\ atom\ \%$$

$$Fraction\ Total\ Enrichment = \frac{\%\ excess\ increase}{\%\ gas\ enrichment\ increase}$$

$$mass\ N\ from\ incubation = \ Fraction\ Total\ Enrichment * total\ N$$

Soil DNA extraction

All steps were completed using manufacturer-provided instructions, apart from the physical disruption step, for which a VWR homogenizer bead mill (VWR, Radnor, PA, USA) was used for five

cycles of 1 minute at a speed of 5.0 m s^{-1}, and the elution step, where two elution cycles using half the recommended elutant volumes each were used. Extractions were purified using the Genomic DNA Cleanup and Concentrator-10 kit (Zymo Research, Irvine, CA, USA) according to manufacturer instructions, except for a two-step elution modification as described above. DNA yield was quantified using a Qubit dsDNA High Sensitivity assay kit (Thermo Fisher Scientific, Waltham, MA, USA).

nifH qPCR

Template DNA was diluted to 2 ng ul^{-1} across samples and used in triplicate with 20 µl reactions with the following reactants: 1 µl PolF and PolR primer each (500 nM in reaction), 10 µl SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 4 µl DNA template, and 4 µl PCR-grade water. Samples were run against a standard curve on a CFX Connect Real-time thermocycler (Bio-Rad, Hercules, CA, USA). The following cycling conditions were used: 98°C for 2 m, followed by 35 cycles of 98°C for 15 s, 68.5°C for 15 s and 72°C for 30 s, followed by a final extension of 72°C for 5 m.

Standards were created using a high-fidelity blunt-end TOPO cloning kit (Thermo Fisher Scientific, Waltham, MA, USA) with a *nifH* gene amplified from *Herbaspirillum seropedicae* (Baldani et al., 1986; ATCC Z 152). Samples were randomized across two reaction plates (94% and 97% efficiency; R^2=0.99) and run against a standard curve on a CFX Connect Real-time thermocycler (Bio-Rad, Hercules, CA, USA). Copy number was calculated on a basis of per ng DNA and per g dry soil.

Soil chemical analysis

Soil texture was calculated by using 0.5 g air-dried and sieved soil suspended in 10 ml of 5% sodium metaphosphate, shaken for 12 hours at 160 rpm. Measurements were made using a particle size laser analyzer in sonication mode. Resulting particle size distributions were aggregated using a modified maximum diameter cut-off of 6.15 µm for clay particles (i.e., 'light-scatter equivalent'; Faé et al., 2019). Soil pH was measured using a 2:1 soil slurry with deionized water. Exchangeable soil acidity was calculated using the Mehlich buffer method (Mehlich, 1976). Total carbon and nitrogen content were measured by dry-combustion and calculated against a casein standard. Total soil Al, Fe, Mo, and V were quantified by the UC Davis Center for Inductively-Coupled Plasma (ICP-MS) Mass Spectrometry on dried,

ground, soil which was microwave-digested with HNO_3. Elemental analysis of plant nutrients including K, Ca, Mg, and orthophosphate P were assessed by Mehlich 3 extracted ICP-MS (Wolf and Beegle, 1995).

NO_3^-, NH_4^+, total EOC and total EON were quantified through the extraction of a 10 g homogenized subsample of field-preserved soil using 0.05 M K_2SO_4 and shaking for 1 hour at 160 rpm. Samples were then filtered and frozen until processed. Inorganic N assays were performed using a 96-well plate format with 4 technical replicates per sample. Methods were adapted from Qiu et al. (1987) for NH_4^+ and Hood-Nowotny et al. (2010) for NO_3^-. Total EON was calculated by first combining 1 ml extract with 0.2 ml of reagent containing 3% (g ml^{-1}) H_3BO_3, 7.6% 5 M (ml ml^{-1}) NaOH, and 5% (g ml^{-1}) $K_2S_2O_4$, and digesting in water baths for 2 hours (Valderrama, 1981). The resulting digestate was assayed using NO_3^- analysis methods described above at a 1:3 dilution. Total EOC was quantified by running extracts on a Shimadzu TOC analyzer (Kyoto, Japan) and calculating concentrations against a standard curve.

Apart from EOC, three additional methods were used to proxy 'microbial accessible' C. First, total organic matter was quantified using the loss on ignition procedure (Schulte and Hoskins, 1995). Next, permanganate-oxidizable carbon (POXC) was quantified using methods described by Culman et al. (2012) with triplicate samples of field-preserved, air-dried soil. Isotopic analysis of soil provided $\delta^{13}C$ natural abundance values, which were used in conjunction with known end-member values of pasture and forest species (de Moraes, 1996) to estimate the percent carbon derived from C3 plants, assumed to be from forest (C_{df} , forest-derived C). Isotopic analysis of soil also provided $\delta^{15}N$ natural abundance Finally, the use of untargeted soil metabolomics was used as an alternative proxy for low molecular weight C compounds (Swenson et al., 2015). This approach has the added benefit of indicating the availability of organic N and P.

Lasso variable selection correlation pre-screening

For Lasso variable selection, the physicochemical parameters listed above in the section 'Physicochemical profiling', as well as rational ratio terms were considered in the initial screening for meaningful relationships with diazotrophic response variable. Aside from those variables listed in the previous section, ratios considered included: total to extractable N (inorganic + organic; TN:TE), total to

extractable organic N (TN:EON), total to inorganic N (TN:In), extractable organic C to N (EOC:EON), POXC to extractable organic N (POXC:EON), and POXC to inorganic N (POXC:In)

Parametric tests and accounting for geospatial heterogeneity: univariate tests and multiple linear regression

Before performing parametric tests, because ANF rate data contained several slightly negative calculated values, all values were adjusted using a constant-additive transformation, calculated as the minimum value plus 0.1 (an arbitrary, small value) before applying a natural log transformation.

The spatial sampling pattern used was selected to potentially explore distance-decay patterns as shown in Rodrigues et al. (2013) and Mirza et al. (2020). However, it was assessed early in data analysis that spatial patterns are not relevant to the activity, chemical, or community measurements in this study at the 100 m scale- with the exception of 0 and 10cm points for some measurements. Some statistical tests, including univariate mixed-model linear regression considered geographic distances between sampling locations, including within- and among sites. Mixed random effects models were checked for adherence to the assumptions, and the significance of LU was tested using a likelihood ratio test with 100 bootstraps.

For Lasso variable selection and multiple linear regression, to avoid 0 and 10cm points adding bias to data linearity, measurements from these two sampling locations were averaged and used as a single data point for each site. After model building for diazotrophic response variables (ANF rate and *nifH* copies), the best model was fit to a generalized linear mixed-model Using LU-nested site grouping as a random effect, in order to check for potential bias associated with significant differences in data spread across site-LU combinations.

Metagenome analysis: assembly-based approach (RA)

Paired-end reads were first trimmed for adaptors and filtered using Cutadapt v.3.5 (Martin, 2011). BBduk (Bushnell, 2014) was further used to trim low-quality sequences. Filtering parameters discarded sequences shorter than 100 bp and with an error rate exceeding 0.1, and 20 was used as a sliding mean quality score cutoff. The FastQC module (Andrews, 2017) was used to compare raw and filtered reads.

MEGAHIT (Li et al., 2015) was used for read assembly with a minimum multiplicity for filtering of 1 and a k-mer size range of 27 to 127, increasing by intervals of 10. Prodigal v.2.6.3 (Hyatt et al., 2010) was used for open reading frames (run in normal mode), and nucleotide to protein translation. Reads were mapped back to assembled contigs using Bowtie2 (Langmead and Salzberg, 2012) in order to calculate read counts. SAMtools (Li et al., 2009) was subsequently used to reconfigure Bowtie2's output to a bam file, followed by HtSeq (htseq-count; Anders et al., 2015) to obtain gene counts from assemblies. Kofamscan (Aramaki et al., 2020), which applies a Hidden Markov Model approach to annotation, was then used in conjunction with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to functionally annotate open reading frames identified by Prodigal, and annotation significance cutoffs were made based on best-matches exceeding threshold minimum scores. Annotations were standardized and aggregated by KO terms, and all sample data were merged to generate a gene abundance table.

Metagenome analysis: gene-targeted approach (GTA)

For this method, annotation was again performed using the KEGG database. A set of 32 'diazotrophic indicator genes' (DIGs), including those encoding Mo- (*nif*), Fe- (*anf*), and V- (*vnf*) containing nitrogenase metalloclusters, transcriptional regulators, and co-factors, as well as nodulation (*nod*) factors were selected to assess both free-living and symbiotic diazotrophic community members. A database was constructed using annotated sequences for each gene available in the KEGG database. To control for genes with a high degree of homology across prokaryotes (for example, *nifA*, a transcriptional regulator), non-diazotrophic organisms were filtered from gene-specific databases by cross-referencing with organisms occurring in the *nifH* database. Then, quality-filtered, but unassembled read pairs were aligned to the constructed database using Bowtie2. A custom script was then used to reformat alignment outputs into a gene/organism count table. Read counts were removed if: (1) it was a singleton organism/gene count, (2) the count-associated organism had less than two gene hits in the database, and (3) the count occurred in only a single sample. An important caveat of this method is that if read alignment scores were equal for multiple organisms, annotation is randomly assigned to one. In conjunction with the more conservative assembly-based approach however, this process serves as a liberal upper bound in gene counts and is additionally fast and computationally-light to perform.

To study taxa-based community structure, count data was Hellinger-transformed, then ordinated using nonmetric multidimensional scaling with the bray distance metric and a maximum of five dimensions. This was performed for taxa based on the inclusion of *nifH*, *nif*K, and *nif*D hits across the data set.

Diazotrophic indicator genes

32 genes were selected as indicators of diazotrophs, that is, genes expected to be present in organisms capable of fixing N_2 gas: The below genes are listed by functional category. Asterisks are next to genes considered the 'minimal core genes' necessary for active fixation based on (Wang et al., 2013)

Functional group	Gene	Functional group	Gene
Sensory/Regulation	*fixK*	**Flavodoxin**	*nifJ*
	nifA		*nifF*
	nifL	**Nitrogenase Stabilization**	*nifW*
	nifHD1	**Alternative Nitrogenase metallocluster formation**	*vnfD*
	nifHD2		*vnfG*
Nitrogenase Metallocluster formation	*nifD**		*vnfH*
	*nifK**		*vnfK*
	*nifH**		*anfG*
	nifU	**nod Sensory/ Regulation**	*nodD*
	nifS	**Plant-targeted Mitotic signal**	*nodA*
Nitrogenase Cofactor Biosynthesis	*nifN**		*nodB*
	*nifB**	**Plant-targeted Nodulation signal**	*nodC*
	*nifE**		*nodE*
	nifZ		*nodF*
	nifQ		
	*nifV**		
	*nifX**		
	nifT		

References

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *bioinformatics* **31**, 166-169.

Andrews, S. (2017). FastQC: a quality control tool for high throughput sequence data. 2010.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252.

Baldani, J., Baldani, V., Seldin, L., and Döbereiner, J. (1986). Characterization of Herbaspirillum seropedicae gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. *International Journal of Systematic and Evolutionary Microbiology* **36**, 86-93.

Bushnell, B. (2014). BBTools software package. *URL http://sourceforge. net/projects/bbmap* **578**, 579.

Culman, S. W., Snapp, S. S., Freeman, M. A., Schipanski, M. E., Beniston, J., Lal, R., Drinkwater, L. E., Franzluebbers, A. J., Glover, J. D., and Grandy, A. S. (2012). Permanganate oxidizable carbon reflects a processed soil fraction that is sensitive to management. *Soil Science Society of America Journal* **76**, 494-504.

de Moraes, J. F., Volkoff, B.C.C.C., Cerri, C.C. and Bernoux, M., (1996). Soil properties under Amazon forest and changes due to pasture installation in Rondônia, Brazil. *Geoderma* **70**, 63-81.

Faé, G. S., Montes, F., Bazilevskaya, E., Añó, R. M., and Kemanian, A. R. (2019). Making soil particle size analysis by laser diffraction compatible with standard soil texture determination methods. *Soil Science Society of America Journal* **83**, 1244-1252.

Hood-Nowotny, R., Umana, N. H.-N., Inselbacher, E., Oswald-Lachouani, P., and Wanek, W. (2010). Alternative methods for measuring inorganic, organic, and total dissolved nitrogen in soil. *Soil Science Society of America Journal* **74**, 1018-1027.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 1-11.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359.

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-6.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

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

Mehlich, A. (1976). New buffer pH method for rapid estimation of exchangeable acidity and lime requirement of soils. *Communications in Soil Science and Plant Analysis* **7**, 637-652.

Mirza, B. S., McGlinn, D. J., Bohannan, B. J., Nüsslein, K., Tiedje, J. M., and Rodrigues, J. L., 2020. . 86(10). (2020). Diazotrophs show signs of restoration in Amazon rain forest soils with ecosystem rehabilitation. *Applied and environmental microbiology,* **86**.

Qiu, X.-c., Liu, G.-P., and Zhu, Y.-Q. (1987). Determination of water-soluble ammonium ion in soil by spectrophotometry. *Analyst* **112**, 909-911.

Rodrigues, J. L., Pellizari, V. H., Mueller, R., Baek, K., Jesus Eda, C., Paula, F. S., Mirza, B., Hamaoui, G. S., Jr., Tsai, S. M., Feigl, B., Tiedje, J. M., Bohannan, B. J., and Nusslein, K. (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc Natl Acad Sci U S A* **110**, 988-93.

Schulte, E., and Hoskins, B. (1995). Recommended soil organic matter tests. *Recommended soil testing procedures for the north eastern USA. northeastern regional publication* **493**, 52-60.

Swenson, T. L., Jenkins, S., Bowen, B. P., and Northen, T. R. (2015). Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biology and Biochemistry* **80**, 189-198.

Valderrama, J. C. (1981). The simultaneous analysis of total nitrogen and total phosphorus in natural waters. *Marine chemistry* **10**, 109-122.

Wang, L., Zhang, L., Liu, Z., Zhao, D., Liu, X., Zhang, B., Xie, J., Hong, Y., Li, P., Chen, S., Dixon, R., and Li, J. (2013). A minimal nitrogen fixation gene cluster from Paenibacillus sp. WLY78 enables expression of active nitrogenase in Escherichia coli. *PLoS Genet* **9**.

Wolf, A., and Beegle, D. (1995). Recommended soil tests for macronutrients: Phosphorus, potassium, calcium, and magnesium. *Recommended soil testing procedures for the northeastern United States. Northeast Regional Bull* **493**, 25-34.

Supplementary Tables and Figures

Supp. Table 2-1: Non-parametric statistics for individual metabolites (external xlsx): **(a)** Results of PERMANOVAs testing the significance of LU in explaining physicochemical as well as full metabolite and identified metabolite only profiles. Mean centroid, groupwise dispersion within LUs was also tested using constrained correspondence permutation tests. **(b)** Results of Kruskall-Wallis non-parametric multiple comparisons tests across sites using a Dunn's *post-hoc* to test for grouping by LU type. Metabolites are ordered by greatest to lowest chi-squared value, including both unknown and known compounds. Values presented for each site (n=7) represents the mean +/- standard error, with Dunn's grouping represented by letters. P-values were adjusted using a Benjamini- Hochberg correction. Those metabolites with strong LU differences are bolded. Red line separates significant from non-significant test statistics.
‡BinBase ID's represent unidentified compounds. More information corresponding to IDs can be found at https://binvestigate.fiehnlab.ucdavis.edu/#/

Supp. Table 2-2: Metabolite peak value correlations with Pasture ANF, ordered by the most significantly negative overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
UDP-N-acetylglucosamine	**-0.75**	**0.00**	**-0.51**	**-0.63**	**Pyrimidine nucleotide sugars**
lyxitol	**-0.59**	**0.01**	**-0.49**	**-0.54**	**Sugar Alcohol**
4-hydroxybenzoic acid	**-0.46**	**0.06**	**-0.62**	**-0.54**	**Ring C**
succinic acid	**-0.57**	**0.01**	**-0.49**	**-0.53**	**Dicarboxylic acid**
sorbitol	**-0.54**	**0.02**	**-0.51**	**-0.53**	**Sugar Alcohol**
stigmasterol	**-0.50**	**0.04**	**-0.49**	**-0.49**	**Steroid**
cholesterol	**-0.42**	**0.09**	**-0.56**	**-0.49**	**Ring C**
phosphate	**-0.54**	**0.02**	**-0.42**	**-0.48**	**phosphate**
oleamide	**-0.47**	**0.05**	**-0.45**	**-0.46**	**Long Chain C**
galactinol	**-0.46**	**0.06**	**-0.44**	**-0.45**	**Disaccharide**
pimelic acid	-0.43	0.07	-0.44	-0.44	Fatty acid
lactulose	-0.45	0.06	-0.42	-0.44	Disaccharide
glutaric acid	-0.36	0.14	-0.51	-0.44	Long Chain C
phthalic acid	-0.35	0.16	-0.47	-0.41	Ring C
3-(4-hydroxyphenyl)propionic acid	-0.34	0.17	-0.45	-0.39	Ring C
inositol-4-monophosphate	-0.40	0.10	-0.37	-0.39	phosphate
cerotinic acid	-0.38	0.12	-0.39	-0.38	Long Chain C
dehydroabietic acid	-0.33	0.18	-0.43	-0.38	Ring C
phytol	-0.44	0.07	-0.31	-0.38	Diterpenoid
1-monostearin	-0.26	0.30	-0.38	-0.32	Long Chain C
glutamic acid	-0.32	0.19	-0.19	-0.26	Amine
capric acid	-0.08	0.75	-0.43	-0.25	Long Chain C
glyceric acid	-0.24	0.34	-0.22	-0.23	C other
glycerol-alpha-phosphate	-0.31	0.21	-0.14	-0.22	phosphate
oxoproline	-0.24	0.33	-0.18	-0.21	Amino Acid
fructose	-0.19	0.45	-0.23	-0.21	Disaccharide
3-aminoisobutyric acid	-0.25	0.31	-0.16	-0.20	Amine
aniline	-0.19	0.44	-0.19	-0.19	Amine
behenic acid	-0.29	0.25	-0.08	-0.18	Long Chain C
beta-gentiobiose	-0.08	0.77	-0.28	-0.18	Disaccharide
adipic acid	-0.13	0.61	-0.23	-0.18	Long Chain C
glycerol	-0.13	0.60	-0.22	-0.18	C other
nonadecanoic acid	-0.20	0.43	-0.14	-0.17	Long Chain C
hydroquinone	-0.25	0.31	-0.08	-0.17	Phenol
melezitose	-0.15	0.54	-0.16	-0.16	Trisaccharide
glucose	-0.24	0.34	-0.07	-0.16	Monosaccharides
serine	-0.21	0.41	-0.10	-0.15	Amino Acid
xylulose	-0.18	0.46	-0.12	-0.15	Monosaccharides
conduritol-beta-epoxide	-0.15	0.56	-0.15	-0.15	Ring C
glycerol-3-galactoside	-0.08	0.75	-0.21	-0.14	Ring C
guanine	-0.18	0.46	-0.09	-0.14	Nucleic Acid
maltose	-0.06	0.81	-0.21	-0.13	Disaccharide

Supp. Table 2-2, contd.: Metabolite peak value correlations with Pasture ANF, ordered from least to most significantly positive overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
adenine	-0.07	0.77	-0.16	-0.12	Nucleic Acid
uracil	-0.01	0.97	-0.20	-0.11	Nucleic Acid
terephthalic acid	-0.18	0.48	-0.03	-0.10	Dicarboxylic acid
proline	-0.09	0.74	-0.09	-0.09	Amino Acid
maltotriitol	0.03	0.91	-0.20	-0.09	Ring C
ribose	0.05	0.86	-0.21	-0.08	Monosaccharides
beta-mannosylglycerate	-0.02	0.94	-0.12	-0.07	Long Chain C
squalene	-0.08	0.74	-0.04	-0.06	Prenol Lipids
urea	-0.04	0.87	-0.08	-0.06	Amino Acid
tagatose	-0.01	0.97	-0.10	-0.06	Monosaccharides
alanine	-0.08	0.76	-0.03	-0.05	Amino Acid
glycine	0.08	0.76	-0.17	-0.05	Amine
4-hydroxycinnamic acid	-0.17	0.49	0.08	-0.05	Ring C
lauric acid	0.05	0.83	-0.13	-0.04	Long Chain C
4',5-dihydroxy-7-glucosyloxyflavanc	-0.17	0.51	0.12	-0.02	Ring C
palmitic acid	0.00	0.99	-0.03	-0.02	Long Chain C
thymine	0.13	0.61	-0.15	-0.01	Nucleic Acid
myo-inositol	0.01	0.99	-0.01	0.00	Ring C
epsilon-caprolactam	0.09	0.73	-0.06	0.01	Ring C
salicylic acid	0.00	1.00	0.03	0.01	Benzoic Acid
sophorose	0.11	0.67	-0.07	0.02	Disaccharide
lignoceric acid	-0.05	0.84	0.09	0.02	Long Chain C
stearic acid	0.05	0.85	0.02	0.03	Long-chain fatty acid
arachidic acid	-0.01	0.97	0.09	0.04	Long Chain C
docosenoic acid	-0.03	0.91	0.12	0.04	Long Chain C
isopentadecanoic acid	0.07	0.80	0.02	0.04	Long Chain C
4-aminobutyric acid	0.03	0.92	0.08	0.05	Amine
phosphoethanolamine	-0.20	0.43	0.32	0.06	phosphate
palmitoleic acid	0.04	0.87	0.09	0.06	Long Chain C
sucrose	0.05	0.86	0.15	0.10	Disaccharide
isoleucine	0.10	0.69	0.09	0.10	Amine
adenosine	-0.07	0.79	0.31	0.12	Nucleic Acid
phytosphingosine	0.08	0.75	0.16	0.12	Amino Alcohol
valine	0.07	0.77	0.17	0.12	Amino Acid
threonine	0.09	0.72	0.16	0.12	Amine
heptadecanoic acid	0.10	0.68	0.17	0.14	Long Chain C
myristic acid	0.22	0.37	0.12	0.17	Long Chain C
tyrosine	0.13	0.60	0.31	0.22	Amino Acid
ethanolamine	0.22	0.38	0.28	0.25	Amine
beta-sitosterol	0.31	0.21	0.22	0.27	Ring C
leucine	0.29	0.24	0.25	0.27	Amine
mannose	0.25	0.31	0.33	0.29	Monosaccharides
phenylalanine	0.42	0.09	0.17	0.30	Amine
xylose	-0.05	0.84	0.70	0.33	Monosaccharides
pentadecanoic acid	0.41	0.09	0.26	0.34	Long Chain C
oleic acid	0.43	0.07	0.37	0.40	Long Chain C
isoheptadecanoic acid	0.39	0.11	0.43	0.41	Long Chain C
erythrose	**0.38**	**0.12**	**0.71**	**0.55**	**Monosaccharides**

Supp. Table 2-2, contd.: Metabolite peak value correlations with Forest ANF, ordered by the most significantly negative overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

ANF Forest					
	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
squalene	-0.27	0.28	-0.37	-0.32	Prenol Lipids
4-hydroxycinnamic acid	-0.19	0.45	-0.36	-0.28	Ring C
lyxitol	-0.07	0.80	-0.56	-0.31	Sugar Alcohol
proline	-0.33	0.18	-0.25	-0.29	Amino Acid
maltotriitol	-0.33	0.18	-0.23	-0.28	Ring C
uracil	-0.08	0.74	-0.45	-0.27	Nucleic Acid
pentadecanoic acid	-0.23	0.36	-0.30	-0.27	Long Chain C
myristic acid	-0.30	0.23	-0.17	-0.24	Long Chain C
urea	-0.12	0.64	-0.29	-0.20	Amino Acid
behenic acid	-0.12	0.64	-0.28	-0.20	Long Chain C
3-aminoisobutyric acid	-0.33	0.19	-0.07	-0.20	Amine
myo-inositol	-0.12	0.64	-0.19	-0.15	Ring C
ethanolamine	-0.12	0.64	-0.17	-0.14	Amine
serine	-0.09	0.72	-0.20	-0.14	Amino Acid
galactinol	-0.20	0.42	-0.07	-0.14	Disaccharide
palmitoleic acid	-0.15	0.54	-0.11	-0.13	Long Chain C
xylose	-0.21	0.40	-0.05	-0.13	Monosaccharides
beta-sitosterol	0.00	1.00	-0.25	-0.12	Ring C
glycerol	-0.20	0.43	-0.04	-0.12	C other
glutamic acid	-0.05	0.83	-0.17	-0.11	Amine
fructose	-0.13	0.60	-0.09	-0.11	Disaccharide
erythrose	-0.14	0.57	-0.07	-0.11	Monosaccharides
alanine	-0.13	0.60	-0.08	-0.11	Amino Acid
guanine	-0.13	0.60	-0.07	-0.10	Nucleic Acid
isoheptadecanoic acid	-0.09	0.73	-0.12	-0.10	Long Chain C
cerotinic acid	-0.13	0.61	-0.06	-0.10	Long Chain C
epsilon-caprolactam	-0.04	0.88	-0.13	-0.09	Ring C
adenine	-0.08	0.76	-0.09	-0.08	Nucleic Acid
conduritol-beta-epoxide	-0.11	0.66	-0.05	-0.08	Ring C
isopentadecanoic acid	-0.08	0.75	-0.08	-0.08	Long Chain C
phytosphingosine	0.10	0.69	-0.25	-0.07	Amino Alcohol
hydroquinone	0.11	0.66	-0.26	-0.07	Phenol
maltose	-0.10	0.70	-0.04	-0.07	Disaccharide
tyrosine	-0.12	0.64	-0.02	-0.07	Amino Acid
adipic acid	0.08	0.75	-0.21	-0.06	Long Chain C
glucose	-0.10	0.68	-0.02	-0.06	Monosaccharides
UDP-N-acetylglucosamin	-0.14	0.57	0.03	-0.05	Pyrimidine nucl. sugars
melezitose	-0.13	0.60	0.03	-0.05	Trisaccharide
4-aminobutyric acid	-0.02	0.94	-0.06	-0.04	Amine
aniline	0.10	0.69	-0.17	-0.04	Amine
heptadecanoic acid	0.01	0.96	-0.08	-0.03	Long Chain C
beta-gentiobiose	-0.02	0.93	-0.03	-0.03	Disaccharide
mannose	0.05	0.83	-0.11	-0.03	Monosaccharides

Supp. Table 2-2, contd.: Metabolite peak value correlations with Forest ANF, ordered from least to most significantly positive overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray.

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
docosenoic acid	-0.13	0.61	0.08	-0.02	Long Chain C
glutaric acid	0.05	0.86	-0.09	-0.02	Long Chain C
sorbitol	-0.05	0.84	0.03	-0.01	Sugar Alcohol
oxoproline	0.02	0.93	-0.04	-0.01	Amino Acid
glycine	-0.01	0.97	0.00	0.00	Amine
dehydroabietic acid	-0.07	0.78	0.07	0.00	Ring C
beta-mannosylglycerate	0.03	0.92	0.00	0.01	Long Chain C
palmitic acid	-0.01	0.97	0.05	0.02	Long Chain C
1-monostearin	-0.06	0.82	0.10	0.02	Long Chain C
arachidic acid	0.12	0.62	-0.05	0.04	Long Chain C
phosphate	0.07	0.79	0.03	0.05	phosphate
glycerol-alpha-phosphate	0.09	0.72	0.02	0.06	phosphate
ribose	0.09	0.72	0.04	0.06	Monosaccharides
lignoceric acid	0.08	0.74	0.06	0.07	Long Chain C
leucine	0.03	0.91	0.11	0.07	Amine
valine	-0.05	0.85	0.19	0.07	Amino Acid
oleamide	0.10	0.68	0.05	0.08	Long Chain C
phosphoethanolamine	0.14	0.57	0.02	0.08	phosphate
pimelic acid	0.02	0.93	0.15	0.08	Fatty acid
stearic acid	0.12	0.62	0.05	0.09	Long-chain fatty acid
inositol-4-monophosphat	0.15	0.55	0.03	0.09	phosphate
phenylalanine	0.13	0.61	0.06	0.09	Amine
4-hydroxybenzoic acid	0.15	0.55	0.04	0.10	Ring C
oleic acid	-0.01	0.96	0.23	0.11	Long Chain C
xylulose	0.15	0.54	0.09	0.12	Monosaccharides
lactulose	0.12	0.64	0.14	0.13	Disaccharide
levoglucosan	0.24	0.33	0.03	0.14	Ring C
sophorose	0.17	0.51	0.12	0.14	Disaccharide
isoleucine	0.07	0.79	0.22	0.15	Amine
cholesterol	0.20	0.43	0.13	0.16	Ring C
glycerol-3-galactoside	0.23	0.35	0.11	0.17	Ring C
thymine	0.19	0.46	0.19	0.19	Nucleic Acid
nonadecanoic acid	0.24	0.34	0.15	0.19	Long Chain C
threonine	0.09	0.72	0.32	0.20	Amine
tagatose	0.30	0.22	0.12	0.21	Monosaccharides
sucrose	0.32	0.20	0.14	0.23	Disaccharide
capric acid	0.26	0.29	0.19	0.23	Long Chain C
salicylic acid	0.29	0.24	0.17	0.23	Benzoic Acid
phthalic acid	0.11	0.66	0.37	0.24	Ring C
stigmasterol	0.37	0.13	0.16	0.26	Steroid
3-(4-hydroxyphenyl)prop	0.21	0.39	0.33	0.27	Ring C
phytol	0.16	0.52	0.41	0.29	Diterpenoid
terephthalic acid	0.37	0.13	0.21	0.29	Dicarboxylic acid
adenosine	0.28	0.26	0.32	0.30	Nucleic Acid
lauric acid	0.43	0.07	0.20	0.31	Long Chain C
glyceric acid	0.32	0.19	0.34	0.33	C other
4',5-dihydroxy-7-glucosyl	0.25	0.31	0.44	0.35	Ring C
succinic acid	**0.52**	**0.03**	**0.42**	**0.47**	**Dicarboxylic acid**

Supp. Table 2-2, contd.: Metabolite peak value correlations with Pasture *nifH* copies, ordered by the most significantly negative overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	Spearman	Est. Pearson	Mean Estimate	Molecular.Class
tagatose	**-0.76**	**0.00**	**-0.70**	**-0.73**	**Monosaccharides**
4-hydroxybenzoic acid	**-0.69**	**0.00**	**-0.51**	**-0.60**	**Ring C**
lyxitol	**-0.61**	**0.01**	**-0.55**	**-0.58**	**Sugar Alcohol**
sorbitol	**-0.55**	**0.02**	**-0.50**	**-0.53**	**Sugar Alcohol**
isopentadecanoic acid	**-0.54**	**0.02**	**-0.40**	**-0.47**	**Long Chain C**
succinic acid	**-0.49**	**0.04**	**-0.44**	**-0.46**	**Dicarboxylic acid**
behenic acid	-0.27	0.28	-0.57	-0.42	Long Chain C
capric acid	-0.47	0.05	-0.36	-0.42	Long Chain C
1-monostearin	-0.38	0.12	-0.45	-0.42	Long Chain C
glycerol	-0.40	0.10	-0.43	-0.41	C other
pimelic acid	-0.43	0.08	-0.37	-0.40	Fatty acid
glucose	-0.40	0.10	-0.37	-0.39	Monosaccharides
cerotinic acid	-0.39	0.11	-0.38	-0.38	Long Chain C
phthalic acid	-0.34	0.17	-0.39	-0.36	Ring C
hydroquinone	-0.38	0.12	-0.32	-0.35	Phenol
stigmasterol	-0.38	0.12	-0.30	-0.34	Steroid
aniline	-0.37	0.13	-0.29	-0.33	Amine
4',5-dihydroxy-7-glucosyloxyflavanone	-0.35	0.15	-0.31	-0.33	Ring C
terephthalic acid	-0.20	0.43	-0.40	-0.30	Dicarboxylic acid
dehydroabietic acid	-0.29	0.25	-0.29	-0.29	Ring C
ethanolamine	-0.26	0.29	-0.27	-0.27	Amine
glycine	-0.28	0.25	-0.22	-0.25	Amine
cholesterol	-0.28	0.26	-0.23	-0.25	Ring C
glutamic acid	-0.23	0.37	-0.28	-0.25	Amine
fructose	-0.28	0.26	-0.22	-0.25	Disaccharide
4-aminobutyric acid	-0.23	0.36	-0.26	-0.25	Amine
3-(4-hydroxyphenyl)propionic acid	-0.24	0.33	-0.24	-0.24	Ring C
glyceric acid	-0.28	0.26	-0.19	-0.24	C other
xylulose	-0.13	0.60	-0.28	-0.21	Monosaccharides
ribose	-0.12	0.64	-0.26	-0.19	Monosaccharides
levoglucosan	-0.12	0.63	-0.24	-0.18	Ring C
epsilon-caprolactam	-0.19	0.46	-0.16	-0.17	Ring C
palmitic acid	-0.04	0.87	-0.24	-0.14	Long Chain C
alanine	-0.17	0.49	-0.08	-0.13	Amino Acid
stearic acid	-0.07	0.79	-0.17	-0.12	Long-chain fatty acid
3-aminoisobutyric acid	-0.14	0.58	-0.08	-0.11	Amine
squalene	-0.04	0.87	-0.17	-0.11	Prenol Lipids
arachidic acid	-0.05	0.84	-0.13	-0.09	Long Chain C

Supp. Table 2-2, contd.: Metabolite peak value correlations with Pasture *nifH* copies, ordered from least to most significantly positive overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
oxoproline	-0.05	0.85	-0.02	-0.03	Amino Acid
lauric acid	0.01	0.97	-0.04	-0.02	Long Chain C
sucrose	-0.09	0.71	0.09	0.00	Disaccharide
palmitoleic acid	0.10	0.68	-0.09	0.01	Long Chain C
phytol	0.00	1.00	0.02	0.01	Diterpenoid
phosphate	0.00	0.99	0.03	0.01	phosphate
maltose	-0.04	0.86	0.09	0.02	Disaccharide
melezitose	0.08	0.76	-0.03	0.02	Trisaccharide
adipic acid	0.00	0.99	0.06	0.03	Long Chain C
mannose	0.10	0.70	-0.01	0.04	Monosaccharides
4-hydroxycinnamic acid	0.07	0.77	0.04	0.06	Ring C
beta-sitosterol	0.08	0.75	0.04	0.06	Ring C
erythrose	0.14	0.58	-0.01	0.07	Monosaccharides
phosphoethanolamine	0.15	0.56	0.00	0.07	phosphate
docosenoic acid	-0.02	0.93	0.18	0.08	Long Chain C
glutaric acid	0.09	0.74	0.08	0.08	Long Chain C
galactinol	0.15	0.55	0.05	0.10	Disaccharide
beta-mannosylglycerate	0.15	0.55	0.06	0.10	Long Chain C
inositol-4-monophosphate	0.04	0.86	0.18	0.11	phosphate
phytosphingosine	0.27	0.27	-0.03	0.12	Amino Alcohol
myristic acid	0.20	0.43	0.09	0.14	Long Chain C
lignoceric acid	0.15	0.54	0.13	0.14	Long Chain C
UDP-N-acetylglucosamine	0.13	0.60	0.16	0.15	Pyrimidine nucl. sugars
adenine	0.14	0.59	0.16	0.15	Nucleic Acid
salicylic acid	0.21	0.41	0.11	0.16	Benzoic Acid
proline	0.16	0.54	0.22	0.19	Amino Acid
tyrosine	0.15	0.56	0.23	0.19	Amino Acid
valine	0.13	0.61	0.27	0.20	Amino Acid
urea	0.25	0.31	0.20	0.22	Amino Acid
beta-gentiobiose	0.24	0.34	0.23	0.23	Disaccharide
lactulose	0.25	0.32	0.24	0.24	Disaccharide
phenylalanine	0.16	0.52	0.39	0.28	Amine
glycerol-3-galactoside	0.24	0.34	0.32	0.28	Ring C
nonadecanoic acid	0.26	0.29	0.32	0.29	Long Chain C
oleic acid	0.38	0.12	0.21	0.30	Long Chain C
glycerol-alpha-phosphate	0.28	0.27	0.39	0.33	phosphate
adenosine	0.40	0.10	0.28	0.34	Nucleic Acid
isoleucine	0.27	0.28	0.42	0.34	Amine
conduritol-beta-epoxide	0.34	0.17	0.35	0.35	Ring C
guanine	0.26	0.29	0.51	0.39	Nucleic Acid
heptadecanoic acid	0.39	0.11	0.40	0.40	Long Chain C
leucine	0.34	0.17	0.47	0.40	Amine
maltotriitol	0.40	0.10	0.46	0.43	Ring C
sophorose	0.47	0.05	0.40	0.44	Disaccharide
uracil	**0.31**	**0.20**	**0.59**	**0.45**	**Nucleic Acid**
thymine	**0.47**	**0.05**	**0.59**	**0.53**	**Nucleic Acid**
isoheptadecanoic acid	**0.59**	**0.01**	**0.54**	**0.57**	**Long Chain C**
myo-inositol	**0.59**	**0.01**	**0.66**	**0.62**	**Ring C**

Supp. Table 2-2, *contd*.: Metabolite peak value correlations with Forest *nifH* copies, ordered by the most significantly negative overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
pimelic acid	**-0.39**	**0.11**	**-0.52**	**-0.46**	**Fatty acid**
glycerol	**-0.33**	**0.19**	**-0.58**	**-0.45**	**C other**
galactinol	**-0.42**	**0.08**	**-0.41**	**-0.41**	**C other**
glycine	-0.35	0.15	-0.43	-0.39	Disaccharide
glyceric acid	-0.14	0.57	-0.62	-0.38	Amine
guanine	-0.43	0.08	-0.32	-0.38	Nucleic Acid
oleamide	-0.27	0.27	-0.46	-0.37	Long Chain C
succinic acid	-0.31	0.21	-0.42	-0.37	Dicarboxylic acid
1-monostearin	-0.44	0.07	-0.28	-0.36	Long Chain C
urea	-0.36	0.14	-0.34	-0.35	Amino Acid
phthalic acid	-0.40	0.10	-0.30	-0.35	Ring C
stearic acid	-0.29	0.25	-0.39	-0.34	Long-chain fatty acid
inositol-4-monophosphate	-0.34	0.16	-0.33	-0.33	phosphate
dehydroabietic acid	-0.18	0.46	-0.40	-0.29	Ring C
melezitose	-0.34	0.17	-0.18	-0.26	Trisaccharide
epsilon-caprolactam	-0.13	0.61	-0.38	-0.25	Ring C
levoglucosan	-0.22	0.37	-0.27	-0.25	Ring C
fructose	-0.17	0.50	-0.33	-0.25	Disaccharide
oxoproline	-0.17	0.51	-0.29	-0.23	Amino Acid
erythrose	-0.26	0.30	-0.19	-0.22	Monosaccharides
sorbitol	-0.05	0.85	-0.38	-0.21	Sugar Alcohol
3-(4-hydroxyphenyl)propionic acid	-0.14	0.59	-0.29	-0.21	Ring C
stigmasterol	-0.17	0.50	-0.23	-0.20	Steroid
phosphate	-0.20	0.42	-0.18	-0.19	phosphate
beta-gentiobiose	-0.20	0.43	-0.18	-0.19	Disaccharide
UDP-N-acetylglucosamine	-0.28	0.26	-0.08	-0.18	Pyrimidine nucl. sugars
adenine	-0.10	0.69	-0.26	-0.18	Nucleic Acid
glycerol-3-galactoside	-0.24	0.34	-0.12	-0.18	Ring C
palmitic acid	-0.22	0.38	-0.14	-0.18	Long Chain C
terephthalic acid	-0.15	0.55	-0.20	-0.17	Dicarboxylic acid
4',5-dihydroxy-7-glucosyloxyflavanone	-0.09	0.72	-0.26	-0.17	Ring C
lauric acid	-0.16	0.52	-0.14	-0.15	Long Chain C
beta-mannosylglycerate	-0.17	0.51	-0.13	-0.15	Long Chain C
salicylic acid	-0.21	0.40	-0.07	-0.14	Benzoic Acid
maltose	-0.05	0.84	-0.23	-0.14	Disaccharide
4-hydroxycinnamic acid	-0.21	0.41	-0.07	-0.14	Ring C
sucrose	-0.27	0.28	0.00	-0.14	Disaccharide
serine	-0.16	0.53	-0.11	-0.13	Amino Acid
phosphoethanolamine	-0.03	0.92	-0.21	-0.12	phosphate
thymine	-0.01	0.97	-0.23	-0.12	Nucleic Acid
uracil	0.05	0.86	-0.28	-0.11	Nucleic Acid
lyxitol	-0.01	0.98	-0.21	-0.11	Sugar Alcohol
hydroquinone	-0.12	0.63	-0.10	-0.11	Phenol

Supp. Table 2-2, *contd*.: Metabolite peak value correlations with Forest *nifH* copies, ordered from least to most significantly positive overall correlation (averaging Pearson and Spearman estimates [both shown]). Any metabolite with average estimates below |0.44| is in gray

	Est. Spearman	pval Spearman	Est. Pearson	Mean Estimate	Molecular.Class
lactulose	-0.20	0.42	-0.01	-0.10	Disaccharide
glycerol-alpha-phosphate	-0.02	0.94	-0.17	-0.10	phosphate
conduritol-beta-epoxide	-0.06	0.81	-0.12	-0.09	Ring C
aniline	-0.09	0.73	-0.06	-0.07	Amine
adenosine	-0.19	0.46	0.04	-0.07	Nucleic Acid
glucose	-0.07	0.79	-0.07	-0.07	Monosaccharides
docosenoic acid	-0.12	0.64	0.01	-0.05	Long Chain C
xylose	0.06	0.81	-0.15	-0.04	Monosaccharides
capric acid	-0.19	0.46	0.11	-0.04	Long Chain C
phytol	-0.13	0.61	0.06	-0.03	Diterpenoid
tagatose	0.00	1.00	-0.06	-0.03	Monosaccharides
maltotriitol	-0.23	0.35	0.21	-0.01	Ring C
mannose	0.12	0.64	-0.12	0.00	Monosaccharides
ethanolamine	0.07	0.80	-0.06	0.00	Amine
palmitoleic acid	-0.25	0.32	0.26	0.01	Long Chain C
4-hydroxybenzoic acid	0.18	0.47	-0.16	0.01	Ring C
cholesterol	0.05	0.84	-0.02	0.01	Ring C
behenic acid	-0.13	0.60	0.16	0.02	Long Chain C
nonadecanoic acid	0.12	0.64	-0.08	0.02	Long Chain C
isoheptadecanoic acid	0.10	0.69	-0.04	0.03	Long Chain C
sophorose	0.17	0.51	-0.10	0.03	Disaccharide
alanine	-0.05	0.84	0.13	0.04	Amino Acid
ribose	0.10	0.69	0.02	0.06	Monosaccharides
isoleucine	-0.01	0.97	0.14	0.07	Amine
oleic acid	-0.04	0.88	0.19	0.07	Long Chain C
xylulose	0.21	0.40	-0.03	0.09	Monosaccharides
4-aminobutyric acid	0.03	0.92	0.17	0.10	Amine
3-aminoisobutyric acid	-0.01	0.96	0.24	0.11	Amine
glutamic acid	0.20	0.43	0.05	0.13	Amine
threonine	0.08	0.76	0.19	0.13	Amine
valine	0.07	0.79	0.22	0.14	Amino Acid
lignoceric acid	-0.01	0.99	0.33	0.16	Long Chain C
phytosphingosine	-0.11	0.65	0.51	0.20	Amino Alcohol
myo-inositol	0.37	0.13	0.03	0.20	Ring C
squalene	0.00	0.99	0.43	0.22	Prenol Lipids
proline	0.28	0.26	0.19	0.23	Amino Acid
glutaric acid	0.12	0.64	0.35	0.24	Long Chain C
tyrosine	0.16	0.54	0.35	0.25	Amino Acid
cerotinic acid	0.30	0.23	0.31	0.30	Long Chain C
leucine	0.27	0.27	0.33	0.30	Amine
myristic acid	-0.08	0.75	0.75	0.33	Long Chain C
isopentadecanoic acid	0.47	0.05	0.22	0.34	Long Chain C
heptadecanoic acid	0.27	0.28	0.53	0.40	Long Chain C
phenylalanine	0.24	0.33	0.60	0.42	Amine
pentadecanoic acid	**0.03**	**0.90**	**0.86**	**0.45**	**Long Chain C**
adipic acid	**0.07**	**0.78**	**0.88**	**0.48**	**Long Chain C**
arachidic acid	**0.30**	**0.23**	**0.81**	**0.56**	**Long Chain C**
beta-sitosterol	**0.41**	**0.09**	**0.80**	**0.60**	**Ring C**

Supp. Table 2-3: Results of ANOVAs of non-significant forest models for *nifH* copies and ANF rate as response variables. Variable selection was performed with **(a)** physicochemical parameters only and **(b)** metabolites as well as physicochemical parameters.

a

	Physicochemical parameters only							
	Variable Stats				Model Stats			
nifH For.	Estimate	Std. Err.	F-val	p-val	F-Stat	p-val	RMSE	*R^2
(Intercept)	2.9E+08	1.7E+08			2.0	0.17	7E+07	.1
Clay +Silt	5.3E+06	3.0E+06	0.22	9.3E-02				
$\delta^{15}N$	-5.2E+07	2.7E+07	3.75	7.2E-02				
ANF For.	Estimate	Std. Err.	F-val	p-val	F-Stat	p-val	RMSE	*R^2
(Intercept)	0.812	0.113			2.5	0.14	0.2	.08
TN/EON	-0.013	0.008	2.45	0.137				

b

	Physicochemical and Metabolites							
	Variable Stats				Model Stats			
ANF For.	Estimate	Std. Err.	F-val	p-val	F-Stat	p-val	RMSE	*R^2
(Intercept)	4.9E-01	1.98E-01			3.2	0.07	0.16	0.2
TN/EON	-1.2E-02	7.7E-03	2.84	0.11		*		
Succinic acid	7.1E-05	3.8E-05	3.50	0.08				

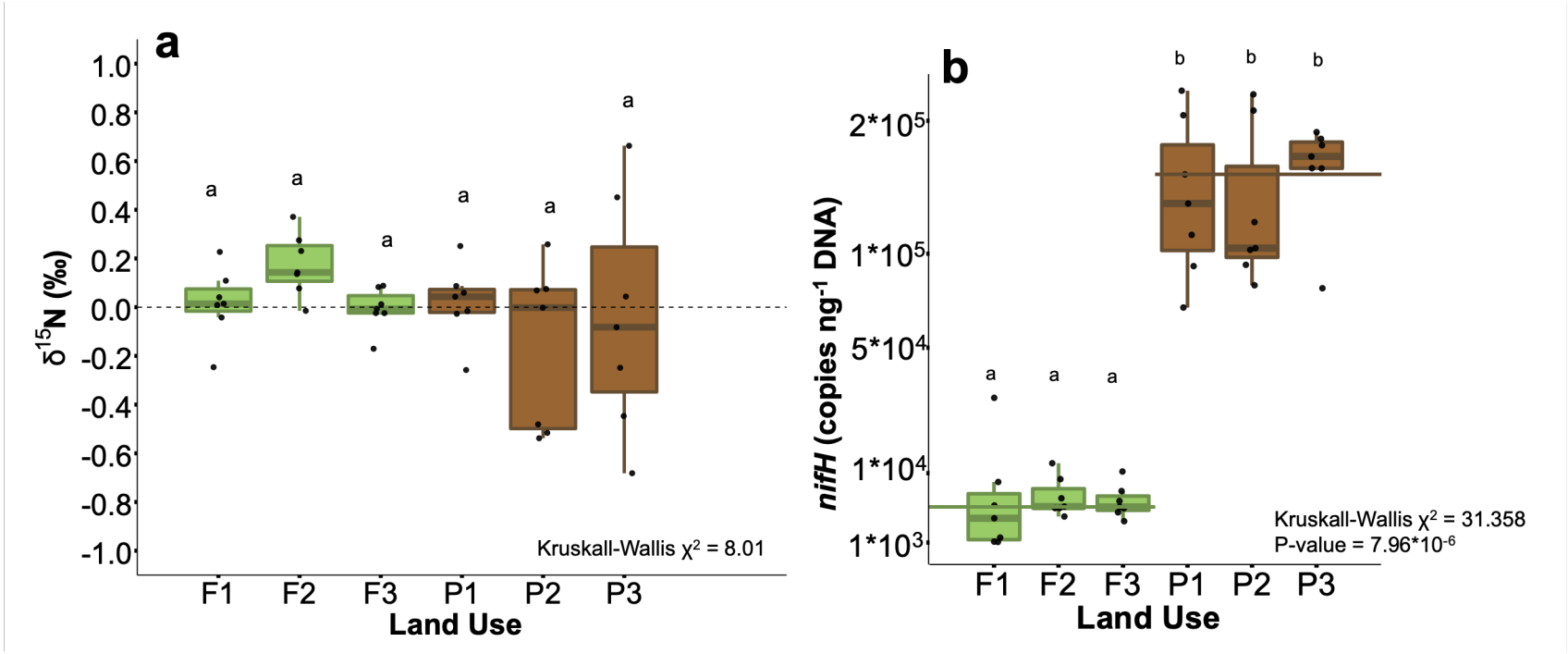
Abbreviations: ANF = asymbiotic nitrogen fixation rate (ng N g^{-1} soil day^{-1}); Clay/Silt (% composition); EOC= Extractable organic C (µg g^{-1} soil); EON = Extractable organic N (µg g^{-1} soil); Fval = F-value; F-stat = F statistic; Metabolites (peak values); Pval = p-value; RMSE = Root mean squared error; Std. Err. = Standard error; TN/EON = Total: extractable organic N (organic + inorganic; µg g^{-1} soil)

Supp. Table 2-4: Gene counts, normalized as counts per million for each DIG, using GTA metagenomic annotation, organized by nitrogen fixation-related function. FC compare forest and pasture replicates using a genewise negative binomial likelihood test. p-values are false discovery rate corrected using the Benjamini-Hochberg method. Bold text represents genes with log_2-fold change greater than 1 and $p < 0.05$.

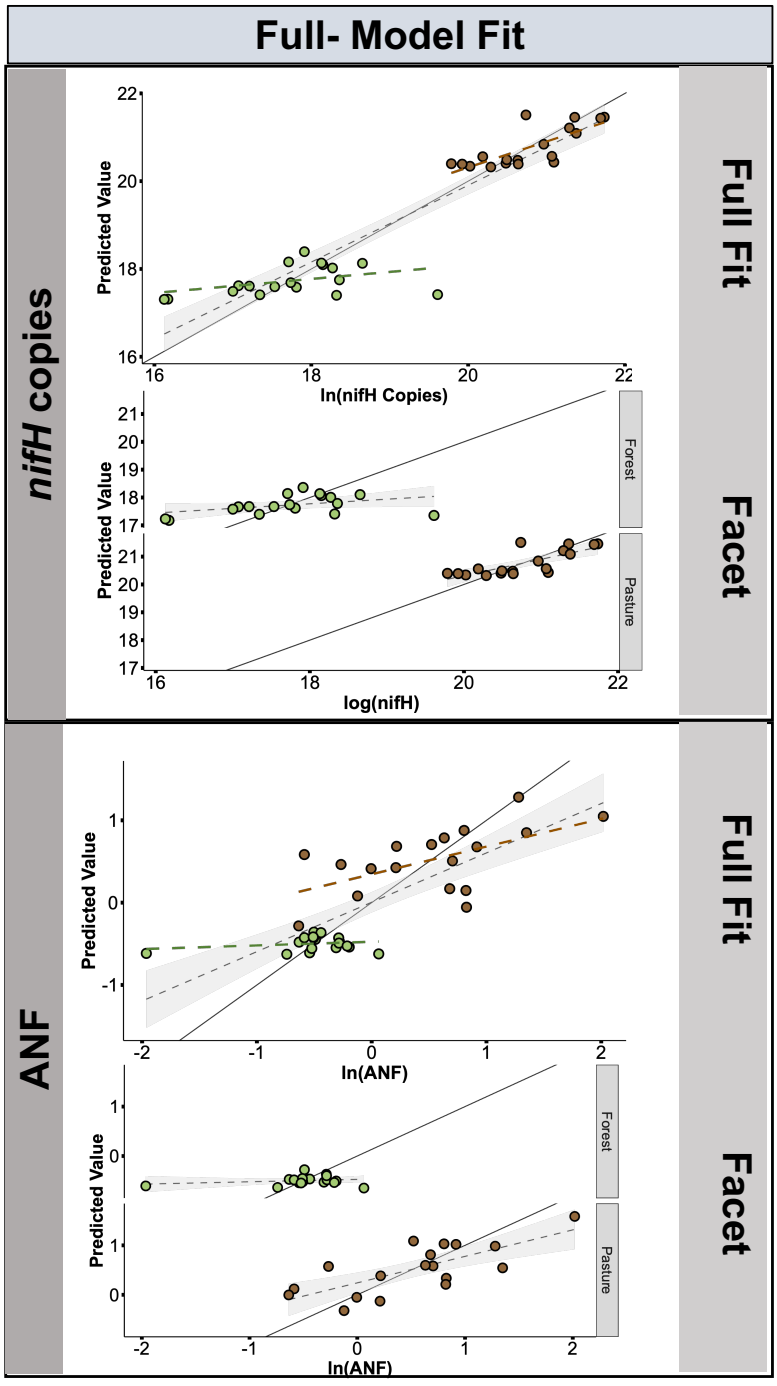
		Forest			Pasture				
Functional group	**Gene**	**1**	**2**	**3**	**1**	**2**	**3**	**Log_2FC**	***p-value**
Sensory/ Regulation	*fixK*	5.05±0.35	7.45±0.34	5.91±0.4	5.9±0.45	4.88±0.30	5.17±0.47	0.22	2.1E-01
	nifA	232.7±11.9	187.1±1.9	329.3±12.	329.4±16.6	286.2±10.3	341.9±9.1	0.50	6.3E-08
	nifL	0.36±0.07	0.25±0.04	0.31±0.04	0.63±0.10	0.44±0.05	0.67±0.07	0.82	1.0E-02
	nifHD1	**0.12±0.03**	**0.07±0.02**	**0.19±0.03**	**0.32±0.04**	**0.29±0.04**	**0.19±0.03**	**1.45**	**1.2E-03**
	nifHD2	**0.08±0.01**	**0.25±0.03**	**0.075±0.02**	**0.26±0.02**	**0.12±0.01**	**0.18±0.04**	**1.72**	**2.0E-04**
Nitrogenase Metallocluster formation	***nifD***	**0.70±0.22**	**0.76±0.08**	**1.17±0.23**	**20.1±1.6**	**16.7±1.6**	**18.9±0.95**	**4.81**	**3.7E-17**
	nifK	**0.60±0.16**	**0.81±0.1**	**0.98±0.20**	**19.8±1.5**	**15.6±1.50**	**17.8±1.3**	**4.99**	**2.9E-17**
	nifH	**1.94±0.11**	**2.42±0.11**	**2.02±0.28**	**15.80±0.9**	**13.89±1.3**	**14.8±0.8**	**3.03**	**3.8E-18**
Nitrogenase Cofactor Biosynthesis	***nifU***	**1.1±0.07**	**0.9±0.05**	**1.4±0.08**	**3.02±0.26**	**2.3±0.2**	**2.5±0.18**	**1.48**	**1.8E-11**
	nifS	56.7±1.7	48.4±1.0	68.9±1.85	80.3±2.1	75.4±1.7	83.3±2.1	0.50	1.6E-09
	nifN	**0.8±0.11**	**0.8±0.07**	**0.7±0.07**	**7.1±0.4**	**5.2±0.4**	**4.9±0.6**	**3.17**	**1.4E-17**
	nifB	**0.7±0.2**	**1.05±0.1**	**1.1±0.2**	**14.5±1.2**	**12.01±0.86**	**12.9±1.1**	**4.27**	**1.1E-15**
	nifE	**0.91±0.35**	**1.08±0.13**	**1.4±0.11**	**20.8±1.28**	**17.57±1.8**	**22.3±1.2**	**4.48**	**4.0E-17**
	nifZ	**0.05±0.02**	**0.1±0.05**	**0.12±0.03**	**1.7±0.2**	**1.18±0.12**	**1.4±0.1**	**4.99**	**1.6E-13**
	nifQ	**0.06±0.02**	**0.18±0.05**	**0.11±0.018**	**1.1±0.131**	**0.64±0.1**	**0.8±0.16**	**4.14**	**2.9E-12**
	nifV	12.4±0.44	8.84±0.2	11.7±0.23	16.8±0.42	14.7±0.5	17.5±0.4	0.45	2.6E-03
	nifX	**0.71±0.05**	**0.8±0.06**	**1.4±0.17**	**2.9±0.22**	**2.15±0.22**	**2.5±0.24**	**2.06**	**3.1E-13**
	nifT	**0.08±0.02**	**0.11±0.02**	**0.14±0.02**	**0.83±0.10**	**0.61±0.08**	**0.95±0.14**	**3.32**	**1.8E-11**
Nitrogenase Stabilization	***nifW***	**.044±0.02**	**0.03±0.02**	**0.11±0.05**	**0.6±0.10**	**0.5±0.08**	**0.63±0.11**	**3.60**	**6.2E-08**
Flavodoxin	***nifJ***	**38.3±2.5**	**42.44±0.8**	**65.9±2.8**	**163.3±4.5**	**147.6±5.5**	**144.8±3.8**	**2.10**	**8.8E-27**
	nifF	**0.13±0.03**	**0.06±0.01**	**0.14±0.03**	**0.29±0.08**	**0.2±0.045**	**0.28±0.08**	**1.17**	**1.3E-02**
Alternative Nitrogenase metallocluster formation	***vnfD***	**0.004±0.004**	**0.02±0.01**	**0±0**	**0.10±0.04**	**0.11±0.03**	**0.10±0.03**	**3.83**	**1.3E-03**
	vnfG	0.14±0.041	0.12±0.03	0.12±0.019	0.15±0.021	0.22±0.032	0.11±0.01	0.11	7.8E-01
	vnfH	**0.003±0.003**	**0.003±0.003**	**0.004±0.004**	**0.024±0.01**	**0.012±0.01**	**0.002±0.002**	**2.42**	**1.6E-02**
	vnfK	0.14±0.02	0.21±0.02	0.09±0.01	0.16±0.03	0.18±0.03	0.28±0.09	0.27	5.4E-01
	anfG	0.003±0.003	0±0	0.01±0.007	0±0	0±0	0.01±0.006	-1.09	3.3E-01
nod Sensory/ Regulation	*nodD*	0.25±0.04	0.22±0.05	0.12±0.06	0.32±0.04	0.21±0.03	0.46±0.03	0.41	3.3E-01
Plant-targeted mitotic signal	*nodA*	0±0	0±0	0±0	0.004±0.004	0.01±0.013	0±0	1.12	3.7E-01
	nodB	0.016±0.007	0.006±0.004	0.002±0.002	0.01±0.007	0.012±0.016	0.02±0.007	-0.46	6.3E-01
Plant-targeted Nodulation signal	***nodC***	**0±0**	**0.003±0.003**	**0±0**	**0.026±0.02**	**0.01±0.006**	**0.024±0.012**	**3.65**	**1.3E-02**
	nodE	4.01±0.37	2.66±0.08	4.12±0.17	4.36±0.08	3.8±0.3	4.7±0.22	0.12	4.1E-01
	nodF	0.06±0.02	0.04±0.02	0.04±0.01	0.12±0.04	0.09±0.03	0.09±0.02	0.93	1.8E-01

Supp. Table 2-5: Results from Spearman correlation analysis between identified *nif* genes and all other genes from RA metagenomic analysis, cross referencing genomic occurrence within diazotrophic genomes **(external .xlsx)**.
[a]p-values were adjusted with a Bonferroni control for multiple comparisons, based on 5,152 comparisons performed.

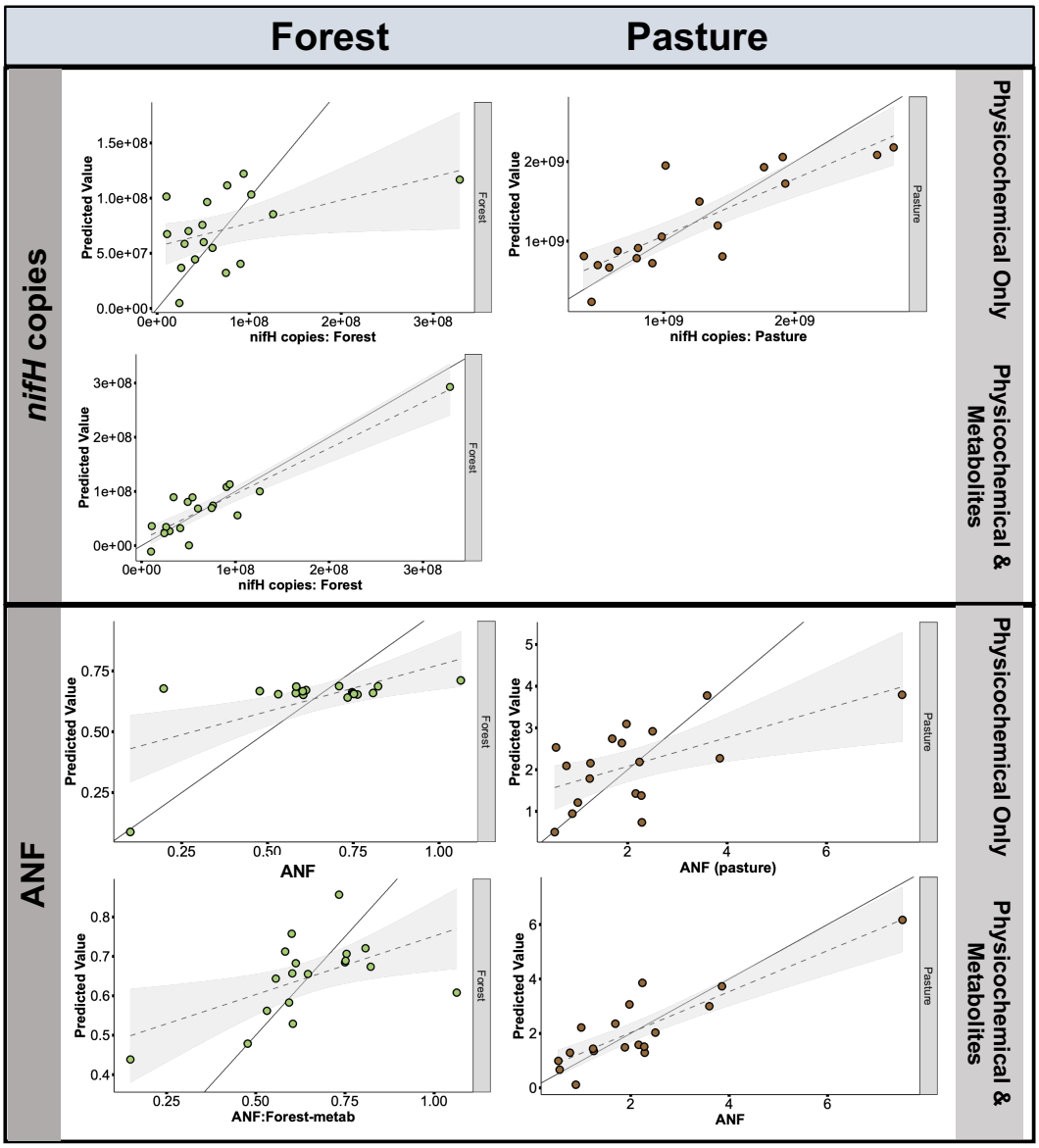
Supp. Figure 2-1: ^{15}N natural abundance from incubations and copy numbers on DNA basis**.** Boxplots represent data for each forest and pasture site sampled, with overlayed individual measurements represented by black points. **(a)** Difference in pre- and post-incubation natural abundance of ^{15}N (i.e., field collected soil versus experimental controls). Permille values in reference to atmospheric standard. Horizontal dashed line indicates no change. Kruskall-Wallis rank sum test and Dunn's *post-hoc* test indicates no significant difference across site groups. **(b)** *nifH* gene copy numbers derived from quantitative polymerase chain reaction, normalized by ng community DNA. Letters above each bar represent Dunn's *post-hoc* grouping.



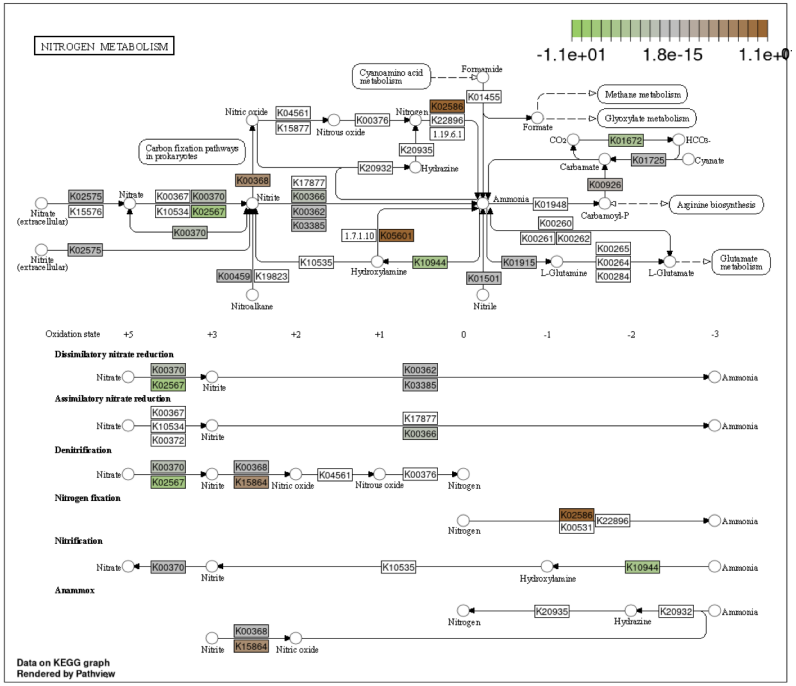
Supp. Figure 2-2: Goodness of fit figures for LU unifying models presented in Figure 2-3. Solid lines represent a perfect fit (e.g., model $R^2 = 1.0$). Plots are overlaid with a dashed linear regression smoothing function as seen in Figure 2-3. For detail, each model is shown together as well as faceted across LUs



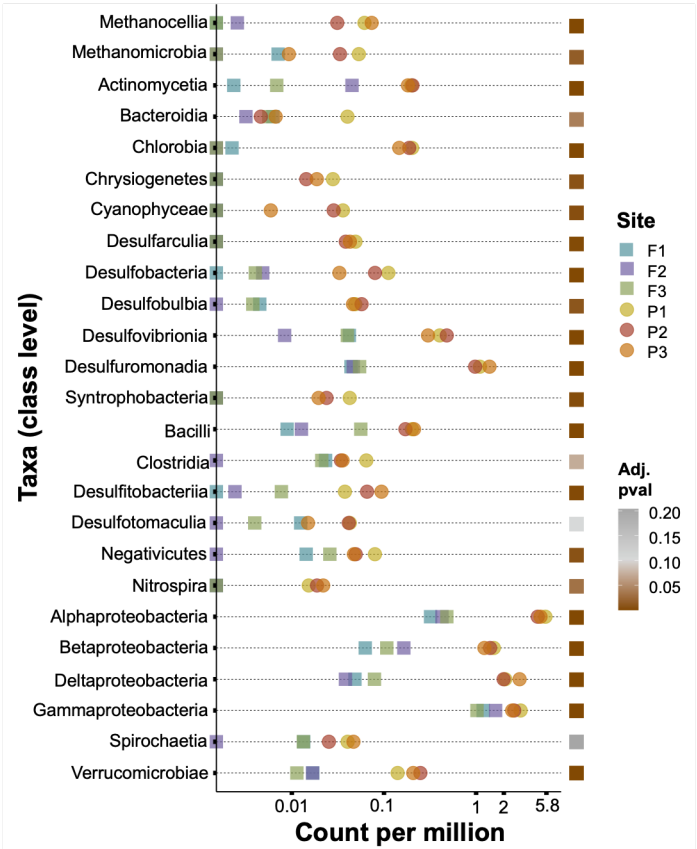
Supp. Figure 2-3: Goodness of fit figures models built separately for each LU, with and without metabolites. Solid lines represent a perfect fit (e.g., model $R^2 = 1.0$). Plots are overlaid with a dashed linear regression smoothing function as seen in Figure 2-4. The missing *nifH* pasture plot indicates inclusion of metabolites in variable selection did not improve model fit



Supp. Figure 2-4: KEGG Nitrogen Metabolism pathways overlaid with genewise negative binomial likelihood $\log_2$ fold-change statistics, derived from assembly-based metagenomic annotation.



Supp. Figure 2-5: Taxonomic composition of GTA-based read counts, filtering for *nifHDK* – positive taxa. Taxonomic assignments of reads are aggregated at the class level. Colored boxes to the right indicate FDR-adjusted p-values based on quasi-likelihood negative binomial tests.



Chapter 3: Transcriptionally active soil diazotroph community composition provides indication of asymbiotic nitrogen fixation rate under land-use change in the Brazilian Amazon

Rachel E. Danielson[1], Jorge L. Mazza Rodrigues[1,2] *

[1] Department of Land, Air, and Water Resources, University of California – Davis, Davis, CA, USA

[2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA USA.

ORCID #

Rachel Danielson: 0000-0001-7044-0747

Jorge L.M. Rodrigues: 0000-0002-6446-6462

***Corresponding author:** jmrodrigues@ucdavis.edu; University of California – Davis, 1110 Plant and Environmental Sciences Building, Davis CA 95616

Abstract

Biological nitrogen fixation is crucial in determining the nitrogen status of many natural and unmanaged landscapes, but it is a highly regulated reaction, given the steep energy input required. In the Amazon Basin of Brazil, the establishment of cattle pasture is a dominant driver of primary forest loss. This land use (LU) change stimulates the augmentation of diazotrophic community abundance and significantly increases soil asymbiotic nitrogen fixation (ANF). While major shifts in the community structure of *potential* diazotrophs (i.e., DNA-based profiles) have also been shown to occur with forest-to-pasture LU change, it is not clear how this functional group response corresponds to the soil prokaryotic community response overall. It is also unknown how the community structure of potential diazotrophs compares to that of metabolically active diazotrophs undergoing nitrogenase-subunit transcription (RNA-based profiles) both within and across LUs. To address these knowledge gaps, we co-extracted RNA and DNA from soils of pastures and primary forests and characterized the total prokaryotic and diazotrophic communities via high-throughput amplicon sequencing. We found that, in comparison to the significant increase in alpha-diversity (via Shannon index) observed for *nifH*-D communities in association with pasture conversion, prokaryotic community alpha-diversity did not change. In contrast, both community profiles exhibited a significant decline in group dispersion (i.e., spatial homogenization) in pastures compared to forests, and both exhibited significant alterations in community composition. Physicochemical conditions associated with compositional shifts were similar for both DNA communities, including pH, clay, sulfur, and inorganic N content. Metrics of community structure across the LU dichotomy differed substantially for *nifH*-R compared to *nifH*-D communities: alpha-diversity was similar in pastures and forests, group dispersion increased in pastures, and community composition showed minimal differentiation between the two. The OTU-specific profiles of active diazotrophs were surprisingly much more similar across LUs than they were to their paired *nifH*-D profiles from the same soil sample. In contrast to *nifH*-D, compositional shifts across *nifH*-R communities were strongly correlated with ANF rates (Pearson's r = 0.88). Further, the relative abundance of 17 out of 882 unique OTUs (dominated by *Bradyrhizobium* and Enterobacteriaceae) within the *nifH*-R community served as a significant indicator of ANF rates. We concluded that while actively *nifH*-transcribing communities showed a surprisingly limited

response to LU change, their community composition was more closely related to functional group activity than potential soil diazotrophs.

Importance

Land use change in the Amazon Basin has profound implications for ecosystem-scale nutrient cycling. In the heavily deforested Amazon state of Rondônia, stimulation of biological nitrogen fixation by free-living diazotrophs is one of largest such biogeochemical impacts of forest conversion to cattle pasture. The magnitude of this effect likely depends on factors including the intensity of management practices and soil physicochemical conditions, given that some pastures remain productive without fertilizers for decades, while others degrade in just a few years. Understanding the interaction of soil diazotrophic communities, their environment, and rates of nitrogen fixation in mature, active pastures is a foundational step in assessing how natural populations of soil diazotrophs may be utilized for sustainable nutrient management in rangeland agriculture. Further, this study evaluates both potential and active asymbiotic diazotrophic communities, an approach that is exceedingly rare in soil systems. This represents a crucial expansion in our understanding of the ecology of this poorly studied functional group.

Keywords: Free-living diazotroph, Asymbiotic diazotroph, Asymbiotic nitrogen fixation, *nifH* community, soil RNA, Brazilian Amazon, land-use change, cattle pasture

Abbreviations: **ANF:** asymbiotic nitrogen fixation; **BH:** Benjamini-Hochberg; **CT:** community type; **F:** forest; **FC:** $\log_2$fold-change; **GD:** geographic distance; **IT:** indicator taxa; **KW:** Kruskall-Wallis**; LU:** land use; ***nifH*-D:** present diazotroph community; ***nifH*-R:** nitrogenase-transcribing diazotroph community; **P:** pasture; **PC:** principal coordinates; **Supp**: Supplementary; **VE:** variation explained

Introduction

Land-use (LU) change for the establishment of cattle pasture in the Amazon Basin is common and ongoing, having claimed an estimated 230,000-275,000 km^2 of primary forest thus far (Tyukavina et al., 2017). This large-scale ecosystem transition has consequences for biodiversity as well as hydrologic, energy, and nutrient cycling both for pastures themselves, as well as for remaining forests in the region (Longo et al., 2020). The N cycle in particular shifts drastically with forest-to-pasture conversion, as evidenced by large declines in rates of mineralization, nitrification, and denitrification, especially as pastures age (Neill et al., 1997; Piccolo et al., 1994a; Piccolo et al., 1994b; Verchot et al., 1999). In aggregate, this suggests LU change to pasture imposes N limitations for both plant and soil microbial community metabolisms (both assimilatory and dissimilatory), as compared to their primary forest counterparts (Dias-Filho et al., 2001). In conjunction with this tightening of the N cycle, biological nitrogen fixation- a microbial process that is crucial in providing new N to natural or unfertilized systems- appears to be upregulated in mature, active pastures, as evidenced by significant increases in soil diazotroph population size (10-30x; Mirza et al., 2014; Chapter 2) and asymbiotic (free-living and root-associated) nitrogen fixation (ANF) rate (47x; Chapter 2; Figure 2-1a,b).

It has also been shown, based on DNA-derived *nifH* gene amplicon sequencing, that pasture conversion drives augmentation of potential diazotroph alpha-diversity, as well as significant compositional shifts (Mirza et al., 2020). We refer to these as 'potential' diazotrophs for several reasons: First, because nitrogenase-encoding genes are remarkably widespread across the prokaryotic phylogenetic tree due to extensive horizontal gene transfer (Koirala and Brozel, 2021; Zehr and Turner, 2001), and genomes containing vestigial nitrogenase marker genes (e.g., *nifH*) do not always contain the minimal gene set required for nitrogen fixation activity (Aramaki et al., 2020; Wang et al., 2013). Second, because those taxa which are genetically capable of N fixation also assume a variety of trophic strategies, whereby acquiring N through ANF may be of variable priority based on environmental conditions and exogenous nutrient concentrations (Le Boulch et al., 2019; Smercina et al., 2019). Therefore, it is not clear whether the substantial changes in the community structure of potential diazotrophs (i.e., *nifH*-D) with LU change (and its associated shifts in physicochemical conditions such as

inorganic N pool size) are unique to this functional group, or are a reflection of how total prokaryotic communities respond.

Further, it is unknown how representative *nifH*-D communities are of those which are not just metabolically active, but also actively transcribing nitrogenase enzyme subunits in preparation for catalyzing the dinitrogen reductase reaction across this LU dichotomy (i.e., *nifH*-R communities). Aside from the reality that just a small fraction of the total microbial soil community is assumed to be metabolically-active at any given time (Kuzyakov and Blagodatskaya, 2015), physiological variation amongst potential diazotrophs may greatly impact the structure of active diazotrophs. For example, active diazotrophs are sensitive to exogenous O_2 concentrations given that the molecule irreversibly inactivates the nitrogenase enzyme (Robson and Postgate, 1980)- but depending on an organism's trophic strategy, O_2 can be an important metabolic electron acceptor (Blagodatskaya and Kuzyakov, 2013). Pure culture studies have shown there to be substantial variation in the O_2 concentration for optimal nitrogenase activity (Khadem et al., 2010; Okon et al., 1976; Volpon et al., 1981) amongst diazotrophic taxa. Additionally, active soil-dwelling diazotrophs are likely to prioritize the uptake of available inorganic and low-molecular-weight organic N sources before investing in the energetically expensive process of ANF (Norman and Friesen, 2017; Smercina et al., 2019). For instance, taxa have been shown to vary in their capacity to fix N_2 in the presence of environmental NH_3 (Dekaezemacker and Bonnet, 2011; Silva et al., 2011), and may be capable of taking up organic sources of N directly (Geisseler et al., 2009). Diazotrophs even exhibit differential tolerance to physicochemical conditions like free aluminum concentration and soil acidity when fixing N_2, an observation particularly relevant to tropical soils (Avelar Ferreira et al., 2012). These conditions impacting nitrogen fixation capacity disproportionately across potential diazotrophs can also shift rapidly at nanoscale soil structures, indicating transient and complex selection mechanisms for active asymbiotic diazotrophs in particular. Taken together, this suggests *nifH*-R profiles may vary significantly from their *nifH*-D reservoir communities within LUs, and that, in consideration of the dramatic soil physicochemical changes observed with forest-to-pasture conversion (Chapter 2, Figure 2-2), *nifH*-R communities could exhibit more drastic LU impacts than *nifH*-D communities.

Understanding how diazotrophic community and environmental dynamics impact ANF- derived N inputs to active pastures is of crucial importance: Fertilizer amendments are uncommon in the region due

to economic constraints, and an appreciable loss of N (or alternatively, a replenishment deficit) from pasture systems ultimately leads to a decline in forage grass productivity, landscape degradation, and potentially abandonment in favor of newly-established pastures (Asner et al., 2004). Given the sensitivities of potential asymbiotic diazotrophs to external conditions as described above, it is of interest to determine whether the community structures of *nifH*-D and *nifH*-R communities are related to functional group activity, particularly because previous work has shown that while an increase in the absolute abundance of *nifH*-D communities (i.e., copies per gram soil) with pasture conversion occurs concomitantly with elevated ANF rates, the two metrics show no meaningful relationship of scale within LUs (Chapter 2, Figure 2-1c).

This study aimed to assess the relationships between total present and active communities across a primary forest-pasture LU dichotomy. Our first aim was to directly compare *nifH*-D to the total present prokaryotic (based on 16S rRNA) community structure across LUs, hypothesizing that *nifH*-D community diversity and structure shifts independently of the 16S community as a whole, in association with soil physicochemical conditions favorable to diazotrophic communities in pastures. Our second aim was to explore how *nifH*-D community structure compares to that of *nifH*-R communities at the time of sampling both within (as paired samples) and across LUs, expecting substantial differences between the two profiles within samples of the two respective LUs. We also expected a more pronounced differentiation in *nifH*-R community structure as compared to *nifH*-D across LUs, based on the significantly higher rates of ANF measured in pastures compared to primary forests. Finally, we sought to ascertain whether *nifH*-D or *nifH*-R community metrics relate to ANF rates, hypothesizing *nifH*-R metrics to serve as stronger indicators of activity.

Materials and Methods

Site Description and Sample Preservation

Soil samples were collected in March and April of 2017 (at the end of the rainy season) near Ariquemes, Rondônia, Brazil. For thorough site descriptions, see Chapter 2 ‘Materials and Methods’. Briefly, six sites, three primary forests (F) and three pastures (P; converted in or around 1972) were

sampled within an established 100 m^2 quadrat. Nucleic acid samples were collected at each corner point of the quadrat (n=24). This sample number was calculated based on US regulatory limitations of live soil transport in a liquid medium. At each sampling location, surface debris was removed and four soil cores 2.5 cm in diameter were extracted (as a block) from 0-10 cm. For RNA preservation, cores were placed in a sterile, RNase-free bag immediately after extraction and homogenized mechanically for 30 s. Two subsamples of 2.5 g were transferred to tubes containing 5 ml LifeGuard Soil Preservation solution (Qiagen, Hilden, Germany), gently shaken, and kept frozen at -20° C. The remaining soil was kept cool (~4-5° C), then sieved to 2 mm (with rocks and roots removed) and either frozen within 24 h or used in ANF rate incubations using $^{15}N_2$- labeled assays (described in Chapter 2, Supplementary [Supp.] Methods). Additional protocols pertaining to soil physicochemical analyses are also described in Chapter 2. Physical measurements included textural size classification. Chemical measurements included C and N pools (combustion-derived), potassium sulfate extractable organic C and N as well as inorganic N (NH_4^+ and NO_3^-), permanganate oxidizable C, loss-on-ignition organic matter, pH, extractable acidity, cation exchange capacity (CEC), ^{15}N and ^{13}C natural abundance, Mehlich-3 extractable acidity, Ca, Mg, K, Cu, Zn, S, and orthophosphate P, as well as nitric acid- digested Al, Fe, Mo, and V.

DNA and RNA isolation and processing

RNA and DNA were co-extracted from four 1.25 g replicate aliquots of LifeGuard-preserved soil using the Qiagen (Hilden, Germany) RNeasy PowerSoil Total RNA and DNA elution kits, respectively. Following elution, community DNA and RNA were quantified using Qubit dsDNA and RNA High-sensitivity assays, respectively (Invitrogen, Waltham, MA), and replicate DNA extractions were pooled unless yields were unusually low (i.e., less than 20% of average yield). DNA was then purified using a Genomic DNA Cleanup and Concentrator-10 kit (Zymo Research, Irvine, CA, USA).

For RNA, trace genomic DNA was removed using the TURBO DNA-free kit (Invitrogen), then samples were purified and concentrated using the Qiagen MinElute Cleanup kit using a modified two-step elution at 55% the total recommended elutant volume each. A NanoDrop Spectrophotometer (ThermoFisher, Waltham, MA) was used to verify the RNA quality using 260/280 and 260/230 ratios. After quality was checked, concentrations were standardized, and reverse transcription of 200 ng RNA was

performed in 20 µL reactions using SuperScript IV single-strand reverse transcriptase (Invitrogen) following manufacturer instructions, with the following gene-specific conditions: The *nifH3* primer (5' ATRTTRTTNGCNGCRTA 3'; Zani et al., 2000) was used for gene-specific reverse transcription at a concentration of 150 nM in reaction and an incubation temperature of 52.5 °C, as previously described by Calderoli et al. (2017). cDNA synthesis failed for one pasture sample due to insufficient quantity and quality of RNA, so 23 rather than 24 samples were utilized for downstream processing and analysis. Negative control reactions were run to ensure the absence of DNA. Both the 16S and *nifH* genes were quantified via qPCR using gene-specific primers but unfortunately, attempts to amplify *nifH* transcripts from cDNA using the qPCR-specific master mix resulted in a degree of primer dimer formation which interfered with accurate quantification.

Amplicon sequencing of the *nifH* gene was performed for both community DNA and cDNA. The gene was targeted using nifH1 (5'-TGYGAYCCNAARGCNGA-3') and nifH2 (5'-ADNGCCATCATYTCNCC-3' ; Zehr and McReynolds, 1989) primers. nifH1 included an inline 12 bp barcode on the 5' end, specific to each sample. PCR reactions were performed in triplicate in 40 µl volumes, including 4.5 µl cDNA template (or 4 µl DNA template at 2 ng μl^{-1}), 20 µl HotStarTaq mastermix, 4 µl each of forward and reverse primer (1.0 µM in reaction), and 7.5 µl PCR-grade water (7 µl for DNA reactions). The thermocycler was run in accordance with HotStarTaq recommendations: 95°C for 15 min, 40 cycles of 94°C for 0:45 s, 61°C for 1:00, and 72°C for 1:00, followed by a final extension at 72°C for 10 min. The pooled library was submitted to the UC Davis DNA Technologies Core facility (Davis, CA) for PippinHT size selection (Sage Science, Beverly, MA), Illumina (San Diego, CA) sequence adapter ligation, and final sequencing on an Illumina MiSeq 250 bp paired-end run. The 16S rRNA sequence library was prepared using Illumina-specific primers as described in Caporaso et al. (2011) and run on a separate MiSeq lane.

High-throughput sequencing and bioinformatic pipeline

Sequence processing for DNA- (*nifH*-D) and RNA- (*nifH*-R) based communities was performed using a modified version of the pipeline described in Gaby et al. (2018). Modifications included using DADA2 (Callahan et al., 2016) for sequence denoising, and utilizing a Hidden Markov Model built with

nifH sequences to perform frameshift correction and to filter non-*nifH* protein sequences, as described in Angel et al. (2018a). Sequences were clustered at 95% similarity with dual-strand search, and clusters with five or fewer sequences were removed. A custom script was used to configure a final taxonomic table using percent ID cutoffs specified in Gaby et al. (2018): 91.9% for species, 88.1% for genus, and 75% for family. 16S sequence processing was conducted entirely in DADA2 using the same settings described for the *nifH* pipeline. 16S-D ASVs were annotated taxonomically in DADA2 using the silva_nr99_v138.1_train_set.fa reference database. Additional procedural detail concerning DNA/RNA isolation and processing, reverse-transcription, qPCR, and amplicon library preparation and sequence processing are available in **Supp. Methods**.

Statistical Procedures

Shannon diversity was calculated for 16S-D (DNA), *nifH*-D and *nifH*-R (RNA) sequence profiles using the 'vegan' package (Oksanen et al., 2013) in R. Although sequences were processed differently for 16S and *nifH*-D communities, comparisons of 16S clusters vs ASVs have shown both are sensitive to differences in alpha- and beta- diversity across environments (Chiarello et al., 2022). Since the goal of diversity metric measurements in this study was to compare relative differences across LUs rather than absolute value differences between community profiles, the use of differential taxa definitions should not impact the conclusions of the analysis made. Differences across sites and LUs were tested using a Kruskall-Wallis (KW) rank sum test with a Dunn's *post-hoc* to test for LU grouping, and a Benjamini-Hochberg (BH) correction for multiple comparisons. Diversity was then correlated with physicochemical variables (with natural-log transformation [ln] for variables with 10x or greater LU differences) as well as prokaryotic and diazotrophic indicators (gene copies in ng^{-1} DNA and ANF rate). A significant Pearson's r cutoff of |0.55| was established using BH p-value correction, but relationships were further scrutinized using a simultaneous Spearman's ρ cutoff > |0.45| to account for potential outlier effects driving Pearson values, as well as visual trend inspection.

For community-level compositional analysis, dissimilarity was calculated by repeatedly (n=100) rarefying sample profiles to an even depth and applying a Hellinger transformation as well as a Bray-Curtis distance metric. Samples were ordinated using unconstrained principal coordinates (PC) and

partial-constrained distance-based redundancy analysis (using the *capscale* function in 'vegan'). The first two PC- based ordinal axes were plotted to assess sample variation across LUs for each community. To statistically test for the degree of in-group variation, multivariate dispersion, a measure of beta-diversity (Anderson et al., 2006) was calculated for each community with respect to site and LU. Permutation and Kruskall-Wallis rank sum (plus Dunn's *post-hoc*) tests were used to determine significance. PERMANOVAs (using function *adonis2* in 'vegan') were used to test for the effect of LU change on community composition by first accounting for the potentially confounding effect of variation in geographic distance (GD) amongst sampling locations within and across LUs. Total, shared, and non-redundant variation explained (VE) was then calculated. Physicochemical variables were tested for non-redundant VE by first accounting for LU and GD, and the 'vegan' function *envfit* was used to assess correlations with community structure. The DNA- and RNA-based *nifH* communities were also considered together (*nifH* Total) for community- level composition testing to determine whether community type (CT; *nifH*-D vs. *nifH*-R communities) had a larger or smaller effect as compared to LU. Similar analyses as those described above were performed for *nifH* Total profiles, including beta-dispersion, environmental correlation, and PERMANOVA group testing, with the added test of CT effect, as well as quantifying pairwise distance across LU and CT groupings.

The remaining analyses focused on just the *nifH*-based communities. Initial *nifH*-specific comparisons were made at the OTU level. Two effects and four total comparisons were tested: To assess LU effect, differences between F and P communities were tested for (1) DNA (F-DNA and P-DNA) and (2) RNA- based communities (F-RNA and P-RNA), respectively. To test for pairwise (i.e., within sample) CT effect, differences within DNA- and RNA- based communities were assessed for (3) F (F-DNA and F-RNA) and (4) P (P-DNA and P-RNA), respectively. For each comparison, 'co-occurring' (i.e., testable) OTUs were determined based on their presence in two or more samples in each treatment, and presence in ca. 1/3 of samples across treatments. This largely eliminated rare OTUs, retaining on average 91$\pm$0.05% of OTU abundance across samples. Differential abundance was tested using negative quasi-likelihood binomial tests as implemented in the R package 'edgeR' (Robinson et al., 2010), accounting for sitewise replicate nesting. P-values were adjusted using a BH correction. Within each pair of effect tests (LU or CT), OTUs common to both comparisons were assessed for agreement or disagreement in

significance and trend, based on p- values and directionality of log_2-fold change (FC) values. Total OTUs were aggregated to the family level (most conservative classification based on percent identity in annotation procedure) and were assessed statistically if they were present in ca. 1/3 of samples and accounted for at least 1000 total sequences in the data set. The same four differential abundance tests were made as previously performed at the OTU level, using an adjusted p-value cutoff of < 0.05 and an FC cutoff of > |0.5|.

Finally, indicator analysis was performed to identify any OTUs significantly associated with ANF rate in the *nifH*-R community (based on previous findings from community-level composition). Methods were adapted from those described in Smercina et al. (2021). First, samples were binned into rate categories (ng N fixed g $soil^{-1}$ day^{-1}) including ND (no detection; 0 or negative rates measured; n=3), X-Low (x<0.1; n=5), Low (0.1<=x<0.3; n=6), Medium (0.3<=x<1.5; n=5), and High (x>=1.5; n=4). Bins were set based on natural break points as assessed by a rate-index plot (**Supp. Figure 3-1**) while attempting to balance the number of samples in each bin. The analysis was performed using the 'indicspecies' package function *multipatt* (De Caceres et al., 2016) with 10,000 permutations and a maximum of 3 rate-bin associations allowed. Significant indicator OTUs (IT) were selected based on the following association combinations: High, Medium, High + Medium, and High + Medium + Low. ITs were then assessed for their impact on community structure, their overall relative abundance by rate group, and their aggregate relative abundance relationship with ANF process rates.

Results

Community diversity and compositional patterns across prokaryotes and diazotrophs

Shannon diversity indices differed significantly across sites for *nifH*-D (KW χ^2= 19.8; p-value [p-val] = 0.001) and marginally for 16S-D (KW χ^2= 9.4; p-val = 0.09), but only *nifH*-D reflected a meaningful LU effect in these differences, with mean (± standard error) values of 2.5 ± 0.24 for forest (F) compared to 4.4 ± 0.11 for pasture (P) soils, respectively (**Figure 3-1a**). This difference in LU trend between *nifH*-D and 16S-D communities is also reflected by absolute qPCR-derived quantification of the two genes across LUs (**Supp. Figure 3-2**). In contrast, to *nifH*-D, *nifH*-R communities did not exhibit a significant site

effect, nor an appreciable difference in F (3.2 ± 0.09) and P (3.6 ± 0.15) mean values. Additional metrics including Chao1, evenness, and inverse Simpson index reflected similar trends for *nifH*-D and *nifH*-R communities (**Supp. Figure 3-3**). Of the 26 physicochemical parameters and diazotrophic indicators (gene copy numbers and ANF measurements) correlated against Shannon diversity, a total of 9 showed a significant correlation with *nifH*-D communities (Pearson and Spearman values > |0.5|; **Supp. Table 3-1a,b**). Physicochemical variables including NH_4^+ and pH correlated positively with *nifH*-D diversity, while ln(NO_3^-), inorganic: total N (In:TN), S, clay, and $\delta^{15}N$ correlated negatively (**Figure 3-1b**). Some of these correlations showed no clear scaling within LUs (e.g., NH_4^+, ln(NO_3^-), In:TN), while others were driven mainly by the diversity of F communities (e.g., pH, clay). Diazotrophic indicators including ln(*nifH* copies); (Pearson's r = 0.8) and ANF rate (r = 0.57) both correlated positively with *nifH*-D diversity, but without clear linear scaling within LUs. *nifH*-R communities reflected similar but non-significant (e.g., Pearson's r < 0.55, Spearman's ρ < 0.45) trends with several of the *nifH*-D - correlated variables, but contrastingly, 16S-D community diversity did not display correlation coefficients greater than |.25| for any of the above variables, nor did it display any other significant associations. Diversity did not correlate significantly between any of the three community profiles surveyed (16S, *nifH*-D and *nifH*-R; **Supp. Table 3-1a,b**).

For both DNA-based communities, approximately half of the total community compositional variation explained (VE) could be visualized on the first two principal coordinate (PC) axes (45.9% for 16S and 47% for *nifH*; **Figure 3-2a, b**), and both communities reflected a high degree of separation between LUs along the first axis. Surprisingly, the VE by geographic distance (GD) and LU for community composition of 16S-D and *nifH*-D communities were nearly identical: Together they accounted for ca. 45% of the compositional variation, with LU contributing significantly after partitioning GD and shared variance (approximately 12%; F-stat = 4.17; p-val < 0.002; **Table 3-1**). Dispersion of 16S-D communities was modestly higher among F compared to P soils (constrained correspondence permutation F-stat = 7.6; p-val= 0.01). At the site level, however, 16S-D dispersion did not group significantly by LU. Dispersion of *nifH*-D communities with respect to LU reflected a similar trend of dissimilarity-based homogenization across LUs (F-stat = 52.1; p-val= $1*10^{-4}$; **Table 3-2)**. However, in contrast to total prokaryotes, *nifH*-D community dispersion at the site level largely grouped by LU, with lower within-site dispersion among P replicates (F-stat = 16.08; p-val= $1*10^{-4}$;)

The first two PCs of the *nifH*-R compositional ordination accounted for a notably lower proportion of total community variation as compared to *nifH*-D (30.8% vs 47%). Separation of *nifH*-R samples by LU was also much less pronounced (**Figure 3-2c**): GD and LU together accounted for 18.5% of community compositional variation, and LU (unique + shared) offered one-quarter the VE of *nifH*-D communities (5.9% F-stat = 1.37; p-val =0.08; **Table 3-1**). In stark contrast to *nifH*-D, which exhibited a major decline in LU-based group dispersion (0.43$\pm$.011 for F samples compared to 0.31$\pm$.014 for P samples), *nifH*-R communities were more dissimilar among P samples (0.36$\pm$.015) than F samples (0.28$\pm$.01; F-stat = 19.5; p-val = $4*10^{-4}$; **Table 3-2**). Site-wise dispersion varied significantly (F-stat = 6.5; p-val = $2.4*10^{-3}$) but was not strictly grouped by LUs. When *nifH*-D and *nifH*-R communities were ordinated together, separation along the first PC (31.1%) was driven mainly by DNA/RNA community type (CT) rather than LU (**Figure 3-2d**). Overall, GD and LU had limited VE over community composition (10.8%), while *nifH*-D vs *nifH*-R CT played a larger role (28.4%; F-stat = 19.2; p-val = 0.001; **Table 3-1**). Dispersion was higher overall for F soils (0.47$\pm$0.02 vs. 0.41$\pm$0.014; F-stat = 10.22; p-val = $2.6*10^{-3}$), with group variance driven by F- DNA (F-stat = 29.03; p-val = $1*10^{-4}$; **Table 3-2**). P- RNA communities were the second most dissimilar, higher than P- DNA and F- RNA. On a pairwise basis, dissimilarity was significantly higher between RNA-based communities and F-DNA compared to P-DNA (KW $\chi 2$ =213; p <0.001; **Figure 3-2e**).

DNA-based (16S and *nifH*) community composition reflected significant correlation and VE (envfit r > 0.5; PERMANOVA p<0.1) with a similar subset of physicochemical and microbial indicator variables, including pH, In:TN, NO_3^-, *nifH* -D Shannon diversity, S, and clay content (with the latter two showing nearly identical directionality, and opposite directionality with pH (**Figure 3-2 a,b**; **Supp. Table 3-2)**. The 16S-based communities additionally reflected a modest relationship (envfit r > 0.3; PERMANOVA p<0.1) with Al, Mg, and K, while *nifH*-D community composition showed a modest association with organic N (ON). In contrast, no physicochemical conditions correlated with *nifH*-R structure (**Figure 3-2c**). However, a key functional group indicator, ANF rate, showed a strong and significant correlation (r = 0.82) with *nifH*-R community structure (**Supp. Table 3-2**). This observation motivated subsequent analysis of RNA community composition in relation to diazotrophic activity, to be discussed below.

Diazotrophic compositional shifts in potential and active communities

Approximately 69% of co-occurring *nifH*-D OTUs (n= 393 of 2,374 total) showed no significant difference in relative abundance between LUs, whereas 27% were enriched in P soils and just 4% were enriched in F soils, with $\log_2$-fold changes (FCs) ranging from 2.5-15 (**Figure 3-3a;** for full statistics, see **Supp. Table 3-3**). RNA communities exhibited far fewer significantly differing OTUs between LUs with just 8.7% and 2.5% of co-occurring OTUs (n = 161 of 882 total) being enriched in P and F, respectively (**Figure 3-3b**). Of those OTUs co-occurring across LUs within DNA and RNA profiles (n=129), just six were commonly enriched in P soils, but no OTUs were found commonly enriched in DNA and RNA communities of F soils (**Supp Figure 3-4a**). When paired F samples were considered between community types (CTs), approximately even proportions of OTUs (~15.8%) were enriched in DNA and RNA, respectively, out of 152 co-occurring OTUs (**Figure 3-3c**). Paired pasture CT profiles showed a much greater proportion of DNA enrichment, 46.5%, compared to 4% for RNA- but those enriched in P-RNA tended to exhibit greater FCs and more significant p-values (**Figure 3-3d**). Of the OTUs common to both F and P CT comparisons (n=140), 11 were similarly enriched in RNA for both F and P diazotrophic communities, accounting for nearly all (85%) OTUs enriched in the latter comparison (**Supp Figure 3-4b**). OTUs commonly enriched in DNA (or alternatively, depleted in RNA; n=12) accounted for approximately half of those in F soils, but just a handful of those enriched in P-DNA.

Aggregating the relative abundance of OTUs annotated to the same taxonomic family revealed that virtually all communities (both *nifH*-D and *nifH*-R) were dominated by Nitrobacteraceae (on average 51.4 $\pm$ 2.8% of sequence abundance; **Figure 3-4**). The only two families exhibiting a significant difference between LUs in *nifH*-R communities (both enriched in P soils) were Leptolyngbaceae (n=10 OTUs) and Ectothiorhodospiraceae (n=16 OTUs, both represented by orange squares). In contrast, 14 families exhibited a significant LU effect in *nifH*-D (represented by orange circles), with just two groups (Clostridaceae FC = 4.9 and Spirochaetaceae FC = 2.2) enriched in F soils (green border). Statistically-robust P enrichment was observed in the high-abundance family Methylobacteriaceae (FC = 4.1; **Supp. Table 3-4**). Paired differential abundance within LUs and across CTs yielded very similar trends among families (represented by triangles), but CT effects were more prolific among taxa in P soils (brown fill). Higher-abundance families tended to be enriched in *nifH*-RNA of paired samples, while lower-abundance

families tended to be depleted in *nifH*-RNA, irrespective of LU. The largest relative abundance difference across CTs was observed within the family Aphanizomenonaceae, which exhibited a relative abundance enrichment in *nifH*-R of more than 100x compared to *nifH*-D. Methylobacteriaceae, Rhodocyclaceae, and Rhodobacteraceae were also significantly enriched in *nifH*-R by more than 10x across both LUs (**Supp. Table 3-4)**. Meanwhile, the most abundant group depleted in *nifH*-R of both F and P soils was the unclassified OTUs (~20x depletion).

ANF process rate indicators

The strong correlation between *nifH*-R community composition and ANF (**Figure 3-2c**) prompted indicator analysis of *nifH*-R OTUs across binned ANF rates. This procedure identified 34 indicator OTUs significantly associated with one, two, or three rate bins. Twelve were associated specifically with high rates (H), while another five were associated with high and medium (H+M), high medium and low (H+M+L), or medium (M) rates (total IT n=17). Across IT, the proportion of total community abundance ranged from 13.3 $\pm$ 5% among high-rate-associated samples to 2.1 $\pm$ 1.2% among no-detection-associated samples (**Figure 3-5a**). These 17 IT impacted community compositional dissimilarity with directional correspondence to increasing ANF rate across samples (**Figure 3-5b**), and their total aggregated relative abundance showed a robust linear correlation with ANF rate (Pearson's r = 0.88; **Figure 3-5c**). On individual bases, the *nifH*-R relative abundance of twelve IT showed significant positive linear correlations (r > 0.5) with ANF rate (**Supp. Figure 3-5**). One IT, annotated to the genus *Bradyrhizobium*, ranked in the top 3 OTUs by relative abundance across *nifH*-R (as well as *nifH*-D) P soil communities, scaling particularly well with rate (r = 0.75). An additional four IT ranked in the top 60 most abundant taxa (considered common) and exerted a strong effect on *nifH*-R community dissimilarity across P samples (see vector length in **Figure 3-5b**). The remaining were considered rare (rank > 100 of 882 OTUs; **Supp. Table 3-5**). At the family level, taxonomic composition of IT included (in order of mean abundance) Nitrospiraceae (n=4), Enterobacteriaceae, Azospriliaceae (n=3), Methylobacteriaceae (n=2), Chromatiaceae, Sphingomonadaceae, Rhizobiaceae, Desulfovibrinaceae (n=2), Frankiaceae, and Ectothiorhodospiraceae.

Discussion

Differential diversity, but similar compositional response to LU change in potential diazotrophs and total prokaryotes

In comparing potential diazotrophic (*nifH*-D) to total prokaryotic (16S-D) community profiles, we identified a disparity in community diversity trends across LUs (**Figure 3-1a**). The substantial increase in the *nifH*-D Shannon index we observed among P soils is consistent with a previous study conducted in the same region of Rondônia (Mirza et al., 2020). Targeting the total prokaryotic profile from the same DNA-based community (i.e., soil extraction) indicated that alpha- diversity of the broader community does not change with LU conversion. 16S-based analyses across the Amazon have produced mixed results concerning the LU effect on prokaryotic community diversity. Several studies have concluded an increase in diversity with pasture (Cerqueira et al., 2018; da C Jesus et al., 2009; Mendes et al., 2015; Navarrete et al., 2015; Ranjan et al., 2015; Rodrigues et al., 2013) while a handful of others, in agreement with this study, observed no overall change (Melo et al., 2021; Pedrinho et al., 2019), likely indicating that a high degree of spatial variability is influencing this metric. Further, while *nifH*-D diversity was linearly associated with several physicochemical parameters, prokaryotic diversity was not related to any environmental measures. Together this suggests that the soil conditions imparted by P conversion selects for diazotroph diversification rather than overall prokaryotic diversification. In this respect, both environments likely provide sufficient resources and dynamic conditions to support similar levels of local prokaryotic diversity (Taylor et al., 2014).

In contrast to their disparate alpha-diversity response, both community profiles exhibited distinct and significant alterations in community composition with LU change- explaining approximately 25.6% and 33.9% of community dissimilarity for 16S-D and *nifH*-D communities, respectively after partitioning GD effect (**Table 3-1**). This is perhaps unsurprising given that studies throughout the Amazon have observed significant compositional dissimilarity across LUs for virtually every microbial group or subgroup probed (see references within Danielson and Mazza Rodrigues, 2022). On the other hand, it was somewhat unexpected that the dissimilarity across samples of both DNA-based communities were associated with a similar subset of physicochemical parameters. pH, clay, and S content (which do not

differ significantly overall between LUs) were associated with both communities, contributing to differentiation *within* LUs. Notably, this subset of conditions was also linearly predictive of *nifH*-D diversity (**Figure 3-1b**). Given the collinearity of vector magnitude and direction in unconstrained ordinal plots, it is not possible to establish the most plausible driving force. pH, for example, can exert selection on prokaryotic communities particularly at lower pHs (Blagodatskaya and Anderson, 1998), but also influences clay surface (in this case, iron and aluminum oxides) charge, in turn impacting nutrient and organic matter adsorption (Sumner, 1963). One such nutrient whose adsorption (and speciation) it impacts is S (Gharmakher et al., 2009), which exhibited the most intra-LU linear correlation with *nifH*-D diversity. S concentration may be related to community composition and diversity through S metabolism: For example, in previous metagenomic profiling of diazotrophic communities across the LU dichotomy, we observed a strong correlation (driven by P soils) between nitrogenase subunit and dissimilatory sulfite reductase genes, indicating S metabolism may serve as an energy conservation mechanism for potential soil diazotrophs (Simon and Kroneck, 2013). Clay content may directly impact prokaryotic composition and potential diazotrophs through several mechanisms as well. First, it is strongly correlated with anoxic soil pore space volume within microaggregate structures, supporting microorganisms carrying out anaerobic metabolisms (Keiluweit et al., 2018). Second, nitrogen fixation potential has shown stimulatory response to iron (oxyhy)droxide clay content, which may form free-radicals and increase the bioavailability of molybdenum, a key metalloid co-factor of the most common nitrogenase enzyme isoform (Yu et al., 2021).

The ratio of inorganic to total N was the most strongly correlated condition separating communities directly by LU type, retaining marginally significant explanatory value after accounting for LU effect (**Supp. Table 3-2**). Considering the substantial shifts to the N cycle imparted by LU change (e.g., Neill, 1995), this is not surprising and indicates the entire prokaryotic community rather than just dissimilatory N cycling functional groups are influenced by shifts in reactive N pools. *nifH*-D community structure was also modestly associated with organic N content, potentially reflecting the variation in the preferential hierarchy of N source (Norman and Friesen, 2017). Prokaryotic community structure was uniquely associated with Mg and K content, potentially indicating sensitivity to plant nutrient status (Carvalho et al., 2009). The association we identified between 16S-D composition and soil Al content is a

common observation amongst Amazon-focused studies, which may represent a gradient of sensitivity to Al toxicity or exchange site occupation amongst taxa in association with pH (Auger et al., 2013; Carvalho et al., 2009).

Finally, we also found that both community profiles exhibited a decline in groupwise dispersion among P samples, despite greater average GD among P as compared to F replicates. This biotic homogenization effect on prokaryotic communities has been shown previously (Mirza et al., 2020; Rodrigues et al., 2013), and we demonstrate here that this trend is more pronounced in *nifH*-D communities, occurring at both the LU and the site level. This effect is presumably borne out of landscape-scale homogenization of the plant community in monoculture pastures compared to florally diverse primary forests (Rodrigues et al., 2013), leading to a homogenization of detrital resources and root architecture (Bacq-Labreuil et al., 2019). This presumption is also supported by reduced dispersion among P-derived untargeted soil metabolomic profiles (**Chapter 2, Figure 2-2b,c**). The elevated effect observed in *nifH*-D compared to prokaryotic communities may be related to the relatively narrower range of trophic strategies (Le Boulch et al., 2019) amongst potential diazotrophs which are promoted by P conditions. Overall, while *nifH*-D diversity reflects a strong and specific favoring of diazotrophs, community composition responds in a remarkably similar manner across DNA communities.

LU effect is diminished in active compared to potential diazotrophs

Given that DNA-based profiling can capture the community structure of dead, inactive, and active microorganisms (Blagodatskaya and Kuzyakov, 2013; Kuzyakov and Blagodatskaya, 2015), the phylogenetic breadth of trophic strategies encompassed by *nifH*-based survey (Koirala and Brozel, 2021; Le Boulch et al., 2019; Yan et al., 2022), and in light of the elevated rates of ANF measured from our survey of P soils (**Chapter 2**), we expected that community profiles of *nifH*-transcribing microorganisms would reflect similar or greater differentiation between LUs as compared to *nifH*-D. Instead, we found that *nifH*-R communities showed minimal distinction with respect to LU. First, there was no augmentation of the diversity index and there was no apparent relationship between *nifH*-R and *nifH*-D diversity. The LU dichotomy we studied imparted large disparities in edaphic conditions like inorganic N pools, micronutrient concentrations, as well as pH across LU types, and n*ifH*-D community diversity and

composition reflected several significant associations with these conditions, in line with other DNA-based studies of soil communities (Han et al., 2019; Hu et al., 2021; Liao et al., 2017; Wang et al., 2017). Although some variables including NH_4^+ and pH did reflect a similar (but non-significant) trend with *nifH*-R diversity (**Figure 3-1b**), none of the variables we measured offered appreciable VE to *nifH*-R. Taken together, this suggests that the nitrogenase-transcribing communities we surveyed across LU types were subject to different selective pressures than respective reservoir communities of potential diazotrophs.

Very few studies assessing the community composition of potential and active diazotrophs are focused on upland soils, narrowing our ability to compare our results in a broader context. One study investigating soil diazotrophic communities cultivated with different plant species (somewhat analogous to the LU context of the present study) found transcript quantity and *nifH*-R composition were actually more strongly shaped by plant type (the latter finding based on community clustering) than were potential communities (Bouffaud et al., 2016). Another study along a revegetation gradient found significant differentiation among both *nifH*-D and *nifH*-R communities with revegetated ecosystem age and plant community assemblage, indicating both community types were shaped by the differentiation of environmental factors across the gradient (Wang et al., 2016). The lack of clear environmental relationships with the *nifH*-R communities characterized in the present study suggests elevated importance of parameters that we did not, or could not measure, such as microsite conditions including temperature and O_2 concentration (e.g., Keiluweit et al., 2018; Kuzyakov and Blagodatskaya, 2015). These conditions may be greatly impacted by regional-scale seasonality (e.g., temperature and moisture regimes), owing an explanation to the relative similarity of CTs across the LU replicates we sampled (ca. 15 km range). Across the literature, while the importance of environmental O_2 content on active diazotroph assemblage has been acknowledged (Brown and Jenkins, 2014), C concentration and quality as well as inorganic N have more commonly been identified as key controls, in contrast with findings here (Bürgmann et al., 2005; Gonzalez Perez et al., 2014; Hu et al., 2021; Severin et al., 2015; Wakelin et al., 2010). It is also possible that nutrient flux rather than pool size may provide a greater indication of active communities.

Mirroring compositional patterns, the relative abundance LU shifts of specific OTUs and family-level aggregated taxonomic groups of *nifH*-R communities were minimal compared to *nifH*-D

communities. At the family level, the only consistent trend was P enrichment in Leptolyngbyaceae, a photoautotrophic, cyanobacterial group (Shimura et al., 2015) which presumably colonizes the surface of P soils where solar exposure is elevated compared to F soils. This highlights the importance of future study into the contribution of surface-dwelling and detrital communities to ANF, particularly in the context of variable grazing intensity in P systems, since animal density impacts surface trampling and detrital distribution as well as solar exposure (Nunes et al., 2019; Rauber et al., 2021). The only other family displaying differential abundance across F and P soils of *nifH*-R communities (and not reflected in *nifH*-D community profile shifts) was Ectothiorhodospiraceae (enriched in P), a group typified by anoxygenic, phototrophic purple sulfur bacteria (Imhoff, 2015).

In *nifH*-D communities, notable enrichments not mirrored in *nifH*-R communities included the families Methylobacteriaceae and Methylococcaceae, which are known for metabolizing C1- carbons including methane and methanol (Zhang et al., 2014), and Methylobacteriaceae in particular has oreviously been identified as an important constituent of grassland rhizospheres (Bahulikar et al., 2014). This is interesting given that previous research has revealed an increase in methane efflux concomitant with a decline in the community proportion and relative richness of total methanotrophs in P compared to F soils (Meyer et al., 2017; Meyer et al., 2020). P-DNA was also enriched in Desulfovibrionaceae (anaerobic sulfate-reducers; Galushko et al., 2020)- and Halorhodospiraceae, composed mainly of purple sulfur bacteria (Imhoff et al., 2022). In combination with the P-RNA enrichment of Ectothiorhodospiraceae, this adds more evidence to a potential interaction of S cycling and ANF in P systems.

Dissimilarity between active and potential communities within environments is stronger than LU effect

Active communities of each LU varied significantly from their respective potential diazotroph profiles, but differences were somewhat greater in F communities, where F-RNA composition was actually more similar to P-DNA than F-DNA (**Figure 3-2d,e**). Additionally, while community evenness was similar between P-DNA (0.66 $\pm$ 0.01), P-RNA (0.70 $\pm$ 0.02), and F-RNA (0.65 $\pm$ 0.02), it was substantially lower for F-DNA (0.43 $\pm$ 0.04). A possible explanation for this is that F-DNA communities may contain a

higher proportion of rare organisms bearing *nifH* genes, but either do not transcribe fixation genes under the conditions during which we sampled, or lack the full minimal gene set to actually perform nitrogen fixation (Dos Santos et al., 2012). Despite this difference in the extent of dissimilarity between present and active profiles within LUs, RNA-based profiles were overall much more like each other than they were to present communities of respective LUs. In a study comparing present and active diazotrophic communities in association with variable plant species, Bouffaud et al. (2016) similarly found that RNA-based community structure across plant species clustered more closely with each other than their respective potential, DNA-based communities. Another study of active and present diazotrophic communities along fertilization and seasonal gradients in Brazilian eucalyptus plantations similarly found that RNA -based community profiles were consistently more similar to each other across tested conditions compared to their paired DNA-based profiles (da Silva et al., 2016). In contrast, active communities along a revegetation chronosequence in the Tengger Desert showed greater similarity to corresponding potential communities rather than other active communities along the gradient (Wang et al., 2016).

The most drastic shift we observed between potential and active *nifH*-based communities was the 100x enrichment of a single OTU annotated within the family Aphanizomenonaceae, a group of photoautotrophic cyanobacteria (**Figure 3-4)**. Like the Leptolyngbyaceae enriched across P communities, this is presumably accounted for by cells dwelling at or near the soil surface as crusts (Alteio et al., 2020). Its dominance in the transcriptionally active fraction of diazotrophic communities may indicate an outsized importance of free-living photosynthetic diazotrophs to atmospheric N inputs within these environments. Owing to the ability of these filamentous microorganisms to function in high-O_2 environments by relegating nitrogen fixation to specialized cells (Rascio and La Rocca, 2013) as well as their independence from plant-derived C compounds as an energy source, it is perhaps unsurprising that cyanobacteria in general would account for a major portion of active diazotrophs (Bae et al., 2018; Belnap, 2001). This may in part explain the near-zero rates of ANF measured across F samples: Since rate assay incubations performed on our soils were conducted in the dark, any contributions made by photoautotrophs were missed entirely. Therefore, future investigations would benefit from assays performed under light and dark conditions.

nifH communities in relation to ANF rates: nifH-R indicators

Although the correlation between the *nifH*-D Shannon Diversity Index and ANF rate was technically significant (r = 0.57), this was likely due to collinearity across LUs, since meaningful scaling was not clear within LUs (**Figure 3-1b**). Beyond this, no *nifH*-D structural metrics were found to be associated with activity rates. In contrast, Hsu and Buckley (2009) found diazotroph community structure to be an important factor in driving ANF rates, as compared to land management practices under continuous maize cropping. Another study conducted in a switchgrass agricultural system conversely found a weakly significant relationship between ANF rate and *nifH*-D community composition (Smercina et al., 2021). This may indicate regional variation regarding the explanatory power of DNA-based profiles to process rates. On the other hand, we found that despite the low degree of dissimilarity across *nifH*-R communities of the two LUs, composition reflected a striking positive (Pearson's r=0.82) association with ANF rate (**Figure 3-2c)**. In fact, this seemed to be the only measurement made in this study that reflected any meaningful relationship at all with *nifH*-R. These findings indicate that a large proportion of nitrogenase-transcribing, but not necessarily nitrogenase-expressing diazotrophs are common across soils of both LUs (for reasons unknown), and that taxa driving community differentiation may also be driving process rates.

Based on the above observations, we identified 17 taxa (IT) whose presence was significantly associated with soils boasting medium and high ANF rates, and whose aggregate relative abundance was highly associated with increasing activity (**Figure 3-5**). The most abundant IT was the fourth most abundant *nifH*-R OTU overall (third in P-RNA, eighth in F-RNA), on average comprising 9% of high-rate communities, and 2.1% of ND ('no detection') communities (overall 4.4%). On an individual basis, its relative abundance correlated strongly with ANF rate (Pearson's r = 0.75; **Supp. Figure 3-5**), and it had a substantial impact on community differentiation across *nifH*-R samples. The best-match annotation was made at the genus level within *Bradyrhizobium* (**Supp. Table 3-5**), and a blastN against GenBank matched (99.3% identity) to an environmental sequence previously identified in acidic peatlands of Third Plateau, Parana state, Brazil (Accession KX718944; Etto et al., 2022). Although generally assumed to be obligately symbiotic diazotrophs, evidence is mounting to indicate a prevalence of environmentally-fit, free-living *Bradyrhizobium* species (Sachs et al., 2011; Tao et al., 2021). In fact, several studies have

observed that taxa annotated as *Bradyrhizobium* act as key predictors of community activity as well as a large abundance component of grassland- based potential diazotrophic communities (Bahulikar et al., 2014; Dai et al., 2021; Roley et al., 2018; Smercina et al., 2020). The second most abundant rate IT, annotated to the family Enterobacteriaceae, made up 2.2% of high-rate communities and was absent from ND communities (ranked 31st most abundant overall in *nifH*-R, and 20th in P-RNA). The Enterobacteriaceae indicator had the strongest correlation with ANF rate on an individual basis (r=0.83; **Supp. Figure 3-5**), and a near-identical directionality of association as ANF rate with *nifH*-R community composition (**Figure 3-5b**). Diazotrophic Enterobacteriaceae have been recognized for their ability to fix nitrogen both in the presence and absence of oxygen, and members of the family are known to tolerate stressful environmental conditions and produce a variety of extracellular enzymes including phosphatases (Barraquio et al., 2000). The aggregate abundance of *nifH*-R IT in the potential community also showed a moderate linear association with ANF process rate (Pearson's r = 0.59), as did several IT on an individual basis (**Supp. Figure 3-6**). For example, the *Bradyrhizobium* sp. discussed above was also the third-most abundant *nifH*-D taxa in pastures and had a *nifH*-D abundance-ANF rate Pearson correlation coefficient of 0.53. However, these potentially important taxonomic signals are lost in the noise of the taxa-rich DNA-based profile at the community compositional level. In turn, this substantiates the value of RNA-based profiling to better understand process rates.

Both general and functional group-specific microbial process rates have been shown to be more strongly linked to the structure and diversity of active compared to present communities (Barnard et al., 2015; Bissett et al., 2013). This makes sense given that microbes can respond functionally to favorable conditions on the order of hours (Placella and Firestone, 2013). Although the statistical approach we employed to identify significant taxa does not provide the direct evidence of activity afforded by tracing methods, it revealed that just a handful of taxa may ultimately be driving process rates. Studies utilizing such tracing techniques such as ^{15}N stable isotope probing of RNA (Angel et al., 2018b) and NanoSIMS (Masuda et al., 2020; Woebken et al., 2012) have reached similar conclusions regarding a small active community fraction significantly affecting ANF activity. This may be explained physiologically in part by the fact that transcription of the nitrogenase enzyme does not directly infer active nitrogen fixation. In addition to pre-transcription genetic safeguards, post-translational regulatory strategies including ADP-

ribosylation and P_{II} signal transduction are key to efficient metabolism (Huergo et al., 2012). For example, the DraT-DraG regulatory system activates in response to an increase in the concentration of environmental NH_4^+, depletion of cellular energy levels, or changes to light intensity (in some phototropic organisms), and functions through reversible ribosylation of *nifH*, effectively pausing nitrogen fixation until further notice (Huergo et al., 2012; Masepohl et al., 2002). This stasis may explain the surprising compositional similarity in F and P *nifH*-R communities, despite a general disparity in ANF activity rate. Mechanistically, this suggests that the majority of transcriptionally-active community members may have sensitivity to external N through a high transporter affinity (Darnajoux et al., 2022). The limited number of taxa actively contributing to ANF may have greater NH_4^+ tolerances (Dekaezemacker and Bonnet, 2011), or may be under advantageous microsite conditions (Parkin, 1993) that cannot be directly parsed by bulk soil extraction. Greater temporal resolution and direct probing of the active diazotrophic community are crucial to elucidate this further.

Conclusion

This study surveyed the response of both potential and transcriptionally active asymbiotic soil diazotrophs to LU change in the Brazilian Amazon. We found that the diversity of potential communities increased with cattle pasture conversion, independently of total prokaryotic communities, but that potential community composition was not related to measured diazotroph activity. In contrast, while the diversity of nitrogenase-transcribing diazotrophic communities exhibited no response to LU change, their modest alteration in community composition was highly associated with high vs. low ANF activity rate. A high proportion of cyanobacterial taxa in active community profiles of both LUs suggests a potentially significant contribution of surface-dwelling phototropic diazotrophs to fixed nitrogen inputs across both LU types, which was not accounted for by rate assays incubated in the dark. Future studies should investigate the contribution of this diazotrophic subgroup, particularly in response to variable grazing intensity, since this may have a dramatic impact on soil surface conditions, and the long-term sustainability of pastures.

Funding and Acknowledgments

This work was funded by the National Science Foundation – Dimensions of Biodiversity (DEB 14422214) and U.S. Department of Agriculture – NIFA (Award # 2009-35319-05186). We are grateful to the owners and staff of Agropecuaria Nova Vida for allowing us to sample on their property and assisting us with logistics. Alexandre Pedrinho and Wagner Piccinini were essential in field work preparations and sample collection.

References

Alteio, L. V., Schulz, F., Seshadri, R., Varghese, N., Rodriguez-Reillo, W., Ryan, E., Goudeau, D., Eichorst, S. A., Malmstrom, R. R., Bowers, R. M., Katz, L. A., Blanchard, J. L., and Woyke, T. (2020). Complementary metagenomic approaches improve reconstruction of microbial diversity in a forest foil. *mSystems* **5**.

Anderson, M. J., Ellingsen, K. E., and McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology letters* **9**, 683-693.

Angel, R., Nepel, M., Panholzl, C., Schmidt, H., Herbold, C. W., Eichorst, S. A., and Woebken, D. (2018a). Evaluation of primers targeting the diazotroph functional gene and development of NifMAP - a bioinformatics pipeline for analyzing nifH amplicon data. *Front Microbiol* **9**, 703.

Angel, R., Panholzl, C., Gabriel, R., Herbold, C., Wanek, W., Richter, A., Eichorst, S. A., and Woebken, D. (2018b). Application of stable-isotope labelling techniques for the detection of active diazotrophs. *Environ Microbiol* **20**, 44-61.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252.

Asner, G. P., Townsend, A. R., Bustamante, M. M., Nardoto, G. B., and Olander, L. P. (2004). Pasture degradation in the central Amazon: linking changes in carbon and nutrient cycling with remote sensing. *Global Change Biology* **10**, 844-862.

Auger, C., Han, S., Appanna, V. P., Thomas, S. C., Ulibarri, G., and Appanna, V. D. (2013). Metabolic reengineering invoked by microbial systems to decontaminate aluminum: implications for bioremediation technologies. *Biotechnology advances* **31**, 266-273.

Avelar Ferreira, P. A., Bomfeti, C. A., Lima Soares, B., and de Souza Moreira, F. M. (2012). Efficient nitrogen-fixing Rhizobium strains isolated from amazonian soils are highly tolerant to acidity and aluminium. *World J Microbiol Biotechnol* **28**, 1947-59.

Bacq-Labreuil, A., Crawford, J., Mooney, S. J., Neal, A. L., and Ritz, K. (2019). Cover crop species have contrasting influence upon soil structural genesis and microbial community phenotype. *Sci Rep* **9**, 7473.

Bae, H. S., Morrison, E., Chanton, J. P., and Ogram, A. (2018). Methanogens are major contributors to nitrogen fixation in soils of the Florida everglades. *Appl Environ Microbiol* **84**.

Bahulikar, R. A., Torres-Jerez, I., Worley, E., Craven, K., and Udvardi, M. K. (2014). Diversity of nitrogen-fixing bacteria associated with switchgrass in the native tallgrass prairie of northern Oklahoma. *Appl Environ Microbiol* **80**, 5636-43.

Barnard, R. L., Osborne, C. A., and Firestone, M. K. (2015). Changing precipitation pattern alters soil microbial community response to wet-up under a Mediterranean-type climate. *ISME J* **9**, 946-57.

Barraquio, W. L., Segubre, E. M., Gonzalez, M. A. S., Verma, S. C., James, E. K., Ladha, J. K., and Tripathi, A. K. (2000). Diazotrophic enterobacteria: What is their role in the rhizosphere of rice. *The quest for nitrogen fixation in rice*, 93-118.

Belnap, J. (2001). Factors influencing nitrogen fixation and nitrogen release in biological soil crusts. *In* "Biological soil crusts: structure, function, and management", pp. 241-261. Springer.

Bissett, A., Richardson, A. E., Baker, G., Kirkegaard, J., and Thrall, P. H. (2013). Bacterial community response to tillage and nutrient additions in a long-term wheat cropping experiment. *Soil Biology and Biochemistry* **58**, 281-292.

Blagodatskaya, E., and Kuzyakov, Y. (2013). Active microorganisms in soil: Critical review of estimation criteria and approaches. *Soil Biology and Biochemistry* **67**, 192-211.

Blagodatskaya, E. V., and Anderson, T.-H. (1998). Interactive effects of pH and substrate quality on the fungal-to-bacterial ratio and qCO2 of microbial communities in forest soils. *Soil Biology and Biochemistry* **30**, 1269-1274.

Bouffaud, M. L., Renoud, S., Moenne-Loccoz, Y., and Muller, D. (2016). Is plant evolutionary history impacting recruitment of diazotrophs and nifH expression in the rhizosphere? *Sci Rep* **6**, 21690.

Brown, S. M., and Jenkins, B. D. (2014). Profiling gene expression to distinguish the likely active diazotrophs from a sea of genetic potential in marine sediments. *Environ Microbiol* **16**, 3128-42.

Bürgmann, H., Meier, S., Bunge, M., Widmer, F., and Zeyer, J. (2005). Effects of model root exudates on structure and activity of a soil diazotroph community. *Environmental Microbiology* **7**, 1711-1724.
Calderoli, P. A., Collavino, M. M., Behrends Kraemer, F., Morras, H. J. M., and Aguilar, O. M. (2017). Analysis of nifH-RNA reveals phylotypes related to Geobacter and Cyanobacteria as important functional components of the N2 -fixing community depending on depth and agricultural use of soil. *Microbiologyopen* **6**.
Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-3.
Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences* **108**, 4516-4522.
Carvalho, J. L. N., Carlos Eduardo Pelegrino, C., Feigl, B. J., Píccolo, M. d. C., Godinho, V. d. P., Herpin, U., and Cerri, C. C. (2009). Conversion of Cerrado into agricultural land in the south-western Amazon: Carbon stocks and soil fertility. *Scientia Agricola* **66**, 233-241.
Cerqueira, A. E. S., Silva, T. H., Nunes, A. C. S., Nunes, D. D., Lobato, L. C., Veloso, T. G. R., De Paula, S. O., Kasuya, M. C. M., and Silva, C. C. (2018). Amazon basin pasture soils reveal susceptibility to phytopathogens and lower fungal community dissimilarity than forest. *Applied Soil Ecology* **131**, 1-11.
Chiarello, M., McCauley, M., Villeger, S., and Jackson, C. R. (2022). Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLoS One* **17**, e0264443.
da C Jesus, E., Marsh, T. L., Tiedje, J. M., and de, S. M. F. M. (2009). Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J* **3**, 1004-11.
da Silva, M. d. C. S., Mendes, I. R., Paula, T. d. A., Dias, R. S., de Paula, S. O., Silva, C. C., Bazzolli, D. M. S., and Kasuya, M. C. M. (2016). Expression of thenifHgene in diazotrophic bacteria inEucalyptus urograndisplantations. *Canadian Journal of Forest Research* **46**, 190-199.
Dai, X., Song, D., Guo, Q., Zhou, W., Liu, G., Ma, R., Liang, G., He, P., Sun, G., Yuan, F., and Liu, Z. (2021). Predicting the influence of fertilization regimes on potential N fixation through their effect on free-living diazotrophic community structure in double rice cropping systems. *Soil Biology and Biochemistry* **156**.
Danielson, R. E., and Mazza Rodrigues, J. L. (2022). Impacts of land-use change on soil microbial communities and their function in the Amazon Rainforest. Vol. 175, pp. 179-258.
Darnajoux, R., Reji, L., Zhang, X. R., Luxem, K. E., and Zhang, X. (2022). Ammonium Sensitivity of Biological Nitrogen Fixation by Anaerobic Diazotrophs in Cultures and Benthic Marine Sediments. *Journal of Geophysical Research: Biogeosciences* **127**.
De Caceres, M., Jansen, F., and De Caceres, M. M. (2016). Package 'indicspecies'. *indicators* **8**.
Dekaezemacker, J., and Bonnet, S. (2011). Sensitivity of N_2 fixation to combined nitrogen forms (NO_3^- and NH_4^+) in two strains of the marine diazotroph Crocosphaera watsonii (Cyanobacteria). *Marine Ecology Progress Series* **438**, 33-46.
Dias-Filho, M. B., Davidson, E. A., and Carvalho, C. (2001). Linking Biogeochemical Cycles to Cattle Pasture Management and Sustainability. *The biogeochemistry of the Amazon Basin.*, 84-105.
Dos Santos, P., Fang, Z., Mason, S., Setubal, J., and Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**.
Etto, R. M., Jesus, E. C., Cruz, L. M., Schneider, B. S. F., Tomachewski, D., Urrea-Valencia, S., Goncalves, D. R. P., Galvao, F., Ayub, R. A., Curcio, G. R., Steffens, M. B. R., and Galvao, C. W. (2022). Influence of environmental factors on the tropical peatlands diazotrophic communities from the Southern Brazilian Atlantic Rain Forest. *Lett Appl Microbiol* **74**, 543-554.
Gaby, J. C., Rishishwar, L., Valderrama-Aguirre, L. C., Green, S. J., Valderrama-Aguirre, A., Jordan, I. K., and Kostka, J. E. (2018). Diazotroph community characterization via a high-throughput *nifH* amplicon sequencing and analysis pipeline. *Appl Environ Microbiol* **84**.
Galushko, A., Kuever, J., and Whitman, W. B. (2020). Desulfovibrionaceae. *In* "Bergey's Manual of Systematics of Archaea and Bacteria", pp. 1-13.
Geisseler, D., Horwath, W. R., and Doane, T. A. (2009). Significance of organic nitrogen uptake from plant residues by soil microorganisms as affected by carbon and nitrogen availability. *Soil Biology and Biochemistry* **41**, 1281-1288.

Gharmakher, H. N., Machet, J.-M., Beaudoin, N., and Recous, S. (2009). Estimation of sulfur mineralization and relationships with nitrogen and carbon in soils. *Biology and Fertility of Soils* **45**, 297-304.
Gonzalez Perez, P., Ye, J., Wang, S., Wang, X., and Huang, D. (2014). Analysis of the occurrence and activity of diazotrophic communities in organic and conventional horticultural soils. *Applied Soil Ecology* **79**, 37-48.
Han, L. L., Wang, Q., Shen, J. P., Di, H. J., Wang, J. T., Wei, W. X., Fang, Y. T., Zhang, L. M., and He, J. Z. (2019). Multiple factors drive the abundance and diversity of the diazotrophic community in typical farmland soils of China. *FEMS Microbiol Ecol* **95**.
Hsu, S. F., and Buckley, D. H. (2009). Evidence for the functional significance of diazotroph community structure in soil. *ISME J* **3**, 124-36.
Hu, T., Wang, X., Zhen, L., Gu, J., Song, Z., Sun, W., and Xie, J. (2021). Succession of diazotroph community and functional gene response to inoculating swine manure compost with a lignocellulose-degrading consortium. *Bioresour Technol* **337**, 125469.
Huergo, L. F., Pedrosa, F. O., Muller-Santos, M., Chubatsu, L. S., Monteiro, R. A., Merrick, M., and Souza, E. M. (2012). PII signal transduction proteins: pivotal players in post-translational control of nitrogenase activity. *Microbiology (Reading)* **158**, 176-190.
Imhoff, J. F. (2015). Ectothiorhodospiraceae. *In* "Bergey's Manual of Systematics of Archaea and Bacteria", pp. 1-5.
Imhoff, J. F., Kyndt, J. A., and Meyer, T. E. (2022). Genomic comparison, phylogeny and taxonomic reevaluation of the Ectothiorhodospiraceae and description of Halorhodospiraceae fam. nov. and Halochlorospira gen. nov. *Microorganisms* **10**.
Keiluweit, M., Gee, K., Denney, A., and Fendorf, S. (2018). Anoxic microsites in upland soils dominantly controlled by clay content. *Soil Biology and Biochemistry* **118**, 42-50.
Khadem, A. F., Pol, A., Jetten, M. S. M., and Op den Camp, H. J. M. (2010). Nitrogen fixation by the verrucomicrobial methanotroph 'Methylacidiphilum fumariolicum' SolV. *Microbiology (Reading)* **156**, 1052-1059.
Koirala, A., and Brozel, V. S. (2021). Phylogeny of nitrogenase structural and assembly components reveals new insights into the origin and distribution of nitrogen Fixation across bacteria and archaea. *Microorganisms* **9**.
Kuzyakov, Y., and Blagodatskaya, E. (2015). Microbial hotspots and hot moments in soil: Concept & review. *Soil Biology and Biochemistry* **83**, 184-199.
Le Boulch, M., Dehais, P., Combes, S., and Pascal, G. (2019). The MACADAM database: a MetAboliC pAthways DAtabase for Microbial taxonomic groups for mining potential metabolic capacities of archaeal and bacterial taxonomic groups. *Database (Oxford)* **2019**.
Liao, H., Li, Y., and Yao, H. (2017). Fertilization with inorganic and organic nutrients changes diazotroph community composition and nitrogen fixation rates. *Journal of Soils and Sediments* **18**, 1076-1086.
Longo, M., Saatchi, S., Keller, M., Bowman, K., Ferraz, A., Moorcroft, P. R., Morton, D. C., Bonal, D., Brando, P., Burban, B., Derroire, G., Dos-Santos, M. N., Meyer, V., Saleska, S., Trumbore, S., and Vincent, G. (2020). Impacts of degradation on water, energy, and carbon cycling of the Amazon tropical forests. *J Geophys Res Biogeosci* **125**, e2020JG005677.
Masepohl, B., Drepper, T., Paschen, A., Groß, S., Pawlowski, A., Raabe, K., Riedel, K.-U., and Klipp, W. (2002). Regulation of nitrogen fixation in the phototrophic purple bacterium *Rhodobacter capsulatus*. *Molecular Microbiology and Biotechnology* **4**, 243-248.
Masuda, T., Inomura, K., Takahata, N., Shiozaki, T., Sano, Y., Deutsch, C., Prasil, O., and Furuya, K. (2020). Heterogeneous nitrogen fixation rates confer energetic advantage and expanded ecological niche of unicellular diazotroph populations. *Commun Biol* **3**, 172.
Melo, V. F., Barros, L. S., Silva, M. C. S., Veloso, T. G. R., Senwo, Z. N., Matos, K. S., and Nunes, T. K. O. (2021). Soil bacterial diversities and response to deforestation, land use and burning in North Amazon, Brazil. *Applied Soil Ecology* **158**.
Mendes, L. W., Tsai, S. M., Navarrete, A. A., de Hollander, M., van Veen, J. A., and Kuramae, E. E. (2015). Soil-borne microbiome: linking diversity to function. *Microb Ecol* **70**, 255-65.
Meyer, K. M., Klein, A. M., Rodrigues, J. L., Nusslein, K., Tringe, S. G., Mirza, B. S., Tiedje, J. M., and Bohannan, B. J. (2017). Conversion of Amazon rainforest to agriculture alters community traits of methane-cycling organisms. *Mol Ecol* **26**, 1547-1556.

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

Mirza, B. S., McGlinn, D. J., Bohannan, B. J., Nüsslein, K., Tiedje, J. M., and Rodrigues, J. L., 2020. . 86(10). (2020). Diazotrophs show signs of restoration in Amazon rain forest soils with ecosystem rehabilitation. *Applied and environmental microbiology,* **86**.

Mirza, B. S., Potisap, C., Nusslein, K., Bohannan, B. J., and Rodrigues, J. L. (2014). Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. *Appl Environ Microbiol* **80**, 281-8.

Navarrete, A. A., Tsai, S. M., Mendes, L. W., Faust, K., de Hollander, M., Cassman, N. A., Raes, J., van Veen, J. A., and Kuramae, E. E. (2015). Soil microbiome responses to the short-term effects of Amazonian deforestation. *Mol Ecol* **24**, 2433-48.

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

Neill, C., Piccolo, M.C., Steudler, P.A., Melillo, J.M., Feigl, B.J. and Cerri, C.C., 1 (1995). Nitrogen dynamics in soils of forests and active pastures in the western Brazilian Amazon Basin. *Soil Biology and Biochemistry* **27**, 1167-1175.

Norman, J. S., and Friesen, M. L. (2017). Complex N acquisition by soil diazotrophs: how the ability to release exoenzymes affects N fixation by terrestrial free-living diazotrophs. *ISME J* **11**, 315-326.

Nunes, P. A. d. A., Bredemeier, C., Bremm, C., Caetano, L. A. M., de Almeida, G. M., de Souza Filho, W., Anghinoni, I., and Carvalho, P. C. d. F. (2019). Grazing intensity determines pasture spatial heterogeneity and productivity in an integrated crop-livestock system. *Grassland Science* **65**, 49-59.

Okon, Y., Albrecht, S. L., and Burris, R. (1976). Factors affecting growth and nitrogen fixation of Spirillum lipoferum. *Journal of Bacteriology* **127**, 1248-1254.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2013). Package 'vegan'. *Community ecology package, version* **2**, 1-295.

Parkin, T. (1993). Spatial variability of microbial processes in soil—a review. *Journal of environmental quality* **22**, 409-417.

Pedrinho, A., Mendes, L. W., Merloti, L. F., da Fonseca, M. C., Cannavan, F. S., and Tsai, S. M. (2019). Forest-to-pasture conversion and recovery based on assessment of microbial communities in Eastern Amazon rainforest. *FEMS Microbiol Ecol* **95**.

Piccolo, M. C., Neill, C., and Cerri, C. C. (1994a). Natural abundance of ^{15}N in soils along forest-to-pasture chronosequences in the western Brazilian Amazon Basin. *Oecologia,* **99**, 112-117.

Piccolo, M. C., Neill, C., and Cerri, C. C. (1994b). Net nitrogen mineralization and net nitrification along a tropical forest-to-pasture chronosequence. *Plant and soil* **162**, 61-70.

Placella, S. A., and Firestone, M. K. (2013). Transcriptional response of nitrifying communities to wetting of dry soil. *Appl Environ Microbiol* **79**, 3294-302.

Ranjan, K., Paula, F. S., Mueller, R. C., Jesus Eda, C., Cenciani, K., Bohannan, B. J., Nusslein, K., and Rodrigues, J. L. (2015). Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. *Front Microbiol* **6**, 779.

Rascio, N., and La Rocca, N. (2013). Biological nitrogen fixation.

Rauber, L. R., Sequinatto, L., Kaiser, D. R., Bertol, I., Baldissera, T. C., Garagorry, F. C., Sbrissia, A. F., Pereira, G. E., and Pinto, C. E. (2021). Soil physical properties in a natural highland grassland in southern Brazil subjected to a range of grazing heights. *Agriculture, Ecosystems & Environment* **319**.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.

Robson, R. L., and Postgate, J. R. (1980). Oxygen and hydrogen in biological nitrogen fixation. *Annual Reviews in Microbiology* **34**, 183-207.

Rodrigues, J. L., Pellizari, V. H., Mueller, R., Baek, K., Jesus Eda, C., Paula, F. S., Mirza, B., Hamaoui, G. S., Jr., Tsai, S. M., Feigl, B., Tiedje, J. M., Bohannan, B. J., and Nusslein, K. (2013). Conversion of

the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc Natl Acad Sci U S A* **110**, 988-93.
Roley, S. S., Duncan, D. S., Liang, D., Garoutte, A., Jackson, R. D., Tiedje, J. M., and Robertson, G. P. (2018). Associative nitrogen fixation (ANF) in switchgrass (*Panicum virgatum*) across a nitrogen input gradient. *PLoS One* **13**, e0197320.
Sachs, J. L., Russell, J. E., and Hollowell, A. C. (2011). Evolutionary instability of symbiotic function in Bradyrhizobium japonicum. *PLoS One* **6**, e26370.
Severin, I., Bentzon-Tilia, M., Moisander, P. H., and Riemann, L. (2015). Nitrogenase expression in estuarine bacterioplankton influenced by organic carbon and availability of oxygen. *FEMS Microbiol Lett* **362**.
Shimura, Y., Hirose, Y., Misawa, N., Osana, Y., Katoh, H., Yamaguchi, H., and Kawachi, M. (2015). Comparison of the terrestrial cyanobacterium Leptolyngbya sp. NIES-2104 and the freshwater Leptolyngbya boryana PCC 6306 genomes. *DNA Res* **22**, 403-12.
Silva, K. d., Nóbrega, R. S. A., Lima, A. S., Barberi, A., and Moreira, F. M. d. S. (2011). Density and diversity of diazotrophic bacteria isolated from Amazonian soils using N-free semi-solid media. *Scientia Agricola* **68**, 518-525.
Simon, J., and Kroneck, P. M. (2013). Microbial sulfite respiration. *Adv Microb Physiol* **62**, 45-117.
Smercina, D. N., Evans, S. E., Friesen, M. L., and Tiemann, L. K. (2019). To fix or not To fix: controls on free-living nitrogen fixation in the rhizosphere. *Appl Environ Microbiol* **85**.
Smercina, D. N., Evans, S. E., Friesen, M. L., and Tiemann, L. K. (2020). Impacts of nitrogen addition on switchgrass root-associated diazotrophic community structure and function. *FEMS Microbiol Ecol* **96**.
Smercina, D. N., Evans, S. E., Friesen, M. L., and Tiemann, L. K. (2021). Temporal dynamics of free-living nitrogen fixation in the switchgrass rhizosphere. *GCB Bioenergy* **13**, 1814-1830.
Sumner, M. (1963). Effect of iron oxides on positive and negative charges in clays and soils. *Clay Minerals Bulletin* **5**, 218-226.
Tao, J., Wang, S., Liao, T., and Luo, H. (2021). Evolutionary origin and ecological implication of a unique nif island in free-living Bradyrhizobium lineages. *ISME J* **15**, 3195-3206.
Taylor, D. L., Hollingsworth, T. N., McFarland, J. W., Lennon, N. J., Nusbaum, C., and Ruess, R. W. (2014). A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* **84**, 3-20.
Tyukavina, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Smith-Rodriguez, K., Okpa, C., and Aguilar, R. (2017). Types and rates of forest disturbance in Brazilian Legal Amazon, 2000–2013. *Science Advances* **3**.
Verchot, L. V., Davidson, E. A., Cattânio, H., Ackerman, I. L., Erickson, H. E., and Keller, M. (1999). Land use change and biogeochemical controls of nitrogen oxide emissions from soils in eastern Amazonia. *Global Biogeochemical Cycles* **13**, 31-46.
Volpon, A. G., De-Polli, H., and Döbereiner, J. (1981). Physiology of nitrogen fixation in Azospirillum lipoferum Br 17 (ATCC 29 709). *Archives of Microbiology* **128**, 371-375.
Wakelin, S. A., Gupta, V. V. S. R., and Forrester, S. T. (2010). Regional and local factors affecting diversity, abundance and activity of free-living, N_2-fixing bacteria in Australian agricultural soils. *Pedobiologia* **53**, 391-399.
Wang, J., Bao, J. T., Li, X. R., and Liu, Y. B. (2016). Molecular ecology of *nifH* genes and transcripts along a chronosequence in revegetated areas of the tengger desert. *Microb Ecol* **71**, 150-63.
Wang, L., Zhang, L., Liu, Z., Zhao, D., Liu, X., Zhang, B., Xie, J., Hong, Y., Li, P., Chen, S., Dixon, R., and Li, J. (2013). A minimal nitrogen fixation gene cluster from Paenibacillus sp. WLY78 enables expression of active nitrogenase in Escherichia coli. *PLoS Genet* **9**.
Wang, Y., Li, C., Kou, Y., Wang, J., Tu, B., Li, H., Li, X., Wang, C., and Yao, M. (2017). Soil pH is a major driver of soil diazotrophic community assembly in Qinghai-Tibet alpine meadows. *Soil Biology and Biochemistry* **115**, 547-555.
Woebken, D., Burow, L. C., Prufert-Bebout, L., Bebout, B. M., Hoehler, T. M., Pett-Ridge, J., Spormann, A. M., Weber, P. K., and Singer, S. W. (2012). Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *ISME J* **6**, 1427-39.
Yan, J., Han, X., Lu, X., Chen, X., and Zou, W. (2022). Land use indirectly affects the cycling of multiple nutrients by altering the diazotrophic community in black soil. *J Sci Food Agric* **102**, 3788-3795.

Yu, G. H., Kuzyakov, Y., Luo, Y., Goodman, B. A., Kappler, A., Liu, F. F., and Sun, F. S. (2021). Molybdenum bioavailability and asymbiotic nitrogen fixation in soils are raised by iron (Oxyhydr)oxide-mediated free radical production. *Environ Sci Technol* **55**, 14979-14989.
Zani, S., Mellon, M. T., Collier, J. L., and Zehr, J. P. (2000). Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Applied and Environmental Microbiology* **66**, 3119-3124.
Zehr, J. P., and McReynolds, L. A. (1989). Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium Trichodesmium thiebautii. *Applied and environmental microbiology* **55**, 2522-2526.
Zehr, J. P., and Turner, P. J. (2001). Nitrogen fixation: nitrogenase genes and gene expression. *Methods in Microbiology* **30**, 271-286.
Zhang, X., Xu, S., Li, C., Zhao, L., Feng, H., Yue, G., Ren, Z., and Cheng, G. (2014). The soil carbon/nitrogen ratio and moisture affect microbial community structures in alkaline permafrost-affected soils with different vegetation types on the Tibetan plateau. *Res Microbiol* **165**, 128-39.

Tables and Figures

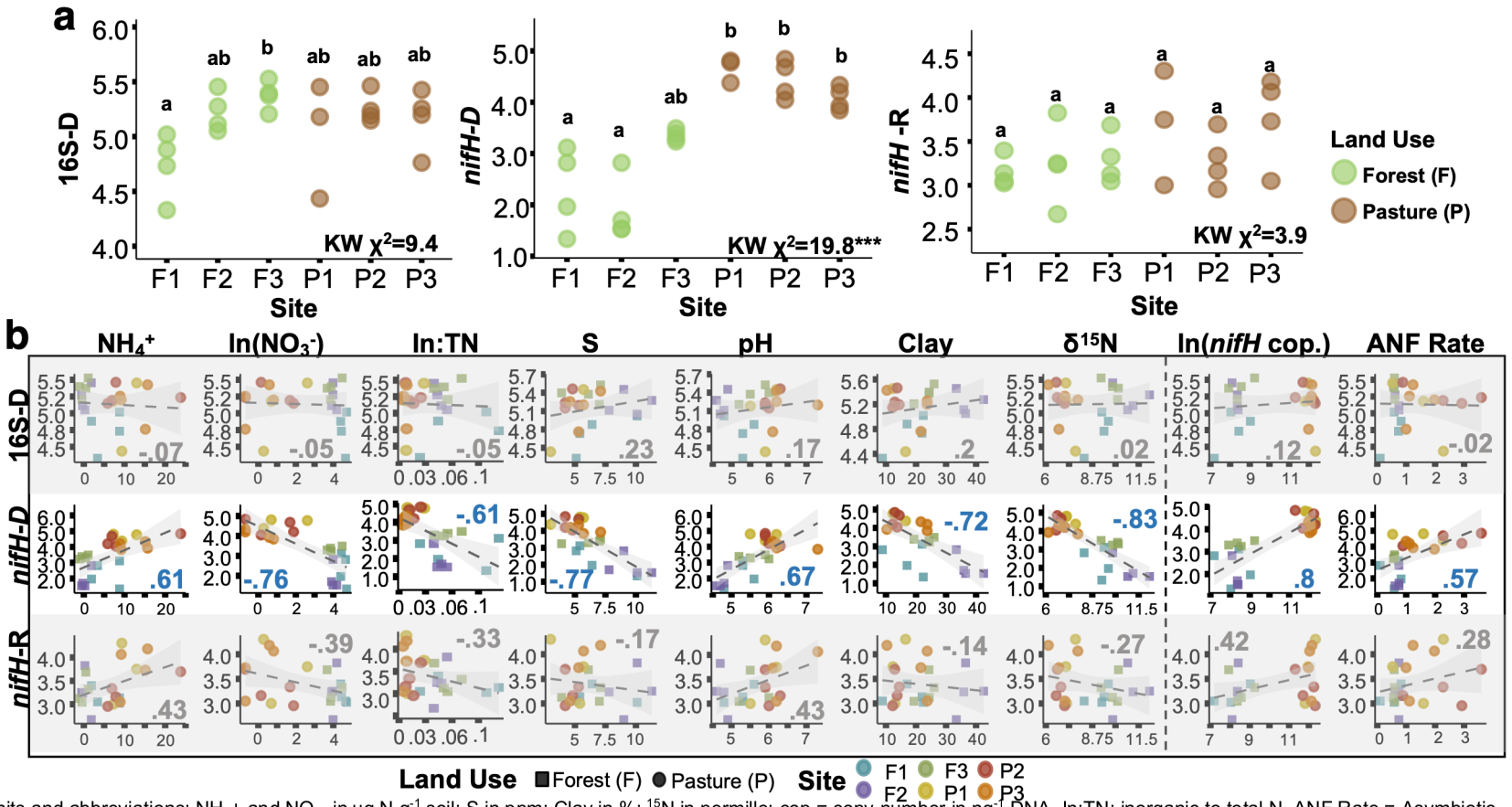
Table 3-1: PERMANOVA testing of LU effect after accounting for GD for 16S, *nifH*-D, *nifH*-R, and total *nifH* communities to determine the variation explained (VE). Community type (CT: DNA vs RNA) effect was also tested within the total *nifH* community.

Total VE	**LU + GD**		**LU | GD**		
	Total	**Shared**	**VE**	**F-statistic**	**p-value**
16S-D	0.44	0.136	0.122	4.17	0.001***
***nifH*-D**	0.46	0.221	0.118	4.18	0.002***
***nifH*-R**	0.185	0.028	0.059	1.37	0.08*
***nifH* Total**	0.108	-0.004	0.035	1.67	ns
Total VE	**CT + LU + GD**		**CT | GD+LU**		
	Total	**Shared**	**VE**	**F-statistic**	**p-value**
***nifH* Total**	.393	0.00	0.284	19.19	0.001***

Table 3-2: Group dispersion calculated based on median centroid distance. LU and site were tested for 16S, *nifH*-D, and *nifH*-R communities, and LU-community type combinations were tested across combined *nifH* communities ('*nifH* Total'). Values presented represent mean (± standard error) distances. F-statistics are derived from permuted constrained correspondence analysis, and χ^2 statistics are derived from Kruskall-Wallis rank sum tests. Letters represent *post-hoc* Dunn's test grouping.

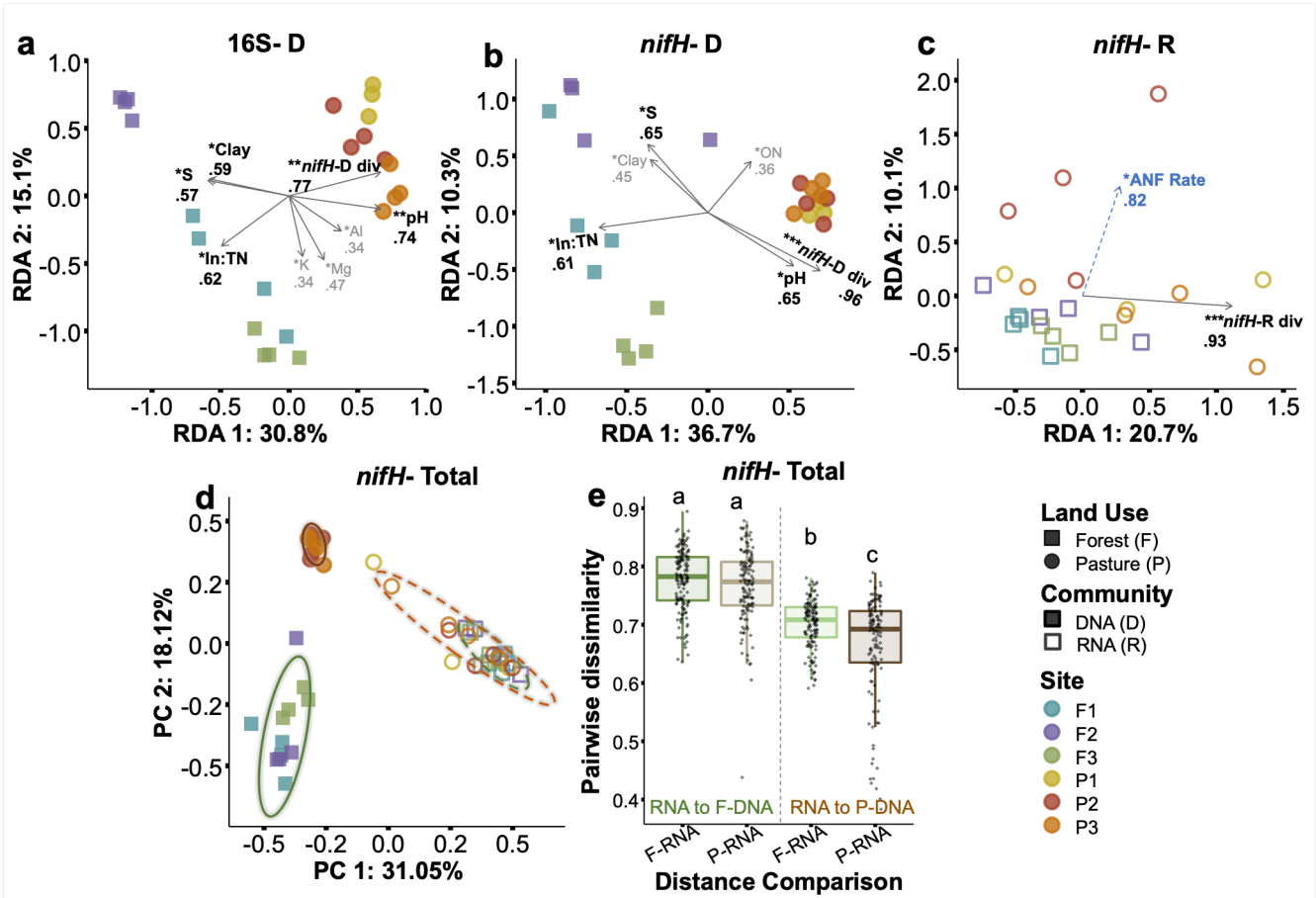
	LU				Site							
	F	**P**	**F-stat (p-val)**	**X^2 (p-val)**	**F1**	**F2**	**F3**	**P1**	**P2**	**P3**	**F-stat (p-val)**	**X^2 (p-val)**
16S-D	0.46 ±.011	0.42 ±.012	7.56 (0.01)	6.37 (0.012)	0.43 ±.016 **a**	0.27 ±.011 **b**	0.34 ±.019 **ab**	0.35 ±.008 **ab**	0.40 ±.014 **a**	0.34 ±.031 **ab**	9.14 ($5*10^{-4}$)	15.18 (0.01)
***nifH*-D**	.43 ±.01	.31 ±.014	52.09 ($1*10^{-4}$)	13.2 ($2.8*10^{-4}$)	0.45 ±.009 **a**	0.35 ±.036 **abc**	0.34 ±.008 **ab**	0.23 ±.005 **d**	0.28 ±.006 **bcd**	0.26 ±.024 **cd**	16.08 ($1*10^{-4}$)	17.97 ($2.9*10^{-3}$)
***nifH*-R**	0.28 ±.01	0.36 ±.015	19.5 ($4*10^{-4}$)	11.5 ($7.1*10^{-4}$)	0.28 ±.015 **ab**	0.30 ±.023 **ab**	0.26 ±.013 **a**	0.41 ±.025 **c**	0.33 ±.019 **abc**	0.34 ±.024 **bc**	6.5 ($2.4*10^{-3}$)	14.3 (0.01)
	F	**P**	**F-stat (p-val)**	**X^2 (p-val)**	**F-DNA**	**P-DNA**	**F-RNA**	**P-RNA**			**F-stat (p-val)**	**X^2 (p-val)**
***nifH* Total**	0.47 ±.02	0.41 ±.014	10.22 ($2.6*10^{-3}$)	7.43 ($6.4*10^{-3}$)	0.43 ±.010 **a**	0.31 ±.014 **b**	0.28 ±.015 **b**	0.36 ±.015 **c**			29.03 ($1*10^{-4}$)	29.6 (0.00)

Figure 3-1: Shannon alpha diversity indices for **(a)** DNA-based 16S (16S-D), DNA-based *nifH* (*nifH*-D) and RNA-based *nifH* (*nifH*-R) community profiles. KW χ2 values represents Kruskall-Wallis Rank sum test statistics (***p-value < 0.001). Letters represent Dunn's *post-hoc* grouping to determine LU significance. **(b)** Shannon diversity metrics correlated against physicochemical variables and microbial indicators (right of dotted line). Variables which did not meet significance cutoffs with any diversity metric are not shown (n=17). Pearson r values are shown in either gray (below significance) or blue (above significance). Full Pearson and Spearman values are available in Supplementary Table 3-1a,b.



Units and abbreviations: NH_4+ and NO_3- in µg N g^{-1} soil; S in ppm; Clay in %; ^{15}N in permille; cop = copy number in ng^{-1} DNA. ln:TN: inorganic to total N. ANF Rate = Asymbiotic N_2 fixation rate in ng N g^{-1} soil day^{-1}. ln() =natural-log transformation.

Figure 3-2: Community composition represented by the first two ordinal axes of unconstrained redundancy analysis for **(a)** 16S- DNA profiles **(b)** *nifH*- DNA profiles, and **(c)** *nifH*-RNA profiles. Plots are overlayed with vectors of significantly correlated physicochemical or diazotrophic indicator measurements (dashed, blue line) after accounting for the variation explained by LU effect, separate from community composition. Variables with 0.35< r <0.5 are in reduced font size. Asterisks correspond to significance of VE (*p<0.1;**p<0.01; ***p<0.005) **(d)** *nifH*-D and *nifH*-R ordinated together ('Total'), overlayed with 90% confidence ellipses for each LU/CT grouping. **(e)** Pairwise dissimilarity between *nifH*-D and *nifH*-R community profiles for each group combination.



Abbreviations: In:TN = inorganic to total N ratio; div = Shannon diversity index; ANF rate = asymbiotic nitrogen fixation rate; Al= Aluminum content; S = total sulfur content.

Figure 3-3: Differential abundance analysis at the OTU level showing mean relative abundance compared to log_2-fold change for 4 comparisons. Point size corresponds to BH- corrected p-values: **(a)** LU effect on F (green circles) vs. P (brown circles) *nifH*-D communities and **(b)** LU effect on F vs pasture *nifH*-R communities. Comparison of significant trends was also made within **(c)** paired CT effect for forest *nifH*-D (filled green squares) vs *nifH*-R (open green squares), and **(d)** paired CT effect for pasture *nifH*-D (filled brown) vs *nifH*-R (open brown). Symbol size corresponds to increasingly significant p-value, and OTUs with no significant change are in gray.

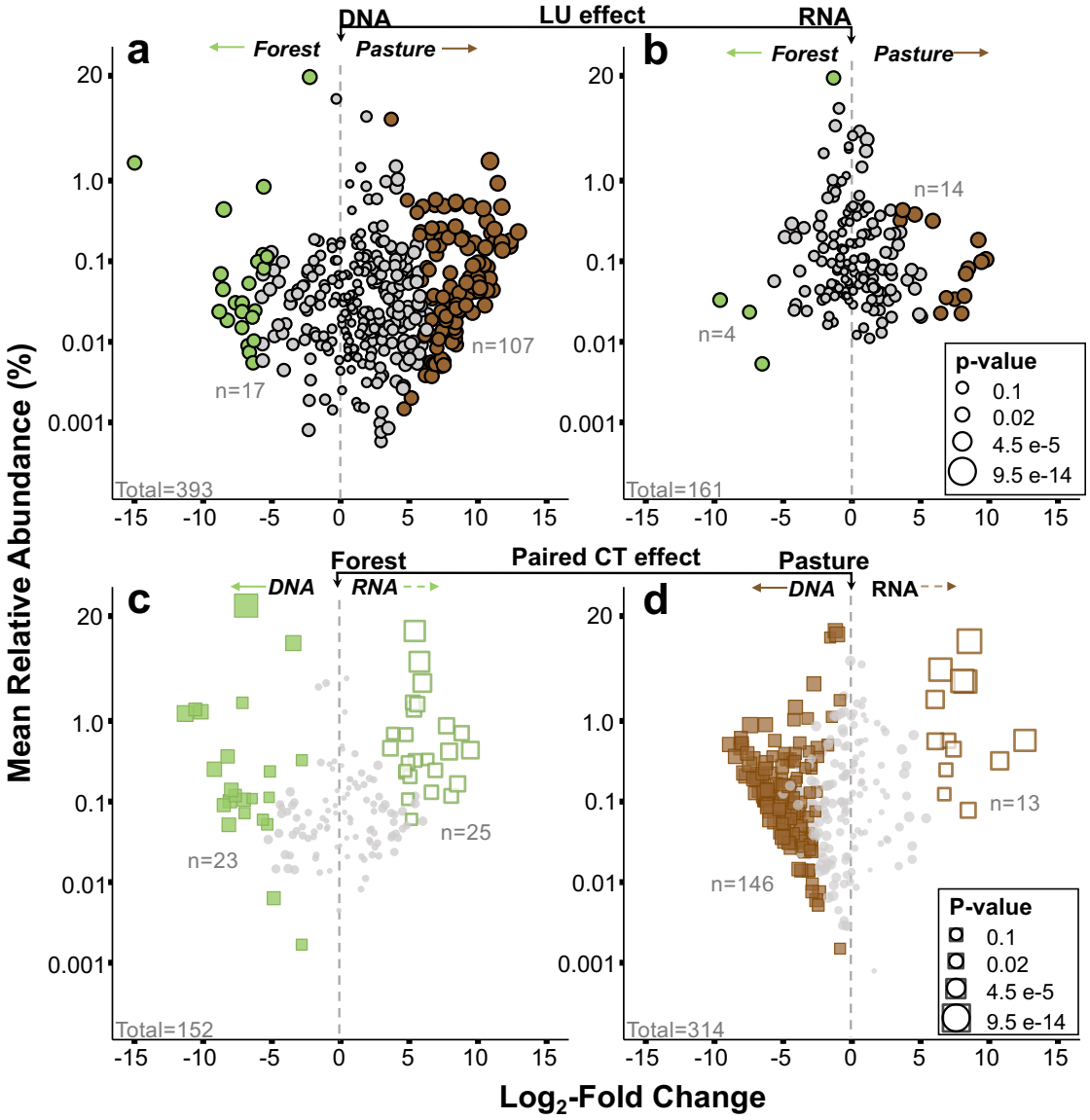


Figure 3-4: Relative abundance of taxa aggregated at the family level for *nifH*-D (closed circles) and *nifH*-R (open squares) across F (green) and P (brown) samples. Symbols to the right indicate the significance of tests for LU effect on *nifH*-D and *nifH*-R communities (orange circle/ box with trend denoted by border color), and CT effect for F and P based on paired sample profiles (green/brown triangle, with direction indicating enrichment (upward, red border) or depletion (downward, blue border) in *nifH*-R communities.

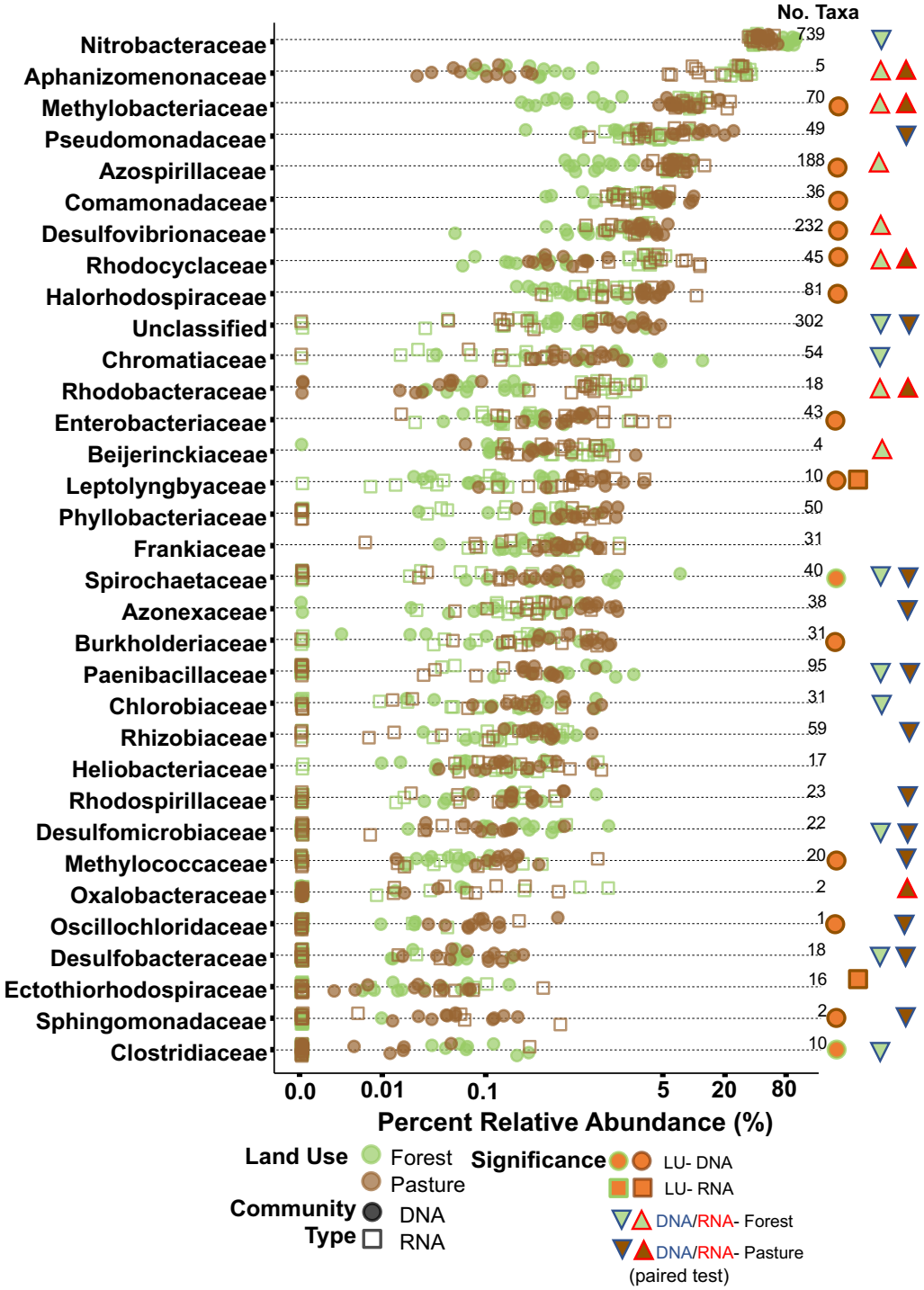
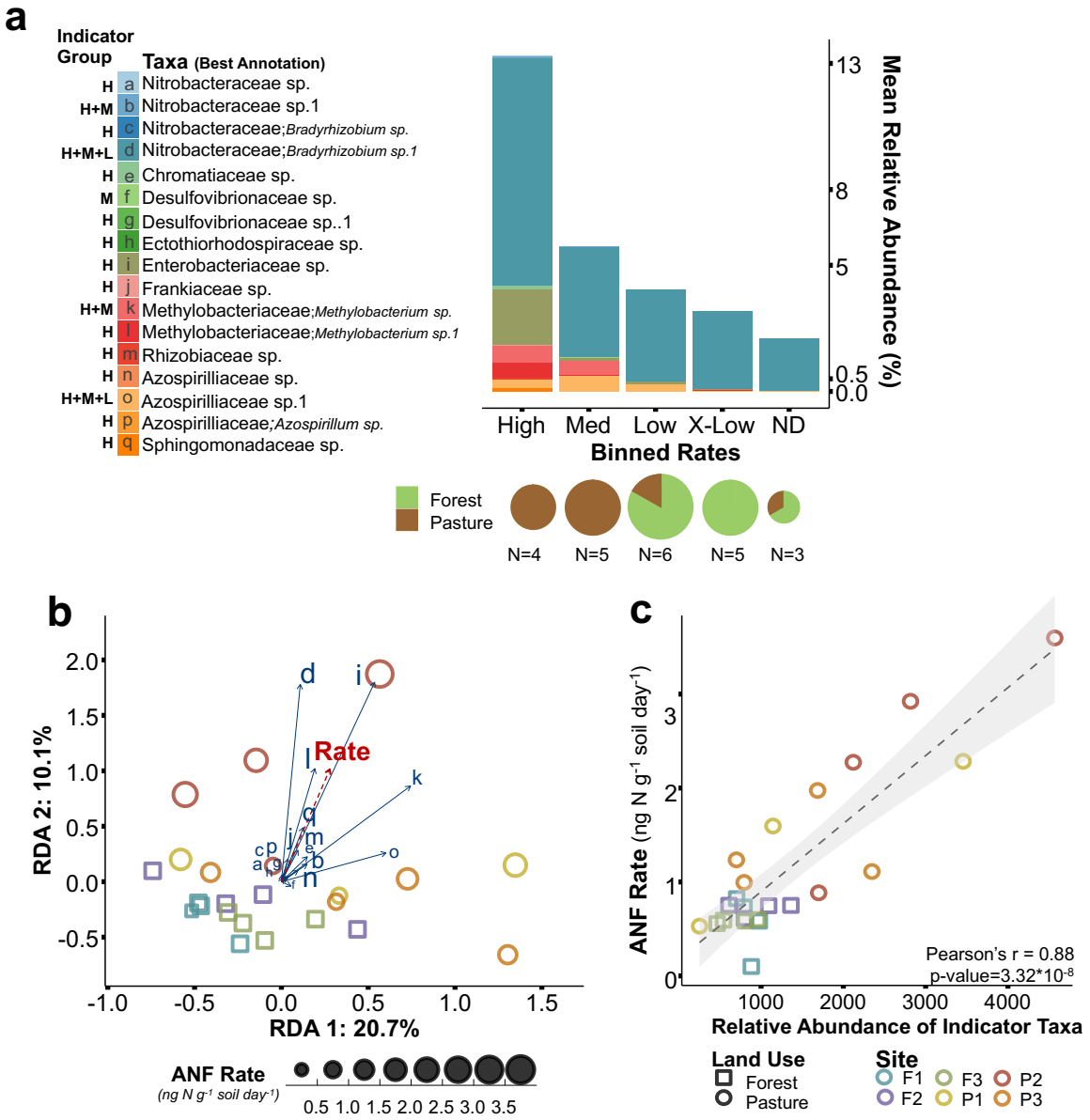


Figure 3-5: IT (derived from nifH-R communities) associated with increasing rates of ANF. Taxa significantly associated with High (H; n=12), Medium (M; n=1), High and Medium (H+M; n=2), and High, Medium, and Low (H+M+L; n=2) were included. **(a)** Relative abundance of each IT in each rate group. **(b)** bi-plot vectors of IT in community ordination (blue) compared to direction of envfit-derived ANF rate correlation (red dashed). Letters correspond to the taxa legend in **(a)**, and letter sizes correspond with level of increasing significance. **(c)** Summed relative abundance of IT against ANF rate, fit with a linear model trendline and 90% confidence interval.



Supplementary Methods

Site spacing

Decimal degrees represent location of quadrat origin points

	F1	F2	F3	P1	P2	P3
Latitude	-10.1405830	-10.1335190	-10.1424440	-10.1703610	-10.1630000	-10.2043970
Longitude	-62.9000000	-62.8882140	-62.8819440	-62.8326110	-62.8363890	-62.7741315

DNA and RNA isolation and processing

Nucleic acid extraction was performed according to manufacturer instructions. Steps were performed with user-provided biotechnology grade phenol: chloroform: isoamyl alcohol (25:24:1). We made a modification to the elution step, whereby elutant was added to spin columns twice at 55% of the recommended volume each in order to increase yield. Average DNA yield from replicate extractions was 55.5 $\pm$ 3.5 ng μl^{-1} (100 µl total), while average RNA yield was 8.8 ng μl^{-1} (100 µl total). While processing RNA using the Qiagen MinElute Cleanup kit, we again added elutant to spin columns twice at 55% of the recommended volume each in order to increase yield. NanoDrop was used to verify RNA quality using 260/280 and 260/230 ratios, ensuring peak values were sufficiently high as compared to spectral noise.

SuperScript IV nifH-targeted reverse-transcription

All work areas were thoroughly cleaned with RNAse Zap, and all consumables including tips and tubes were certified RNAse-free throughout the steps. A second set of reactions was prepared with water rather than transcriptase enzyme to serve as a negative control. Quantitative PCR using 16S primers was performed on negative controls to ensure samples were DNA-free. qPCR reactions were 20 ul in volume containing 1 µl 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') 16S primers each (250 nM in reaction; Caporaso et al., 2011), 10 µl SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 2 µl RT-derived cDNA template (negative controls), and 6 µl PCR-grade water. Reactions were run against positive controls (16S standard curve) on a CFX Connect Real-time thermocycler (Bio-Rad, Hercules, CA, USA). The following cycling conditions were used: 95°C for 3 m, followed by 32 cycles of 95°C for 15 s, 52°C for 20 s

and 72°C for 30 s, followed by a final extension of 72°C for 5 m. Samples with critical threshold values above 29 (water-control value) were considered DNA-free.

Reverse transcriptase primer nifH3**:** 5'- ATRTTRTTNGCNGCRTA -3' position:494-478 *nifH* coverage:100% (Zani et al., 2000)

Quantitative PCR

Template DNA was diluted to 2 ng ul^{-1} and was used in triplicate 20 µl reactions for 16S and nifH gene quantification, using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA). See the table below for reaction specifications and primers used:

Gene	**Primer set**	**Primer reference**	**Primer Vol each (conc.)**	**Vol Sso**	**Vol DNA**	**Vol PCR-H_2O**
nifH	PolF: TGCGAYCCSAARGCBGACTC PolR: ATSGCCATCATYTCRCCGGA	(Poly et al., 2001)	1 µl (500nM)	10 µl	4 µl	4 µl
16S	515F: GTGCCAGCMGCCGCGGTAA 806R: GGACTACHVGGGTWTCTAAT	(Caporaso et al., 2011)	1 µl (250nM)	10 µl	1 µl	7 µl

qPCR was performed using a CFX Connect Real-time thermocycler (Bio-Rad, Hercules, CA, USA) using the following protocols:

Gene	Start	Cycle no.	Denature	Anneal	Extend	Final Extend
nifH-D	98°C, 2m	35	98°C, 15s	68.5°C, 15s	72°C, 30s	72°C, 5m
16S-D	98°C, 2m	32	98°C, 10s	52°C, 10s	72°C, 15s	72°C, 5m

Standards were created using a high-fidelity blunt-end TOPO cloning kit (Thermo Fisher Scientific, Waltham, MA, USA) with 16S and *nifH* genes amplified from *Herbaspirillum seropedicae* (Baldani et al., 1986; ATCC Z 152). Samples and standards for each gene were run on a 96-well reaction plate (90% and 95% efficiency for *nifH* and 16S, respectively; R^2=0.99) and run against standard curves. Copy number was calculated on a basis of per ng DNA.

Amplicon sequencing

Following amplification, *nifH* amplicon samples were inspected visually with gel electrophoresis for correct amplification before combining replicates, followed by quantification of each sample using the Qubit DNA High-sensitivity assay (Invitrogen, Waltham, MA). Samples were then pooled to 50 ng each, and purified across 4 replicate Qiaquick PCR Purification (Qiagen Inc., Valencia, CA, USA) cleanups and re-pooled. The purified libraries were re-quantified and imaged using gel electrophoresis to ensure high product-to-non-specific amplification ratio. PippinHT size selection (Sage Science, Beverly, MA, USA) removed all amplicons outside the 280-500 bp range.

Primer name	**sequence**	**position**	***nifH* coverage**	**reference**
ZM1	TGYGAYCCNAARGCNGA	115-131	95%	(Zehr and McReynolds, 1989)
ZM2	ADNGCCATCATYTCNCC	460-476	94%	(Zehr and McReynolds, 1989)

16S-D was amplified using the 515F/806R primer pair, with 12bp Golay barcodes inserted in the reverse primer. Reactions were run using 20 µl, including 10 µl Phusion High-Fidelity PCR Master Mix (New England Biolabs, Ipswich, MA, USA), 1 µl each forward and reverse primer (500 µl in reaction), 2 µl template DNA (2ng μl^{-1} concentration) and 6 µl PCR-grade water. Reactions were run in triplicate under the following thermocycler reaction: 98°C for 30 s, 27 cycled of 98°C for 10 s, 50°C for 30 s, 72°C for 15 s, and a final extension at 72°C for 7 min. Downstream preparation of the 16S library followed the same methods as described above for the *nifH* library.

Primer name	**sequence**	**Barcoded?**	**reference**
515F	5'-GTGCCAGCMGCCGCGGTAA-3'	No	(Gilbert et al., 2014)
806R	5'-GGACTACHVGGGTWTCTAAT-3'	Yes	(Gilbert et al., 2014)

Sequence processing

All sequence processing steps were run on a machine with 196 GB RAM and 64 processing cores. Sequencing of the *nifH* gene yielded 8.2 million reads across 46 samples. Quality filtering and trimming were performed on paired reads using sickle, with a quality score cutoff of 25 and a minimum length cutoff of 125 bp. This step removed 11.6% of reads. Since adapter ligation prior to sequencing results in approximately 50% of amplicons in the wrong sequencing orientation, demultiplexing needed to be performed twice. Demultiplexing was performed using sabre (https://github.com/najoshi/sabre), with

barcodes targeted in the forward and reverse reads separately, and 1 mismatch of the 12 bp barcode allowed. Across the two attempts at demultiplexing, 75.1% of sequences were retained. Since reads were unmerged at this stage, this resulted in two read-pair libraries per sample. To consolidate these, sequences with barcodes identified in the read 2 sequences were relabeled and joined with sequences with barcodes identified in read 1 sequences. The converse was performed from unbarcoded reads, making sure to retain read order in consolidated read-pair files. Since the primers used were highly degenerate to maximize the sequence diversity captured, a custom script was used to identify and remove forward and reverse primer sequences, taking all possible primer sequences into account, and discarding sequences lacking a possible primer configuration. On average, 94.5% of remaining sequences met this requirement.

The next sequence processing steps mainly followed the protocol described by Gaby et al. (2018), with the inclusion of elements from Angel et al. (2018), and an added step to denoise sequencing errors. Following this initial reformatting phase, denoising was accomplished using DADA2 (Callahan et al., 2016). Samples were trimmed to 200 bp in read 1 and 180 bp in read 2, merged with a maximum mismatch set to 10, and screened for chimeric sequences. Approximately 63% of sequences were successfully processed.

The below table shows the percent of sequences removed by each successive DADA2 step, and the absolute percent of sequences remaining at the end of processing.

	Filter and Trim	**Denoising**	**Merging**	**Chimera Removal**	**Percent Left (absolute)**
Average	28%	1.31%	1.48%	2.47%	62.9%
SE	0.94%	0.06%	0.11%	0.55%	1.56%

The resulting amplicon sequence variant (ASV) table from this step was reformatted to a fasta file and hmmsearch (http://hmmer.org/) was used with the option -domtblout and an e-value threshold of 0.001 against the reference model hmm_nuc_1160_nifH.hmm (Angel et al., 2018) to filter non-*nifH* sequences. This identified 72 erroneous ASVs out of 47,487, which were subsequently removed. Framebot (Wang et al., 2013) was used to perform frameshift correction using the reference database nifH_prot_ref.fasta (Angel et al., 2018).

Finally, sequences were clustered to OTUs using vsearch (Rognes et al.) using the option -cluster_fast, with an identification cutoff of 95% and dual-strand consideration (Gaby et al., 2018). All

clusters with 4 or fewer counts were removed, resulting in 2,374 OTU clusters. A nucleotide-based local BLAST was performed using seqDatabase.fasta (Gaby et al., 2018) as the reference database with an e-value cutoff of 0.01. A custom script was used to configure a final taxonomic table using percent ID cutoffs specified in Gaby et al. (2018): 91.9% for species, 88.1% for genus, and 75% for family.

16S data was processed through the DADA2 pipeline utilizing the R environment with default settings, aside from the following modifications: During filter and trim, forward and reverse reads were truncated at 200 and 150 bp, respectively, and a maximum error value of 2 was used. During the merge step, a maximum mismatch of 10 was allowed.

The below table shows the percent of sequences removed by each successive DADA2 step, and the absolute percent of sequences remaining at the end of processing.

	Filter and Trim	Denoising	Merging	Chimera Removal	Percent Left (absolute)
Average	7.42%	6.52%	40.1%	15.6%	42.9%
SE	0.2%	0.13%	0.44%	0.69%	0.3%

References

Angel, R., Nepel, M., Panholzl, C., Schmidt, H., Herbold, C. W., Eichorst, S. A., and Woebken, D. (2018). Evaluation of primers targeting the diazotroph functional gene and development of NifMAP - a bioinformatics pipeline for analyzing nifH amplicon data. *Front Microbiol* **9**, 703.

Baldani, J., Baldani, V., Seldin, L., and Döbereiner, J. (1986). Characterization of Herbaspirillum seropedicae gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. *International Journal of Systematic and Evolutionary Microbiology* **36**, 86-93.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-3.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences* **108**, 4516-4522.

Gaby, J. C., Rishishwar, L., Valderrama-Aguirre, L. C., Green, S. J., Valderrama-Aguirre, A., Jordan, I. K., and Kostka, J. E. (2018). Diazotroph community characterization via a high-throughput *nifH* amplicon sequencing and analysis pipeline. *Appl Environ Microbiol* **84**.

Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC biology* **12**, 1-4.

Poly, F., Monrozier, L. J., and Bally, R. (2001). Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil. *Research in microbiology* **152**, 95-103.

Rognes, T., Flouri, T., Nichols, B., and Quince, C. Mah e, F.(2016). VSEARCH: A versatile open source tool for metagenomics. PeerJ.

Wang, Q., Quensen III, J. F., Fish, J. A., Kwon Lee, T., Sun, Y., Tiedje, J. M., and Cole, J. R. (2013). Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *MBio* **4**, e00592-13.

Zani, S., Mellon, M. T., Collier, J. L., and Zehr, J. P. (2000). Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Applied and Environmental Microbiology* **66**, 3119-3124.

Zehr, J. P., and McReynolds, L. A. (1989). Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium Trichodesmium thiebautii. *Applied and environmental microbiology* **55**, 2522-2526.

Supplementary Tables and Figures

Supp. Table 3-1: Full correlation of measured environmental variables, activity and community size, and Shannon diversity indices across 23 F and P samples. Values represent **(a)** Pearson's r and **(b)** Spearman's ρ estimates. Plots shown in Figure 3-1b were selected based on FDR-adjusted p-values < 0.05, which correspond to r and ρ cutoffs of > 0.55 and > 0.45, respectively.

(a) Pearson	NH4+	log.NO3-	OrganicN	ln.TN	POXC	TC	Mo	V	Fe	S	P	pH	Clay	Sand	δ15N	ExtractN	K	Mg	Ca	Acidity	CEC	Z	Cu	Al	log.*nifH*.copies	ANF rate	log.16S.copies	Shannon *nifD*	Shannon 16S	Shannon *nifR*
NH4+	1.00																												
log.NO3-	-0.73	1.00																											
Organic N	0.43	-0.57	1.00																										
ln.TN	-0.34	0.77	-0.50	1.00																									
POXC	0.13	-0.11	0.47	-0.41	1.00																								
TC	0.04	0.03	0.29	-0.29	0.96	1.00																							
Mo	-0.30	0.65	-0.31	0.64	0.14	0.28	1.00																						
V	0.24	-0.55	0.71	-0.52	0.48	0.42	-0.26	1.00																					
Fe	0.01	-0.33	0.57	-0.35	0.43	0.42	-0.08	0.93	1.00																				
S	-0.40	0.56	-0.10	0.47	-0.18	-0.10	0.18	-0.18	0.02	1.00																			
P	-0.12	0.06	-0.06	-0.05	-0.04	0.03	0.12	0.10	0.34	0.25	1.00																		
pH	0.50	-0.66	0.35	-0.59	0.21	0.11	-0.46	0.58	0.35	-0.69	-0.11	1.00																	
Clay	-0.55	0.42	-0.14	0.39	-0.35	-0.27	0.08	-0.02	0.16	0.70	0.06	-0.53	1.00																
Sand	0.42	-0.18	-0.05	-0.18	0.28	0.26	0.16	-0.19	-0.28	-0.56	0.06	0.28	-0.93	1.00															
δ15N	-0.79	0.80	-0.47	0.63	-0.32	-0.18	0.44	-0.38	-0.15	0.66	0.03	-0.73	0.81	-0.61	1.00														
ExtractN	0.29	-0.25	0.91	-0.18	0.56	0.43	-0.03	0.60	0.54	0.11	-0.05	0.12	-0.04	-0.08	-0.26	1.00													
K	-0.01	0.28	-0.11	0.32	-0.09	-0.12	0.12	-0.16	-0.17	-0.03	0.06	0.23	-0.33	0.30	-0.06	0.02	1.00												
Mg	0.06	-0.11	0.36	-0.20	0.46	0.36	-0.25	0.46	0.38	-0.13	-0.15	0.52	-0.22	0.00	-0.26	0.40	0.48	1.00											
Ca	0.13	-0.21	0.28	-0.38	0.63	0.67	-0.01	0.70	0.58	-0.25	-0.13	0.60	-0.21	0.08	-0.27	0.26	-0.05	0.50	1.00										
Acidity	-0.42	0.61	-0.15	0.38	0.21	0.29	0.52	-0.37	-0.15	0.63	0.21	-0.87	0.40	-0.19	0.58	0.12	-0.21	-0.30	-0.34	1.00									
CEC	-0.18	0.25	0.19	-0.09	0.79	0.87	0.35	0.43	0.47	0.22	0.02	-0.01	0.07	-0.05	0.14	0.38	-0.11	0.37	0.73	0.38	1.00								
Z	0.23	-0.38	0.39	-0.42	0.30	0.22	-0.33	0.65	0.51	-0.18	-0.08	0.73	-0.19	0.01	-0.35	0.26	0.12	0.64	0.75	-0.60	0.34	1.00							
Cu	0.32	-0.56	0.37	-0.52	0.11	0.04	-0.30	0.61	0.59	-0.31	0.51	0.68	-0.34	0.23	-0.54	0.15	0.11	0.28	0.39	-0.58	-0.03	0.67	1.00						
Al	0.23	-0.38	0.39	-0.42	0.30	0.22	-0.33	0.65	0.51	-0.18	-0.08	0.73	-0.19	0.01	-0.35	0.26	0.12	0.64	0.75	-0.60	0.34	1.00	0.67	1.00					
log.*nifH*.copies	0.66	-0.88	0.51	-0.72	0.03	-0.13	-0.68	0.43	0.18	-0.56	-0.08	0.75	-0.51	0.27	-0.84	0.18	-0.06	0.17	0.18	-0.72	-0.34	0.38	0.55	0.38	1.00				
corr_rate	0.58	-0.64	0.04	-0.45	-0.15	-0.20	-0.55	0.02	-0.18	-0.47	-0.19	0.41	-0.37	0.25	-0.54	-0.21	-0.32	-0.17	0.07	-0.49	-0.33	0.10	0.13	0.10	0.66	1.00			
log.16S.copies	0.43	-0.63	0.38	-0.66	0.23	0.17	-0.36	0.34	0.23	-0.56	0.17	0.55	-0.58	0.42	-0.60	0.19	0.04	0.20	0.25	-0.42	-0.05	0.29	0.53	0.29	0.59	0.39	1.00		
Shannon *nifD*	0.61	-0.76	0.19	-0.62	0.05	-0.08	-0.46	0.20	-0.05	-0.77	-0.22	0.72	-0.67	0.47	-0.83	-0.13	0.00	0.14	0.15	-0.67	-0.34	0.30	0.42	0.30	0.80	0.57	0.45	1.00	
Shannon 16S	-0.07	-0.05	-0.10	-0.05	-0.43	-0.51	-0.41	-0.03	-0.03	0.23	0.29	0.17	0.20	-0.22	0.02	-0.24	0.29	0.11	-0.20	-0.24	-0.34	0.27	0.32	0.27	0.12	-0.02	-0.19	0.13	1.00
Shannon *nifR*	0.43	-0.39	0.28	-0.33	0.09	0.04	-0.31	0.28	0.14	-0.17	-0.22	0.43	-0.14	0.01	-0.27	0.15	-0.04	0.30	0.31	-0.44	0.00	0.51	0.32	0.51	0.42	0.28	0.35	0.24	-0.13	1.00

Supp. Table 3-1, *contd.*:

(b) Spearman	NH4+	log.NO3-	OrganicN	In.TN	POXC	TC	Mo	V	Fe	S	P	pH	Clay	Sand	δ15N	ExtractN	K	Mg	Ca	Acidity	CEC	Z	Cu	Al	log.*nifH*.copies	ANF rate	log.16S.copies	Shannon *nifD*	Shannon 16S	Shannon *nifR*
NH4+	1.00																													
log.NO3-	-0.58	1.00																												
Organic N	0.57	-0.45	1.00																											
In.TN	-0.57	0.80	-0.67	1.00																										
POXC	0.45	-0.35	0.61	-0.72	1.00																									
TC	0.38	-0.18	0.61	-0.57	0.88	1.00																								
Mo	-0.49	0.73	-0.35	0.62	-0.36	-0.25	1.00																							
V	0.23	-0.25	0.74	-0.57	0.68	0.74	-0.24	1.00																						
Fe	0.05	-0.08	0.57	-0.37	0.55	0.70	-0.01	0.92	1.00																					
S	-0.37	0.57	-0.08	0.51	-0.14	0.02	0.36	0.11	0.16	1.00																				
P	-0.28	0.43	-0.05	0.24	0.00	0.16	0.32	0.03	0.08	0.51	1.00																			
pH	0.59	-0.67	0.43	-0.69	0.45	0.25	-0.54	0.30	0.13	-0.60	-0.32	1.00																		
Clay	-0.59	0.36	-0.19	0.44	-0.30	-0.11	0.24	0.10	0.24	0.53	0.07	-0.46	1.00																	
Sand	0.43	-0.16	0.00	-0.23	0.06	-0.11	-0.02	-0.24	-0.32	-0.45	0.01	0.34	-0.93	1.00																
δ15N	-0.81	0.69	-0.44	0.69	-0.46	-0.30	0.67	-0.22	-0.03	0.58	0.21	-0.77	0.75	-0.58	1.00															
ExtractN	0.32	0.04	0.79	-0.23	0.49	0.61	0.02	0.61	0.60	0.18	0.05	0.09	-0.01	-0.12	-0.10	1.00														
K	0.05	0.22	-0.18	0.29	-0.12	-0.20	0.26	-0.21	-0.19	0.01	0.04	0.35	-0.24	0.31	-0.16	0.00	1.00													
Mg	0.15	-0.11	0.17	-0.28	0.57	0.48	-0.18	0.43	0.40	-0.10	-0.28	0.52	-0.15	0.03	-0.27	0.26	0.51	1.00												
Ca	0.40	-0.42	0.34	-0.64	0.72	0.61	-0.30	0.47	0.45	-0.50	-0.40	0.69	-0.39	0.24	-0.52	0.27	0.20	0.81	1.00											
Acidity	-0.43	0.65	-0.16	0.53	-0.06	0.06	0.56	-0.07	0.06	0.66	0.44	-0.84	0.31	-0.28	0.60	0.23	-0.20	-0.28	-0.46	1.00										
CEC	-0.12	0.37	0.31	-0.02	0.52	0.67	0.28	0.53	0.60	0.57	0.29	-0.24	0.31	-0.37	0.33	0.61	-0.06	0.37	0.24	0.56	1.00	0.12								
Z	0.28	-0.46	0.24	-0.52	0.51	0.31	-0.32	0.38	0.32	-0.29	-0.33	0.76	-0.27	0.17	-0.50	0.06	0.44	0.82	0.81	-0.54	0.12	1.00								
Cu	0.46	-0.78	0.45	-0.70	0.40	0.14	-0.53	0.36	0.17	-0.43	-0.36	0.86	-0.34	0.23	-0.63	0.00	0.20	0.45	0.57	-0.76	-0.25	0.80	1.00							
Al	0.28	-0.46	0.24	-0.52	0.51	0.31	-0.32	0.38	0.32	-0.29	-0.33	0.76	-0.27	0.17	-0.50	0.06	0.44	0.82	0.81	-0.54	0.12	1.00	0.80	1.00						
log.*nifH*.copies	0.59	-0.71	0.48	-0.68	0.34	0.16	-0.77	0.24	-0.01	-0.53	-0.37	0.73	-0.42	0.24	-0.73	0.07	-0.01	0.20	0.38	-0.75	-0.41	0.36	0.67	0.36	1.00					
corr_rate	0.49	-0.69	0.33	-0.62	0.28	0.24	-0.74	0.12	0.00	-0.39	-0.23	0.37	-0.30	0.11	-0.56	-0.06	-0.39	-0.08	0.24	-0.47	-0.26	0.13	0.39	0.13	0.75	1.00				
log.16S.copies	0.55	-0.64	0.45	-0.70	0.43	0.35	-0.34	0.17	0.09	-0.55	-0.18	0.59	-0.55	0.38	-0.60	0.13	0.01	0.25	0.58	-0.51	-0.14	0.47	0.58	0.47	0.51	0.50	1.00			
Shannon *nifD*	0.62	-0.80	0.15	-0.62	0.22	-0.05	-0.67	-0.05	-0.29	-0.66	-0.49	0.67	-0.60	0.45	-0.78	-0.32	-0.01	0.11	0.34	-0.68	-0.60	0.38	0.66	0.38	0.74	0.58	0.50	1.00		
Shannon 16S	-0.13	-0.18	-0.16	-0.07	-0.04	-0.11	-0.17	0.03	-0.04	0.16	0.26	0.32	0.16	-0.10	-0.09	-0.35	0.32	0.19	-0.03	-0.25	-0.05	0.35	0.34	0.35	0.03	-0.15	-0.12	0.17	1.00	
Shannon *nifR*	0.39	-0.31	0.34	-0.26	0.38	0.34	-0.24	0.32	0.20	-0.01	-0.14	0.33	-0.04	-0.04	-0.19	0.14	-0.05	0.28	0.32	-0.26	0.08	0.32	0.46	0.32	0.37	0.24	0.30	0.21	-0.04	1.00

Supp. Table 3-2: Environmental variables tested against community composition **(left)** results from PERMANOVAs run to test the variation in community composition explained by physicochemical variables after accounting for independent effects of LU and GD on physicochemical variables. **(right)** Results of envfit tests of each variable. Both tests were used to select meaningful variables.

16S-D Community

16S	PERMANOVA Test						Envfit Test					
	Variable	Sum Squares	R2	F	Pr(>F)			MDS1	MDS2	r2	Pr(>r)	
1	NH4+	0.16	0.03	0.95	0.465		NH4+	0.79	0.62	0.51	0.004	**
2	ln(NO3-)	0.13	0.02	0.77	0.658		ln(NO3-)	-0.79	-0.61	0.80	0.001	**
3	OrganicN	0.38	0.07	2.46	0.013		OrganicN	0.72	0.70	0.27	0.042	*
4	In.TN	0.26	0.05	1.64	0.083		In.TN	-0.80	-0.60	0.62	0.001	**
5	POXC	0.27	0.05	1.67	0.055		POXC	0.78	-0.63	0.09	0.41	
6	TC	0.28	0.05	1.74	0.037		TC	0.42	-0.91	0.04	0.82	
7	Mo	0.19	0.03	1.13	0.272		Mo	-0.78	-0.63	0.35	0.017	*
8	V	0.23	0.04	1.39	0.175		V	0.94	0.35	0.19	0.116	
9	Fe	0.23	0.04	1.44	0.142		Fe	0.84	0.54	0.03	0.747	
10	S	0.36	0.06	2.34	0.027		S	-0.98	0.21	0.57	0.001	**
11	P	0.21	0.04	1.30	0.173		P	-0.51	0.86	0.11	0.34	
12	pH	0.41	0.07	2.65	0.009		pH	0.99	-0.14	0.74	0.001	**
13	Clay	0.37	0.06	2.36	0.019		Clay	-0.98	0.19	0.59	0.001	**
14	Sand	0.19	0.03	1.15	0.297		Sand	0.98	-0.22	0.26	0.054	.
15	δ15N	0.21	0.04	1.29	0.177		δ15N	-0.93	-0.36	0.84	0.001	**
16	ExtractN	0.43	0.08	2.84	0.012		ExtractN	0.72	0.69	0.04	0.675	
17	K	0.28	0.05	1.78	0.081		K	0.22	-0.98	0.34	0.014	*
18	Mg	0.32	0.06	2.04	0.025		Mg	0.47	-0.88	0.47	0.004	**
19	Ca	0.23	0.04	1.38	0.136		Ca	0.76	-0.65	0.18	0.087	.
20	CEC	0.29	0.05	1.83	0.047		CEC	-0.40	-0.92	0.08	0.457	
21	Cu	0.23	0.04	1.44	0.125		Cu	0.97	0.25	0.33	0.01	**
22	Al	0.31	0.05	1.97	0.048		Al	0.82	-0.57	0.34	0.009	**
23	ln(nifH copies)	0.16	0.03	0.93	0.476		ln(nifH copies	0.79	0.62	0.93	0.001	**
24	ANF	0.17	0.03	1.04	0.347		ANF	0.64	0.77	0.40	0.008	**
25	ln(16S copies)	0.21	0.04	1.29	0.199		ln(16S copies	0.97	0.25	0.50	0.004	**
26	nifD div	0.46	0.08	3.09	0.004		nifD div	0.97	0.25	0.77	0.001	**
27	Prok div	0.27	0.05	1.71	0.069		Prok div	0.52	-0.86	0.00	0.985	
28	RNA div	0.12	0.02	0.73	0.697		RNA div	0.90	0.44	0.19	0.133	

Supp. Table 3-2, *contd.:*

***nifH*-D community**

nifH-D	PERMANOVA Test						Envfit Test					
	Var	Sum Squares	R2	F	Pr(>F)			MDS1	MDS2	r2	Pr(>r)	
1	NH4+	0.10	0.02	0.75	0.636		NH4+	0.93	0.36	0.49	0.002	*
2	ln(NO3-)	0.06	0.01	0.50	0.952		ln(NO3-)	-0.98	-0.22	0.82	0.001	*
3	OrganicN	0.28	0.06	2.35	0.033		OrganicN	0.51	0.86	0.36	0.017	*
4	ln.TN	0.21	0.05	1.70	0.1		ln.TN	-0.98	-0.19	0.61	0.001	*
5	POXC	0.21	0.05	1.72	0.062		POXC	0.89	-0.46	0.01	0.931	
6	TC	0.21	0.05	1.70	0.061		TC	-0.86	-0.50	0.00	0.975	
7	Mo	0.16	0.04	1.26	0.223		Mo	-0.95	-0.31	0.41	0.007	*
8	V	0.13	0.03	0.99	0.369		V	0.81	0.59	0.18	0.136	
9	Fe	0.13	0.03	1.00	0.372		Fe	0.35	0.93	0.06	0.556	
10	S	0.30	0.07	2.56	0.016		S	-0.52	0.85	0.65	0.001	*
11	P	0.13	0.03	0.99	0.415		P	-0.14	0.99	0.10	0.386	
12	pH	0.24	0.05	1.95	0.064		pH	0.75	-0.66	0.65	0.001	*
13	Clay	0.23	0.05	1.86	0.069		Clay	-0.60	0.80	0.45	0.001	*
14	Sand	0.14	0.03	1.14	0.239		Sand	0.46	-0.89	0.22	0.073	.
15	δ15N	0.16	0.04	1.29	0.186		δ15N	-0.99	0.15	0.76	0.001	*
16	ExtractN	0.31	0.07	2.66	0.016		ExtractN	0.13	0.99	0.18	0.143	
17	K	0.14	0.03	1.14	0.274		K	-0.17	-0.99	0.23	0.079	.
18	Mg	0.20	0.05	1.66	0.077		Mg	0.18	-0.98	0.17	0.154	
19	Ca	0.13	0.03	1.06	0.351		Ca	0.54	-0.84	0.06	0.606	
20	CEC	0.20	0.04	1.62	0.119		CEC	-0.95	0.32	0.08	0.446	
21	Cu	0.10	0.02	0.75	0.672		Cu	0.98	-0.21	0.30	0.019	*
22	Al	0.14	0.03	1.13	0.302		Al	0.67	-0.75	0.19	0.111	
23	ln(nifH copies)	0.13	0.03	1.06	0.312		ln(nifH copies	0.98	0.21	0.88	0.001	*
24	ANF	0.07	0.02	0.57	0.883		ANF	0.94	0.33	0.42	0.006	*
25	ln(16S copies)	0.17	0.04	1.35	0.191		ln(16S copies	1.00	-0.05	0.31	0.024	*
26	nifD div	0.44	0.10	3.98	0.003		nifD div	0.80	-0.59	0.96	0.001	*
27	Prok div	0.13	0.03	1.03	0.349		Prok div	0.75	-0.66	0.04	0.689	
28	RNA div	0.09	0.02	0.71	0.693		RNA div	0.96	0.26	0.12	0.297	

Supp. Table 3-2, *contd.:*

***nifH*-R community**

nifH-R	PERMANOVA Test						Envfit Test					
	Var	Sum Squares	R2	F	Pr(>F)			MDS1	MDS2	r2	Pr(>r)	
1	NH4+	0.12	0.05	1.22	0.188		NH4+	0.54	0.84	0.50	0.001	**
2	ln(NO3-)	0.10	0.04	1.01	0.396		ln(NO3-)	-0.54	-0.84	0.48	0.004	**
3	OrganicN	0.09	0.04	0.86	0.647		OrganicN	0.91	-0.42	0.11	0.276	
4	In.TN	0.07	0.03	0.69	0.888		In.TN	-0.61	-0.79	0.26	0.044	*
5	POXC	0.08	0.03	0.76	0.721		POXC	0.30	-0.95	0.02	0.767	
6	TC	0.08	0.03	0.75	0.737		TC	0.03	-1.00	0.03	0.702	
7	Mo	0.08	0.03	0.74	0.846		Mo	-0.62	-0.78	0.27	0.058	.
8	V	0.11	0.05	1.09	0.314		V	0.89	-0.45	0.10	0.334	
9	Fe	0.11	0.05	1.09	0.29		Fe	0.45	-0.89	0.09	0.398	
10	S	0.08	0.03	0.73	0.856		S	-0.35	-0.94	0.19	0.105	
11	P	0.12	0.05	1.14	0.27		P	-0.93	-0.38	0.02	0.764	
12	pH	0.09	0.04	0.85	0.643		pH	0.85	0.53	0.24	0.077	.
13	Clay	0.10	0.04	0.95	0.485		Clay	-0.37	-0.93	0.13	0.262	
14	Sand	0.10	0.04	1.02	0.372		Sand	0.19	0.98	0.07	0.494	
15	δ15N	0.10	0.04	1.01	0.411		δ15N	-0.45	-0.89	0.32	0.018	*
16	ExtractN	0.07	0.03	0.71	0.932		ExtractN	0.34	-0.94	0.12	0.259	
17	K	0.11	0.04	1.05	0.342		K	-0.13	-0.99	0.12	0.269	
18	Mg	0.10	0.04	0.95	0.492		Mg	0.43	-0.90	0.26	0.041	*
19	Ca	0.10	0.04	0.97	0.452		Ca	0.90	-0.44	0.07	0.415	
21	CEC	0.07	0.03	0.72	0.885		CEC	-0.16	-0.99	0.09	0.34	
23	Cu	0.13	0.05	1.28	0.184		Cu	1.00	0.09	0.16	0.146	
24	Al	0.12	0.05	1.17	0.263		Al	0.90	-0.43	0.29	0.051	.
25	ln(nifH copies)	0.09	0.04	0.87	0.643		ln(nifH copies	0.58	0.82	0.42	0.004	**
26	ANF	0.15	0.06	1.53	0.07		ANF	0.27	0.96	0.82	0.001	**
27	ln(16S copies)	0.11	0.04	1.04	0.351		ln(16S copies	0.77	0.64	0.18	0.141	
28	nifD div	0.08	0.03	0.77	0.83		nifD div	0.37	0.93	0.35	0.013	*
29	Prok div	0.11	0.05	1.11	0.273		Prok div	-0.74	0.67	0.01	0.946	
30	RNA div	0.36	0.15	4.09	0.001		RNA div	1.00	-0.08	0.93	0.001	**

Supp. Table 3-3: Output results from OTU-level edgeR differential abundance testing **(external .xlsx)** for the four comparisons made: (1) LU DNA: F vs P, (2) LU RNA F vs P, (3) Paired F DNA vs RNA, and (4) Paired P DNA vs RNA. P-value are adjusted for multiple comparisons using the BH method. OTUs tested for each comparison were present in at least 2 samples of each treatment, and 8 (~1/3) samples overall. Significance determined at p.adjusted $\leq$ 0.05.

Supp. Table 3-4: edgeR output results for family-level community abundance changes, corresponding to annotations in Figure 3-4. The sign of LU log_2-fold change indicates P enrichment if positive and F enrichment if negative. The sign in paired-sample comparisons indicates RNA enrichment if positive and RNA depletion if negative. CPM = counts per million. p-value are BH-adjusted for multiple comparisons.

LU:DNA	logFC	logCPM	F	p-value*
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Nitrobacteraceae	-0.68	19.30	20.94	4.57E-04
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	1.40	16.06	3.28	1.44E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Methylobacteriaceae	**4.07**	**15.46**	**25.83**	**1.97E-04**
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	**2.26**	**15.33**	**12.91**	**4.61E-03**
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Azospirillaceae	**2.98**	**15.19**	**24.11**	**2.22E-04**
Proteobacteria;Gammaproteobacteria;Chromatiales;Halorhodospiraceae	**2.62**	**14.51**	**15.29**	**2.18E-03**
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae	**1.91**	**14.48**	**7.04**	**3.54E-02**
Unclassified;Unclassified;Unclassified;Unclassified;Unclassified	0.53	14.15	0.68	5.24E-01
Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae	0.94	14.14	1.26	3.70E-01
Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae	**-2.24**	**13.08**	**7.11**	**3.54E-02**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Phyllobacteriaceae	0.83	12.75	1.00	4.28E-01
Firmicutes;Bacilli;Bacillales;Paenibacillaceae	-1.74	12.70	4.09	1.04E-01
Cyanobacteria;Cyanophyceae;Synechococcales;Leptolyngbyaceae	**2.41**	**12.65**	**6.04**	**4.76E-02**
Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae	2.24	12.63	3.65	1.26E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae	**2.63**	**12.40**	**6.70**	**3.82E-02**
Actinobacteria;Actinobacteria;Frankiales;Frankiaceae	1.23	12.26	2.13	2.53E-01
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae	2.11	12.21	7.53	6.91E-02
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae	**2.38**	**12.12**	**8.00**	**3.15E-02**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae	1.31	11.97	1.52	3.55E-01
Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae	0.74	11.91	0.41	6.00E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Rhizobiaceae	0.05	11.66	0.00	9.49E-01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae	-1.21	11.25	1.29	3.70E-01
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	0.55	11.18	0.32	6.35E-01
Cyanobacteria;Nostocales;Nostocaceae;Aphanizomenonaceae	-0.73	11.11	0.61	5.24E-01
Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae	0.84	10.79	0.60	5.24E-01
Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae	**2.43**	**9.90**	**5.02**	**3.46E-02**
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae	0.17	9.40	0.02	9.25E-01
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	-1.39	9.40	1.29	3.70E-01
Chloroflexi;Chloroflexia;Chloroflexales;Oscillochloridaceae	**10.49**	**9.20**	**32.13**	**1.21E-04**
Firmicutes;Clostridia;Clostridiales;Clostridiaceae	**-4.96**	**8.65**	**5.91**	**4.76E-02**
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae	**8.75**	**8.53**	**17.75**	**1.42E-03**
Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae	-0.62	8.38	0.19	7.12E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	-6.56	6.56	3.02	1.58E-01

Supp. Table 3-4, *contd.*:

LU:RNA	logFC	logCPM	F	p-value*
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Nitrobacteraceae	0.30	18.54	5.28	2.58E-02
Cyanobacteria;Nostocales;Nostocaceae;Aphanizomenonaceae	-0.98	17.77	3.30	8.66E-02
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Methylobacteriaceae	-0.21	16.68	0.45	5.05E-01
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Azospirillaceae	-0.33	16.01	0.87	3.54E-01
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae	0.45	15.43	0.82	3.70E-01
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	0.60	15.28	1.19	2.80E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	0.18	14.82	0.10	7.50E-01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae	0.13	14.63	0.06	8.15E-01
Proteobacteria;Gammaproteobacteria;Chromatiales;Halorhodospiraceae	1.44	14.28	3.62	6.30E-02
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	-1.00	13.71	1.64	2.07E-01
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae	2.30	13.06	2.71	1.13E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae	-0.94	12.92	0.99	3.24E-01
Cyanobacteria;Cyanophyceae;Synechococcales;Leptolyngbyaceae	**4.05**	**12.27**	**8.33**	**5.75E-03**
Actinobacteria;Actinobacteria;Frankiales;Frankiaceae	2.07	12.24	2.67	1.08E-01
Unclassified;Unclassified;Unclassified;Unclassified;Unclassified	1.12	11.82	0.46	5.02E-01
Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae	1.69	11.62	0.95	3.34E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Phyllobacteriaceae	1.91	11.57	0.58	4.53E-01
Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae	1.67	11.48	0.91	3.44E-01
Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae	-0.44	11.46	0.08	7.79E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae	1.39	11.43	0.51	4.78E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	3.28	10.60	2.14	1.50E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Rhizobiaceae	-2.06	10.59	0.82	3.71E-01
Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae	2.56	9.91	1.05	3.10E-01
Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae	1.47	9.81	0.39	5.33E-01
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	5.64	9.70	3.94	5.27E-02
Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae	0.32	9.62	0.02	8.96E-01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae	-0.35	8.55	0.02	9.00E-01
Firmicutes;Bacilli;Bacillales;Paenibacillaceae	2.50	8.52	0.81	3.73E-01
Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae	9.15	8.33	5.44	2.47E-02
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae	6.52	8.24	2.98	9.25E-02
Chloroflexi;Chloroflexia;Chloroflexales;Oscillochloridaceae	8.25	7.21	3.92	5.46E-02
Firmicutes;Clostridia;Clostridiales;Clostridiaceae	0.00	7.16	0.00	1.00E+00
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae	4.34	6.92	1.30	2.60E-01

Supp. Table 3-4, *contd.*:

CT: Forest	logFC	logCPM	F	p-value*
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Nitrobacteraceae	**-1.20**	**19.44**	**7.82**	**1.17E-02**
Cyanobacteria;Nostocales;Nostocaceae;Aphanizomenonaceae	**6.66**	**17.09**	**132.15**	**8.96E-25**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Methylobacteriaceae	**3.47**	**15.67**	**48.79**	**9.04E-11**
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Azospirillaceae	**2.47**	**15.24**	**50.20**	**7.99E-11**
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	0.56	14.87	2.61	1.77E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	0.47	14.51	1.86	2.65E-01
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae	**3.57**	**14.21**	**80.49**	**2.62E-16**
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae	**1.59**	**14.12**	**15.87**	**2.25E-04**
Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae	**-3.76**	**13.52**	**16.83**	**1.52E-04**
Proteobacteria;Gammaproteobacteria;Chromatiales;Halorhodospiraceae	0.67	13.25	1.43	3.06E-01
Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae	**-6.61**	**13.18**	**32.84**	**8.66E-08**
Unclassified;Unclassified;Unclassified;Unclassified;Unclassified	**-3.01**	**12.91**	**11.78**	**1.70E-03**
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	**3.94**	**12.89**	**49.50**	**8.20E-11**
Firmicutes;Bacilli;Bacillales;Paenibacillaceae	**-9.73**	**12.69**	**43.88**	**6.94E-10**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae	**1.90**	**12.44**	**7.70**	**1.17E-02**
Actinobacteria;Actinobacteria;Frankiales;Frankiaceae	0.43	12.14	0.31	6.11E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Phyllobacteriaceae	-2.55	11.66	4.81	5.04E-02
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae	0.04	11.57	0.00	9.64E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Rhizobiaceae	-0.90	11.20	0.68	4.82E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae	-2.00	10.90	1.50	3.06E-01
Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae	0.94	10.90	1.22	3.44E-01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae	**-8.63**	**10.89**	**24.45**	**4.29E-06**
Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae	**-3.30**	**10.73**	**10.62**	**2.88E-03**
Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae	0.80	10.66	0.60	4.95E-01
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	-4.27	10.56	5.24	4.15E-02
Cyanobacteria;Cyanophyceae;Synechococcales;Leptolyngbyaceae	-1.31	10.18	1.44	3.06E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	2.76	9.81	1.83	2.65E-01
Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae	-1.35	8.99	0.87	4.29E-01
Firmicutes;Clostridia;Clostridiales;Clostridiaceae	**-8.31**	**8.98**	**34.18**	**5.29E-08**
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae	**-6.25**	**8.18**	**17.60**	**1.13E-04**
Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae	**-5.08**	**8.17**	**7.64**	**1.17E-02**
Chloroflexi;Chloroflexia;Chloroflexales;Oscillochloridaceae	-1.11	6.20	0.29	6.11E-01
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae	-0.84	4.95	0.57	4.95E-01

Supp. Table 3-4, *contd.*:

CT:Pasture	logFC	logCPM	F	p-value*
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Nitrobacteraceae	0.08	18.77	0.80	4.10E-01
Cyanobacteria;Nostocales;Nostocaceae;Aphanizomenonaceae	**7.82**	**16.80**	**136.59**	**3.71E-06**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Methylobacteriaceae	**1.00**	**16.65**	**21.02**	**4.26E-05**
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	**-1.09**	**16.14**	**14.72**	**8.21E-04**
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Azospirillaceae	0.36	15.99	1.74	2.21E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	-0.78	15.42	5.11	4.48E-02
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae	**3.44**	**15.20**	**59.60**	**4.61E-12**
Proteobacteria;Gammaproteobacteria;Chromatiales;Halorhodospiraceae	-0.73	15.02	3.44	1.01E-01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae	-0.37	14.80	1.00	3.63E-01
Unclassified;Unclassified;Unclassified;Unclassified;Unclassified	**-2.95**	**13.79**	**12.37**	**1.81E-03**
Cyanobacteria;Cyanophyceae;Synechococcales;Leptolyngbyaceae	-0.06	13.33	0.01	9.32E-01
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae	0.53	13.24	0.45	5.19E-01
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Phyllobacteriaceae	-3.43	12.70	2.96	1.14E-01
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	**6.43**	**12.70**	**42.74**	**3.91E-09**
Proteobacteria;Betaproteobacteria;Rhodocyclales;Azonexaceae	**-1.58**	**12.69**	**7.94**	**1.30E-02**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae	1.11	12.65	3.06	1.14E-01
Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae	-1.37	12.61	2.97	1.14E-01
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae	-1.44	12.51	3.28	1.06E-01
Actinobacteria;Actinobacteria;Frankiales;Frankiaceae	-1.01	12.44	1.80	2.21E-01
Firmicutes;Clostridia;Clostridiales;Heliobacteriaceae	0.67	11.47	0.53	4.95E-01
Chlorobi;Chlorobia;Chlorobiales;Chlorobiaceae	-4.09	11.27	4.56	5.80E-02
Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae	**-3.67**	**11.26**	**6.45**	**2.38E-02**
Proteobacteria;Alphaproteobacteria;Hyphomicrobiales;Rhizobiaceae	**-3.19**	**11.07**	**5.95**	**2.96E-02**
Firmicutes;Bacilli;Bacillales;Paenibacillaceae	**-5.85**	**11.00**	**12.18**	**1.81E-03**
Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	**-4.08**	**10.55**	**8.82**	**8.81E-03**
Proteobacteria;Gammaproteobacteria;Methylococcales;Methylococcaceae	**-4.51**	**10.17**	**6.94**	**1.94E-02**
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae	**-6.01**	**9.34**	**13.08**	**1.63E-03**
Chloroflexi;Chloroflexia;Chloroflexales;Oscillochloridaceae	**-7.93**	**9.23**	**23.35**	**1.71E-05**
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae	**-6.41**	**9.12**	**10.88**	**3.23E-03**
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae	**-4.48**	**9.08**	**7.21**	**1.80E-02**
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	**6.37**	**8.99**	**12.81**	**1.64E-03**
Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae	-2.42	8.37	1.75	2.21E-01
Firmicutes;Clostridia;Clostridiales;Clostridiaceae	-4.00	7.49	3.46	1.01E-01

Supp. Table 3-5: Sequence and statistical output of ANF rate IT derived from the nifH-R community. Significance values are derived from multipatt testing (R package' indicspecies'). are ordered from most to least abundant.

No.	% Abund	F rank	P rank	Cluster	Tax Annotation	Indicator statistic	Indicator pvalue
1	4.41	8	3	cluster I	Bradyrhizobium sp.1	0.88	1.88E-02
TCGACCCGCCTTATCCTCCACGCCAAGGCCCAGGACACCATCCTGAGCCTCGCCGCCGAGCAAGGCAGCGTCGAGGACCTCGAGATCGAGGACGTGATGAAGATCGGCTACAAGGAC ATCAAGTGCGTCGAGTCTGGTGGTCCTGAGCCGGGAGTGGGCTGCGCGGGCCGCGGCGTGATCACCTCGATCAACTTCCTCGAAGAAAACGGCGCATACGAGGACATCGACTATGTG TCCTACGACGTGCTCGGCGACGTGGTGTGCGGCGGCTTCGCCATGCCCATCCGCGAGAACAAGGCGCAGGAAATCTACATCGTTATGTCC							
2	4.30E-01	137	20	cluster I	Enterobacteriaceae sp.	0.95	1.60E-03
TCGACTCGACTGATGCTCCACGAGAAGGCGCAGAACACGATCATGCACCTGGCGGCCGAGGCCGGCGGGGTGGAAGACCTCGAGCTCGAGCAGGTGATGAAGGTCGGCTATGGCGG CGTCAAGTGCGTCGAGTCGGGTGGTCCCGAGCCGGGCGTCGGTTGCGCCGGACGGGGCGTCATCACGGCCATCAACTTCCTTGAAGAAGAAGGGGCCTATGACGAGGATCTGAACTT CGTGTTCTATGACGTCCTTGGCGACGTCGTCTGCGGCGGCTTCGCCATGCCGATTCGCGAGAAGAAGGCCCAGGAGATCTACATCGTGTGCTCC							
3	2.80E-01	67	38	cluster I	Azospirillaceae sp.1	0.92	9.20E-03
TCGACCCGGCTCATCCTCCACGCCAAGGCGCAGGACACGGTCCTCAGTCTCGCGGCGGCTGCCGGCTCGGTCGAAGACCTCGAGATCGAGGACGTGATGAAGGTCGGCTTCAAGGGT ATTCGCTGCGTCGAATCGGGCGGTCCGGAGCCGGGTGTCGGTTGCGCGGGACTCGGCGTCATCACCTCCATCAACTTCCTGGAAGAGAACGGCGCGTACGATGACGTCGACTACGTC TCCTACGACGTGCTCGGTGACGTGGTGTGCGGCGGCTTCGCCATGCCCATCCGCGAGAACAAGGCCCAGGAGATTTACATCGTGATGTCG							
4	2.40E-01	284	36	cluster I	Methylobacterium sp.	0.81	5.07E-02
TCGACCCGCCTCATTCTTCACGCGAAGGCGCAGGACACCGTGCTGAGCCTCGCGGCGGCGGCCGGTACGGTCGAAGACCTCGAGCTCGATGACGTTCTCAAGGTCGGATACAAGGGC ATCAAGTGCGTCGAATCGGGCGGCCCGGAGCCCGGCGTCGGATGCGCTGGCCGCGGCGTCATCACGTCCATCAACTTCCTCGAGGAGAACGGCGCGTACGACGACGTGGACTACGT CTCGTATGACGTCCTCGGAGACGTCGTCTGCGGCGGCTTCGCGATGCCGATCCGCGAGAACAAGGCCCAGGAGATCTACATCGTCATGTCC							
5	1.10E-01	251	59	cluster I	Methylobacterium sp.1	0.97	5.00E-04
TCCACTCGCCTGATCCTCCACTCCAAGGCGCAGGACACCGTGCTCAGCCTCGCCGCCGCCGCCGGCTCGGTCGAAGACCTCGAGCTCGAAGACGTGCTCAAGGTCGGCTACAAGGGC ATCCGCTGCGTCGAATCAGGCGGTCCCGAGCCGGGCGTCGGTTGCGCAGGCCGCGGCGTCATCACGTCGATCAACTTCCTCGAGGAAAACGGCGCCTACGACGACGTCGACTACGTC TCGTACGACGTCCTCGGCGACGTGGTCTGCGGCGGCTTCGCCATGCCCATTCGCGAAAACAAGGCCCAGGAAATCTATATCGTCATGTCC							
6	4.00E-02	243	137	cluster I	Chromatiaceae sp.	0.80	1.11E-02
TCGACGCGCCTGATGCTTCACGAAAAGGCCCAGAACACCATCATGCACCTGGCCGCTGAAGCCGGCGGCGTTGAAGACCTGGAGCTCGAAGACGTGATGAAGGTCGGCTATGGCAGC ACCAAATGCGTGGAGTCCGGCGGACCTGAACCCGGCGTCGGCTGTGCGGGCCGGGGCGTCATCACCGCCATCAACTTCCTCGAAGAGGAAGGCGCTTACGACAAAGACTTGAACTTC GTGTTCTATGACGTTCTGGGTGATGTCGTGTGCGGCGGCTTCGCCATGCCGATTCGCGAAAAGAAAGCCCAAGAAATCTACATCGTCGTCTCC							
7	2.50E-02	559	148	cluster I	Sphingomonadaceae sp.	1.00	2.00E-04
TCGACCCGTCTCCTCCTCGGCGGGCTTTCACAAAAGACCGTCCTCGATACGCTCAGGAGTGAGGGTGAGGACCTCGACCTCGATGACGTAATGAAGATAGGTTTCCAGGGTACGCGGT GCGTCGAATCGGGAGGTCCCGAGCCGGGCGTTGGCTGCGCGGGCAGGGGAATCATCACCTCGATCAACCTTCTGGAGCAGCTCGGCGCTTACTCCGAGAGCATCGGCCTCGATTACG CCTTTTACGACGTGCTCGGCGACGTCGTCTGTGGTGGCTTCGCCATGCCGATCCGGGACGGCAAGGCCAAGGAGATCTATATCGTCGTTTCC							
8	1.30E-02	578	223	cluster I	Frankiaceae sp.	0.82	7.80E-03
TCGACGCGGCTGATCCTCCATTCCAAGGCCCAGACGTCGGTGCTCGCCACTGCGGCCGAGCGCGGCTCCGTGGAGGACGTCGAGCTGGAGGACGTGCTGCTGACCAGCCAGTCGGG CATCCGTTGCGTCGAATCAGGCGGCCCCGAACCCGGCGTCGGTTGCGCCGGGCGCGGCGTGATCACCGCGATCACGTATCTCGAGGAGAACGGCGCCTACGAGGATCTCGACTACG CGTGCTACGACGTCCTCGGGGACGTCGTCTGCGGCGGGTTCGCGATGCCGATCCGGCAGGGCAAGGCCCAGGAGATCTACATCGTGACCTCG							

Abbreviations: % Abund = average abundance across *nifH*-R samples; rank = ordered abundance rank; Tax annotation= taxonomic best-hit from local blastn

Supp. Table 3-5 *contd.*

No.	% Abund	F rank	P rank	Cluster	Tax Annotation	Indicator statistic	Indicator pvalue
9	1.00E-02	595	261	cluster I/IA	Nitrobacteraceae sp.1	0.75	2.78E-02

TCCACCCGTCTGATCCTGCACGCCAAGGCCCAGGCAACGGTCATGGACCTGGTGCGCGAGCGCGGCACCGTCGAGGACCTGGAGCTAGAAGATGTGCTCAAGGTCGGTTACGGCGAC
GTGCGCTGCGTGGAGTCCGGCGGACCCGAGCCGGGTGTGGGCTGCGCCGGCCGCGGCGTCATCGCCGCCATCAACTTTCCGGAGGAAAACGGTGCCTACACCCCTGACCTGGATTT
TGTGTTCTACGACGTACTGGGGGACGTTGTCTGCGGCGGGTTCGCCATGCCGATCCGCGAGAACAAGGCCGAGGAGATCTATATCGTCTGTTCC

10	8.30E-03	602	281	cluster III	Desulfovibrionaceae sp.	0.77	2.67E-02

TCCACGCGACTGCTCCTCGGGGGACTGGCGCAGAAGAGCGTGCTGGATACGCTGCGCGAAGAGGGCGAGGACGTGGATCTGGAGGACATTCGCTCCCAAGGGTTCGGCAAGACCCT
CTGCGTGGAATCGGGTGGTCCGGAGCCCGGCGTAGGCTGCGCCGGCCGTGGCATCATCACCAGCATCAACCTGCTCGAGCAGCTTGGCGCCTACGACGAGGACAATGCCCTGGACTA
CGTCTTCTACGACGTGCTGGGCGATGTCGTCTGCGGCGGGTTCGCCATGCCGATTCGCGAGGGCAAGGCCGAAGAGATCTACATTGTCTGCTCC

11	8.20E-03	396	298	cluster I	Rhizobiaceae sp.	0.71	3.47E-02

TCCACCCGCCTGATCCTGAACGCCAAGGCACAGGACACGGTGCTGAGCCTCGCCGCGGAGGCTGGTTCGGTTGAGGATCTGGAGATCGAGGATGTGCTGAAGGTTGGCTACAAGAAC
ATCAAATGCGTTGAGTCCGGCGGACCGGAACCGGGCGTGGGCTGCGCGGGTCGGGGCGTCATCACGTCCATCAATTTCCTGGAAGAGAACGGCGCGTATGACGATGTCGATTACGTC
TCCTATGACGTTCTTGGCGACGTCGTCTGCGGCGGGTTCGCGATGCCGATCCGTGAGAACAAGGCGCAGGAAATCTATATTGTGATGTCC

12	3.50E-03	647	384	cluster I	Azospirillaceaee sp.	0.82	6.20E-03

TCCACCCGTCTGATCCTGAACGCGAAAGCCCAGGACACGGTGTTGAGCTTGGCGGCGGCCGCCGGCTCGGTCGAGGATCTCGAGCTCGATATGGTGCTCAAGACCGGCTACAAGGGG
ATCAAGTGCGTCGAGTCGGGAGGCCCGGAGCCCGGCGTGGGCTGCGCCGGCCGCGGCGTCATCACTTCGATCAACTTCCTGGAAGAGAACGGCGCCTATGAGGACATCGACTACGTG
TCCTACGACGTGCTCGGCGACGTCGTCTGCGGCGGTTTTGCGATGCCGATCCGCGAGAACAAGGCGCAGGAGATCTACATCGTCATGTCC

13	3.00E-03	656	398	cluster I/IA	Bradyrhizobium sp.	0.71	3.64E-02

TCCACCCGTCTGATTCTTCATGCCAAGGCCCAGGCAACGGTCATGGATATGGTGCGCGAGCTGGGCACGGTCGAGGACCTGGAGCTGACTGATGTGCTCAAGGTAGGCTACGGCGAT
GTGAAATGCGTAGAGTCGGGTGGGCCTGAGCCGGGCGTCGGCTGTGCCGGCCGCGGCGTCATCGCCGCCATCAACTTCCTGAAGGAGAACGGCGCCTACACCCCCGATCTGGATTTC
GTCTTCTATGACGTTCTCGGCGACGTTGTCTGCGGCGGGTTCGCCATGCCGATCCGCGAGGACAAGGCCGATGAGATCTACATCGTCTGCTCC

14	1.40E-03	704	496	cluster III	Desulfovibrionaceae sp.1	0.71	3.64E-02

TCGACGCGGCTGCTGCTGGGCGGCCTGGCCCAGAGGACAGTGCTGGACACGCTGCGGGAGGAAGGCGAGGAGGTGGATCTTTGCGACATCCGCCGCCAGGGCTTCGGCAGCACGC
TTTGCACCGAGAGCGGCGGCCCCGAGCCTGGCGTCGGCTGCGCCGGACGCGGCATCATTACCTCCATCAACCTGCTGGAGCAGTTGGGGGCTTATGACGTGAGCGAGAATATCGATT
ACGTCTTCTACGACGTGCTGGGCGACGTGGTCTGCGGCGGGTTTGCCATGCCCATCCGCGACGGTAAGGCGCAGGAGATATACATCGTCTGTAGC

15	1.10E-03	722	528	cluster I	Ectothiorhodospiraceae sp.	0.71	3.27E-02

TCGACGCGGCTGATGCTCCACGAGAAGGCGCAGAACACGATCATGCACCTGGCGGCCGAGGCCGGCGGCGTCGAGGACCTCGAACTCGATCAAGTGCTCAAGGTCGGCTACGGCGC
GGTTAAGTGCGTCGAGTCCGGCGGCCCCGAGCCGGGCGTCGGCTGCGCCGGCCGCGGTGTCATCACGGCGATCAACTTCCTCGAGGAAGAAGGCGCCTACGAAGAGGACCTGAACT
ACGTCTTCTACGACGTGCTGGGTGACGTCGTCTGCGGCGGGTTCGCCATGCCGATCCGCGAGAAGAAGGCCCAGGAGATCTACATCGTGGTCTCC

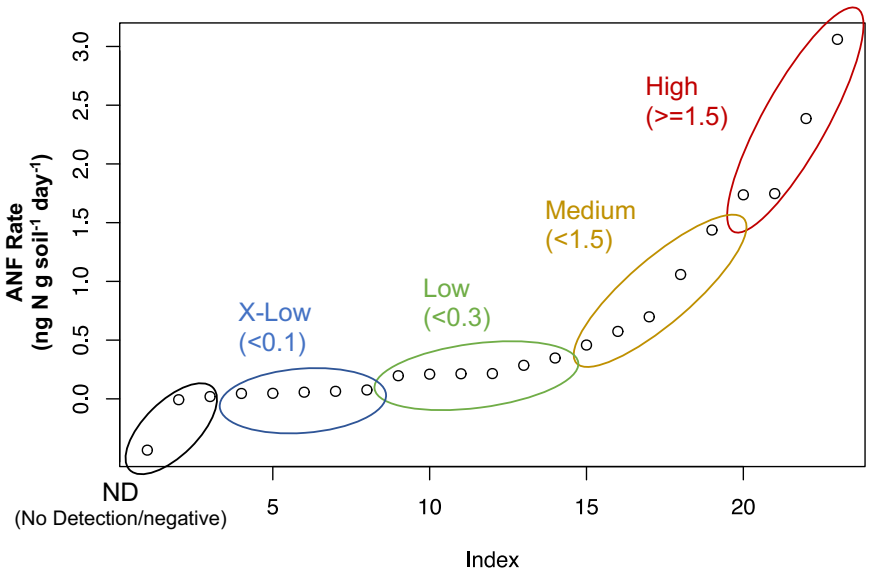
16	5.10E-03	782	618	cluster I	Azospirillum sp.	0.85	3.80E-03

TCGACCCGCCTGATCCTGAACGCGAAAGCTCAGGACACGGTGCTGAGCCTCGCCGCGGCCTCGGGCTCGGTCGAGGATCTGGAGCTCGAGATGGTCCTCAAGACCGGTTACAAGGGG
ATCAAGTGCGTCGAGTCGGGCGGCCCGGAGCCGGGCGTCGGTTGCGCCGGCCGCGGCGTCATCACCTCGATCAACTTCCTCGAAGAGAACGGCGCGTACGAGGACATCGATTACGT
GTCCTATGACGTGCTGGGCGACGTGGTGTGCGGCGGCTTCGCGATGCCGATCCGTGAGAACAAGGCCCAGGAAATTTACATCGTCATGTCC

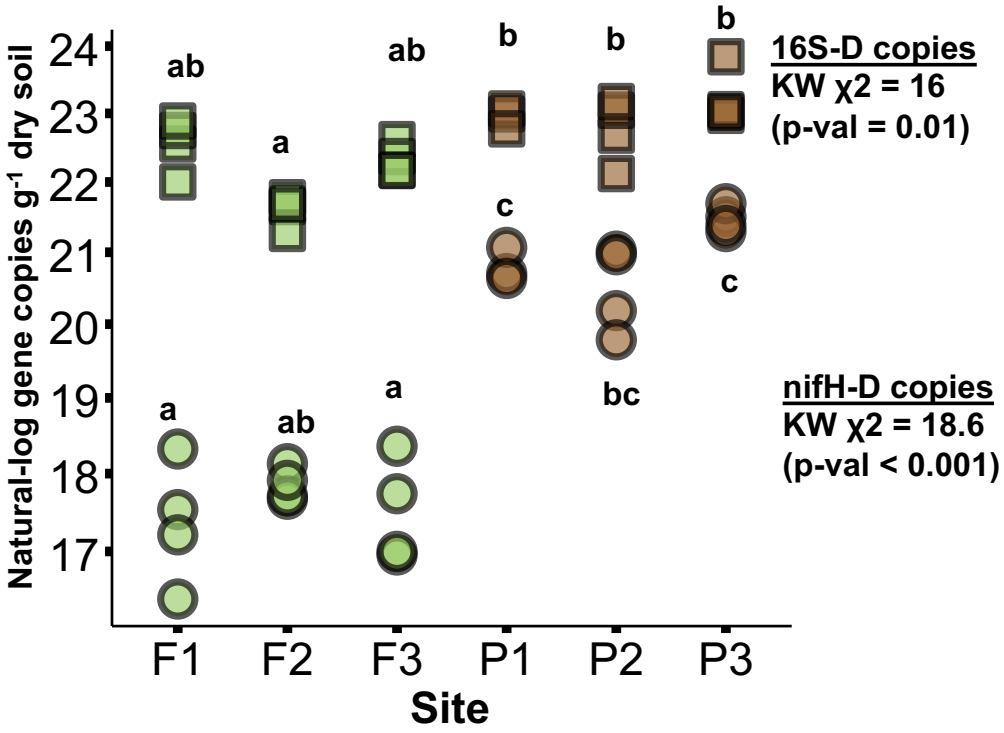
17	2.90E-04	834	684	cluster I	Nitrobacteraceae sp.	0.71	3.47E-02

TCGACCCGCCTGATCCTGCACGCTAAGGCACAGGATACCATCCTGAGCCTGGCAGCGCAGAAAGGCTCCGTTGAGGACCTGGAAATTGAAGAGGTCATGAAGCACGGCTACAAAACAA
TCCGCTGCGTTGAGTCGGGCGGCCCGGAGCCGGGCGTCGGGTGCGCTGGACGCGGCGTCATCACCTCAATCAACTTCCTGGAAGAAAACGGCGCTTACGAAGATATCGACTATGTCT
CATACGACGTGCTCGGCGATGTAGTATGCGGCGGTTTTGCTATGCCGATCCGCGAAAACAAGGCGCAGGAGATCTACATTGTCATGTCG

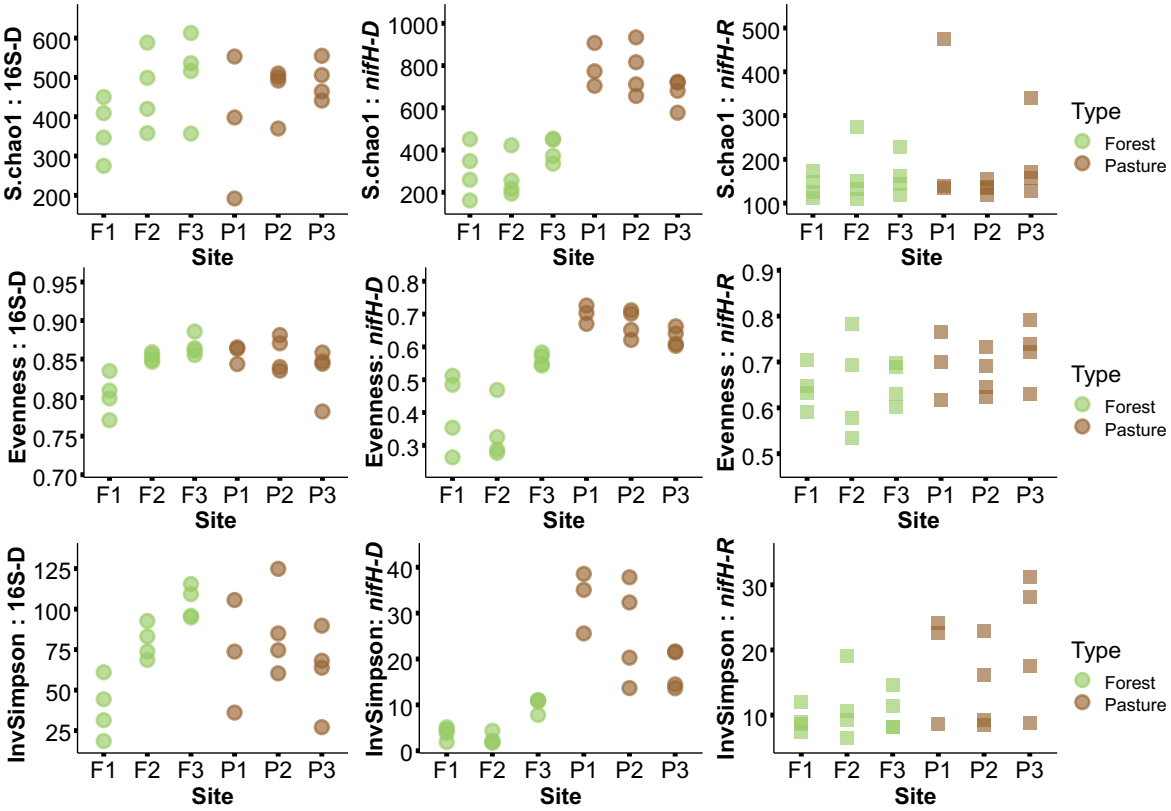
Supp. Figure 3-1: Scheme of rate binning for *nifH*-R-rate indicator species analysis



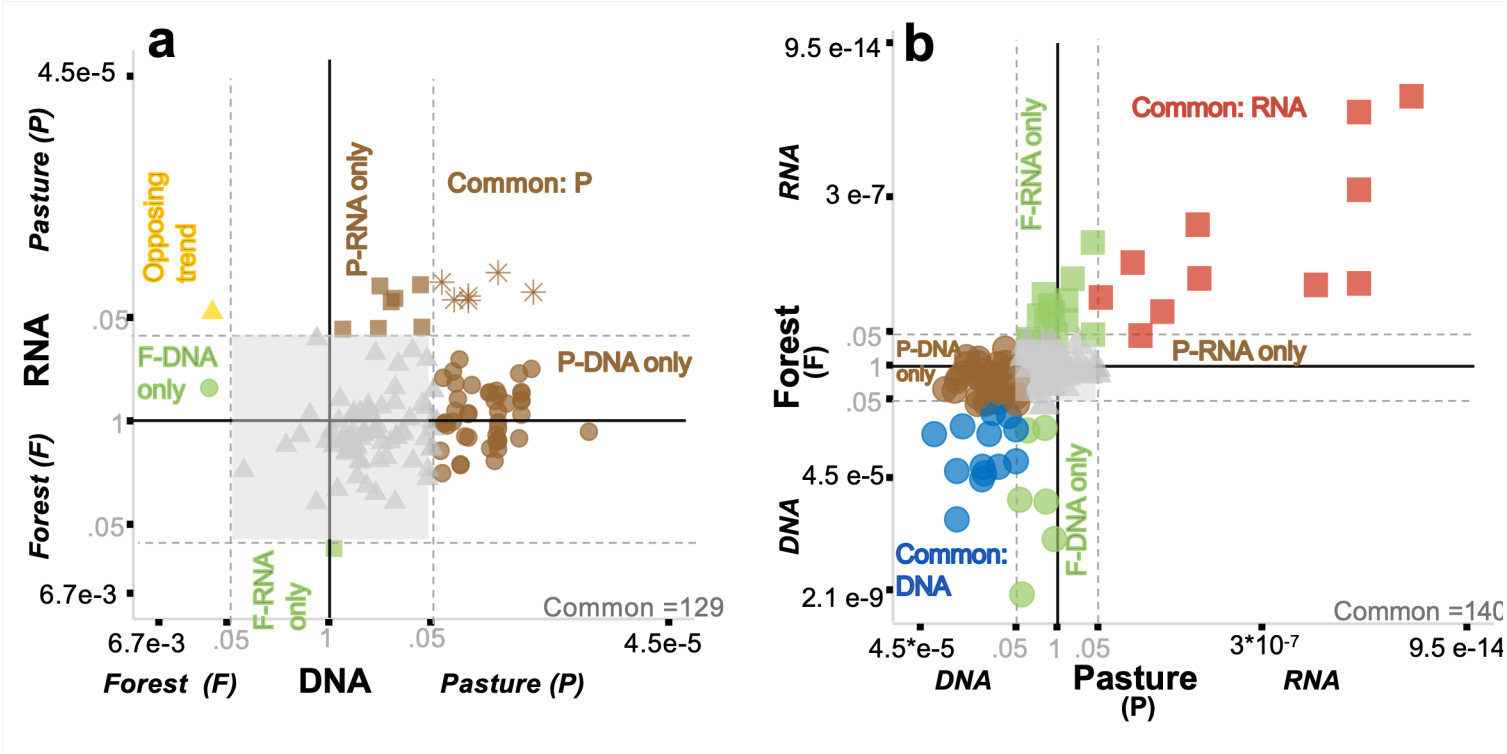
Supp. Figure 3-2: Absolute gene quantification via qPCR for 16S-D and *nifH*-D communities (natural log scale). Statistics to the right represent non-parametric KW tests on differences among sites, with a Dunn's post-hoc to test for grouping by LU.



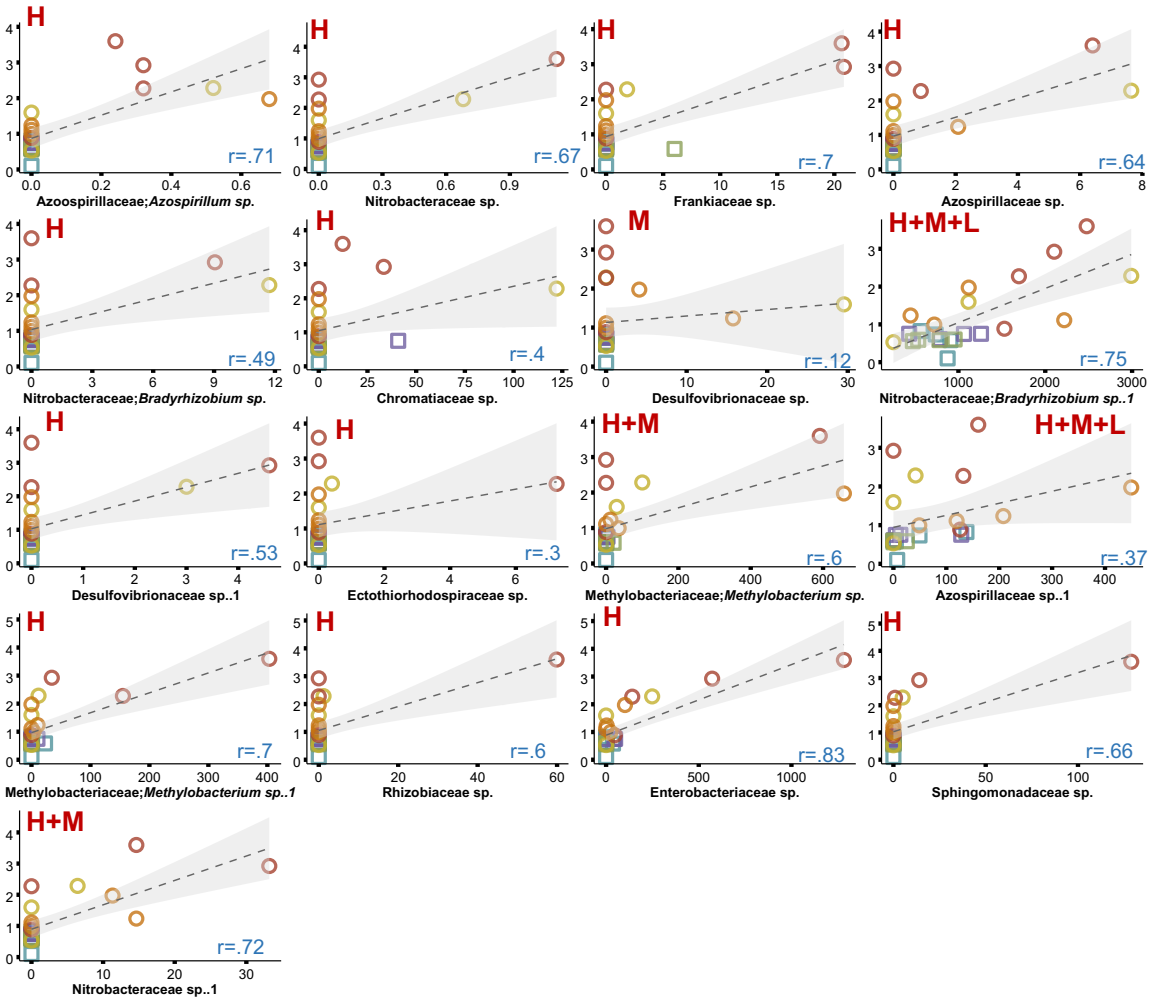
Supp. Figure 3-3: Alternate diversity metrics including Chao1, Evenness, and Inverse Simpson for the three community profiles surveyed.



Supp. Figure 3-4: Comparing OTU enrichment trends across **(a)** LU effects (DNA compared to RNA-based communities) and **(b)** CT effects (Forest compared to Pasture paired DNA and RNA communities). Non-significant OTUs are shaded in gray (p-value > 0.05).



Supp. Figure 3-5: ANF rate bin IT identified from *nifH*-R communities correlated individually against rate. Red letters in upper right represent the bin IT are associated with (H= high, M = medium, L=low, see Supp. Figure 3-1), and blue values in the lower left represent the Pearson correlation estimate.



Supp. Figure 3-6: Individual correlations of *nifH*-R-derived IT relative abundance within *nifH*-D communities, against ANF rate. Blue values in the lower left represent the Pearson correlation estimate. The larger bottom figure shows aggregated relative abundance of *nifH*-R IT within the *nifH*-D community against ANF rate.