MOLECULAR INVESTIGATION OF COMPOSITION AND FUNCTION OF THE SOIL MICROBIOME IN RESTORED FEDERAL WETLANDS FROM LANDSCAPE TO LAB

by

Nicholas Reed Alexander

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Molecular Biosciences

Middle Tennessee State University

May 2025

Dissertation Committee:

Dr. Donald Walker, Research Advisor

Dr. Frank Bailey, Chair

Dr. Ryan Otter

Dr. Allison Veach

Dr. Jeffrey Walck

ABSTRACT

Microorganisms have contributed directly to the development of contemporary conditions that sustain life on Planet Earth, through the modification of oceanic redox states, control of organic carbon flux in aquatic and terrestrial environments, and the cycling of key nutrients such as Nitrogen (N). Anthropogenic disturbances to terrestrial ecosystems have both direct and indirect impacts on the balance of biogeochemical cycling through changes in land-use, land management practices, and land degradation. Historic trends in global fertilizer use have indicated steady increases in both demand and application that has led to the pollution of waterways in a process known as eutrophication. Conservation research has shown that the restoration of global wetlands can serve to reduce the impacts of eutrophication by sequestering runoff nutrients such as N and providing unique conditions for biogeochemical cycling. Building predictive models of biogeochemical cycling in restored wetlands is a major focus for ecosystem ecologists and microbial ecologists alike and will serve to inform conservation land-use and best management practices. The microorganisms that perform biogeochemical cycling in soils offer direct and unique insight into nutrient process rates and the potential to mitigate eutrophication on a global-scale. The overall objective of my work is to understand how the soil microbiome influences N cycling across the landscape and in the lab. Previous work has suggested that microorganisms may be used as bioindicators of biogeochemical processes rates, therefore, I am interested in understanding how taxonomic composition and functional gene presence can be used to predict N cycling in restored wetlands. To address these objectives, I sampled soil bacterial assemblages in restored federal wetlands in Kentucky and Tennessee and utilized DNA sequencing to characterize the soil microbiome compositionally and functionally across the landscape, and the soil microbiome response to NH_4-N in the lab. Results indicate that both functional gene copy numbers

and microbiome composition can be used to predict N-flux rates on the landscape using machine learning and deep learning techniques. Additionally, I tested for an effect of NH_4-N concentration on microbiome community assembly and found that carbon source, rather than N concentration, had the greatest impact on stabilized communities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: LEVERAGING FINE-SCALE VARIATION AND HETEROGENEITY OF THE WETLAND SOIL MICROBIOME TO PREDICT NUTRIENT FLUX ON THE LANDSCAPE

N. Reed Alexander[1], Robert S. Brown[2], Shrijana Duwadi[2], Spencer G. Womble[2], David W. Ludwig[3], Kylie C. Moe[1], Justin N. Murdock[2], Joshua L. Phillips[3], Allison M. Veach[4], Donald M. Walker[1]

[1] Middle Tennessee State University, Department of Biology, Murfreesboro, TN 37132, USA

[2] Tennessee Tech University, Department of Biology, Cookeville, TN 38505, USA

[3] Middle Tennessee State University, Department of Computer Science, Murfreesboro, TN 37132, USA

[4] The University of Texas at San Antonio, Department of Biology, Health, and the Environment, San Antonio, Texas, USA

Alexander *et al.* Leveraging Fine-Scale Variation and Heterogeneity of the Wetland Soil Microbiome to Predict Nutrient Flux on the Landscape. *Microbial Ecology*, 2025;**88**:22.

DOI: 10.1007/s00248-025-02516-1 ;

Abstract

Shifts in agricultural land use over the past two hundred years have led to a loss of nearly 50% of existing wetlands in the United States, and agricultural activities contribute up to 65% of the nutrients that reach the Mississippi River Basin, directly contributing to biological disasters such as the hypoxic Gulf of Mexico “Dead” Zone. Federal efforts to construct and restore wetland habitats have been employed to mitigate the detrimental effects of eutrophication, with an

emphasis on restoration of ecosystem services such as nutrient cycling and retention. Soil microbial assemblages drive biogeochemical cycles and offer a unique and sensitive framework for the accurate evaluation, restoration, and management of ecosystem services. The purpose of this study was to elucidate patterns of soil bacteria within and among wetlands by developing diversity profiles from high-throughput sequencing data, link functional gene copy number of nitrogen cycling genes to measured nutrient flux rates collected from flow-through incubation cores, and to predict nutrient flux using microbial assemblage composition. Soil microbial assemblages showed fine-scale turnover in soil cores collected across the topsoil horizon (0 - 5cm; top vs bottom partitions) and were structured by restoration practices on the easements (tree planting, shallow water, remnant forest). Connections between soil assemblage composition, functional gene copy number, and nutrient flux rates show the potential for soil bacterial assemblages to be used as bioindicators for nutrient cycling on the landscape. In addition, predictive accuracy of flux rates was improved when implementing deep learning models that paired connected samples across time.

Introduction

Eutrophication of global watersheds is a growing threat to both fresh and saline environments, affecting the conservation of threatened species, water quality, and food availability (Glibert 2020). The primary cause of this degradation is nutrient over-enrichment, attributed to agriculture and the excess discharge of nitrogen (N) and phosphorus (P) (Glibert *et al.* 2014). The excess nutrient pollution is the prime cause of algal blooms, which deplete oxygen in water bodies – a condition known as hypoxia – ultimately leading to widespread ecosystem degradation (Nazari-Sharabian *et al.* 2018). Recent federal efforts in the United States directed at the sequestration of runoff nutrients have been implemented through the restoration of wetlands from previously

degraded agricultural land. One notable effort is the Wetland Reserve Program (WRP), introduced under the 1990 Farm Bill, which provided a framework for land owners to collaborate with non-profits and researchers to abate dissolved nutrient loads from agricultural runoff (Comín *et al.* 1997).

Numerous ecosystem services have been restored as part of this program including, nutrient cycling and retention, and the reestablishment of wildlife habitat (Mitsch and Gosselink 2000; Zedler and Kercher 2005; Mitsch *et al.* 2013). Restored wetlands have shown patterns of biogeochemical cycling including increased rates of carbon (C) sequestration and reduction in greenhouse gas emission (Hemes *et al.* 2019), increased P sorption through modification of site hydrology over the course of restoration (Xu *et al.* 2022), and localized shifts in functional N-cycling gene abundances at the watershed-scale (Kasak *et al.* 2021). As restoration efforts continue, there is a growing need to develop predictive models of landscape-level nutrient cycling. These models can help evaluate restoration success, guide adaptive management strategies, and inform future restoration efforts (Correa-Garcia *et al.* 2023).

The soil microbiome – a collection of bacteria, fungi, protists, archaea, and algae in an environment – plays a multifunctional role in wetland ecosystems. It directly contributes to terrestrial fertility (Chaparro *et al.* 2012), soil formation (Sokol *et al.* 2022), and collectively drives the biogeochemical cycling of N, P, and organic C (Falkowski *et al.* 2008; Lappalainen *et al.* 2016). The soil microbiome offers a unique framework to explore the predictability of nutrient cycling at the landscape scale. Microbial assemblage variation has been linked to differences in climatic, edaphic, and biotic parameters, and the resulting multivariate niche space generated from the interactions between these influences (O'Brien *et al.* 2016; Correa-Garcia *et al.* 2023). Shifts in soil microbial assemblage structure and their associated functions have been shown to respond

to several factors including hydrologic connectivity among sites (Tomasek *et al.* 2017), wetting and drying regimes (Veach and Zeglin 2020), land-use practices (Rasche and Cadisch 2013; Tomasek *et al.* 2017; Dai *et al.* 2020), soil-depth (Tang *et al.* 2018), and historic landscape use such as fertilizer amendments (Creamer *et al.* 2016; Dai *et al.* 2020). Ecological restoration practice has been shown to respond to and positively impact soil biodiversity, and measured metabolic function correlates with restoration age (Cookson *et al.* 2007; Fernández-Martínez *et al.* 2017; Dai *et al.* 2020; Wang *et al.* 2021), suggesting restoration practices can improve biogeochemical cycling rates via the microbiome. Microbes respond to abiotic variation present within their environment in both spatially and temporally dynamic ways, affecting microbial activity and function. These interactions have a direct influence on the physical and chemical properties of soils to shape wetland ecosystem services (Urakawa and Bernhard 2017).

Ecological studies have historically focused on determining patterns of community composition and their distribution across the landscape (Schneider 2001). The observed patterns in these studies are defined and constrained by the *scale* at which the study is conducted (Wiens 1989; Levin 1992; Schneider 2001). *Ecological scale* is defined by three axes: *spatial*, *temporal*, and *phylogenetic* (Graham *et al.* 2018; Ladau and Eloe-Fadrosh 2019), each of which are composed of two elements, *grain* and *extent* (O'Neill *et al.* 1989; Wiens 1989). *Spatial scale* is determined by the distance, area, or volume being studied, *temporal scale* refers to the time period encompassing sampling effort, while *phylogenetic scale* is defined by the taxonomic, functional, or trophic group(s) being investigated (Vellend 2010; Ladau and Eloe-Fadrosh 2019). *Grain* refers to the breadth of the smallest sampling unit, while *extent* refers to the scope of all samples collected in the study (O'Neill *et al.* 1989; Ladau and Eloe-Fadrosh 2019). To detect non-spurious patterns the *scale* of a study must align with the *domain* of interest. Studies conducted at fine-scales may

be useful for elucidation of mechanisms, while broad-scale studies may reveal generalizations about the *domain* in question. Patterns that are consistent across multi-scalar studies are more likely to be reflective of true ecosystem-wide processes-pattern relationships.

Soils along with the microorganisms that inhabit them, exhibit vertical stratification, due to both formation processes and contemporary soil conditions (Hartemink *et al.* 2020; Ivanova *et al.* 2020; Naumova *et al.* 2021). Trends in microbiome composition and function across soil depth have been well documented (Naylor *et al.* 2022). Past studies have shown that soil microbiomes exhibit significant millimeter-scale differences by depth, in addition to quantified abiotic differences at the same scale (Cai *et al.* 2022). Given that abiotic factors, taxonomic composition and abundance, and functional nutrient cycling genes have been shown to vary as a result of soil depth (Tang *et al.* 2018; Naylor *et al.* 2022; Qian *et al.* 2023), soil strata must be considered to build predictive models of nutrient cycling using microbiome composition (Young *et al.* 2019).

Soils represent a multi-scalar continuum of ecological gradients and conditions rather than a single homogeneous environment (Fierer 2017). The composition and diversity of bacterial assemblages within soils are shaped by a variety of biotic and abiotic factors including soil pH, redox potential, nutrient availability, temperature, soil type and structure, and the presence or absence of other members of the soil microbiome consortia. These factors can create soil conditions that may vary on the millimeter-scale (Fierer 2017; Furtak and Gałązka 2019). Bacterial assemblages have been shown to respond to environmental changes with shifts in both α- and β-diversity because of both deterministic and stochastic processes such as environmental filtering, diversification, dispersal limitation, and drift (Vellend 2010). This sensitivity to environmental factors makes bacterial assemblages valuable bioindicators of soil function.

The goal of this study was to elucidate fine-scale patterns of bacterial assemblage composition and function within the surface soil microbiome of federally restored WRP wetlands. Specifically my objectives were to: 1) use 16S rRNA metabarcoding and amplicon sequencing to characterize microbial assemblage diversity across restoration practices and strata within the topsoil horizon (0 - 5cm); 2) use quantitative PCR (qPCR) to determine whether measured nutrient turnover is predictive of functional nitrogen cycling gene abundance, 3) to develop predictive models of nutrient turnover at the landscape scale using bacterial microbiome composition. I hypothesized that: 1) soil bacterial assemblages within restored ecosystems vary by soil horizon and restoration practice, 2) nutrient turnover is positively correlated to qPCR log copy number, and 3) machine learning can link bacterial assemblage composition to measured nutrient turnover.

Methods

Field sampling and experimental design

This study was carried out on WRP (n = 8) easements across the Mississippi Valley Loess Plain in western Kentucky (n = 4) and western Tennessee (n = 4), USA, during the summer seasons (June, July, August) of 2019 and 2020 (Fig. 1). Wetland easements were selected based on restoration management practices conducted at the sites, such that each easement contained all three focal habitat types: constructed shallow water areas (excavated ponds and flashboard risers), tree plantings, and remnant forest. Shallow water habitats were characterized by standing water and included a mix of emergent and floating herbaceous vegetation. Tree plantings on easements were established within one to four years of entering the restoration program and contained a mix of bottomland hardwood saplings that included a variety of oaks (*Quercus* sp.) and occasionally bald cypress (*Taxodium distichum*). Remnant forest habitats were mature bottomland hardwood species that formed a complete canopy and an open understory, likely due to frequent flooding.

Sampling transects within habitat restoration treatments were taken to ensure vegetation surrounding the sampling location was homogeneous among soil cores across a particular habitat.

Soil samples for microbial analysis were collected using homemade pre-sterilized soil corers (herein called sample rods) consisting of a disposable spatula (VWR International, Radnor, PA; 310 mm in length) inserted into an autoclaved (2 hr) aluminum sleeve (Fig. 2a). At each sampling location, a sample rod was inserted approximately 5 cm into the ground (Fig. 2b). Soils collected within the spatula (1 cm diameter, ~ 5 cm of sediment) were then separated into three even "partitions" (~ 1.6 cm each; Fig. 2e, f), partitioned according to the soil depth profile (top/surface, middle, and bottom). Soil samples were stored independently in individual, pre-sterilized 15 ml centrifuge tubes (VWR International, Radnor, PA) and placed on ice within 30 minutes of collection and transported to Middle Tennessee State University (MTSU; Murfreesboro, TN) within six hours of collection in the field where they were frozen to -80 C until further processing.

Thirty pairs of larger soil cores (n = 10/habitat; 7.62 cm diameter x 15 cm depth; Fig. 2c, d, g) were collected from each easement, within 15 cm of microbial soil cores. The large soil cores were transported on ice to Tennessee Tech University's Water Center Lab (Cookeville, TN) post-collection. One large soil core was used for flow-through incubation trials (Appendix A Supplemental Methods), while the second was used to measure total C, total N, and extractable phosphorus (Appendix A Supplemental Methods).

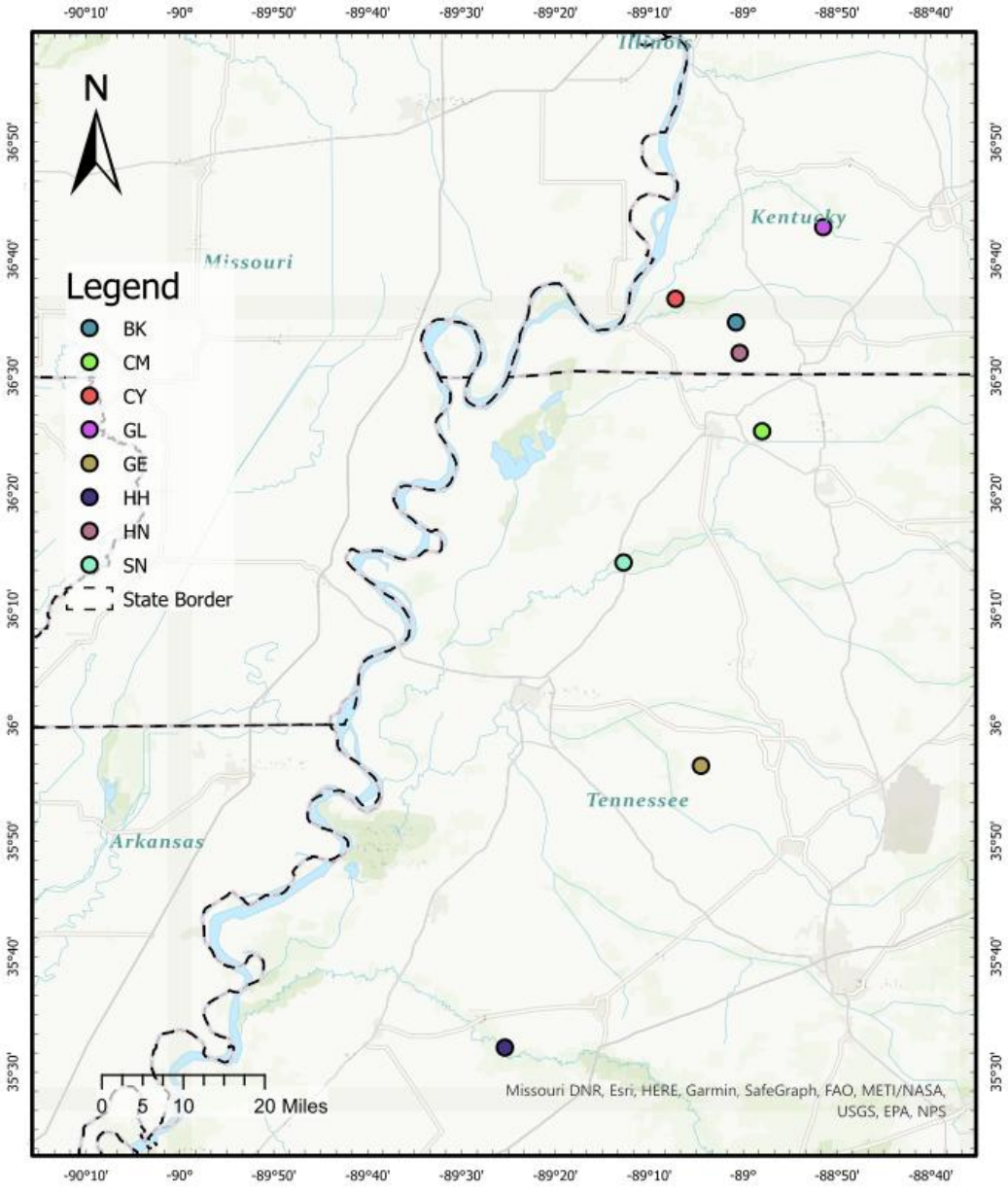


Figure 1. Federally restored wetland monitoring sites in Kentucky and Tennessee, USA. Point locations of field sampling sites are depicted with colored circles. Soil samples were collected from BK, GE, GL, and SN in 2019, and samples from CM, HH, HN, and SN were collected in 2020

Figure 2. Field sampling methodology. Microbial soil sampling rods consist of an aluminum rod exterior with an VWR spatula inside (a). Sterilized sampling rods are used to collect soil samples in the field (b). Soil sampling rods are compared to incubation core collection devices (c, d). Soil sampling rods collect a consistent 5cm soil sample (e) which is aseptically partitioned in the field (f) into top, middle and bottom partitions (g, right)

Flow-through incubation experiment

A parallel flow-through incubation design (Fig. 3, 4a, b) was selected to measure dissolved nutrient retention rates of NH_4, NO_2, NO_3, and PO_4, and dissolved gas retention rates of N_2 and O_2 from intact soil cores (Miller-Way and Twilley 1996; Groffman *et al.* 2006; Nifong *et al.* 2022). Synthetic water (Appendix Table C1) (Brown 2023) was developed and maintained for all experimental incubations (Speir *et al.* 2017). This experiment was designed to measure maximal NO_3 and PO_4 flux rates within the inundated experimental cores (Appendix A Supplemental Methods).

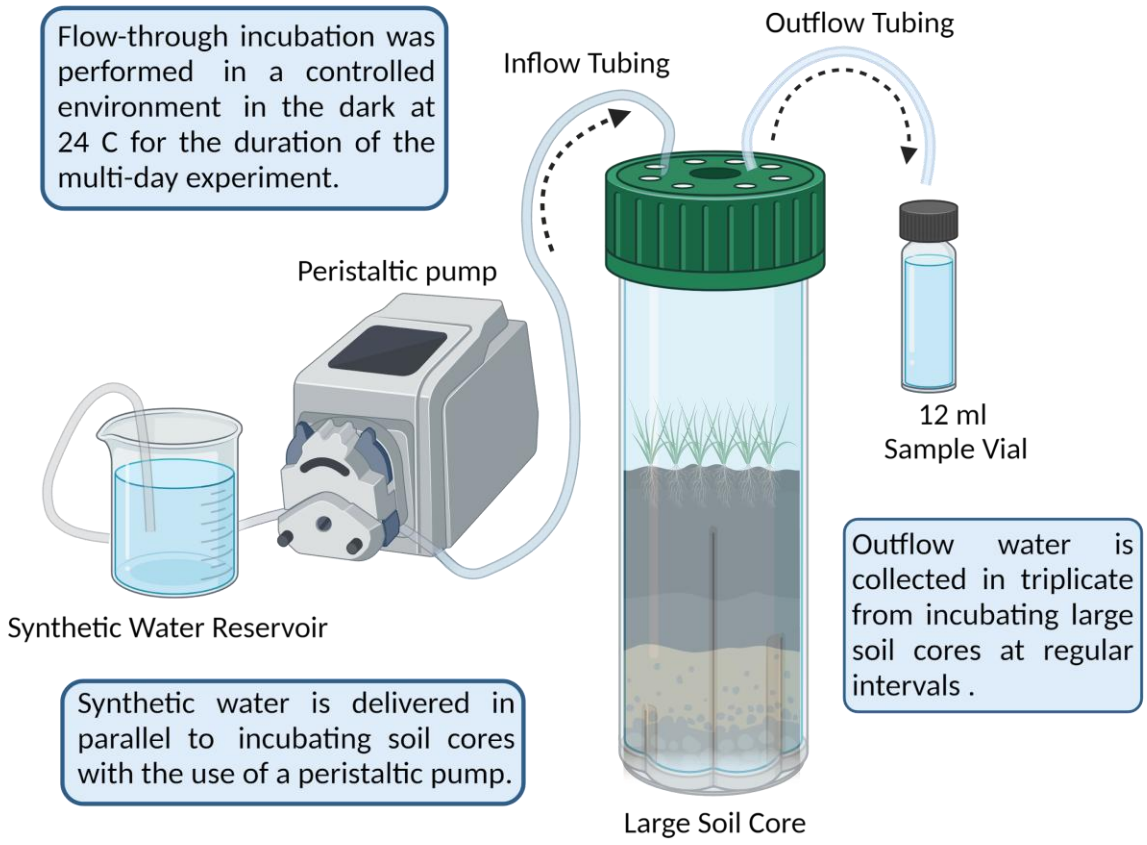


Figure 3. Conceptual Diagram of Flow-Through Incubation Experiments. Synthetic water is pumped into large soil cores and outflow is collected from tubing into a 12ml sample vial. Created in https://BioRender.com

Nutrient flux rates were calculated for each dissolved nutrient or gas using Equation 1 (Appendix A). Positive values represent release from the soil to the water column, while negative values represent uptake from the water column by the soil. Time-weighted average flux rates (Appendix A Equation 2) were calculated for each core that represented the average flux rate of a given nutrient over the course of the experiment.

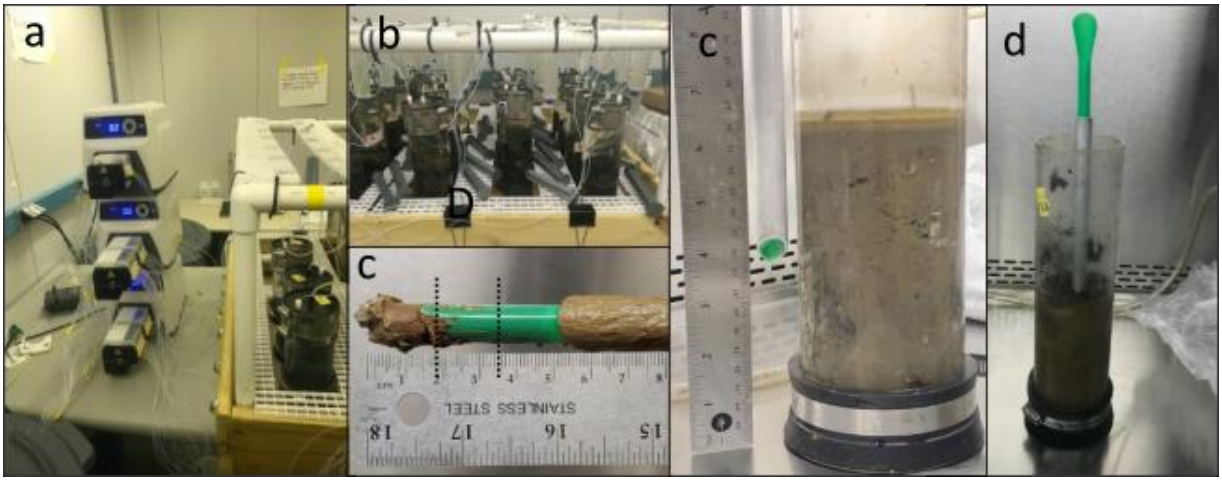


Figure 4 Flow-through lab incubation experiments. Peristaltic pumps (a) are used to deliver nutrient rich water in parallel to incubation cores (b). After core incubations (c), incubation cores are moved to adjacent BSL2 hoods where they are sampled using microbial soil sampling rods (d) and partitioned (e) identically to field collected soils

Post-incubation soil sampling

Post-incubation, surface water was siphoned off and large soil cores were transported to a BSL2 hood (Fig. 4c, d). Microbial soil cores were then collected from the large soil cores and partitioned using an identical sampling methodology described above for field collection using sampling rods (Fig. 4d, e), with the exception that ~ 1.25 g of soil from top and bottom partitions was aseptically transferred to 2 ml microcentrifuge tubes filled with 900 µL of sterile millipore

water (2-hr autoclaved) then placed on ice. All samples were kept on ice while being transported to MTSU where they were stored at -80 C until further processing.

Soil DNA extraction

Soil samples collected in the field (n = 366), post-lab incubation (n = 366), and control swabs (n = 28) were removed from -80 C storage and thawed on ice for two hours. Approximately 1.25 g of soil was added into a 2 mL microcentrifuge tube pre-filled with 900 μL of sterile millipore water (2-hr autoclaved) for field collected samples. Soil samples and control swabs were then homogenized using a FastPrep FP120 (Thermo LabSystems, Waltham, MA). Following sample homogenization, 200 μL of soil slurry (~ 0.25 g) or 200 μl of homogenized control swab supernatant was transferred to a 96-well plate and DNA extracted using the Qiagen DNeasy PowerSoil HTP 96 kit. A single well was left blank on each 96-well plate and treated equally throughout the extraction process to control for contamination in downstream bioinformatics analyses (n = 8).

Library preparation and high-throughput sequencing

Libraries were prepared according to the Illumina 16S Metagenomic Sequencing Library Preparation protocol for the Illumina MiSeq instrument (n = 8 independent runs). DNA was PCR amplified in 25 μL reactions by adding 2 uL DNA, 12.5 uL MCLAB I-5™ 2X Hi-Fi PCR Master Mix, 1 μL 515F primer (10 μM), 1 μL 806R primer (10 μM) (Caporaso *et al.* 2010), and 8.5 μL PCR water. The V4 region of the 16S rRNA gene was dual indexed according to Kozich *et al.* (2013). Samples were purified using HighPrep PCR Clean Up magnetic beads (MagBio Genomics Inc., Gaithersburg, MD) after initial amplicon PCR and indexing PCR steps. The concentration of PCR products was determined using a Quantus Fluorometer (Promega Corporation, Madison, WI). Samples were diluted in 1% Tris buffer solution and normalized to 4.0 nanomolar concentration

prior to sample pooling, spiking of 10% PhiX control, and loading on an Illumina MiSeq v2 flow cell and sequenced using a 500-cycle reagent kit (2 × 250 bp paired-end reads).

Bioinformatics

Raw sequencing data was demultiplexed, aligned, and processed using the MiSeq SOP bioinformatics pipeline (Kozich *et al.* 2013) using mothur v1.45.3 (Schloss *et al.* 2009; Schloss 2020) to quality control, process, and cluster amplicon sequence data into operational taxonomic units (OTUs) at 97 % sequence similarity (Grajal-Puche *et al.* 2020). The SILVA v132 reference database was used to assign taxonomy to OTUs (Quast *et al.* 2012). Raw sequence reads were deposited in the National Center for Biotechnology Information: BioProject PRJNA1129551. All mothur code is included in Appendix B.

Contaminating sequence reads were identified with algorithmic detection and removed (in R package '*decontam*') using a probability threshold of 0.45 (Davis *et al.* 2018). Post-decontamination, rare OTUs were removed if they did not meet a minimum abundance threshold of 10 reads across all samples (Brown *et al.* 2015; Cao *et al.* 2021). Samples were arranged by sequencing depth, and rarefied (Sequencing Depth: 15,268 sequence reads / sample) to normalize coverage across samples and produce the final dataset used for downstream statistical analysis. The final OTU abundance table contained 51,402 OTUs representing microbial assemblages from 703 samples. All R code is included in Supplemental File B at https://doi.org/10.1007/s00248-025-02516-1.

Quantitative PCR

Quantitative PCR (qPCR) was used to determine the abundance of functional genes responsible for the final steps of nitrification (nitritation: $NO_2^- \rightarrow NO_3^-$) and denitrification ($N_2O \rightarrow N_2$) using degeneratative primers developed by Keeley *et al.* (2020) targeting the bacterial

periplasmic nitrite oxidoreductase (nxrB; Peri nxrB RK) and two markers for nitrous oxide reductase (nosZ I, nosZ-AI RK; nosZ II, nosZ-BII RK), independently targeting both Clade I and Clade II nitrous oxide reducers (Hallin *et al.* 2018), respectively. Reactions for all markers were run in triplicate on an AriaMx Real-Time PCR System (Agilent, Santa Clara, CA) following protocols described in Keeley *et al.* (Keeley *et al.* 2020) (Appendix A Supplemental Methods).

Statistical Analysis

Measures of assemblage diversity

Bacterial taxonomic diversity (α-diversity) and composition (β-diversity) was compared for soil sample partition (Top, 0.00 - 1.67 cm; Bottom, 3.33 - 5.00 cm) collected in the field and after incubation treatment. Metadata that included habitat restoration practices (Remnant Forest, Shallow Water, and Tree Planting) was recorded at each easement. Statistical analysis was performed in R version 4.4.0 (R Core Team, 2022). Taxonomic diversity was assessed by Hill diversity measures (D) of two orders (l = 1, Richness; l = 0, Hill-Shannon; in R package '*hillR*'), using the generalized mean as defined by Equation 3 (Appendix A) (Hill 1973; Bullen 2003; Jost 2006; Roswell *et al.* 2021).

Modeling assemblage diversity

To evaluate the relationship between soil bacterial assemblage alpha (α) diversity and abiotic soil parameters, generalized additive mixed effects models (GAMMs, with R package '*mgcv*') were utilized to detect nonlinear and non-parametric trends, while accounting for spatial autocorrelation and random effects (Wood 2003, 2011, 2017; Wood *et al.* 2016). I tested for the effect of soil N concentration on assemblage Richness and Hill-Shannon diversity. The relationship between soil N and measures of α-diversity was evaluated using a non-parametric smoothing term and cubic regression basis function. For all models, soil partition by treatment

group (field collected or lab incubated), and habitat restoration practice were included as factor smooths of soil N, and an additional smoothing factor was added for the three way interaction of both factors, and soil N, using cubic regression basis functions. Random effect terms included point locations, state (KY, TN) of collection, sampling year, easement site, and core ID. Distribution families and link functions for candidate models were assessed using AIC selection criteria, where a model was considered better-fitting if its AIC was at least two integer values below that of an alternative model (Pedersen *et al.* 2019), resulting in selection of a location-scaled t distribution with an identity link function. Basis dimensions of each smoothing term were evaluated and the final model was constructed using restricted maximum likelihood and final fit was then validated by examining the plot of residuals (Simpson 2018).

Investigating taxonomic composition

Bacterial assemblage composition across taxonomic groups was investigated by aggregating OTUs to the Phylum level, except for the Phylum Proteobacteria, which was aggregated to the Class level, followed by conversion to relative abundance. A Kruskal-Wallis test (from R package '*stats*') was used to make comparisons between soil partition and habitat restoration separately to determine if significant differences were present across factors, followed by a Dunn's test (from R package '*rstatix*') with post-hoc Bonferroni correction to simultaneously assess pairwise differences.

Beta diversity

Beta (β) diversity across assemblages was evaluated using two dissimilarity metrics, Bray-Curtis (from R package '*vegan*'), an abundance based metric that is influenced by differences in assemblage richness, and the Raup-Crick metric (from R package '*vegan*'), which accounts for differences in richness by generating a null expectation of the number of shared OTUs between

sites (Chase *et al.* 2011). A permutational test of multivariate dispersion was conducted separately across habitats, and sample partitions, to determine homogeneity of variance, followed by Tukey's honestly significant difference (from R package '*stats*') to test if significant differences in dispersion were detected. Permutational multivariate ANOVAs (PerMANOVAs) were used to determine if differences in multivariate centroid position were present between groups (in R package '*vegan*'). Assemblage composition was visualized using non-metric multidimensional scaling (NMDS) of Bray-Curtis (K(3) = 0.098) and Raup-Crick (K(3) = 0.022) dissimilarities (with R package '*ggplot2*'). To investigate the effect of soil N concentration and richness on assemblage composition, distance-based redundancy analysis (dbRDA, from R package '*vegan*') was conducted.

Modeling nutrient flux with qPCR copy number

GAMM models were employed to investigate the potential non-linear relationships between time-weighted average nutrient flux values, and log transformed qPCR copy numbers of functional N cycling genes (using R package '*mgcv*'). Log transformed qPCR values of nosZ-A and nosZ-B were modeled independently for dinitrogen gas (N_2) flux using soil partition and habitat restoration as both parametric and smoothing terms. Log-transformed qPCR values of periplasmic nxrB were constructed identically to those of nosZ-A and nosZ-B, differentially evaluating its predictability from both nitrite (NO_2^-) and nitrate (NO_3^-) flux rates. Non-parametric smoothing terms were included for the nutrient flux rate of interest (NO_2^-, and NO_3^- respectively), and its interaction by factor terms (Appendix A Supplemental Methods).

Predicting nutrient flux with bacterial assemblage composition

Machine learning models were used to generate independent predictions of nutrient flux values from bacterial assemblage composition for each soil partition. Bacterial OTUs were

aggregated according to genera for each assemblage (Armour *et al.* 2022) and columns with non-zero sums were retained. Machine learning models were constructed using the '*caret*' package in R (Kuhn 2008) to implement the Generalized Linear Models with Elastic Net regularization (GLMNET) (Friedman *et al.* 2010).

Predictive accuracy was evaluated using values of RMSE followed by conversion to normalized root mean squared errors (NRMSE) by dividing values of RMSE by the standard deviation (SD) of each respective nutrient response variable (D. N. Moriasi *et al.* 2007; Ranatunga *et al.* 2016). Values of NRMSE standardized by SD (0 – 1) represent the ratio of explained vs unexplained variation by the regression against the response variable (Otto, 2019). Lower values of NRMSE represent better predicting models, where the proportion of explained variation is significantly greater than the proportion of unexplained variation, where a value of 0 represents a model that perfectly predicts the response variable (D. N. Moriasi *et al.* 2007; Ranatunga *et al.* 2016).

Alongside GLMNET, we also employed SetBERT, a recent set transformer-based architecture designed to process entire samples (i.e. sets of DNA sequences) (In Review). SetBERT is capable of making more informed predictions by leveraging sequence interactions, as opposed to most models that process DNA sequences in isolation. Using a pre-trained SetBERT model, we fine-tuned it for NO2 flux rate prediction using two different approaches. The first approach trains four independent SetBERT models, each predicting the flux rate from a given field or lab sample of a particular partition. The input sample is embedded and passes through a single dense layer. The second approach trains two SetBERT models, where each model predicts the flux rate given a field sample and its corresponding lab sample pair for a partition. The field and lab samples are embedded using the same SetBERT model, and the embeddings are concatenated and

passed through a dense layer to predict the flux rate. These models were trained with MSE loss using the Adam optimizer with a learning rate of 1e-4.

Results

Factors influencing patterns of assemblage composition in field and lab collected soil samples

The number of *effective commonly abundant* OTUs decreased with increasing N concentration (Fig. 5a, Table 1). Significant partial effects were detected for the smooths of soil N by sample partition (Fig. 5b), and soil N by habitat type (Fig. 5c), but not for the two factor partition-habitat interaction smooth of Soil N (Table 1). Partial effects smooth differences for partitions revealed that *effective* OTUs decline with sample depth, while the incubation experiment significantly increased *effective* OTUs for top partitions along the range of Soil N between 0.1% - 0.3% (Fig. 5b). Partial effects smooth differences for habitats showed that at the lowest range of soil N similar numbers of *effective* OTUs were detected. As soil N increased from 0.15 - 0.3%, shallow water habitats exhibit an increase in detected effective OTUs (Fig. 5c). Across the highest range of soil N concentration, all habitats experienced a significant decline in *effective* OTUs though the differences were not significantly different from one another (Fig 5c). Additional analysis of trends in richness can be found in the supplement content (Appendix E Supplemental Results, Appendix Table C2, Appendix Fig. D1).

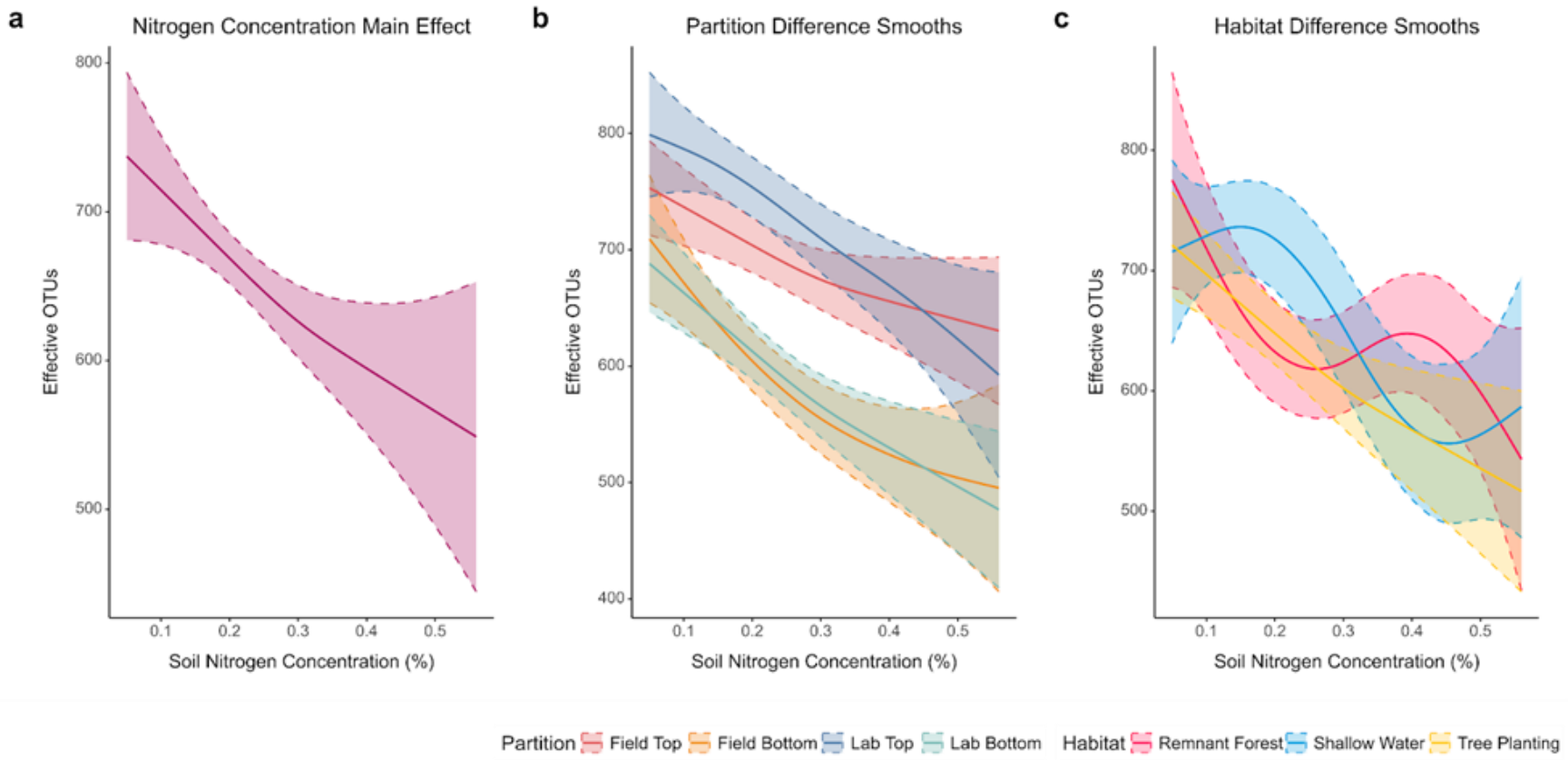


Figure 5. Effective OTUs Decline with Increased Soil Nitrogen Concentration. Significant GAMM model smooths are shown with Bayesian confidence intervals for (a) the Main Effect of Soil N concentration on effective OTUs colored in purple. Partial Effects Difference Smooths depict the differences in estimated effective OTUs between the main effect and smooth effect for each level colored by (b) partition and (c) habitat along the gradient of soil N

Table 1. GAMM results of Hill-Shannon Diversity model. ***SE***; Standard Error, ***e.d.f.***; expected degrees of freedom, ***Ref d.f.***; reference degrees of freedom.

Component	**Term**	**Estimate**	***SE***	***T***	***P***	
A. parametric coefficients	(Intercept)	658.917	98.100	6.717	0.0000	***
Component	**Term**	***e.d.f.***	***Ref d.f.***	***F***	***P***	
B. smooth terms	s(Soil Nitrogen)	1.289	5	54.103	0.0001	***
	s(Soil Nitrogen, Partition)	6.287	30	114.971	0.0000	***
	s(Soil Nitrogen, Habitat)	3.594	20	69.109	0.0037	**
	s(Soil Nitrogen, Partition, Habitat)	0.014	60	0.015	0.3213	
	s(Latitude, Longitude)	4.162	99	392.975	0.1338	
	s(Site)	0.001	7	0.001	0.3233	
	s(State)	0.911	1	20.580	0.0000	***
	s(Year)	0.001	1	0.000	0.4408	
	s(Sample ID)	87.759	146	254.699	0.0000	***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.461, Deviance explained 0.523

-REML : 3687.976, Scale est: 1.000, N: 559

Family: Scaled t(5.532,123.715), Link function: Identity

Assemblage relative abundance

An assessment of taxonomic relative abundance by sample partition (Appendix Table C3) showed several statistically significant differences among partitions, within taxonomic groups (Fig. 6, Appendix Table C4). Significant differences were detected among partitions within Acidobacteria, Gammaproteobacteria, Actinobacteria, Deltaproteobacteria, Chloroflexi, Bacteriodetes, Gemmatimonadetes, among the Unclassified Bacteria, and including Nitrospirae, Rokubacteria, Firmicutes, Latesibacteria, and Cyanobacteria (Appendix Table C4). No significant differences were detected among soil partitions within Alphaproteobacteria, Verrucomicrobia, or Plantomycetes, or rare taxa classified as "Other". Several significant differences between field top and field bottom samples persisted after the flow-through incubation experiment, with the exception of Firmicutes (Table 2) which showed no difference after incubation. Comparisons between field top and lab top samples indicated a statistically significant shift in the relative abundance of Gammaproteobacteria, Actinobacteria, Chloroflexi, and Bacteriodetes, suggesting that average assemblage composition of field top samples changes over the course of the flow-through incubation experiment (Table 2). The only detected significant difference between field bottom and lab bottom partitions was the relative abundance of Nitrospirae, which showed significant enrichment in lab bottom partitions (Table 2).

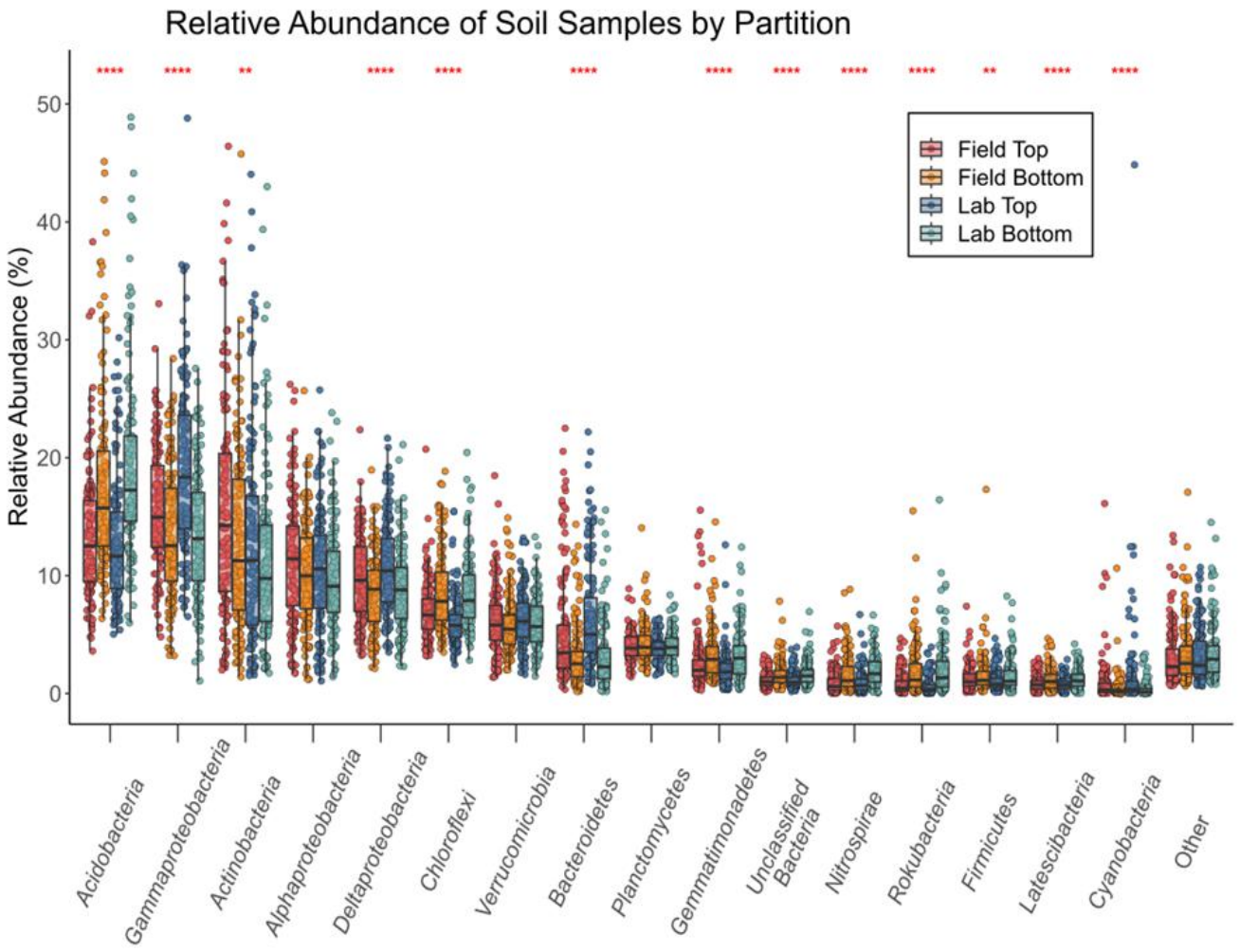


Figure 6. Assemblage Relative Abundance by Partition. Box plots of relative abundance (%) by Phylum, and Class (for Phylum Proteobacteria) colored by soil sample partition. Red asterisks indicate a significant difference detected by Kruskal-Wallis test between partitions by taxonomic group

Table 2. Dunn's Test results of relative abundance by pairwise partitions (Fig. 6). Significant *p-values* are depicted in bold.

	Field Top - Field Bottom		**Field Top - Lab Top**		**Field Top - Lab Bottom**		**Field Bottom - Lab Top**		**Field Bottom - Lab Bottom**		**Lab Top - Lab Bottom**	
Taxonomic Group	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***
Acidobacteria	5.271	**0.0000**	-1.069	1.0000	7.034	**0.0000**	-6.345	**0.0000**	1.776	0.4542	8.107	**0.0000**
Gammaproteobacteria	-3.191	**0.0085**	3.590	**0.0020**	-3.293	**0.0059**	6.773	**0.0000**	-0.114	1.0000	-6.863	**0.0000**
Actinobacteria	-2.275	0.1374	-3.055	**0.0135**	-4.319	**0.0001**	-0.765	1.0000	-2.045	0.2452	-1.294	1.0000
Alphaproteobacteria	-1.597	0.6612	-1.054	1.0000	-2.909	**0.0218**	0.550	1.0000	-1.313	1.0000	-1.868	0.3707
Deltaproteobacteria	-2.827	**0.0282**	2.155	0.1871	-2.078	0.2260	4.979	**0.0000**	0.735	1.0000	-4.221	**0.0001**
Chloroflexi	4.002	**0.0004**	-3.811	**0.0008**	4.262	**0.0001**	-7.806	**0.0000**	0.274	1.0000	8.053	**0.0000**
Verrucomicrobia	-1.332	1.0000	1.144	1.0000	-0.131	1.0000	2.474	0.0802	1.192	1.0000	-1.267	1.0000
Bacteroidetes	-4.501	**0.0000**	3.385	**0.0043**	-4.550	**0.0000**	7.881	**0.0000**	-0.065	1.0000	-7.918	**0.0000**
Planctomycetes	1.214	1.0000	-1.054	1.0000	0.100	1.0000	-2.266	0.1409	-1.105	1.0000	1.146	1.0000
Gemmatimonadetes	4.286	**0.0001**	-1.943	0.3122	4.106	**0.0002**	-6.230	**0.0000**	-0.164	1.0000	6.042	**0.0000**
Unclassified Bacteria	4.634	**0.0000**	-0.196	1.0000	5.484	**0.0000**	-4.837	**0.0000**	0.863	1.0000	5.687	**0.0000**
Nitrospirae	4.926	**0.0000**	0.373	1.0000	7.887	**0.0000**	-4.563	**0.0000**	2.969	**0.0179**	7.530	**0.0000**
Rokubacteria	6.155	**0.0000**	-1.640	0.6061	6.716	**0.0000**	-7.800	**0.0000**	0.582	1.0000	8.356	**0.0000**
Firmicutes	2.097	0.2160	-1.988	0.2807	1.487	0.8221	-4.082	**0.0003**	-0.600	1.0000	3.464	**0.0032**
Latescibacteria	3.219	**0.0077**	-0.411	1.0000	3.970	**0.0004**	-3.633	**0.0017**	0.760	1.0000	4.384	**0.0001**
Cyanobacteria	-4.184	**0.0002**	1.404	0.9621	-3.841	**0.0007**	5.590	**0.0000**	0.327	1.0000	-5.241	**0.0000**
Other	1.152	1.0000	0.961	1.0000	2.198	0.1677	-0.196	1.0000	1.046	1.0000	1.247	1.0000

Taxonomic relative abundance assessed by habitat type (Appendix Table C5) revealed statistically significant differences across multiple taxonomic groups (Fig. 7, Appendix Table C6). Statistically significant differences among habitat types were detected within Acidobacteria, Actinobacteria, Alphaproteobacteria, Deltaproteobacteria, Chloroflexi, Bacteriodetes, Planctomycetes, Gemmatimonadetes, Unclassified Bacteria, Nitrospirae, Firmicutes, Cyanobacteria, Spirochaetes, and rare taxa classified as "Other" (Appendix Table C6). More significant differences across taxonomic groups occurred between remnant forest and shallow water habitats (n = 12/17), and shallow water and tree planting habitats (n = 12/17), than between remnant forest and tree planting habitats (n = 4/17), indicating that remnant forest and tree planting habitats are more similar in bacterial assemblage taxonomic composition than either are to shallow water habitat (Table 3).

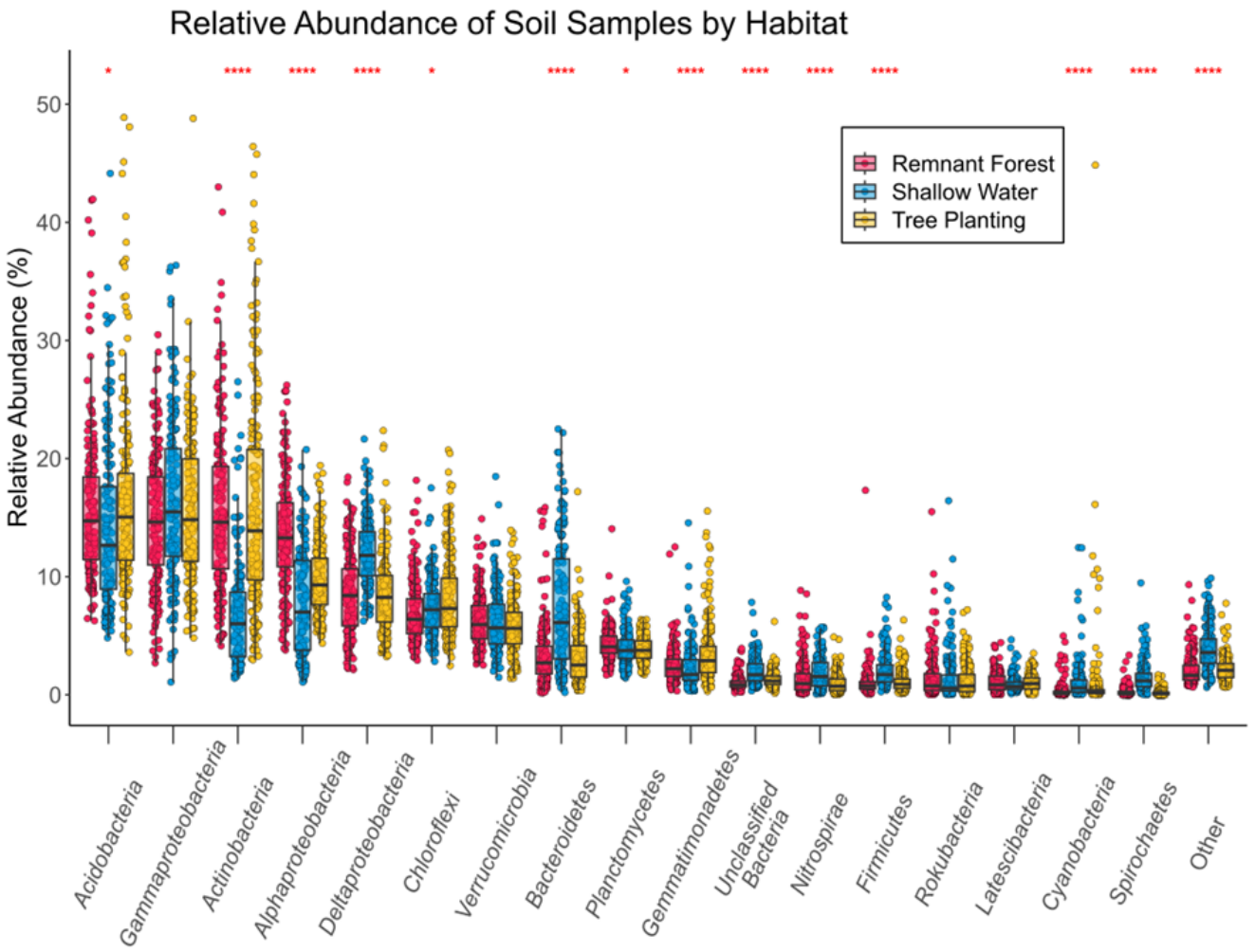


Figure 7. Assemblage Relative Abundance by Habitat. Box plots of relative abundance (%) by Phylum, and Class (for Phylum Proteobacteria) colored by habitat type. Red asterisks indicate a significant difference as detected by Kruskal-Wallis test between habitats by taxonomic group

Table 3. Dunn's Test results of relative abundance by pairwise habitats (Fig. 7). Significant *p values* are depicted in bold.

	Remnant Forest - Shallow Water		**Remnant Forest - Tree Planting**		**Shallow Water - Tree Planting**	
Taxonomic Group	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***
Acidobacteria	-2.612	**0.0270**	0.846	1.0000	3.476	**0.0015**
Gammaproteobacteria	2.082	0.1119	1.475	0.4208	-0.699	1.0000
Actinobacteria	-11.830	**0.0000**	-0.616	1.0000	11.458	**0.0000**
Alphaproteobacteria	-11.098	**0.0000**	-7.890	**0.0000**	3.699	**0.0007**
Deltaproteobacteria	9.649	**0.0000**	-0.388	1.0000	-10.204	**0.0000**
Chloroflexi	2.290	0.0660	3.794	**0.0004**	1.325	0.5552
Verrucomicrobia	-0.404	1.0000	-2.195	0.0844	-1.705	0.2646
Bacteroidetes	7.986	**0.0000**	-1.020	0.9226	-9.119	**0.0000**
Planctomycetes	-2.564	**0.0310**	-3.359	**0.0023**	-0.627	1.0000
Gemmatimonadetes	-2.379	0.0521	4.225	**0.0001**	6.497	**0.0000**
Unclassified Bacteria	10.844	**0.0000**	5.021	**0.0000**	-6.205	**0.0000**
Nitrospirae	4.324	**0.0000**	-2.364	0.0542	-6.684	**0.0000**
Firmicutes	10.093	**0.0000**	2.017	0.1310	-8.337	**0.0000**
Rokubacteria	-1.126	0.7806	-0.940	1.0000	0.241	1.0000
Latescibacteria	-2.217	0.0799	-0.056	1.0000	2.205	0.0824
Cyanobacteria	10.044	**0.0000**	4.031	**0.0002**	-6.346	**0.0000**
Spirochaetes	12.777	**0.0000**	-1.945	0.1553	-14.891	**0.0000**
Other	11.937	**0.0000**	2.351	0.0562	-9.893	**0.0000**

Beta diversity

Variation partitioning of assemblage composition assessed by Bray-Curtis dissimilarity (Fig. 8a, b, Appendix Table C7) showed significant effects of easement site, habitat (Fig. 8b), partition (Fig. 8a), richness, and soil N concentration, on the soil microbiome. Pairwise PERMANOVA of partitions showed significant differences for all but field bottom / lab bottom (Appendix Table C8). Independent pairwise PERMANOVA analysis of habitat showed significant differences across all combinations of habitat type comparisons (Appendix Table C8).

Variation partitioning of assemblage composition assessed using Raup-Crick dissimilarity (Fig. 8d, e, Appendix Table C9) revealed significant effects of easement site, habitat (Fig. 8e), partition (Fig. 8d), richness, but not soil N concentration. Pairwise PERMANOVA of partitions showed significant differences for all but field top / field bottom and field bottom / lab bottom (Appendix Table C8). Independent pairwise PERMANOVA of habitat showed significant differences across all combinations of habitat type (Appendix Table C8). Similarly, permutational tests of homogeneity of dispersion for both Bray-Curtis and Raup-Crick dissimilarities can be found in Supplementary Results.

The Bray-Curtis dbRDA (Fig. 8c; adj. $R^2 = 0.0726$) analysis revealed that both soil N concentration (PERMANOVA, SS = 2.180, F = 8.848, $P < 0.001$) and assemblage richness (PERMANOVA, SS = 8.533, F = 34.641, $P < 0.001$) significantly contributed to variation in assemblage composition. The richness vector was parallel to CAP1, indicating that increase in richness increased along the CAP1 axis. I found that soil N concentration had a weaker relationship than that of richness, based on the length of the vector, but was still significant and samples with increasing soil N concentration having a more negative CAP2 value. The Raup-Crick dbRDA (Fig. 8e; adj. $R^2 = 0.047$) analysis indicated that only assemblage richness (PERMANOVA, SS =

13.78, $F = 22.624$, $P < 0.001$) significantly contributed to the variation in assemblage composition and increased along the CAP1 axis.

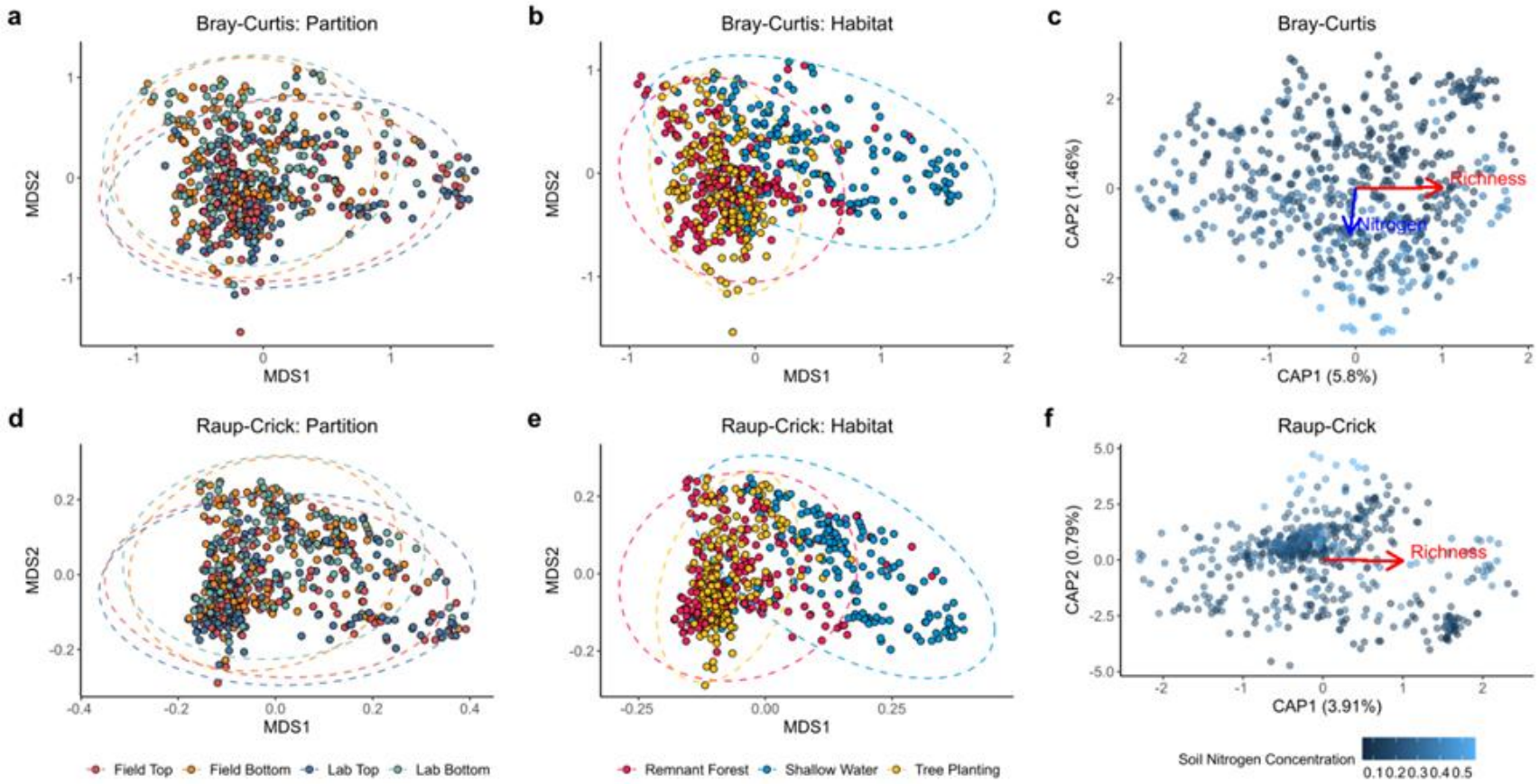


Figure 8. Ordinations of Assemblage Beta Diversity. The first two dimensions of NMDS plots of Bray-Curtis β-diversity index colored by partition (a) and habitat type (b), contrasted by plots of Raup-Crick β-diversity colored by partition (d) and habitat (e) are shown. Constrained ordinations of Bray-Curtis (c) and Raup-Crick (f) dissimilarities are depicted with sample N concentration in blue gradient, with significant loadings for Richness (red) and N concentration (blue)

Prediction of functional nitrogen cycling gene abundance

Log copy numbers of nozZ-A were substantially enriched for field bottom, lab top, and lab bottom samples as compared to field top samples (Appendix Table C10). While the main effect of N_2 average flux on nosZ-A copies was not significant, the three-way interaction smooth was significant for remnant forest in lab tops (Fig. 9a), as was for tree planting in lab bottoms (Fig. 9c). While log copy numbers of nosZ-B did not differ significantly among partitions (Appendix Table C11), significant smooth terms were detected for the three-way smooth interaction of N_2, partition, and habitat factors that included remnant forest in field tops (Fig. 9b), and tree planting in lab bottoms (Fig. 9d). Additional significant parametric effects can be found in Supplementary Results, in addition to trends of other functional gene markers.

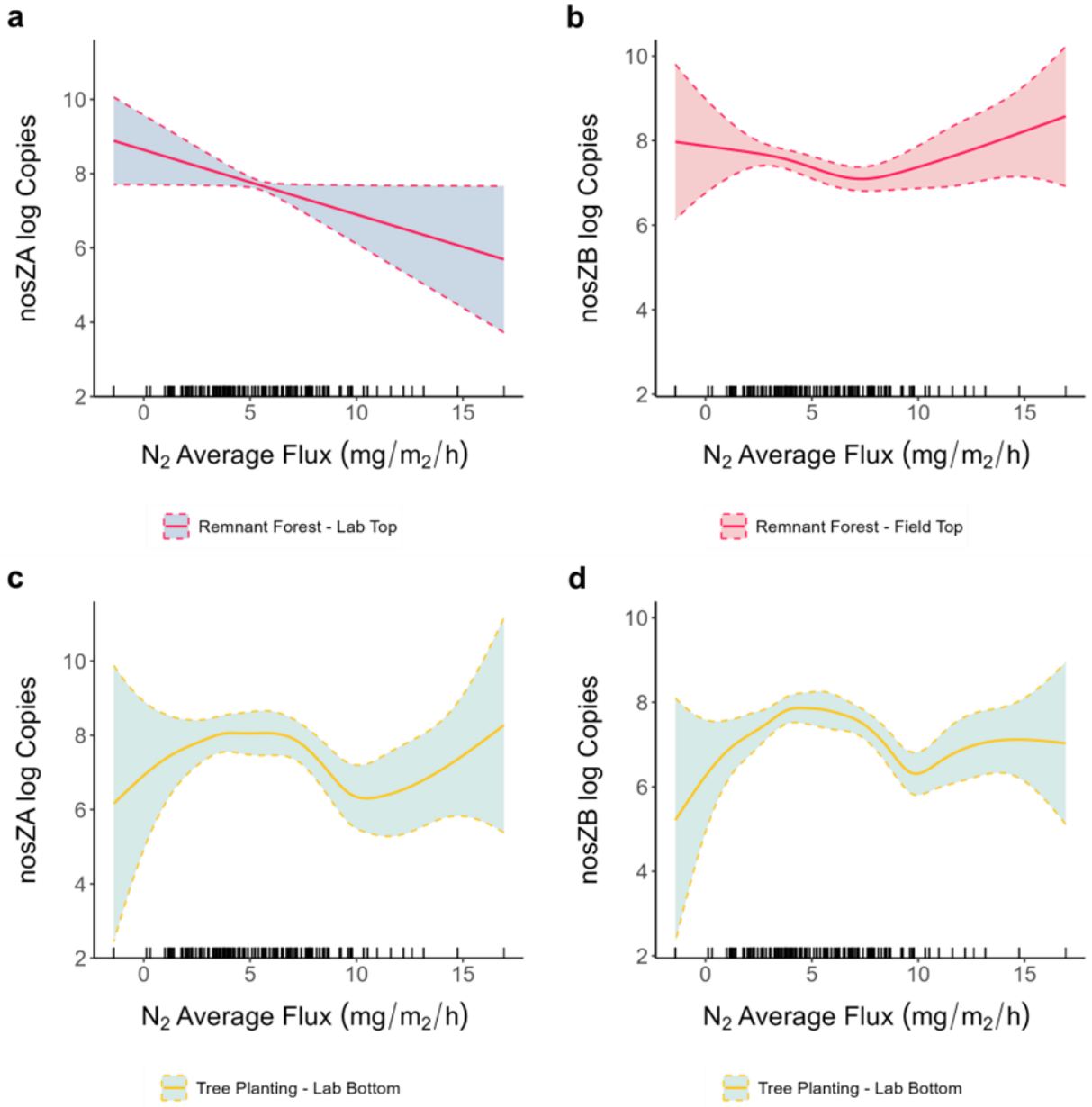


Figure 9. Trends in nosZ-A and nosZ-B Gene Abundance Estimated with N_2 Flux Rates. Significant GAMM model smooths depicted with Baysian confidence intervals and rug plots of sampling effort are shown for the three-way interactions of N_2 predicting noszA log copies with remnant forest lab tops (a), nosZB log copies with remnant forest field tops (b), nosZA log copies with tree planting lab bottoms (c), and nosZB log copies with tree planting lab bottoms (d)

Machine Learning and Deep Neural Networks

GLMNET models showed significant differential predictive accuracy of soil partitions across nutrient flux rates (Fig. 10a). Field and lab top samples produced the greatest predictive accuracy for both N_2 and NO_2 flux rates. GLMNET models for NH_4 and NO_3 produced poorly fitting models regardless of the soil partition that was selected. The best fitting GLMNET model was that of PO_4 predicted by field bottom partitions. SetBERT flux rate predictions of NO_2 (Fig. 10b)indicated a shift in predictive accuracy with lab bottom partitions showing the greatest predictive accuracy among independent models (0.4732). Using combined models that linked partitions (top vs bottom) across incubation treatment, setBERT was able to outperform all independent partition models of NO_2 flux rates, and indicated that the greatest predictive accuracy of NO_2 flux rates was generated using bottom samples from field and lab (0.294).

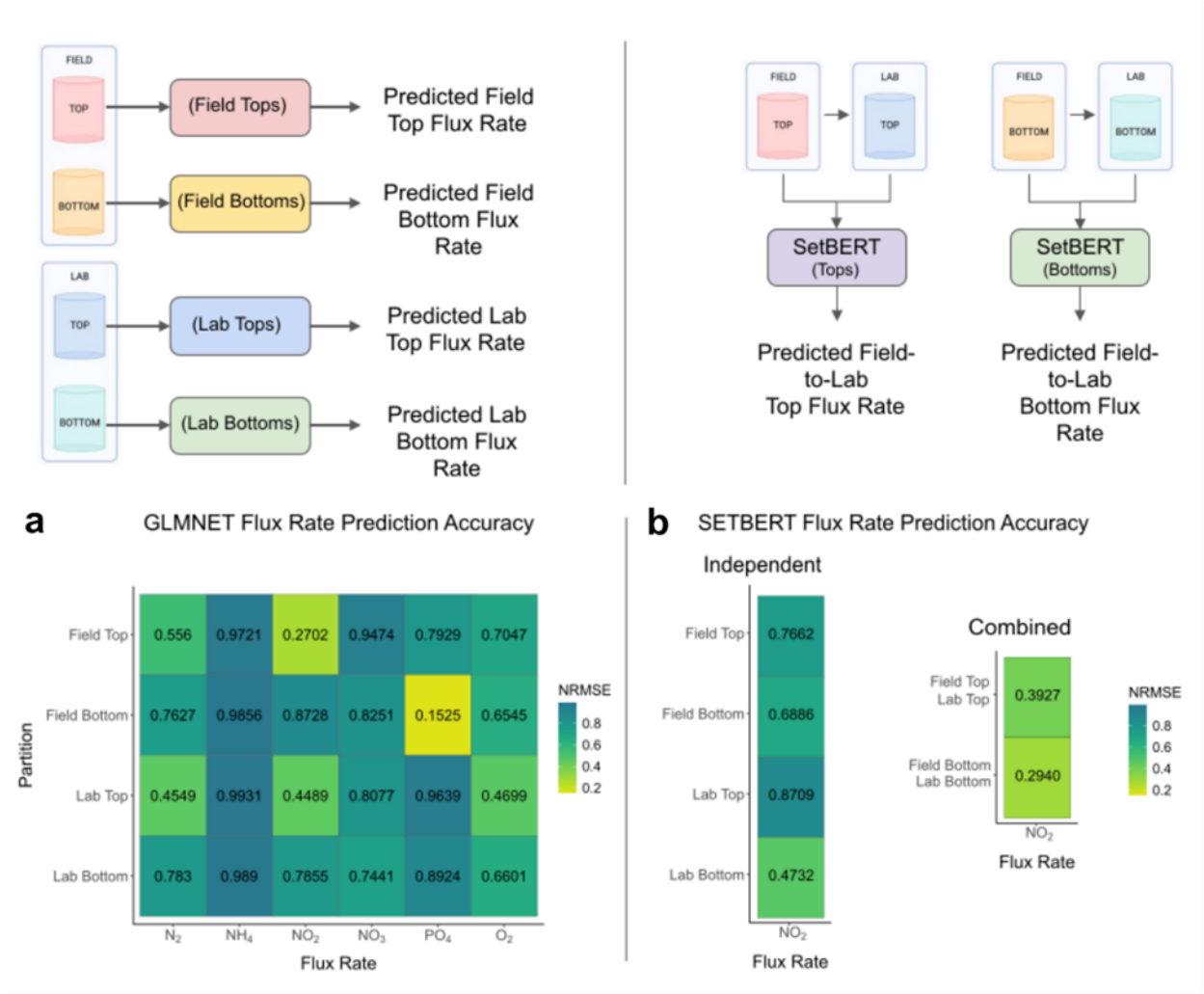


Figure 10. Machine Learning and Deep Neural Network prediction of nutrient flux rates. Accuracy of independent GLMNET models predicting nutrient flux rates from partition microbial assemblage OTU counts aggregated to genera (a). SetBERT model accuracy (b) of predicted NO_2 model accuracy for independent partitions (right), and combined top or bottom samples (left). Model accuracies are presented as NRMSE values and colored from dark green to yellow depicting increasing accuracy as values approach 0. Created in https://BioRender.com

Discussion

Soil microbiomes contribute directly to the multifunctionality of wetland ecosystems and offer a unique framework for building predictive models of biogeochemical processes such as nutrient cycling across the landscape. They can also be used as a conservation tool to inform future management and conservation decisions. This study focused on a fine-scale and multidisciplinary approach to surveying soil bacterial assemblage composition on the landscape. It explored the relationships between three key restoration practices commonly found in restored wetlands – remnant forest, shallow water, and tree plantings, at both the soil surface and subsurface, and assessed bacterial response before and after inundation with nutrient rich water in a controlled lab experiment. This study's major findings were 1) soil bacterial assemblage composition was structured by N concentration, soil depth, and habitat type, 2) functional N cycling gene copy numbers could be predicted by measured N flux rates, 3) the composition of individual soil partitions could be used to predict N flux rates and 4) N flux rate prediction could be further improved by accounting for temporal variation in soil partitions. Together, these findings provide evidence for the utility of soil bacterial assemblages to serve as bioindicators of nutrient cycling in remediated wetland ecosystems.

Our GAMM models of both Richness (Appendix E Supplemental Results) and Hill-Shannon Diversity (Fig. 5a) indicate that bacterial assemblage diversity decreased with N concentration in sampled soils. These results are consistent with previous studies that have shown bacterial phylotype (Yuan *et al.* 2017), relative abundance (Li *et al.* 2016), and plant diversity (Midolo *et al.* 2019) respond similarly to increasing N concentrations at the landscape scale. Our model also revealed an increased measured assemblage diversity in top soil samples following the lab incubation experiment, particularly along the low to medium range of N concentration

suggesting a rapid shift in the number of commonly abundant OTUs. Restoration practices at the landscape scale responded differentially to increasing N concentration, with remnant forest and shallow water habitats exhibiting an offset sinusoidal pattern along the measured N gradient. This differential response of restoration practices to N concentration may be reflective of differences in environmental filtering (Thakur and Wright 2017; Furtak and Gałązka 2019; Kasak *et al.* 2021) across microhabitats on the landscape and its interaction with N concentration. This differential response was strongly linked to moisture gradients (Tomasek *et al.* 2017; Veach and Zeglin 2020), soil formation processes (Sokol *et al.* 2022), and soil structure (Tang *et al.* 2018). The depression in commonly abundant OTUs observed in tree planting habitats along the N concentration gradient may be attributed to both disturbance and succession ecology (Connell and Slatyer 1977; McCook 1994), or may also be in response to historical agricultural practices (Rasche and Cadisch 2013; Tomasek *et al.* 2017; Hemes *et al.* 2019; Dai *et al.* 2020) and their lasting effect on microbial diversity (Gupta *et al.* 2022).

Significant differences in relative abundance of sampled microbial assemblages were observed when the phylogenetic scale was sub-divided to phylum and class level (Fig. 6 & 7). These differences align with the potential environmental heterogeneity among soil partitions (Fig. 6), controlled lab incubations (Fig. 6), and restoration practices (Fig. 7). These findings are consistent with previous literature which highlights microbial compositional differences driven by soil depth (Fierer 2017; Tang *et al.* 2018; Cai *et al.* 2022; Naylor *et al.* 2022), moisture availability (Tomasek et al., 2017; Veach and Zeglin, 2020), and microhabitats (Mueller *et al.* 2015; Fierer 2017; Wang *et al.* 2021). The importance of spatial and temporal scale for detecting domain specific patterns is further highlighted by the enrichment of Nitrospirae taxa, which represented the only significant difference in relative abundance detected between field bottom and lab bottom

partitions (Fig. 6). Members of the phylum Nitrospirae represent a diverse range of aerobic chemolithotrophs that are involved in key nutrient cycling pathways, including nitrification (Paul and Frey 2023).

Soil assemblage β-diversity as assessed using the abundance based Bray-Curtis metric, was primarily influenced by soil partition (Fig. 8a), and habitat type (Fig. 8b). A strata-specific response was observed for field and lab top samples, consistent with changes in α-diversity described earlier. These results suggest that top soils may experience rapid shifts in abundance during the incubation experiments (Urakawa and Bernhard 2017). Raup-Crick dissimilarity revealed significant turnover between partitions (Fig. 8d) and habitat types (Fig. 8e). However, pairwise PERMANOVA results no longer showed significant differences between field collected and lab incubated top or bottom samples. Taken in tandem, this may indicate that the assemblage turnover observed during lab incubation in top samples was primarily driven by shifts in abundance, while community compositions remain relatively stable.

Constrained ordinations of Bray-Curtis (Fig. 8c) indicated that the N concentration on the landscape significantly contributed to differences in abundance across the landscape (Fig. 8c), while Raup-Crick analysis revealed factors other than N concentration are key contributors to measured richness (Fig. 8f). Past studies have shown significant shifts in β-diversity associated with N amendment on the landscape (Mueller *et al.* 2015; Tang *et al.* 2018), and our results highlight the importance of using traditional metrics alongside null models to determine whether β-diversity shifts are significantly related to trends in measured α-diversity. Patterns observed across α- and β-diversity also support the hypothesis that soil bacterial assemblage composition differs by soil horizon and restoration practice on the landscape.

Smooth trends for both remnant forest and tree planting habitats were consistent for both nosZ-A and nosZ-B log copy numbers. These findings highlight the importance of habitat in making predictions and suggest that either molecular marker may be estimated using N_2 gas flux, and also that functional gene abundance is differentially affected by habitats. The three-way interaction among nutrient flux rates, habitats, and partitions observed for both N_2 gas GAMMs and the NO_2 GAMM underscores the importance of accounting for both fine-scale (partitions) and environmental (habitats) heterogeneity when developing predictive models (Young *et al.* 2019). While these results indicate that habitat may be an important indicator of functional gene copy numbers, without accounting for fine-scale spatial variation, significant smooth trends could not be detected. It is important to note that GAMM model estimates may be erroneous outside of the sampled effort, so estimates extended to the upper and lower bounds of flux rates should be interpreted with caution.

GLMNET results indicate that soil assemblage composition offers a meaningful and wide ranging ability to predict nutrient flux. GLMNET results revealed that the poorest flux rate predictions were for ammonia and nitrite, likely due to water saturation with those compounds and that the turnover of microbial assemblages and transformation chemical compounds may likely be decoupled in time. Interestingly, both top partitions were better predictors of dinitrogen gas and NO_2 flux, with differential predictive accuracy between field and lab collected samples. These distinct compositional differences in assemblages, based on α- and β-diversity metrics, may well be the driving factor in the predictive accuracy of these assemblages. Further exploration with Deep Neural Networks using setBERT showed paired setBERT models revealed that taking partition into consideration and spanning the temporal scale of lab incubations produced better predictions than either partition alone. The turnover that occurs as a result of the incubation

experiment is important in driving increased accuracy of predictions. Both GLMNET and setBERT results support the hypothesis that machine learning can link soil bacterial composition to nutrient cycling rates, with varying degrees of success.

Despite showing the utility of bacterial assemblage composition to build predictive models of nutrient cycling on the landscape, all microbial consortia (e.g. fungi, protists, viruses) of the soil microbiome were not fully explored, and differential patterns of abundance, composition, and response to abiotic factors have been shown for fungi as well (Mueller *et al.* 2015). We acknowledge these limitations and future research should consider the multiple signals that are provided by other members of the soil microbiome, in addition to coupling patterns of distribution with multiple abiotic factors instead of a single measure.

Conclusion

The soil microbiome can be used to build predictive models of nutrient flux rates, however it is important to consider the major factors that drive microbial assemblage composition and account for these when building modeling approaches. When the ecological scale (spatial, temporal, and phylogenetic) matches the domain of interest, the best predictive accuracy is revealed. Predictive accuracy of biogeochemical processes and other landscape level functions of the microbiome will only be improved by embracing the complexity of the soil microbiome, and collecting measures that span the greatest variation of both abiotic and biotic factors. Of particular interest to future researchers is the ability to design studies that not only span the temporal scale, but do so in a multi-scalar manner. Deep learning methods and their modeling utility will drive a step-function change in predictive accuracy of landscape-functions utilizing soil microbiome composition, as they may incorporate data that spans the entirety of ecological scale to produce better fitting models and bring a picture of microbial dynamics into focus.

Acknowledgments

This work was supported by the Natural Resource Conservation Services sub-award 220858, National Science Foundation grants EF-2125065 and CAREER 2236580 to D. Walker, the Molecular Biosciences Program at MTSU, and National Science Foundation grant DEB 1933925 to J. Phillips. The authors would like to thank Alexander Romer, Steven Levenhagen for their technical assistance on this project. Any opinions, findings, and conclusions or recommendations expressed in this material are my own and do not necessarily reflect the views of the National Science Foundation, or Natural Resource Conservation Service.

References

Armour CR, Topçuoğlu BD, Garretto A *et al.* A Goldilocks Principle for the Gut Microbiome: Taxonomic Resolution Matters for Microbiome-Based Classification of Colorectal Cancer. 2022;**13**, DOI: https://doi.org/10.1128/mbio.03161-21.

Brown RS. Assessing Nutrient Retention and Removal Measurements Among Restored Floodplain Wetlands. 2023.

Brown SP, Veach AM, Rigdon-Huss AR *et al.* Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecol* 2015;**13**:221–5.

Bullen P. The power means. *Handb Means Their Inequalities* 2003:175–265.

Cai Y-J, Liu Z-A, Zhang S *et al.* Microbial community structure is stratified at the millimeter-scale across the soil–water interface. *ISME Commun* 2022;**2**:53.

Cao Q, Sun X, Rajesh K *et al.* Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. *Front Microbiol* 2021;**11**:607325.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.

Chaparro JM, Sheflin AM, Manter DK *et al.* Manipulating the soil microbiome to increase soil health and plant fertility. *Biol Fertil Soils* 2012;**48**:489–99.

Chase JM, Kraft NJB, Smith KG *et al.* Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* 2011;**2**:art24.

Comín FA, Romero JA, Astorga V *et al.* Nitrogen removal and cycling in restored wetlands used as filters of nutrients for agricultural runoff. *Water Sci Technol* 1997;**35**:255–61.

Connell JH, Slatyer RO. Mechanisms of succession in natural communities and their role in

community stability and organization. *Am Nat* 1977;**111**:1119–44.

Cookson WR, Osman M, Marschner P *et al.* Controls on soil nitrogen cycling and microbial community composition across land use and incubation temperature. *Soil Biol Biochem* 2007;**39**:744–56.

Correa-Garcia S, Constant P, Yergeau E. The forecasting power of the microbiome. *Trends Microbiol* 2023;**31**:444–52.

Creamer RE, Hannula SE, Leeuwen JPV *et al.* Ecological network analysis reveals the inter-connection between soil biodiversity and ecosystem function as affected by land use across Europe. *Appl Soil Ecol* 2016;**97**:112–24.

D. N. Moriasi, J. G. Arnold, M. W. Van Liew *et al.* Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans ASABE* 2007;**50**:885–900.

Dai Z, Liu G, Chen H *et al.* Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. *ISME J* 2020;**14**:757–70.

Davis NM, Proctor DM, Holmes SP *et al.* Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;**6**:226.

Falkowski PG, Fenchel T, Delong EF. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* 2008;**320**:1034–9.

Fernández-Martínez MA, Pointing SB, Pérez-Ortega S *et al.* Functional ecology of soil microbial communities along a glacier forefield in Tierra del Fuego (Chile). *Int Microbiol* 2017:161–73.

Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* 2017;**15**:579–90.

Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;**33**, DOI: 10.18637/jss.v033.i01.

Furtak K, Gałązka A. Edaphic Factors And Their Influence On The Microbiological Biodiversity Of The Soil Environment. *Postępy Mikrobiol - Adv Microbiol* 2019;**58**:375–84.

Glibert PM. Harmful algae at the complex nexus of eutrophication and climate change. *Harmful Algae* 2020;**91**:101583.

Glibert PM, Maranger R, Sobota DJ *et al.* The Haber Bosch–harmful algal bloom (HB–HAB) link. *Environ Res Lett* 2014;**9**:105001.

Graham CH, Storch D, Machac A *et al.* Phylogenetic scale in ecology and evolution. *Glob Ecol Biogeogr* 2018;**27**:175–87.

Grajal-Puche A, Murray CM, Kearley M *et al.* Microbial Assemblage Dynamics Within the American Alligator Nesting Ecosystem: a Comparative Approach Across Ecological Scales. *Microb Ecol* 2020;**80**:603–13.

Groffman PM, Altabet MA, Böhlke JK *et al.* METHODS FOR MEASURING DENITRIFICATION: DIVERSE APPROACHES TO A DIFFICULT PROBLEM. *Ecol Appl* 2006;**16**:2091–122.

Gupta A, Singh UB, Sahu PK *et al.* Linking Soil Microbial Diversity to Modern Agriculture Practices: A Review. *Int J Environ Res Public Health* 2022;**19**:3141.

Hallin S, Philippot L, Löffler FE *et al.* Genomics and Ecology of Novel N2O-Reducing Microorganisms. *Trends Microbiol* 2018;**26**:43–55.

Hartemink AE, Zhang Y, Bockheim JG *et al.* Soil horizon variation: A review. *Advances in Agronomy*. Vol 160. Elsevier, 2020, 125–85.

Hemes KS, Chamberlain SD, Eichelmann E *et al.* Assessing the carbon and climate benefit of restoring degraded agricultural peat soils to managed wetlands. *Agric For Meteorol* 2019;**268**:202–14.

Hill MO. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 1973;**54**:427–32.

Ivanova EA, Pershina EV, Shapkin VM *et al.* Shifting prokaryotic communities along a soil formation chronosequence and across soil horizons in a South Taiga ecosystem. *Pedobiologia* 2020;**81–82**:150650.

Jost L. *Entropy and diversity*. *Oikos* 2006;**113**:363–75.

Kasak K, Espenberg M, Anthony TL *et al.* Restoring wetlands on intensive agricultural lands modifies nitrogen cycling microbial communities and reduces N2O production potential. *J Environ Manage* 2021;**299**:113562.

Keeley RF, Rodriguez-Gonzalez L, Class USFG *et al.* Degenerate PCR primers for assays to track steps of nitrogen metabolism by taxonomically diverse microorganisms in a variety of environments. *J Microbiol Methods* 2020;**175**:105990.

Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol* 2013;**79**:5112–20.

Kuhn M. Building Predictive Models in *R* Using the **caret** Package. *J Stat Softw* 2008;**28**, DOI: 10.18637/jss.v028.i05.

Ladau J, Eloe-Fadrosh EA. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol* 2019;**27**:662–9.

Lappalainen HK, Kerminen V-M, Petäjä T *et al.* Pan-Eurasian Experiment (PEEX): towards a holistic understanding of the feedbacks and interactions in the land–atmosphere–ocean–society continuum in the northern Eurasian region. *Atmospheric Chem Phys* 2016;**16**:14421–61.

Levin SA. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology* 1992;**73**:1943–67.

Li J-G, Shen M-C, Hou J-F *et al.* Effect of different levels of nitrogen on rhizosphere bacterial community structure in intensive monoculture of greenhouse lettuce. *Sci Rep* 2016;**6**:25305.

McCook L. Understanding ecological community succession: causal models and theories, a review. *Vegetatio* 1994;**110**:115–47.

Midolo G, Alkemade R, Schipper AM *et al.* Impacts of nitrogen addition on plant species richness and abundance: A global meta-analysis. *Glob Ecol Biogeogr* 2019;**28**:398–413.

Miller-Way T, Twilley R. Theory and operation of continuous flow systems for the study of benthic-pelagic coupling. *Mar Ecol Prog Ser* 1996;**140**:257–69.

Mitsch WJ, Bernal B, Nahlik AM *et al.* Wetlands, carbon, and climate change. *Landsc Ecol* 2013;**28**:583–97.

Mitsch WJ, Gosselink JG. The value of wetlands: importance of scale and landscape setting. *Ecol Econ* 2000;**35**:25–33.

Mueller RC, Belnap J, Kuske CR. Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. *Front Microbiol* 2015;**6**:157127.

Naumova NB, Belanov IP, Alikina TY *et al.* Undisturbed Soil Pedon under Birch Forest: Characterization of Microbiome in Genetic Horizons. *Soil Syst* 2021;**5**:14.

Naylor D, McClure R, Jansson J. Trends in Microbial Community Composition and Function by Soil Depth. *Microorganisms* 2022;**10**:540.

Nazari-Sharabian M, Ahmad S, Karakouzian M. Climate Change and Eutrophication: A Short Review. *Eng Technol Appl Sci Res* 2018;**8**:3668–72.

Nifong RL, Taylor JM, DeVilbiss S. Spatial and temporal patterns of benthic nutrient cycling define the extensive role of internal loading in an agriculturally influenced oxbow lake. *Biogeochemistry* 2022;**159**:413–33.

O'Brien SL, Gibbons SM, Owens SM *et al.* Spatial scale drives patterns in soil bacterial diversity: Spatial scale drives soil diversity. *Environ Microbiol* 2016;**18**:2039–51.

O'Neill RV, Johnson AR, King AW. A hierarchical framework for the analysis of scale. *Landsc Ecol* 1989;**3**:193–205.

Otto SA. *How to Normalize the RMSE*. https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/

Paul E, Frey S. *Soil Microbiology, Ecology and Biochemistry*. Elsevier, 2023.

Pedersen EJ, Miller DL, Simpson GL *et al.* Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 2019;**7**:e6876.

Qian L, Yu X, Gu H *et al.* Vertically stratified methane, nitrogen and sulphur cycling and coupling mechanisms in mangrove sediment microbiomes. *Microbiome* 2023;**11**:71.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2012;**41**:D590–6.

R Core Team. R: A Language and Environment for Statistical Computing. 2022.

Ranatunga T, Tong STY, Yang YJ. An approach to measure parameter sensitivity in watershed hydrological modelling. *Hydrol Sci J* 2016:1–17.

Rasche F, Cadisch G. The molecular microbial perspective of organic matter turnover and nutrient cycling in tropical agroecosystems - What do we know? *Biol Fertil Soils* 2013;**49**:251–62.

Roswell M, Dushoff J, Winfree R. A conceptual guide to measuring species diversity. *Oikos* 2021;**130**:321–38.

Schloss PD. Reintroducing mothur: 10 years later. *Appl Environ Microbiol* 2020;**86**:e02343-19.

Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.

Schneider DC. The Rise of the Concept of Scale in Ecology. *BioScience* 2001;**51**:545.

Simpson G. Introducing gratia. *Bottom Heap* 2018.

Sokol NW, Slessarev E, Marschmann GL *et al.* Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* 2022;**20**:415–30.

Speir SL, Taylor JM, Scott JT. Seasonal Differences in Relationships between Nitrate Concentration and Denitrification Rates in Ditch Sediments Vegetated with Rice Cutgrass. *J Environ Qual* 2017;**46**:1500–9.

Tang Y, Yu G, Zhang X *et al.* Changes in nitrogen-cycling microbial communities with depth in temperate and subtropical forest soils. *Appl Soil Ecol* 2018;**124**:218–28.

Thakur MP, Wright AJ. Environmental filtering, niche construction, and trait variability: the missing discussion. *Trends Ecol Evol* 2017;**32**:884–6.

Tomasek A, Staley C, Wang P *et al.* Increased Denitrification Rates Associated with Shifts in Prokaryotic Community Composition Caused by Varying Hydrologic Connectivity. *Front Microbiol* 2017;**8**:2304.

Urakawa H, Bernhard AE. Wetland management using microbial indicators. *Ecol Eng* 2017;**108**:456–76.

Veach AM, Zeglin LH. Historical Drought Affects Microbial Population Dynamics and Activity During Soil Drying and Re-Wet. *Microb Ecol* 2020;**79**:662–74.

Vellend M. Conceptual Synthesis in Community Ecology. *Q Rev Biol* 2010;**85**:183–206.

Wang C, Wang S, Wu B *et al.* Ecological restoration treatments enhanced plant and soil microbial diversity in the degraded alpine steppe in Northern Tibet. *Land Degrad Dev* 2021;**32**:723–37.

Wiens JA. Spatial Scaling in Ecology. *Funct Ecol* 1989;**3**:385.

Wood SN. Thin-plate regression splines. *J R Stat Soc B* 2003;**65**:95–114.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc B* 2011;**73**:3–36.

Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC, 2017.

Wood SN, N., Pya *et al.* Smoothing parameter and model selection for general smooth models (with discussion). *J Am Stat Assoc* 2016;**111**:1548–75.

Xu G, Ren Y, Yue M *et al.* Phosphorus sorption capacity in soils from freshwater restored coastal wetlands increased with restoration age. *Geoderma* 2022;**422**:115926.

Young JM, Skvortsov T, Kelleher BP *et al.* Effect of soil horizon stratigraphy on the microbial ecology of alpine paleosols. *Sci Total Environ* 2019;**657**:1183–93.

Yuan X, Knelman JE, Wang D *et al.* Patterns of soil bacterial richness and composition tied to plant richness, soil nitrogen, and soil acidity in alpine tundra. *Arct Antarct Alp Res* 2017;**49**:441–53.

Zedler JB, Kercher S. WETLAND RESOURCES: Status, Trends, Ecosystem Services, and Restorability. *Annu Rev Environ Resour* 2005;**30**:39–74.

CHAPTER II: COMMUNITY ASSEMBLY IN MINIMAL MEDIA REVEALS STABILIZATION IN GLUCOSE BUT NOT ACROSS AMMONIUM CONCENTRATION GRADIENT

[1]N. Reed Alexander, [1]Kylie C. Moe, [1]Donald Walker

[1]Department of Biology, Middle Tennessee State University, 1301 E Main St, Murfreesboro, TN 37132

Abstract

Predicting microbial processes of the soil microbiome in the presence of rising nutrient concentrations on the landscape is confounded by the heterogeneity of soil conditions that drive differences in microbial community composition. While established literature indicates that variation of soil microbiome composition in the field is linked to a number of abiotic and biotic parameters on the landscape such as pH and moisture availability, limited experiments have investigated the impact of nutrient gradients in the field on microbiome variation in the lab. Enrichment culture experiments using synthetic minimal media have been used to study the effects of carbon sources and organic acids on bacterial community assemblage and have shown the emergence of stabilized community states after successive transfers. To fill a knowledge gap, I sought to use soils from wetland mesocosms as source inocula to study the effects of a nitrogen gradient on bacterial community assembly using enrichment culture techniques and media with elevated concentrations of ammonium. Synthetic media containing glucose resulted in stabilized microbial communities, but did not provide evidence for an effect of ammonium concentration on the final community states. In addition, evidence for alternative stabilized community states was

observed for soils grown in the presence of glucose compared to the M9 (glucose negative) controls. This study contributed to our understanding of the assembly of complex soil bacterial communities in the lab setting and provides a foundation for further work investigating factors that influence microbial community stability and diversity.

Introduction

Food security and soil nutrient management is a growing global concern which can be mediated through appropriate conservation practices focused on how the soil microbiome can influence soil fertility (Chaparro *et al.*, 2012), nutrient availability (Falkowski *et al.*, 2008; Lappalainen *et al.*, 2016), and remediation of previously degraded landscapes (Robinson *et al.*, 2023; Prasad *et al.*, 2024). Eutrophication of global watersheds has a direct impact on water quality and food availability, as excess agricultural fertilizer runoff contributes to increased concentrations of nitrogen (N) and phosphorus (P) in the environment (Glibert, 2020). The influence of fertilizer amendments on the soil microbiome must be investigated in order to appropriately direct management strategies and conservation practices that promote healthy soils capable of sustaining agricultural productivity in the face of global climate change (Glibert *et al.*, 2014).

The soil microbiome plays a significant role shaping both the environments they inhabit and the outcomes of a multitude of ecosystem processes they contribute to (Falkowski *et al.*, 2008). The ability to predict these processes remains limited despite the increase in accumulated microbiome data (Hall *et al.*, 2018). Recent work posits that microbial ecology and the environmental outcomes soil microbiomes contribute to can be better understood by delineating microbial characteristics into *microbial membership*, measured by sequencing phylogenetic markers; *community properties*, such as microbial community biomass or quantification of functional gene abundance; direct measures of *microbial processes*, such as nutrient cycling rates;

and investigating the connections between these three categories (Hall *et al.*, 2018). In order to link microbial communities to the processes they contribute to, we must first improve our understanding of how spatial and environmental gradients influence microbial community composition and assembly (Martiny *et al.*, 2011; Goldford *et al.*, 2018).

Environmental heterogeneity complicates studying soil microbiomes and their influence on environmental process rates *in situ*, as soil microbial community composition and function are directly influenced by the inherent properties of the environments they inhabit, climatic conditions they experience, and anthropogenic factors such as land-use practices (Rasche and Cadisch, 2013; O'Brien *et al.*, 2016; Gupta *et al.*, 2022). In order to appropriately predict microbial community membership and its influence on microbial processes in the environment, experiments must be designed that reduce the influence of environmental heterogeneity, while investigating key gradients that contribute to community composition and variation. Soil conditions including pH, temperature, and nutrient availability have all been shown to influence microbial community membership (Fierer, 2017; Dai *et al.*, 2020; Islam *et al.*, 2020), which can be controlled in lab-based experiments. The soil microbiome consists of bacteria, fungi, archaea, protists, and viruses, which all interact and contribute to the collective metabolism that defines specific microbial processes in the environment (Hall *et al.*, 2018). Predictive models of microbial processes can be improved by studying the forces that drive specific assembly patterns and directly influence the composition and variation of the soil microbiome (Balser *et al.*, 2013; Urakawa and Bernhard, 2017; Correa-Garcia *et al.*, 2023).

High-throughput methodologies have been developed to investigate the ecological processes that contribute to microbiome composition and assembly in lab settings, known as enrichment community experiments (Estrela *et al.*, 2021a). The foundation of these experiments

is the use of synthetic media, whose nutrient compositions can be controlled, quantified, and manipulated to investigate the influence of carbon source (Goldford *et al.*, 2018), organic acids (Estrela *et al.*, 2021b), and multifactorial combinations of co-limiting nutrients (Bello *et al.*, 2020) on the composition of microbial communities.

Community based enrichment culture experiments have revealed the existence of rules of community assembly which can be experimentally tested, quantified, and verified (Estrela *et al.*, 2021a, 2022). These experiments have shown that microbial communities seeded from diverse inocula sources exhibit convergence of family-level composition measured by relative abundance, and that the landscape of community structure can be navigated such that the trajectory of convergence is dependent on the nutrient composition of the medium (Goldford *et al.*, 2018; Bello *et al.*, 2020; Estrela *et al.*, 2022). Another notable observation from these experiments is the emergent metabolic organization of these communities, in which composition of stabilized communities may be directed by nutrient utilization and potential cross-feeding on secondary metabolites produced by other community members (Goldford *et al.*, 2018). However, other experiments indicate the potential persistence of alternative stable states of family-level convergence and functional outcomes based on initial abundances of microbes in source inocula (Bittleston *et al.*, 2020).

Soil amendments typically contain N in the form of ammonium-nitrate (NH_4NO_3) to improve agricultural output, which is cycled by bacteria into bioavailable forms (Glibert *et al.*, 2014). Ammonium (NH_4) and nitrate (NO_3) are transformed by functionally distinct bacteria in separate steps of the N-cycle (Jetten, 2008), and bacterial community composition has been shown to respond differentially to these nutrient amendments (Zhang *et al.*, 2021). The study presented in the previous chapter (Alexander *et al.* 2025) suggests that overall alpha diversity declines with

increasing N in field-collected and lab-incubated wetland soils (Chapter 2), and NH_4-N has been shown to be the dominant predictor of bacterial composition in some soil environments (Nie *et al.*, 2018). The goal of this study was to investigate the influence of NH_4-N concentration on bacterial community assembly using enrichment community experiments that utilized synthetic media representing a gradient of increasing NH_4 concentration. My primary objective was to characterize bacterial community composition using 16S rRNA metabarcoding and amplicon sequencing, and determine if final stabilized microbial communities would be altered along the supplied gradient. I hypothesized that: 1) ammonium concentration would alter stabilized communities in a deterministic manner, 2) ammonium concentration would be negatively correlated to alpha diversity of stabilized communities, and 3) source inocula grown in media that lacked a carbon source would exhibit stochastic loss of community members, resulting in greater dispersion of samples over time.

Methods

Mesocosm construction and sample collection

Wetland mesocosms containing a monoculture of rice cutgrass (*Leersia oryzoides*) were constructed in June 2021 and maintained outside at Tennessee Tech University's Shipley Farm, where they served as the source inocula of microbial communities. Mesocosms (n = 12) were established in 379 L poly stock tanks (Behlen Country, Baker City, OR) according to established methods (Tyler *et al.*, 2012; Taylor *et al.*, 2015) by filling with ~ 30 cm of silica sand and ~ 15 cm of wetland soil sourced from a stream restoration site by the West Tennessee River Basin Authority. Mesocosm topsoil was planted with *L. oryzoides* collected from a riparian buffer near Little Creek (Cookeville, TN) in June, 2021 (Womble, 2023). Wetland mesocosms were allowed

to establish over a one-month period, followed by a flooding regime consisting of alternating three-day periods of flooding and drying for one month, beginning in July.

Soil samples were collected from the center of each mesocosm (n = 12) using pre-sterilized sample rods (1 cm diameter, ~ 5 cm depth; described in chapter II) at the conclusion of the one-month flooding regime in August 2021. Samples were added to individual pre-sterilized 15 mL centrifuge tubes (VWR International, Radnor, PA) containing a 15% glycerol solution. Samples were placed on ice upon collection and transported to Middle Tennessee State University (Murfreesboro, TN) where they were frozen at -80º C until further processing. A biological control sample was also collected during soil sampling using a sterile sampling rod that was dipped into and then removed from the 15% glycerol solution of an independent pre-sterilized 15 mL centrifuge tube and treated identically to soil samples throughout sample collection. Four glycerol-preserved samples consisting of three soil samples chosen using a random number generator (MES 7, MES 19, MES 22) and one biological negative control (glycerol flask with no soil) were selected for downstream experimentation.

Experimental Media and Plate Preparation

Experimental media (n = 8) was prepared using Difco M9 Minimal Salts (BD Biosciences, Milpitas, CA) and trace minerals ($MgSO_4$ and $CaCL_2$) and increasing concentrations of ammonium chloride (NH_4Cl) according to Table 4. For each concentration of NH_4Cl (1X-4X), media was prepared both with a single carbon source (0.2% glucose) and without a carbon source as a control. Stock 100 mL solutions of the 8 different culture media types were prepared concurrently and filter-sterilized with a 0.2 μm filter (Nalgene Rapid-Flow, Thermo Scientific). Each media type (500 μl) was added to four sterile 2 ml microcentrifuge tubes, one for each combination of media and source inocula (n = 32). Culture media was then added to sterile non-treated 96-well flat

bottomed assay plates (Costar; Corning Inc, Kennebunk, ME) in 250 µl aliquots using liquid handling robots (Integra LifeScience, Princeton, NJ) according to Fig. 11, the lid set in place, then edges sealed with parafilm (n = 16 total plates). All microcentrifuge tubes and experimental plates were frozen at -20 °C prior to the start of experimentation. All media and plate preparation was conducted in a BSL-2 hood.

Table 4. Stock solution volumes used to prepare the eight experimental media types. 1X AMO represents a 1X M9 solution with the addition of trace minerals, magnesium sulfate ($MgSO_4$) and calcium chloride ($CaCL_2$). Six additional media types are generated with increasing volumes of dissolved ammonium chloride (NH_4Cl) such that 2X, 3X, and 4X solutions represent the concentration of ammonium relative to stock 1X M9. GLU represents the addition of 0.2% glucose as the sole carbon source, control media is also prepared for each concentration of NH_4Cl that does not include the glucose carbon source. All stock solutions were filter sterilized prior to media preparation.

Media Type	5x M9 (ml)	10X NH_4Cl (ml)	10% Glucose (ml)	1M $MgSO_4$ (µL)	1M $CaCl_2$ (µL)	ddH_2O (ml)
1X AMO	20	0	0	200	10	79.79
1X AMO + GLU	20	0	2	200	10	77.79
2X AMO	20	10	0	200	10	69.79
2X AMO + GLU	20	10	2	200	10	67.79
3X AMO	20	20	0	200	10	59.79
3X AMO + GLU	20	20	2	200	10	57.79
4X AMO	20	30	0	200	10	49.79
4X AMO + GLU	20	30	2	200	10	47.79

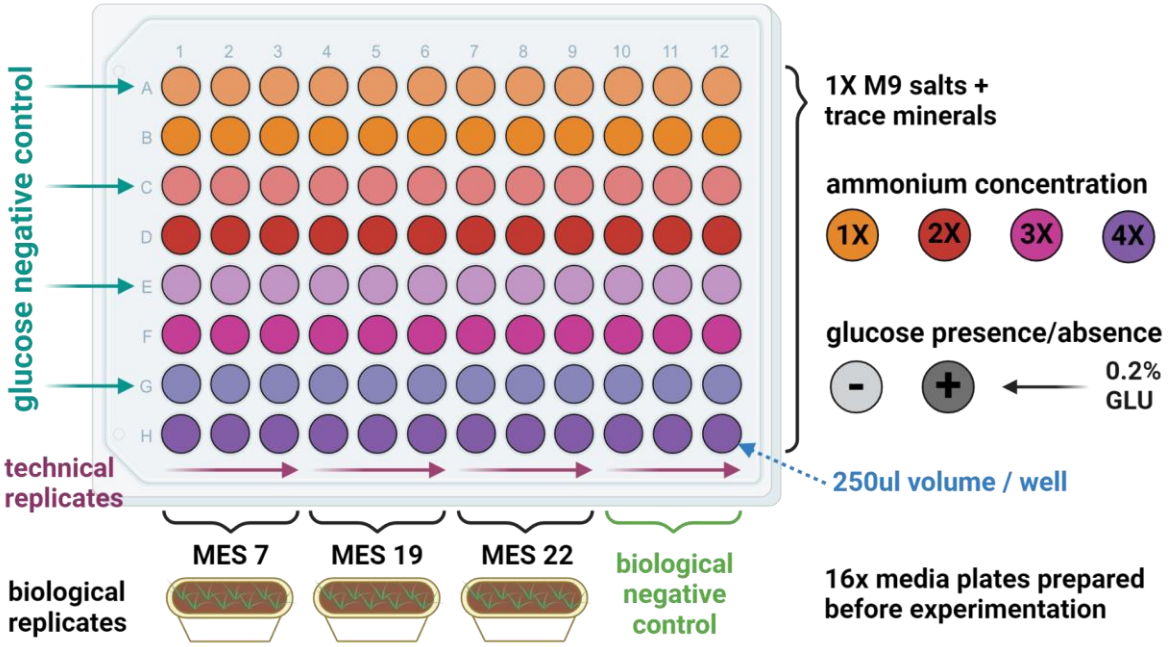


Figure 11. Experimental design and 96-well plate layout of culture media. M9 minimal culture media is depicted in colors based on increasing concentration of ammonium (1X - 4X; down the plate) with alternating rows of glucose negative controls. Technical replicates are shown across columns in sets of three for each biological replicate and one biological negative control.

Source Inoculum Preparation

Glycerol preserved soil samples were allowed to thaw to room temperature prior to homogenization using a digital vortexer (3000 rpm, 40s; VWR International, Radnor PA). As this study focused on bacterial assemblages, fungal species were eliminated from the soil samples, using a 1 L stock solution of phosphate buffered saline (1X PBS, 10 g/L; Ward's Science+, Rochester, NY) with dissolved crystalline cycloheximide (200 µg/ml, 0.2 g/L; VWR Life Science, Radnor PA) that was prepared by filter-sterilization (0.2 µm filter; Nalgene Rapid-Flow, Thermo Scientific). A 5 ml subsample of each source inoculum was then added to individual pre-sterilized 250 ml erlenmeyer flasks (n = 4) containing 50 ml of cycloheximide supplemented 1X PBS solution. Flasks were then covered with aluminum foil, swirled to mix, and incubated at room

temperature in the dark for 48 hours. Supernatant was collected from each flask after incubation and placed into independent 50 ml falcon tubes to serve as the microbial source inoculum. All transfers were conducted in a BSL-2 hood.At the conclusion of the 48 hour incubation period, microcentrifuge tubes containing media (n = 32) and one experimental plate were thawed at room temperature. Microbial source inocula (n = 4) were inverted to mix and 40 µl of each was used to inoculate 500 µl of each media type (n = 32; 4x microbial inocula & 8x culture media). Microcentrifuge tubes were inverted to mix and 2 µl aliquoted to inoculate three independent wells of the respective media type, generating three technical replicates, for each microbial source based on the experimental plate design (Fig. 11). The remainder of inoculated media in microcentrifuge tubes was then transferred to a separate sterile 96-well plate in three 100 µl aliquots using an identical plate layout and to serve as the initial experimental time point and spun down at 14,000 RPM for 10 minutes before freezing at -80 °C. To preserve microbial source inocula, sterile 40% glycerol aliquots equal to the remaining volumes were added to both the microbial source inocula and the inoculated microcentrifuge tubes prior to inverting to mix and freezing at -80 °C. The inoculated experimental plate was sealed with parafilm and moved to incubate in the dark at 30 °C for 48 hrs under static conditions (ECHOtherm, Torrey Pines Scientific, Carlsbad, CA).

Serial passaging and experimental measures

The first 96-well experimental plate was removed from the incubator after 48 hours and allowed to cool to room temperature in a BSL2-hood to reduce condensation on the lid. Measurements of turbidity (OD620) were collected using a CLARIOstar microplate reader (BMG Labtech, Cary, NC). While the incubated plate was equilibrating to room temperature, a fresh 96-well experimental plate (containing media) was thawed to room temperature. Each well of the actively growing plate was then homogenized by mixing 10 times using a Integra Mini 96 liquid

handling robot and 2 µl from each well was used to inoculate the freshly thawed plate of new media. The newly inoculated plate was sealed in parafilm and placed into the incubator representing the second passage event. Following processing of the fresh plate of media, 100 µl from the incubated plate was added to a sterile 96-well plate and spun down (14,000 RPM; 10 min). A fluorometric quantitative ammonium assay (Sigma-Aldrich, Burlington, MA) was performed using 40 µl of supernatant collected from the spun down plate and measured on a CLARIOstar microplate reader. The pelleted cells were then frozen at -80 °C for DNA processing. Enrichment communities at the end of the incubation period were then preserved by homogenizing the samples in a sterile 40% glycerol solution and frozen down at -80 °C. This process was repeated for 13 serial passage events over a 26 day period producing 14 plates of microbial samples (n = 1,344).

DNA extraction, library preparation, and high-throughput sequencing

Four time points were selected for processing including the initial microbial source inocula and the microbial communities collected after the first, third, and fifth serial passage events (n = 384). Selected plates were DNA extracted using the DNeasy PowerSoil HTP 96 kit (Qiagen, Hilden, Germany), and a single well was left blank on each 96-well plate and treated equally throughout the extraction process to control for contamination in downstream bioinformatics. Libraries were prepared according to Illumina 16S Metagenomic Sequencing Library Preparation protocol for the MiSeq instrument (Illumina, San Diego, CA). Extracted DNA was PCR amplified using 2 µL of DNA in 25 µl reactions with 12.5 uL MCLAB I-5™ 2X Hi-Fi PCR Master Mix, 1 µL 515F primer (10 µM), 1 µL 806R primer (10 µM) (Caporaso *et al.* 2010) and 8.5 µL of PCR water. Sample plates were prepared in pairs by dual indexing the V4 region of 16S rRNA (Kozich *et al.* 2013) with unique index combinations for each plate. Purification of samples was performed

using HighPrep PCR Clean Up magnetic beads (MagBio Genomics Inc., Gaithersburg, MD) after both amplicon PCR and indexing PCR. Sample concentration of PCR products was measured using a Quantus Fluorometer (Promega Corp., Madison, WI), and libraries were normalized to 4.0 nanomolar concentration using 1% Tris buffer prior to sample pooling and spiking of 10% PhiX control. Plate pairs of processed libraries were loaded on an Illumina MiSeq V2 flow cell and sequenced using a 500-cycle reagent kit (2 x 250 bp paired-end reads) in two sequencing runs.

Bioinformatics

Raw sequences were demultiplexed, aligned, and processed using the MiSeq SOP bioinformatics pipeline (Kozich *et al.* 2013) using mothur v1.48.0 (Schloss *et al.*, 2009; Schloss, 2020) for quality control, processing, and clustering of amplicon sequence data into operational taxonomic units (OTUs) at 97% sequence similarity (Grajal-Puche *et al.* 2020). The SILVA v138 reference database was used to assign taxonomy to OTUs (Quast *et al.* 2012). Prior to decontamination, rare OTUs were removed if they did not meet a minimum abundance threshold of 10 reads across all samples (Brown *et al.* 2015; Cao *et al.* 2021). Contaminating sequences were identified with prevalence based algorithmic detection with a probability threshold of 0.262 using the *decontam* package in R (Davis *et al.* 2018). After decontamination, rare OTUs were removed if they did not meet an abundance threshold of 10 reads across all samples (Brown *et al.* 2015). Samples were arranged according to sequencing depth and rarefied to normalize sample coverage, and samples that fell below the rarefaction depth were sampled with replacement (25,520 sequence reads / sample). After preliminary analysis, biological negative controls were removed from the sample pool for more accurate statistical measures between treatments. The final OTU abundance table contained 3,626 bacterial OTUs from 288 samples.

Statistical analysis

Modeling Assemblage Diversity and Taxonomic Composition

Bacterial taxonomic diversity (α-diversity) and composition (β-diversity) were compared across carbon source (glucose presence/absence), passage event, and ammonium concentration. Statistical analysis was performed in R version 4.4.2 (R Core Team 2024). Taxonomic diversity was assessed with measures of OTU richness. The relationship between enrichment community α-diversity and media treatments was evaluated using generalized additive mixed models (GAMM, with R package '*mgcv*') to model nonlinear trends while accounting for random effects (Wood 2011, 2017). The main effect of passage events on community richness was tested for, while including factor smoothing terms for ammonium concentration, glucose, and the three-way interaction between the two factor smooths across passage events. Random effects terms were included for biological replicates and well ID, to control for variation in inoculum source, and sample location across 96-well experimental plates respectively. A location-scaled t-distribution with an identity link function was selected after AIC selection (Pedersen *et al.* 2019). Basis dimensions for each smoothing term were evaluated and the final model was fit using restricted maximum likelihood, followed by validation of residual plots (Simpson 2018). Enrichment community composition across taxonomic groups was investigated by aggregating OTUs to the family level. To make comparisons between glucose treatment and passage events a Kruskal-Wallis test (R package '*stats*') was conducted, followed by a Dunn's test (R package '*rstatix*') with post-hoc Bonferroni correction to assess pairwise differences simultaneously.

Beta Diversity

β-diversity across enrichment communities was evaluated using three dissimilarity metrics (R package '*vegan*') (Oksanen *et al.* 2007). Bray-Curtis was used as an abundance based metric

which is influenced by differences in richness. To account for differences in alpha diversity, the Raup-Crick metric was utilized, as it provides a null expectation of shared OTUs to account for differences in Richness between communities (Chase *et al.* 2011). The Sørensen metric which is a presence/absence form of Bray-Curtis was also used. To determine homogeneity of variance, a permutational test of multivariate dispersion was conducted across treatment groups, followed by Tukey's honestly significant difference to test for significant differences in dispersion. Permutational multivariate ANOVAs (PerMANOVAs) were conducted to test for differences in multivariate centroid position between glucose treatment and time (R package '*vegan*'). Post-hoc pairwise PerMANOVAs were conducted for the factor combination of glucose treatment and passage event to further test for differences in multivariate centroid position after a significant interaction was detected. Assemblage composition was visualized using non-metric multidimensional scaling (NMDS) of Bray-Curtis, Raup-Crick, and Sørensen dissimilarities in the first two dimensions (R package '*ggplot2*').

Results

Influence of media type on enrichment community richness across passage events

On average, measured richness declined over the course of the experiment (Fig. 12, Table 5, GAMM, Deviance explained = 66%). Significant partial effects were detected for the smooths of passage events (GAMM, *e.d.f.* = 2.895, $F = 365.388$, $P < 0.0001$), time by glucose treatment (Fig. 12a, GAMM, *e.d.f.* = 2.803, $F = 216.063$, $P < 0.0001$), and the two factor glucose-ammonium interaction smooths with time (Fig. 12c, GAMM, *e.d.f.* = 2.471, $F = 7.466$, $P < 0.05$), but not for the smooth of ammonium treatment by time (Fig. 12b). Partial effects smooth differences for glucose treatment revealed that OTU richness declined between the initial time point and the 48 hour sampling event for both glucose positive and glucose negative treatments. Glucose negative

treatments experienced continued declines in richness across subsequent sampling events, while the richness of glucose positive treatments was maintained between ~175 - 200 OTUs for the remainder of the experiment (Fig. 12a). The significant two factor interaction of glucose-ammonium along the duration of the experiment indicated that increasing richness was maintained at greater concentrations of ammonium for glucose positive samples, and that richness declined quicker as ammonium concentration increased in glucose negative samples (Fig. 12c).

Table 5. GAMM results of Richness model. ***SE***; Standard Error, ***e.d.f.***; expected degrees of freedom, ***Ref d.f.***; reference degrees of freedom.

Component	**Term**	**Estimate**	***SE***	***T***	***P***	
A. parametric coefficients	(Intercept)	176.911	7.906	22.378	0.0000	***
Component	**Term**	***e.d.f.***	***Ref d.f.***	***F***	***P***	
B. smooth terms	s(Time)	2.895	3	365.388	0.0000	***
	s(Time, Ammonium)	2.455	9	6.222	0.0670	.
	s(Time, Glucose)	2.803	3	216.063	0.0000	***
	s(Time, Ammonium, Glucose)	2.471	9	7.466	0.0221	*
	s(Source Inoculum)	1.846	2	38.725	0.0000	***
	s(Well ID)	14.681	71.000	19.640	0.0296	*

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.681, Deviance explained 0.660

-REML : 1443.713, Scale est: 1.000, N: 288

Family: Scaled t (12.954 , 30.679), Link function: Identity

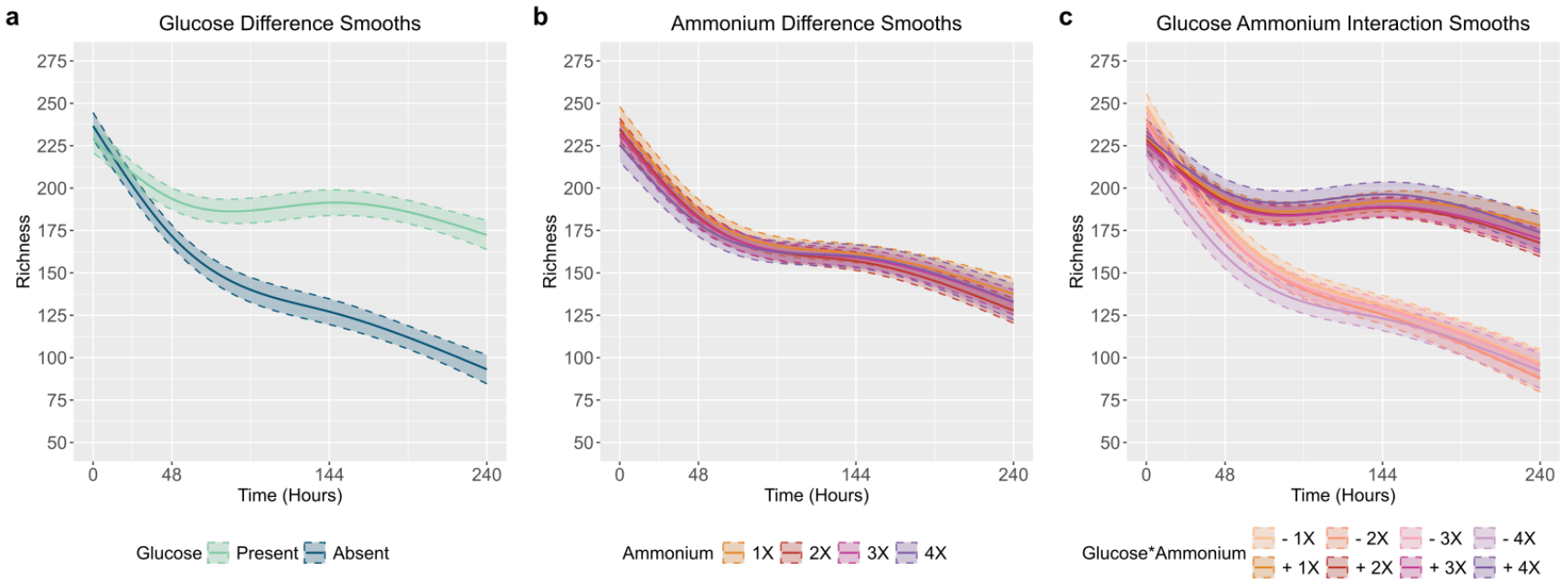


Figure 12. OTU Richness declines over the course of passage events. Significant GAMM model Partial Effects Difference smooths are shown with Bayesian confidence intervals for carbon source (glucose 0.2%) colored by presence and absence (a) and ammonium treatments colored by concentration (b), and two-factor interaction smooths glucose and ammonium treatment combinations (c).

Assemblage Relative Abundance

Taxonomic relative abundance by glucose treatment and passage event showed statistically significant differences across all groups at the taxonomic rank of family for the global Kruskal-Wallis test (Fig. 13, Table 6). No family level differences were detected between glucose positive and glucose negative samples for the initial time point, indicating that all samples started with similar proportions of family composition (Table 7). At the first transfer (48hr), differences in relative abundance by glucose treatments became evident for a majority of bacterial families (n = 7/11 bacterial families), and these differences were maintained for the third (144hr; n = 7/11 bacterial families) and fifth (240hr n = 6/11 bacterial families) transfers (Table 7). Glucose negative samples became dominated by *Pseudomonadaceae* after the first transfer and remained consistently dominant over the remaining passages.

Comparing glucose positive treatments between time points, samples experienced significant shifts in all family groups except for *Pseudomonadaceae* (n = 10/11 bacterial families) between the initial time point and the first transfer (48hrs). Between the first transfer (48hr) and third transfer (144hr), the relative abundance of only two families shifted significantly: *Pseudomonadaceae* and *Aeromonadaceae*. Glucose positive samples between the third (144hr) and fifth (240hr) passage event experienced no significant differences in relative abundance of family composition. The family level composition of final glucose positive samples was dominated by *Pseudomonadaceae* and *Enterobacteriaceae*, with the exception of one biological sample (MES 19) which maintained an elevated relative abundance of *Aeromonadaceae* between transfers compared to other glucose positive samples. Additional pairwise comparisons of relative abundance of family composition are available in Table 7.

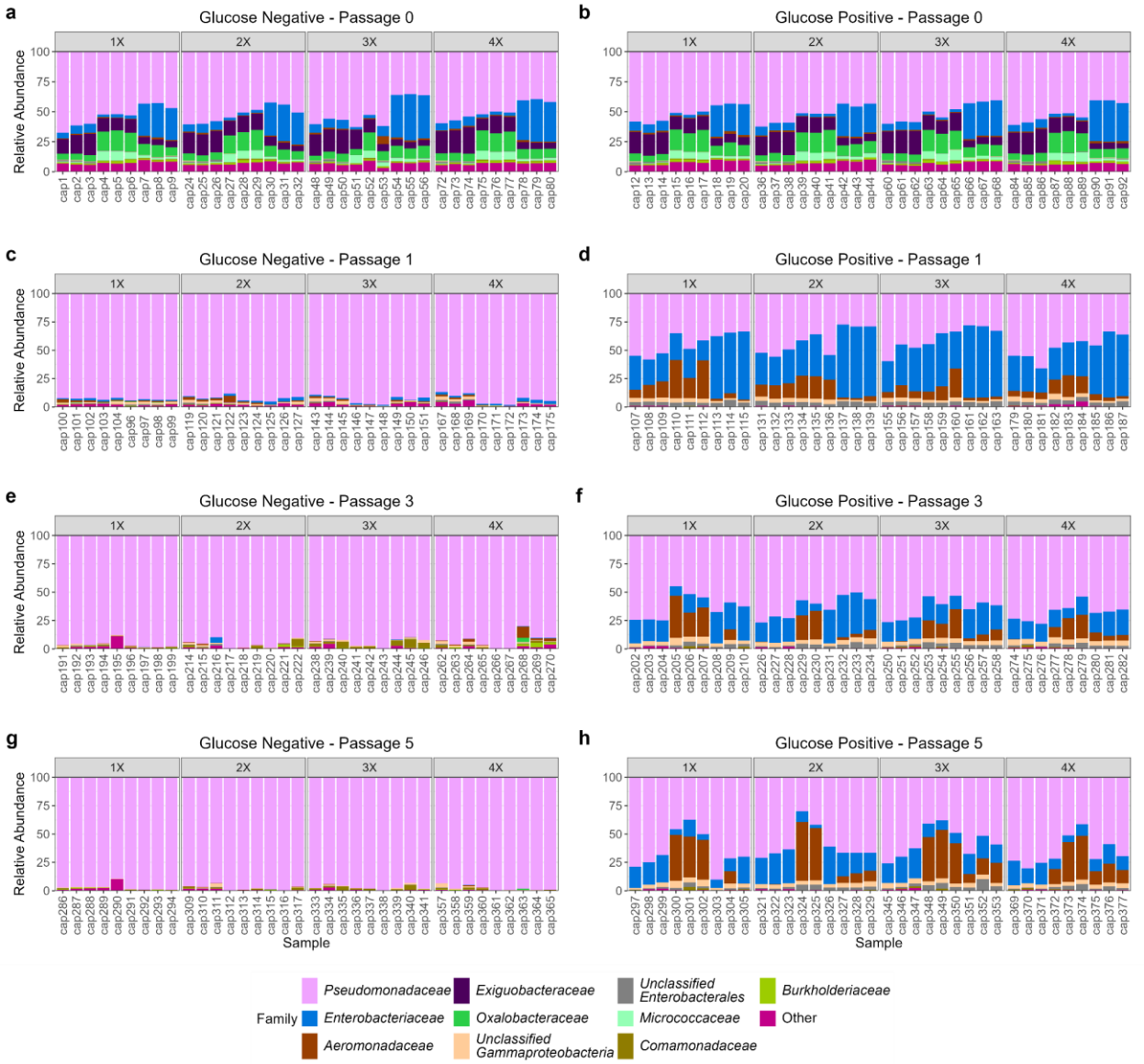


Figure 13. Taxonomic relative abundance by sample. Passage events are separated by rows (passages 0 - 5), and grouped columns within passage events represent ammonium concentration of media (1X-4X). Within each ammonium treatment level samples are arranged in triplicate for each biological replicate, representing MES 7, MES19, and MES22 respectively. Enrichment communities grown in media without a carbon source are shown on the left (a, c, e, g), and communities grown with a carbon source (Glucose 0.2%) are shown on the right (b, d, f, h).

Table 6. Summary statistics of relative abundance for family level taxonomic groups. Mean and standard deviation are calculated based on samples aggregated by glucose treatment and passage event. ***SD*;** Standard Deviation.

Glucose - Passage	Negative - 0		Positive - 0		Negative - 1		Positive - 1		Negative - 3		Positive - 3		Negative - 5		Positive - 5	
Taxonomic Group	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Pseudomonadaceae	51.452	8.461	51.152	7.003	92.962	2.775	43.041	10.466	94.531	4.248	64.273	9.047	97.229	2.270	61.928	14.255
Enterobacteriaceae	13.845	13.068	12.524	11.619	1.558	0.841	39.099	15.125	0.203	0.854	19.880	7.194	0.010	0.022	16.085	7.514
Aeromonadaceae	1.135	1.006	1.091	0.321	1.093	1.024	10.461	8.597	0.630	1.705	8.086	10.047	0.070	0.151	14.685	17.037
Exiguobacteraceae	11.617	5.509	11.931	5.256	0.001	0.003	0.001	0.002	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Oxalobacteraceae	9.505	4.295	10.531	4.740	0.069	0.114	0.003	0.004	0.112	0.610	0.000	0.000	0.028	0.152	0.000	0.001
Unclassified Gammaproteobacteria	0.216	0.126	0.205	0.159	1.448	0.885	2.872	1.016	0.903	0.752	4.109	0.869	0.569	0.803	3.792	0.897
Unclassified Enterobacterales	0.663	0.572	0.695	0.644	0.445	0.517	3.638	0.687	0.168	0.302	2.478	1.364	0.083	0.142	2.448	1.884
Micrococcaceae	2.973	2.701	3.105	2.720	0.006	0.010	0.001	0.002	0.007	0.033	0.000	0.000	0.002	0.013	0.000	0.001
Comamonadaceae	0.359	0.261	0.285	0.197	0.193	0.167	0.057	0.099	1.403	1.940	0.343	0.610	0.691	0.972	0.353	0.637
Burkholderiaceae	1.331	0.520	1.474	0.630	0.095	0.064	0.005	0.008	0.538	0.744	0.001	0.003	0.226	0.209	0.002	0.010
Other	6.904	1.244	7.008	1.534	2.129	1.257	0.822	0.841	1.505	1.963	0.832	0.542	1.092	1.709	0.707	0.466

Table 7.1. Dunn's Test results of relative abundance of family level taxonomic groups. Comparisons are made for pairwise combinations of glucose treatments and passage events. Significant *P* values are depicted in bold.

	Pseudomonadaceae		*Enterobacteriaceae*		*Aeromonadaceae*		*Exiguobacteraceae*	
Pairwise Comparison	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***
-_0 / +_0	-0.149	1.000	-0.390	1.000	0.294	1.000	0.159	1.000
-_0 / -_1	6.964	**0.000**	-4.073	**0.001**	-0.427	1.000	-7.544	**0.000**
-_0 / +_1	-1.673	1.000	4.206	**0.001**	4.384	**0.000**	-7.224	**0.000**
-_0 / -_3	7.830	**0.000**	-6.472	**0.000**	-3.445	**0.016**	-9.217	**0.000**
-_0 / +_3	2.727	0.179	1.490	1.000	0.902	1.000	-9.397	**0.000**
-_0 / -_5	8.987	**0.000**	-7.432	**0.000**	-5.495	**0.000**	-9.397	**0.000**
-_0 / +_5	2.099	1.000	0.730	1.000	1.212	1.000	-9.397	**0.000**
+_0 / -_1	7.112	**0.000**	-3.683	**0.006**	-0.721	1.000	-7.703	**0.000**
+_0 / +_1	-1.524	1.000	4.596	**0.000**	4.091	**0.001**	-7.383	**0.000**
+_0 / -_3	7.978	**0.000**	-6.082	**0.000**	-3.739	**0.005**	-9.377	**0.000**
+_0 / +_3	2.875	0.113	1.880	1.000	0.609	1.000	-9.556	**0.000**
+_0 / -_5	9.136	**0.000**	-7.042	**0.000**	-5.789	**0.000**	-9.556	**0.000**
+_0 / +_5	2.247	0.690	1.120	1.000	0.918	1.000	-9.556	**0.000**
-_1 / +_1	-8.636	**0.000**	8.279	**0.000**	4.812	**0.000**	0.320	1.000
-_1 / -_3	0.866	1.000	-2.399	0.461	-3.018	0.071	-1.673	1.000
-_1 / +_3	-4.237	**0.001**	5.563	**0.000**	1.330	1.000	-1.853	1.000
-_1 / -_5	2.024	1.000	-3.358	**0.022**	-5.068	**0.000**	-1.853	1.000
-_1 / +_5	-4.865	**0.000**	4.803	**0.000**	1.640	1.000	-1.853	1.000
+_1 / -_3	9.502	**0.000**	-10.677	**0.000**	-7.830	**0.000**	-1.993	1.000
+_1 / +_3	4.399	**0.000**	-2.716	0.185	-3.482	**0.014**	-2.173	0.834
+_1 / -_5	10.660	**0.000**	-11.637	**0.000**	-9.880	**0.000**	-2.173	0.834
+_1 / +_5	3.771	**0.005**	-3.476	**0.014**	-3.172	**0.042**	-2.173	0.834
-_3 / +_3	-5.103	**0.000**	7.962	**0.000**	4.348	**0.000**	-0.179	1.000
-_3 / -_5	1.158	1.000	-0.960	1.000	-2.050	1.000	-0.179	1.000
-_3 / +_5	-5.731	**0.000**	7.202	**0.000**	4.658	**0.000**	-0.179	1.000
+_3 / -_5	6.260	**0.000**	-8.921	**0.000**	-6.398	**0.000**	0.000	1.000
+_3 / +_5	-0.628	1.000	-0.760	1.000	0.310	1.000	0.000	1.000
-_5 / +_5	-6.889	**0.000**	8.161	**0.000**	6.708	**0.000**	0.000	1.000

Table 7.2. Dunn's Test results of relative abundance of family level taxonomic groups. Comparisons are made for pairwise combinations of glucose treatments and passage events. Significant *P* values are depicted in bold.

	Oxalobacteraceae		*Unclassified Gammaproteobacteria*		*Unclassified Enterobacterales*		*Micrococcaceae*	
Pairwise Comparison	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***
-_0 / +_0	0.286	1.000	-0.342	1.000	-0.024	1.000	-0.039	1.000
-_0 / -_1	-4.240	**0.001**	4.625	**0.000**	-1.595	1.000	-6.127	**0.000**
-_0 / +_1	-7.234	**0.000**	7.361	**0.000**	6.082	**0.000**	-8.057	**0.000**
-_0 / -_3	-8.262	**0.000**	2.740	0.172	-3.843	**0.003**	-7.873	**0.000**
-_0 / +_3	-9.485	**0.000**	9.620	**0.000**	4.148	**0.001**	-9.535	**0.000**
-_0 / -_5	-8.941	**0.000**	1.262	1.000	-4.613	**0.000**	-9.146	**0.000**
-_0 / +_5	-9.345	**0.000**	9.003	**0.000**	3.829	**0.004**	-9.350	**0.000**
+_0 / -_1	-4.527	**0.000**	4.967	**0.000**	-1.571	1.000	-6.088	**0.000**
+_0 / +_1	-7.520	**0.000**	7.704	**0.000**	6.106	**0.000**	-8.019	**0.000**
+_0 / -_3	-8.549	**0.000**	3.082	0.058	-3.819	**0.004**	-7.835	**0.000**
+_0 / +_3	-9.772	**0.000**	9.962	**0.000**	4.172	**0.001**	-9.497	**0.000**
+_0 / -_5	-9.227	**0.000**	1.605	1.000	-4.589	**0.000**	-9.108	**0.000**
+_0 / +_5	-9.631	**0.000**	9.345	**0.000**	3.853	**0.003**	-9.312	**0.000**
-_1 / +_1	-2.993	0.077	2.737	0.174	7.677	**0.000**	-1.930	1.000
-_1 / -_3	-4.022	**0.002**	-1.885	1.000	-2.248	0.688	-1.746	1.000
-_1 / +_3	-5.245	**0.000**	4.995	**0.000**	5.743	**0.000**	-3.408	**0.018**
-_1 / -_5	-4.701	**0.000**	-3.362	**0.022**	-3.018	0.071	-3.019	0.071
-_1 / +_5	-5.104	**0.000**	4.378	**0.000**	5.424	**0.000**	-3.223	**0.035**
+_1 / -_3	-1.029	1.000	-4.622	**0.000**	-9.925	**0.000**	0.184	1.000
+_1 / +_3	-2.252	0.682	2.258	0.670	-1.934	1.000	-1.478	1.000
+_1 / -_5	-1.707	1.000	-6.099	**0.000**	-10.695	**0.000**	-1.089	1.000
+_1 / +_5	-2.111	0.974	1.642	1.000	-2.253	0.679	-1.293	1.000
-_3 / +_3	-1.223	1.000	6.880	**0.000**	7.991	**0.000**	-1.662	1.000
-_3 / -_5	-0.679	1.000	-1.477	1.000	-0.770	1.000	-1.273	1.000
-_3 / +_5	-1.082	1.000	6.263	**0.000**	7.672	**0.000**	-1.477	1.000
+_3 / -_5	0.544	1.000	-8.358	**0.000**	-8.760	**0.000**	0.389	1.000
+_3 / +_5	0.141	1.000	-0.617	1.000	-0.318	1.000	0.185	1.000
-_5 / +_5	-0.404	1.000	7.741	**0.000**	8.442	**0.000**	-0.204	1.000

Table 7.3. Dunn's Test results of relative abundance of family level taxonomic groups. Comparisons are made for pairwise combinations of glucose treatments and passage events. Significant *P* values are depicted in bold.

	Comamonadaceae		*Burkholderiaceae*		*Other*	
Pairwise Comparison	***H-value***	***P-value***	***H-value***	***P-value***	***H-value***	***P-value***
-_0 / +_0	-0.594	1.000	0.232	1.000	-0.041	1.000
-_0 / -_1	-2.092	1.000	-5.481	**0.000**	-4.361	**0.000**
-_0 / +_1	-5.536	**0.000**	-9.196	**0.000**	-8.156	**0.000**
-_0 / -_3	0.713	1.000	-3.081	0.058	-6.843	**0.000**
-_0 / +_3	-3.966	**0.002**	-10.101	**0.000**	-7.618	**0.000**
-_0 / -_5	-0.803	1.000	-4.439	**0.000**	-7.887	**0.000**
-_0 / +_5	-3.630	**0.008**	-10.141	**0.000**	-8.334	**0.000**
+_0 / -_1	-1.498	1.000	-5.713	**0.000**	-4.320	**0.000**
+_0 / +_1	-4.942	**0.000**	-9.428	**0.000**	-8.115	**0.000**
+_0 / -_3	1.307	1.000	-3.313	**0.026**	-6.802	**0.000**
+_0 / +_3	-3.372	**0.021**	-10.333	**0.000**	-7.577	**0.000**
+_0 / -_5	-0.209	1.000	-4.671	**0.000**	-7.846	**0.000**
+_0 / +_5	-3.036	0.067	-10.373	**0.000**	-8.293	**0.000**
-_1 / +_1	-3.444	**0.016**	-3.715	**0.006**	-3.795	**0.004**
-_1 / -_3	2.805	0.141	2.400	0.460	-2.482	0.366
-_1 / +_3	-1.873	1.000	-4.620	**0.000**	-3.258	**0.031**
-_1 / -_5	1.289	1.000	1.042	1.000	-3.526	**0.012**
-_1 / +_5	-1.538	1.000	-4.661	**0.000**	-3.974	**0.002**
+_1 / -_3	6.249	**0.000**	6.115	**0.000**	1.313	1.000
+_1 / +_3	1.570	1.000	-0.905	1.000	0.538	1.000
+_1 / -_5	4.733	**0.000**	4.757	**0.000**	0.269	1.000
+_1 / +_5	1.906	1.000	-0.945	1.000	-0.178	1.000
-_3 / +_3	-4.679	**0.000**	-7.020	**0.000**	-0.775	1.000
-_3 / -_5	-1.517	1.000	-1.358	1.000	-1.044	1.000
-_3 / +_5	-4.343	**0.000**	-7.060	**0.000**	-1.491	1.000
+_3 / -_5	3.162	**0.044**	5.662	**0.000**	-0.269	1.000
+_3 / +_5	0.336	1.000	-0.040	1.000	-0.716	1.000
-_5 / +_5	-2.827	0.132	-5.703	**0.000**	-0.447	1.000

Beta Diversity

Variation partitioning of enrichment community composition assessed by Bray-Curtis (K(4), stress = 0.055; Fig. 14a, Table 8), Raup-Crick (K(4), stress = 0.039; Fig. 14b, Table 9), and Sørensen dissimilarities (K(4), stress = 0.067; Fig. 14c, Table 10), showed significant effects of glucose treatment, passage event, and their interaction. Multivariate dispersion across all metrics was not significantly different between glucose-positive and glucose-negative samples measured at the initial time point (Table 11; -_0/+_0; P_{Bray} = 1.000, P_{Raup} = 0.990, $P_{Sør}$ = 0.999). After the first passage event, glucose negative samples showed a significant increase in multivariate dispersion for Bray-Curtis, but not Raup-Crick or Sørensen metrics (Table 11; -_0/-_1; P_{Bray} < 0.05, P_{Raup} = 1.000, $P_{Sør}$ = 0.058). Multivariate dispersion increased significantly for all metrics in glucose negative samples after the third passage event (Table 11; -_1/-_3; P_{Bray} < 0.001, P_{Raup} < 0.005, $P_{Sør}$ < 0.001), and did not significantly shift after the fifth passage (Table 11; -_3/-_5; P_{Bray} = 0.477, P_{Raup} = 0.538, $P_{Sør}$ = 0.840). Significant differences in multivariate dispersion in glucose positive samples were observed for the Bray-Curtis and Sørensen metrics after the first passage event (Table 11; +_0/+_1; P_{Bray} < 0.05, P_{Raup} = 0.857, $P_{Sør}$ < 0.001) and multivariate dispersion remained constant for the remainder of the experiment across all metrics.

Pairwise PerMANOVA results showed no difference in multivariate centroid position across all groups for Bray-Curtis, Raup-Crick, and Sørensen metrics at the initial time point (Table 12; -_0 / +_0; P_{Bray} = 0.938, P_{Raup} = 0.792, $P_{Sør}$ = 0.965), indicating that average community composition was equal at the start of the experiment. Between the initial measure and the first passage event (48hr), both glucose positive (Table 12; +_0 / +_1; P_{Bray} < 0.005, P_{Raup} < 0.005, $P_{Sør}$ < 0.005) and glucose negative (Table 12; -_0 / -_1; P_{Bray} < 0.005, P_{Raup} < 0.005, $P_{Sør}$ < 0.005) treatments showed a significant difference in multivariate centroid position across all metrics and

occupied significantly different multivariate space from one another (Table 12; -_1 / +_1 ; P_{Bray} < 0.005, P_{Raup} < 0.005, $P_{Sør}$ < 0.005). Between the first (48hr) and third (144hr) passage events, multivariate centroid position was significantly different for glucose positive (Table 12; +_1 / +_3; P_{Bray} < 0.005, P_{Raup} < 0.05, $P_{Sør}$ < 0.005) and glucose negative samples (Table 12; -_1 / -_3; P_{Bray} < 0.005, P_{Raup} < 0.005, $P_{Sør}$ < 0.005) across all metrics. Between the third passage (144hr) and fifth (240hr) passage events, multivariate centroid position was not significantly different across any metric for glucose positive samples (Table 12; +_3 / +_5; P_{Bray} = 0.080, P_{Raup} = 0.913, $P_{Sør}$ = 0.163), and for glucose negative samples, multivariate centroid position was significantly different for only the Sørensen metric (Table 12; -_3 / -_5; P_{Bray} = 0.079, P_{Raup} = 0.304, $P_{Sør}$ < 0.05).

Table 8. Bray-Curtis PERMANOVA Results. ***d.f.***; degrees of freedom, ***SS***; Sum of Squares, R^2; R squared, ***pseudoF***; test statistic.

Main Effects	***d.f.***	***SS***	R^2	***pseudoF***	***P-Value***
Glucose	1	6.982	0.133	56.183	**0.001**
Passage	1	7.932	0.152	63.822	**0.001**
Ammonium	3	0.362	0.007	0.970	0.353
Glucose : Passage	1	2.431	0.0466	19.559	**0.001**
Glucose : Ammonium	3	0.245	0.005	0.657	0.725
Passage : Ammonium	3	0.280	0.005	0.752	0.592
Glucose : Passage : Ammonium	3	0.167	0.003	0.449	0.933
Residual	272	33.804	0.646		
Total	287	52.203	1.000		

Significance values based on 999 permutations

Table 9. Raup-Crick PERMANOVA Results. ***d.f.***: degrees of freedom, ***SS***; Sum of Squares, R^2; R squared, ***pseudoF***; test statistic.

Main Effects	***d.f.***	***SS***	R^2	***pseudoF***	***P-Value***
Glucose	1	11.906	0.219	211.196	**0.001**
Passage	1	22.121	0.407	392.595	**0.001**
Ammonium	3	0.250	0.005	1.480	0.216
Glucose : Passage	1	5.012	0.092	88.941	**0.001**
Glucose : Ammonium	3	0.228	0.004	1.135	1.00
Passage : Ammonium	3	0.102	0.002	0.602	0.640
Glucose : Passage : Ammonium	3	0.191	0.003	1.128	1.00
Residual	272	15.326	0.646		
Total	287	54.298	1.000		

Significance values based on 999 permutations

Table 10. Sørensen PERMANOVA Results. ***d.f.***: degrees of freedom, ***SS***; Sum of Squares, R^2; R squared, ***pseudoF***; test statistic.

Main Effects	***d.f.***	***SS***	R^2	***pseudoF***	***p-value***
Glucose	1	6.752	0.061	20.495	**0.001**
Passage	1	6.543	0.060	19.860	**0.001**
Ammonia	3	1.133	0.010	1.146	0.110
Glucose : Passage	1	2.702	0.025	8.200	**0.001**
Glucose : Ammonium	3	1.120	0.010	1.133	0.118
Passage : Ammonium	3	1.00	0.009	1.012	0.307
Glucose : Passage : Ammonium	3	0.989	0.009	1.000	0.335
Residual	272	89.615	0.816		
Total	287	109.854	1.000		

Significance values based on 999 permutations

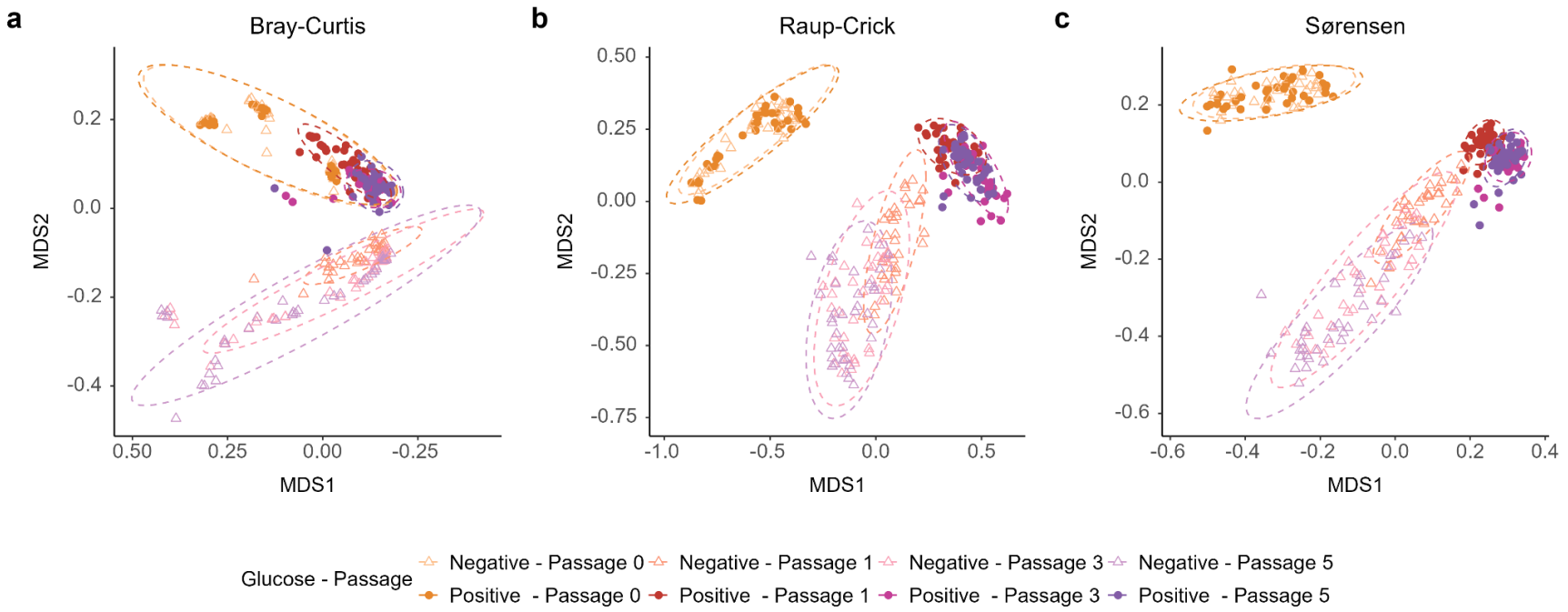


Figure 14. Ordinations of assemblage beta diversity. The first two dimensions of NMDS plots of Bray-Curtis (a), Raup-Crick (b), and Sorensen (c) β-diversity indexes. Samples are colored by media type and passage event. Glucose positive samples are depicted with filled circles and glucose negative samples are depicted with unfilled triangles.

Table 11. Tukey's Honestly Significant Test Results. Significant adjusted p-values are depicted in bold.

	Bray - Curtis		Raup - Crick		Sørensen	
Pairwise Comparison	**Difference**	***p***	**Difference**	***p***	**Difference**	***p***
-_0 / +_0	-0.0074	1.000	0.0070	0.980	0.0038	0.999
-_0 / -_1	-0.0848	**0.026**	-0.0020	1.000	-0.0209	0.058
-_0 / +_1	0.0957	**0.006**	-0.0033	0.999	-0.0598	**0.000**
-_0 / -_3	0.0420	0.736	0.0300	**0.002**	0.0248	**0.010**
-_0 / +_3	-0.0952	**0.00**6	-0.0033	0.999	-0.0394	**0.000**
-_0 / -_5	0.0941	**0.008**	0.0441	**0.000**	0.0348	**0.000**
-_0 / +_5	-0.0527	0.460	-0.0027	1.000	-0.0316	**0.000**
+_0 / -_1	-0.0774	0.060	0.0090	0.924	-0.0247	**0.010**
+_0 / +_1	-0.0883	**0.016**	-0.0103	0.857	0.0636	**0.000**
+_0 / -_3	0.0494	0.547	0.0230	**0.041**	0.0210	0.055
+_0 / +_3	-0.0879	**0.018**	-0.0103	0.855	-0.0433	**0.000**
+_0 / -_5	0.1014	**0.003**	0.0371	**0.000**	0.0310	**0.003**
+_0 / +_5	-0.0453	0.654	-0.0097	0.888	-0.0354	**0.000**
-_1 / +_1	-0.0109	0.999	0.0013	1.000	-0.0390	**0.000**
-_1 / -_3	0.1268	**0.000**	0.0320	**0.001**	0.0457	**0.000**
-_1 / +_3	-0.0105	0.999	-0.0013	1.000	-0.0186	0.136
-_1 / -_5	0.1788	**0.000**	0.0461	**0.000**	0.0556	**0.00**0
-_1 / +_5	0.0321	0.919	-0.0008	1.000	-0.0107	0.782
+_1 / -_3	0.1377	**0.000**	0.0333	**0.000**	0.0846	**0.000**
+_1 / +_3	0.0005	1.000	0.0000	1.000	0.0204	0.070
+_1 / -_5	0.1898	**0.000**	0.0474	**0.000**	0.0946	**0.000**
+_1 / +_5	0.0430	0.712	0.0005	1.000	0.0282	**0.002**
-_3 / +_3	-0.1373	**0.000**	-0.0333	**0.000**	-0.0642	**0.000**
-_3 / -_5	0.0520	0.477	0.0141	0.538	0.0110	0.840
-_3 / +_5	-0.0947	**0.007**	0.0327	**0.000**	-0.0564	**0.000**
+_3 / -_5	0.1893	**0.000**	0.0474	**0.000**	0.0742	**0.000**
+_3 / +_5	0.0426	0.723	0.0006	1.000	0.0078	0.950
-_5 / +_5	-0.1467	**0.000**	0.0468	**0.000**	-0.0664	**0.000**

Table 12. Pairwise PERMANOVA results for Bray-Curtis, Raup-Crick, and Sørensen. Significant adjusted *p*-values are depicted in bold. ***d.f.***; degrees of freedom, R^2; R squared, ***P***; P-value.

		Bray - Curtis		**Raup - Crick**		**Sørensen**	
Pairwise Comparison	***d.f.***	R^2	***p***	R^2	***p***	R^2	***p***
-_0 / +_0	1, 70	0.002	0.938	0.162	0.792	0.013	0.965
-_0 / -_1	1, 70	0.497	**0.001**	0.998	**0.001**	0.152	**0.001**
-_0 / +_1	1, 70	0.494	**0.001**	0.999	**0.001**	0.205	**0.001**
-_0 / -_3	1, 70	0.344	**0.001**	0.992	**0.001**	0.138	**0.001**
-_0 / +_3	1, 70	0.485	**0.001**	0.999	**0.001**	0.206	**0.001**
-_0 / -_5	1, 70	0.344	**0.001**	0.986	**0.001**	0.144	**0.001**
-_0 / +_5	1, 70	0.448	**0.001**	0.999	**0.001**	0.199	**0.001**
+_0 / -_1	1, 70	0.508	**0.001**	0.998	**0.001**	0.150	**0.001**
+_0 / +_1	1, 70	0.515	**0.001**	0.999	**0.001**	0.203	**0.001**
+_0 / -_3	1, 70	0.353	**0.001**	0.991	**0.001**	0.135	**0.001**
+_0 / +_3	1, 70	0.501	**0.001**	0.999	**0.001**	0.204	**0.001**
+_0 / -_5	1, 70	0.352	**0.001**	0.985	**0.001**	0.142	**0.001**
+_0 / +_5	1, 70	0.462	**0.001**	0.999	**0.001**	0.197	**0.001**
-_1 / +_1	1, 70	0.583	**0.001**	0.999	**0.001**	0.109	**0.001**
-_1 / -_3	1, 70	0.082	**0.002**	0.750	**0.001**	0.064	**0.001**
-_1 / +_3	1, 70	0.394	**0.001**	0.999	**0.001**	0.106	**0.001**
-_1 / -_5	1, 70	0.186	**0.001**	0.795	**0.001**	0.089	**0.001**
-_1 / +_5	1, 70	0.359	**0.001**	0.998	**0.001**	0.103	**0.001**
+_1 / -_3	1, 70	0.375	**0.001**	0.985	**0.001**	0.167	**0.001**
+_1 / +_3	1, 70	0.276	**0.001**	0.755	**0.011**	0.040	**0.001**
+_1 / -_5	1, 70	0.386	**0.001**	0.982	**0.001**	0.186	**0.001**
+_1 / +_5	1, 70	0.252	**0.001**	0.879	0.995	0.040	**0.001**
-_3 / +_3	1, 70	0.272	**0.001**	0.988	**0.001**	0.154	**0.001**
-_3 / -_5	1, 70	0.035	0.079	0.055	0.304	0.019	**0.043**
-_3 / +_5	1, 70	0.255	**0.001**	0.986	**0.001**	0.145	**0.001**
+_3 / -_5	1, 70	0.324	**0.001**	0.983	**0.001**	0.171	**0.001**
+_3 / +_5	1, 70	0.033	0.080	0.775	0.913	0.016	0.163
-_5 / +_5	1, 70	0.302	**0.001**	0.981	**0.001**	0.161	**0.001**

Significance values based on 999 permutations

Discussion

Enrichment culture experiments allow for a high-throughput design for the study of microbial community assembly, and provide a platform for the exploration of the impacts of nutrient gradients and other environmental conditions in highly controlled settings. These experiments can be used to establish stabilized microbial communities that can offer insights into assembly patterns based on a nutrient of interest, and serve as community models for further experimentation. This study focused on exploring the impacts of a nitrogen concentration gradient on soil inocula sourced from established wetland mesocosms through an enrichment culture experiment in media with elevated concentrations of ammonium in the presence or absence of glucose. This study's major findings were that 1) ammonium concentration had limited impact on final community composition, 2) the presence of glucose had the greatest impact on enrichment communities over the course of the experiment, and 3) the establishment of stabilized microbial communities in glucose positive media and the potential for alternative stable states.

Previous research has indicated that NH_4-N addition can reduce the relative abundance of some bacterial phyla, while maintaining consistent bacterial community α-diversity (Weng *et al.* 2023). The GAMM model of richness (Table 5) and PERMANOVA results for all dissimilarity metrics (Table 8, 9, 10) indicate that there was no direct effect of NH_4-N concentrations on both average community richness (Fig. 12b) and average community composition. Despite these findings, the GAMM model of richness indicates that a significant interaction effect between passage events and glucose-ammonium concentrations was present (Fig. 12c), and that stable richness was positively correlated with NH_4-N concentration when glucose was present and negatively correlated with NH_4-N concentration when glucose was absent. The maintenance of elevated bacterial community richness in response to increasing NH_4-N observed in this study

suggests that the microbial community may be resistant and resilient to increases in NH_4-N concentrations and is consistent with other work that has shown that long-term NH_4-N addition on the landscape has a limited impact on soil microbiome composition (Frey *et al.* 2020). Furthermore, impacts of NH_4-N on bacterial relative abundance may be reduced during wet seasons on the landscape (Nie *et al.* 2018), which may be why no significant differences were detected in the liquid based enrichment experiment. While the NH_4-N gradient had no direct effect on community composition, the presence of the significant interaction term in the GAMM model highlights the potential of NH_4-N to act as a limiting factor for maintenance of α-diversity when a carbon source is present, and act as an environmental filter when additional carbon is not present in the media (Yan *et al* 2019).

The presence of a carbon source in synthetic media was the most significant driver of both taxonomic and structural community differences throughout the course of the experiment. The GAMM model of OTU richness (Table 5, Figure 12a) indicates that the number of observed OTUs declined during the first 48 hours of the experiment. Enrichment communities that lacked a carbon source experienced greater decline in the first 48 hours of the experiment compared to those with glucose added to the media. Communities growing in media without additional carbon continued to experience declining OTU richness throughout the remainder of the experiment while communities grown in the presence of glucose maintained OTU richness through the fifth passage event. Soil microbiome diversity and biogeochemical cycling is dependent on the availability of carbon sources (Sokol *et al.* 2022), and the results of the community enrichment experiment indicate that carbon source is a limiting factor for the maintenance of bacterial α-diversity. It should be noted that control media that did not directly contain glucose still contained a carbon source in the form of the bacterial biomass that was added as the inoculation source (Goldford *et*

al. 2018; Sokol *et al.* 2022), and this initial bacterial biomass carbon was sufficient to maintain bacterial existence. Carbon sources are required for the maintenance of soil bacterial biodiversity whether it is provided directly from the environment or from the presence of dead microorganisms (Sokol *et al.* 2022). Bacterial community composition has been shown to respond to environmental availability of carbon sources in addition to altering biogeochemical process rates of denitrification (Hang *et al.* 2016; Pan *et al.* 2023) and nitrification (Steuernagel *et al.* 2018).

The presence of glucose in synthetic media had the greatest impact on relative abundance measures in this enrichment experiment. Glucose impacted β-diversity by significantly altering the taxonomic composition (Fig. 13) and community structure (Fig. 14) of enrichment communities. Significant differences in relative abundance of enrichment communities were observed at the family level (Fig. 13) that aligned with the decline in OTU richness observed for samples grown in media absent of a carbon source (Fig. 12a). No significant relative abundance differences were detected for glucose positive samples between the third and fifth passage events, indicating they had reached a stabilized community state. These results are consistent with other enrichment culture experiments that have suggested that microbial communities typically converge to stabilized states between the fourth and sixth passage event (Goldford *et al.* 2018; Estrela, Sánchez and Rebolleda-Gómez 2021). Final stabilized microbial communities grown in media with glucose contained consistent relative abundances of *Pseudomonadaceae* and *Enterobacteriaceae*, however a single soil source inoculum also maintained elevated levels of *Aeromonadaceae* across transfers. The dominance of *Pseudomonadaceae* and *Enterobacteriaceae* in glucose media has been observed in other enrichment culture experiments (Goldford *et al.* 2018; Estrela *et al.* 2022). The authors have suggested that *Enterobacteriaceae* are glucose specialists that use respiro-fermentative metabolism and produce organic acids as secondary metabolites which can then be

metabolized by *Pseudomonadaceae* through cross-feeding (Goldford *et al.* 2018; Estrela *et al.* 2022). In addition, while many observed OTUs that were characterized as *Enterobacteriaceae* remained unclassified at the genus level, the fifth most abundant *Enterobacteriaceae* OTU (OTU00119) was classified as the genus *Kasokonia*, which has been associated with nitrogen fixation and increased crop productivity (Jan-Roblero, Cruz-Maya and Barajas 2020). Members of *Aeromonadaceae* are also considered glucose specialists that produce organic acids as secondary metabolites (Estrela *et al.* 2022), their presence in one biological replicate (MES 19) may be indicative of an alternative stable state. Multistable community trajectories may be explained by differential relative abundance in the initial source inocula, which are evident for MES 19 (Figure 13a, b) or strain-specific adaptations that facilitated their growth and persistence (Estrela *et al.* 2022). Due to experimental constraints, initial inocula are unable to disperse to other wells on the plate and final community composition is highly dependent on the initial community composition. The observed divergence of stabilized communities and the possibility of multiple stable states has been supported by work showing that stable community states are dependent on the species pool of inocula (Hayashi, Fujita and Toju 2024) and may be altered by migration experiments (Estrela *et al.* 2022).

Exploration of enrichment communities using multiple dissimilarity measures allows for further discrimination of subtle changes to community structure throughout the experiment (Anderson, Ellingsen and McArdle 2006). Dissimilarity measures show contrasting effects of glucose on enrichment community structure across passage events (Figure 14, Table 8, 9, 10). Enrichment communities grown in the presence of glucose experienced no significant changes in multivariate dispersion after the first passage event, supporting a rapid convergence in community state (Hayashi, Fujita and Toju 2024). By contrast, a significant difference in all three dispersion

measures was observed for glucose negative samples between the first and third passage events (Table 11), consistent with the decline in richness observed in the GAMM model (Fig. 12b), and shifts in relative abundance observed in visualized taxonomic composition (Fig. 13a, c). Furthermore, average community composition of glucose negative samples was significantly different for the Sørensen metric between the third and fifth passage event (Table 12), but not for Bray-Curtis or Raup-Crick metrics, which may be evidence for the potential stochastic loss of community members due to drift (Anderson, Ellingsen and McArdle 2006; Vellend 2010; Hayashi, Fujita and Toju 2024). The incidence-based sørensen metric indicates that the shift in average community composition between the third and fifth passage event is due to a change in which specific OTUs are present, while the Bray-Curtis metric is heavily influenced by shifts in abundance and showed no change between passage events. These results, paired with a non-significant Raup-Crick measure, suggest that while the proportion of shared OTUs remains the same, variation in average community composition is due to a loss of less abundant taxa (Chase *et al.* 2011; Modin *et al.* 2020; Hayashi, Fujita and Toju 2024).

Multiple lines of evidence suggest that stabilized states were reached at the end of the community enrichment experiment, with limited support for the influence of ammonium concentration on compositional trajectories. While soil bacterial communities are known to be influenced by N concentration on the landscape (Nie *et al.* 2018; Frey *et al.* 2020), the marginal effect of NH_4-N concentration on bacterial communities in this experiment indicates that other abiotic factors such as synthetic media composition (Goldford *et al.* 2018; Estrela *et al.* 2022) or biotic factors such as source inocula (Mueller, Belnap and Kuske 2015) may be influencing assembly of complex soil communities (Fierer 2017). In addition, I acknowledge that these results represent observations from only three source inocula and a single abiotic factor. Further work to

understand the influence of fertilizer amendments on stabilized microbial communities may be improved by increasing the number of biological samples and probing the influence of other soil amendment components and their combined effects on the microbiome.

Conclusion

Enrichment culture experiments can provide new insight into the mechanisms and rules that drive microbial community composition by investigating the effect of nutrient gradients on stabilized community composition. This work supports the utility of synthetic media to investigate the potential influence of nutrient gradients on the bacterial communities in a controlled laboratory setting. Through continued methodological testing of synthetic media types, researchers may gain insight into exactly what nutrients and conditions are driving biodiversity patterns and determine which combination factors lead to specific stabilized states. These high-throughput enrichment culture experiments represent the foundation of future research into the mechanisms underlying community assembly and composition.

Acknowledgements

This work was supported by the Natural Resource Conservation Services sub-award 220858, National Science Foundation grants EF-2125065 and CAREER 2236580 to D. Walker, the Molecular Biosciences Program at MTSU, and National Science Foundation grants REU 175749

References

Alexander NR, Brown RS, Duwadi S *et al.* Leveraging Fine-Scale Variation and Heterogeneity of the Wetland Soil Microbiome to Predict Nutrient Flux on the Landscape. *Microb Ecol* 2025;**88**:22.

Anderson MJ, Ellingsen KE, McArdle BH. Multivariate dispersion as a measure of beta diversity. *Ecol Lett* 2006;**9**:683–93.

Balser TC, Kinzig AP, Firestone MK. 12. Linking Soil Microbial Communities and Ecosystem Functioning. In: Tilman D, Kinzig AP, Pacala S (eds.), *The Functional Consequences of Biodiversity*. Princeton University Press, 2013, 265–93.

Bello MD, Lee H, Goyal A *et al.* A simple linear relationship between resource availability and microbial community diversity. *bioRxiv* 2020:2020–09.

Bittleston LS, Gralka M, Leventhal GE *et al.* Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nat Commun* 2020;**11**:1440.

Brown SP, Veach AM, Rigdon-Huss AR *et al.* Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecol* 2015;**13**:221–5.

Cao Q, Sun X, Rajesh K *et al.* Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. *Front Microbiol* 2021;**11**:607325.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.

Chaparro JM, Sheflin AM, Manter DK *et al.* Manipulating the soil microbiome to increase soil health and plant fertility. *Biol Fertil Soils* 2012;**48**:489–99.

Chase JM, Kraft NJB, Smith KG *et al.* Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* 2011;**2**:art24.

Chase JM, Kraft NJB, Smith KG *et al.* Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* 2011;**2**:art24.

Correa-Garcia S, Constant P, Yergeau E. The forecasting power of the microbiome. *Trends Microbiol* 2023;**31**:444–52.

Dai Z, Liu G, Chen H *et al.* Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. *ISME J* 2020;**14**:757–70.

Davis NM, Proctor DM, Holmes SP *et al.* Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;**6**:226.

Estrela S, Sánchez Á, Rebolleda-Gómez M. Multi-replicated enrichment communities as a model system in microbial ecology. *Front Microbiol* 2021;**12**:657467.

Estrela S, Sanchez-Gorostiaga A, Vila JC *et al.* Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments. *Elife* 2021;**10**:e65948.

Estrela S, Vila JC, Lu N *et al.* Functional attractors in microbial community assembly. *Cell Syst* 2022;**13**:29–42.

Falkowski PG, Fenchel T, Delong EF. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* 2008;**320**:1034–9.

Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* 2017;**15**:579–90.

Frey B, Carnol M, Dharmarajah A *et al.* Only Minor Changes in the Soil Microbiome of a Sub-alpine Forest After 20 Years of Moderately Increased Nitrogen Loads. *Front For Glob Change* 2020;**3**:77.

Glibert PM. Harmful algae at the complex nexus of eutrophication and climate change. *Harmful Algae* 2020;**91**:101583.

Glibert PM, Maranger R, Sobota DJ *et al.* The Haber Bosch–harmful algal bloom (HB–HAB) link. *Environ Res Lett* 2014;**9**:105001.

Goldford JE, Lu N, Bajić D *et al.* Emergent simplicity in microbial community assembly. *Science* 2018;**361**:469–74.

Grajal-Puche A, Murray CM, Kearley M *et al.* Microbial Assemblage Dynamics Within the American Alligator Nesting Ecosystem: a Comparative Approach Across Ecological Scales. *Microb Ecol* 2020;**80**:603–13.

Gupta A, Singh UB, Sahu PK *et al.* Linking Soil Microbial Diversity to Modern Agriculture Practices: A Review. *Int J Environ Res Public Health* 2022;**19**:3141.

Hang Q, Wang H, Chu Z *et al.* Application of plant carbon source for denitrification by constructed wetland and bioreactor: review of recent development. *Environ Sci Pollut Res* 2016;**23**:8260–74.

Hall EK, Bernhardt ES, Bier RL *et al.* Understanding how microbiomes influence the systems they inhabit. *Nat Microbiol* 2018;**3**:977–82.

Hayashi I, Fujita H, Toju H. Deterministic and stochastic processes generating alternative states of microbiomes. *ISME Commun* 2024;**4**:ycae007.

Islam W, Noman A, Naveed H *et al.* Role of environmental factors in shaping the soil microbiome. *Environ Sci Pollut Res* 2020;**27**:41225–47.

Jan-Roblero J, Cruz-Maya JA, Barajas CG. Kosakonia. *Beneficial Microbes in Agro-Ecology*. Elsevier, 2020, 213–31.

Jetten MSM. The microbial nitrogen cycle. *Environ Microbiol* 2008;**10**:2903–9.

Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol* 2013;**79**:5112–20.

Lappalainen HK, Kerminen V-M, Petäjä T *et al.* Pan-Eurasian Experiment (PEEX): towards a holistic understanding of the feedbacks and interactions in the land–atmosphere–ocean–society continuum in the northern Eurasian region. *Atmospheric Chem Phys* 2016;**16**:14421–61.

Martiny JBH, Eisen JA, Penn K *et al.* Drivers of bacterial β-diversity depend on spatial scale. *Proc Natl Acad Sci* 2011;**108**:7850–4.

Modin O, Liébana R, Saheb-Alam S *et al.* Hill-based dissimilarity indices and null models for analysis of microbial community assembly. *Microbiome* 2020;**8**:1–16.

Mueller RC, Belnap J, Kuske CR. Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. *Front Microbiol* 2015;**6**:157127.

Nie Y, Wang M, Zhang W *et al.* Ammonium nitrogen content is a dominant predictor of bacterial community composition in an acidic forest soil with exogenous nitrogen enrichment. *Sci Total Environ* 2018;**624**:407–15.

O'Brien SL, Gibbons SM, Owens SM *et al.* Spatial scale drives patterns in soil bacterial diversity: Spatial scale drives soil diversity. *Environ Microbiol* 2016;**18**:2039–51.

Oksanen J, Kindt R, Legendre P *et al.* The vegan package. *Community Ecol Package* 2007;**10**:719.

Pan Y, Sun R-Z, Wang Y *et al.* Carbon source shaped microbial ecology, metabolism and performance in denitrification systems. *Water Res* 2023;**243**:120330.

Pedersen EJ, Miller DL, Simpson GL *et al.* Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 2019;**7**:e6876.

Prasad S, Dinesh GK, Sinduja M *et al. The Role of Microbes and Microbiomes in Ecosystem Restoration*. Bentham Science Publishers, 2024.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2012;**41**:D590–6.

Rasche F, Cadisch G. The molecular microbial perspective of organic matter turnover and nutrient cycling in tropical agroecosystems - What do we know? *Biol Fertil Soils* 2013;**49**:251–62.

Robinson JM, Hodgson R, Krauss SL *et al.* Opportunities and challenges for microbiomics in ecosystem restoration. *Trends Ecol Evol* 2023;**38**:1189–202.

Schloss PD. Reintroducing mothur: 10 years later. *Appl Environ Microbiol* 2020;**86**:e02343-19.

Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.

Simpson G. Introducing gratia. *Bottom Heap* 2018.

Sokol NW, Slessarev E, Marschmann GL *et al.* Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* 2022;**20**:415–30.

Steuernagel L, De Léon Gallegos EL, Azizan A *et al.* Availability of carbon sources on the ratio of nitrifying microbial biomass in an industrial activated sludge. *Int Biodeterior Biodegrad* 2018;**129**:133–40.

Taylor JM, Moore MT, Scott JT. Contrasting nutrient mitigation and denitrification potential of agricultural drainage environments with different emergent aquatic macrophytes. *J Environ Qual* 2015;**44**:1304–14.

Tyler HL, Moore MT, Locke MA. Influence of three aquatic macrophytes on mitigation of nitrogen species from agricultural runoff. *Water Air Soil Pollut* 2012;**223**:3227–36.

Urakawa H, Bernhard AE. Wetland management using microbial indicators. *Ecol Eng* 2017;**108**:456–76.

Vellend M. Conceptual Synthesis in Community Ecology. *Q Rev Biol* 2010;**85**:183–206.

Weng X, Wang M, Sui X *et al.* High Ammonium Addition Changes the Diversity and Structure of Bacterial Communities in Temperate Wetland Soils of Northeastern China. *Microorganisms* 2023;**11**:2033.

Womble SG. Factors Influencing Nutrietn Retention in Restored Floodplain Wetlands in Western Tennessee and Kentucky, USA. 2023.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc B* 2011;**73**:3–36.

Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC, 2017.

Yan Y, Klinkhamer PGL, Van Veen JA *et al.* Environmental filtering: A case of bacterial community assembly in soil. *Soil Biol Biochem* 2019;**136**:107531.

Zhang HQ, Zhao XQ, Shi Y *et al.* Changes in soil bacterial communities with increasing distance from maize roots affected by ammonium and nitrate additions. *Geoderma* 2021;**398**:115102.

APPENDICES

APPENDIX A: SUPPLEMENTAL METHODS AND EQUATIONS

Synthetic Water Preparation

Synthetic water was prepared to reflect elevated concentrations of NO_3 (~ 10 mg NO_3-N/L) and PO_4 (~ 1mg PO_4-P/L), in addition to trace mineral and micronutrients at historic concentrations from the Obion River and Bayou de Chien, USA shown in (Appenedix Table C1) (Brown 2023). Individual batches of synthetic water were prepared for each experimental incubation using a 135 L volume of deionized (DI) water from a sterile reservoir. Mineral solutions were prepared separately, and trace metals were prepared in concentrated batches of 1L prior to addition into the synthetic water solution (Brown 2023).

Incubation Experiment

Two paired soil cores (7.62 cm diameter x 30 cm depth) were collected from each point location. The first core was placed in an environmental chamber at 24 C for an incubation experiment. The second core was refrigerated at 4 C until processing for total soil C, N, and available soil P.

Synthetic water was pumped in parallel at a fixed flow rate (~ 1.8 mL/min) from the sterile reservoir through standard walled PTFE tubing (Component Supply, Sparta, TN, Inner Diameter [ID]: 0.86 mm) via in-line MasterFlex L/S peristaltic pump (Cole-Palmer, Vernon Hills, IL) with PhthalateFree PVC pump tubing (ISMATEC, Vancouver, WA, ID: 1.52 mm). PVC pump tubing was diverted to inflow tubing destined for individual soil cores or to empty control incubation cores. The inflow tubing length was adjusted to extend through a flangeless-fitting on custom designed lids (Brown 2023) to sit approximately 1 cm above the soil surface of the core. Outflowing water was directed through Tygon E-3603 Lab Tubing (Component Supply, Sparta, TN) into 12 ml Exetainer vials (Labco Exetainer, Lampeter, Wales, UK) through a custom lid and flangeless-fitting, to allow vials to be filled from the bottom. Vials were allowed to overflow by three volumes prior to their collection and the addition of 157 μL of $ZnCl_2$ (50 % w:v) to arrest microbial activity and remove any remaining headspace in the exetainer.

Upon collection, exetainers were sealed and maintained at 4 C until analysis. Membrane Inlet Mass Spectrometry (MIMS) was used to determine N_2 and O_2 dissolved gas concentrations with the Argon method; (Kana *et al.* 1994; Groffman *et al.* 2006) or a AQ400 Discrete Analyzer (Seal Analytical Inc., Mequon, WI) for NH_4, NO_3 + NO_2 and PO_4 concentrations using Seal Analytical standard calorimetric methods. To determine inflow concentrations of dissolved gasses and nutrients, outflow was also collected from each control identical to the above procedure. Flow rates through cores were measured for each sampling time point by measuring the volume of water pumped from the reservoir to the core inlet. Flow-through incubation was run for 24 consecutive hours for samples from the 2019 field season and 48 hours during the 2020 and 2021 field seasons.

Soil Characteristics

Soil characteristics for each point location were determined using the surface soils (0 cm - 5 cm depth) of paired large soil cores as this soil strata has been found to respond to nutrient loading (Morris and Bradley 1999). First, soils were homogenized using a clean putty scraper, followed by drying of a 30 g subsample at 105 C. Dried soils were packed into individual 20 ml scintillation vials and shipped to the Kansas State University's Soil Testing Laboratory where total C and N was measured for each sample using catalytic combustion to CO_2 or N_2 respectively, while soil available P was determined using the Mehlich-3 extraction method (Mehlich 1984).

Post-Incubation Control Sampling

To control for contaminating microbes that might influence downstream analyses, contamination control swabs were collected off the hood surfaces after wiping down the hood with DNA Away and just prior to sampling the cores. Control swabs were immediately transferred to 2 ml microcentrifuge tubes filled with 900 µL of sterile millipore water (2-hr autoclaved)

qPCR Reactions and Functional Gene Copy Numbers

PCR reactions (20 µL volume) for the nxrB amplicon target contained 10 µL PerfeCTa SYBR Green FastMix (Quantabio, Beverly, MA), 0.6 µL forward probe (10 µM), 0.6 µL reverse probe (10 µM), 6.8 µL of PCR grade water, and 2 µL of template DNA. PCR reactions (20 µL volume) for the nosZ-A and nosZ-B amplicon targets contained 10 µL PerfeCTa SYBR Green FastMix, 1 µL forward probe (60 µM), 1 µL reverse probe (60 µM), 6 µL of PCR grade water, and 2 µL of template DNA. Each qPCR plate included three positive controls and a no template negative control run in triplicate. To determine functional gene copy numbers of nxrB, nosZ-A, and nosZ-B present within each sample, a standard curve was created for each marker using a serial dilution of 1 x 10^{10} - 1 amplicon copies of a synthetic gBlock (Integrated DNA Technologies, Coralville, IA) fragment that matched the targeted regions for qPCR in Keeley et al. (2020). A separate equation was calculated for each marker to determine the number of copies per reaction by comparing the average C_t to the known gBlock amplicon copy number. In the following formulas, x represents the average C_t across triplicate technical replicates. The log copy number of periplasmic nxrB gene per qPCR reaction was calculated using the formula y = -0.2679x + 11.916. The log copy number of nosZ-A gene per qPCR reaction was calculated using the formula y = -0.2726x + 12.6656, while the log copy number of nosZ-B gene per qPCR reaction was calculated using the formula y = -0.302x + 15.924.

GLMNET Modeling

The GLMNET algorithm uses a combination of two types of regularization methods to prevent overfitting, ℓ_1 (Lasso) and ℓ_2 and (Ridge) regularization, which is referred to as the Elastic Net. Regularization for GLMNET models is determined by two hyperparameters, lambda (λ) and alpha (α). Here λ defines the magnitude of the regularization penalty, with larger values corresponding to increasing regularization. The mix between ℓ_1 and ℓ_2 regularization is controlled by the α hyperparameter which ranges between 0 and 1, where a value of 0 corresponds to ℓ_2

regularization, a value of 1 implements pure ℓ_1 regularization, and a value between the range corresponds to a mix between the two regularization types.

A hyperparameter grid search with consistent seed values was performed for all models, to ensure the best combination of hyperparameters was selected. The root mean squared error (RMSE) was used for the loss function of all models, and cross validation was performed using the leave-one-out cross validation (LOOCV) (Wong 2017). This method has previously been implemented for machine learning problems when highly dimensional data are coupled with relatively limited sampling effort (Sánchez 2022).

Supplemental Equations

Equation 1: Flux Rate

$$\boldsymbol{F_{i,nc} = \frac{([Core]_{out} - [IN])Q_{core}}{A}}$$

$F_{i,nc}$ represents the nutrient flux (mg m^{-2} h^{-1}) occurring in a given core. Dissolved nutrient or gas concentration is represented by $[Core]_{out}$, and [IN] corresponds to inflow concentrations measured from control cores. Flux rates account for variation in individual core flow rates using the Q_{core} term which represents the outflow rate measured directly from each core. Finally, A represents the core surface area (0.0045604 m^2) that was standardized during core collection.

Equation 2: Time Weighted Average Flux Rates

$$TWA = \left[\left(\frac{Flux_{t1} + Flux_{t2}}{2}\right) * (t2 - t1)\right] + \left[\left(\frac{Flux_{t2} + Flux_{t3}}{2}\right) * (t3 - t2)\right]$$

Time weighted average (TWA) is calculated by the average flux rate between two sampling times multiplied by the elapsed time between sampling. Average flux rates are calculated for all intervals and summed to generate the final TWA measure.

Equation 3: Hill Diversity (Generalized Mean)

$$D = \left(\Sigma_{i=1}^{R} \quad p_i \ \left(\frac{1}{p_i}\right)^{l} \right)^{1/l}$$

The generalized mean was chosen to evaluate α-diversity, in order to provide measures of assemblage diversity and evaluate richness and evenness within a unified framework by regulating the importance of rare vs commonly abundant OTUs across measures (Moreno and Rodríguez 2011; Tuomisto 2012; Roswell, Dushoff and Winfree 2021). *D* is derived from two values, the Richness (*R*) of an assemblage, and p_i, which represents the proportional abundance of OTU_i. The proportional abundance is defined as the abundance of OTU_i divided by the sequence read depth, also commonly referred to as an OTUs relative abundance. The reciprocal of relative abundance ($1/p_i$) is used to generate a rarity value for OTU_i, which is first scaled by exponent *l* (Roswell, Dushoff and Winfree 2021). When $l = 1$, Eq. 2 can be simplified and calculated as *R*, where the relative abundances of OTU_i are not considered, and each OTU contributes a value of 1, referred to as the *arithmetic mean* or *arithmetic rarity scale*. This measure is sensitive to rare OTUs and diversity will increase proportionally to the number of OTUs in an assemblage. Hill-Shannon diversity provides a neutral weight for both rare and abundant OTUs, in that OTU_i is weighted by its proportional difference from the mean, generating the *geometric mean* or *geometric rarity scale*. Hill-Shannon is defined for Equation 2 as the limit as *l* approaches 0, and can be calculated as the exponential of Shannon entropy, and interpreted as the *effective number of common* or *equal OTUs* (Chao, Chiu and Jost 2014; Chao *et al.* 2014). A soil assemblage that contains *R* equally abundant OTUs will have a Hill-Shannon value equal to *R*, while a soil assemblage that contains *R* OTUs with differential abundances will have a Hill-Shannon diversity < *R* that will be reflective of the *effective number of equally proportioned OTUs* in the assemblage.

The *generalized mean* represents a set of equations that varies the *leverage*, of rare relative to abundant elements of a dataset, using the exponent *l*, to both scale the rarity of each element, and back-transform the final value to the units of classification (Roswell, Dushoff and Winfree 2021). Using our OTU matrix, the units of *D* are described as the *effective number of OTUs* (Alberdi and Gilbert 2019), and can be thought of as the number of equally abundant OTUs required to provide an equitable diversity value, for a given scaling or *leverage* parameter, *l*.

Through the use of Hill diversities, we can investigate the influence of ecological gradients on both the number of unique OTUs in an assemblage ($l = 1$), and how an ecological gradient may shift the evenness of a community by altering the *effective number of common OTUs* ($l = 0$) present in an assemblage (Jost 2010; Chao, Chiu and Jost 2014; Chao *et al.* 2014; Alberdi and Gilbert 2019; Roswell, Dushoff and Winfree 2021).

References

Alberdi A, Gilbert MTP. A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol Ecol Resour* 2019;**19**:804–17.

Brown RS. Assessing Nutrient Retention and Removal Measurements Among Restored Floodplain Wetlands. 2023.

Chao A, Chiu C-H, Jost L. Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annu Rev Ecol Evol Syst* 2014;**45**:297–324.

Chao A, Gotelli NJ, Hsieh T *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol Monogr* 2014;**84**:45–67.

Groffman PM, Altabet MA, Böhlke JK *et al.* METHODS FOR MEASURING DENITRIFICATION: DIVERSE APPROACHES TO A DIFFICULT PROBLEM. *Ecol Appl* 2006;**16**:2091–122.

Jost L. The Relation between Evenness and Diversity. *Diversity* 2010;**2**:207–32.

Kana TM, Darkangelo Christina, Hunt MDuane *et al.* Membrane Inlet Mass Spectrometer for Rapid High-Precision Determination of N2, O2, and Ar in Environmental Water Samples. *Anal Chem* 1994;**66**:4166–70.

Mehlich A. "Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Commun Soil Sci Plant Anal* 1984;**15**:1409–16.

Moreno CE, Rodríguez P. Commentary: Do we have a consistent terminology for species diversity? Back to basics and toward a unifying framework. *Oecologia* 2011;**167**:889–92.

Morris JT, Bradley PM. Effects of nutrient loading on the carbon balance of coastal wetland sediments. *Limnol Oceanogr* 1999;**44**:699–702.

Roswell M, Dushoff J, Winfree R. A conceptual guide to measuring species diversity. *Oikos* 2021;**130**:321–38.

Sánchez XR. Machine learning on gut microbiome reveals potential biomarkers for Parkinson's diagnosis. 2022.

Tuomisto H. An updated consumer's guide to evenness and related indices. *Oikos* 2012;**121**:1203–18.

Wong T-T. Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets. *Pattern Recognit* 2017;**65**:97–107.

APPENDIX B: MOTHER COMMANDS

Mothur commands
Linux version
Using Boost,HDF5
mothur v.1.44.3
Last updated: 8/27/20
by
Patrick D. Schloss
Department of Microbiology & Immunology
University of Michigan
http://www.mothur.org
When using, please cite:
Schloss, P.D., et al., Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol, 2009. 75(23):7537-41.
Distributed under the GNU General Public License
Type 'help()' for information on the commands that are available
For questions and analysis support, please visit our forum at https://forum.mothur.org
Type 'quit()' to exit program
[NOTE]: Setting random seed to 19760620.
Interactive Mode
mothur > make.file(inputdir=., prefix=201201_wet_libs1_8, type=gz)
mothur > make.contigs(file=201201_wet_libs1_8.files, processors=30)
mothur > summary.seqs(fasta=current)
mothur > pcr.seqs(fasta=current, group=current, oligos=oligos.txt, pdiffs=1, rdiffs=3)
mothur > summary.seqs(fasta=current)
mothur > screen.seqs(fasta=current, group=current, maxambig=0, maxlength=255, maxhomop=8, minlength=251, processors=30)
mothur > summary.seqs(fasta=current)
mothur > unique.seqs()
mothur > summary.seqs(fasta=current)
mothur > count.seqs(name=current, group=current)
mothur > summary.seqs(fasta=current, count=current)
mothur > count.groups(count=current)
mothur > align.seqs(fasta=current, reference=silva.nr_v132.pcr.align)
mothur > summary.seqs(fasta=current, count=current)
mothur > screen.seqs(fasta=current, count=current, start=1968, end=11550)
mothur > summary.seqs(fasta=current, count=current)
mothur > count.groups(count=current)
mothur > filter.seqs(fasta=current, vertical=T, trump=.)

mothur > summary.seqs(fasta=current, count=current)
mothur > unique.seqs(fasta=current, count=current)
mothur > summary.seqs(fasta=current, count=current)
mothur > pre.cluster(fasta=current, count=current, diffs=2)
mothur > summary.seqs(fasta=current, count=current)
mothur > count.groups(count=current)
mothur > chimera.vsearch(fasta=current, count=current, dereplicate=t, processors=30)
mothur > set.current(processors=30,
fasta=201201_wet_libs1_8.trim.contigs.pcr.good.unique.good.filter.unique.precluster.fasta,count
=201201_wet_libs1_8.trim.contigs.pcr.good.unique.good.filter.unique.precluster.count_table,
accnos=201201_wet_libs1_8.trim.contigs.pcr.good.unique.good.filter.unique.precluster.denovo.
vsearch.accnos)
mothur > remove.seqs(fasta=current, count=current, accnos=current)
mothur > summary.seqs(fasta=current, count=current)
mothur > count.groups(count=current)
mothur > classify.seqs(fasta=current, count=current, reference=silva.nr_v132.pcr.align,
taxonomy=silva.nr_v132.tax, cutoff=80)
mothur > remove.lineage(fasta=current, count=current, taxonomy=current, taxon=Chloroplast-
Mitochondria-unknown-Archaea-Eukaryota)
mothur > summary.seqs(fasta=current, count=current)
mothur > summary.tax(taxonomy=current, count=current)
mothur > cluster.split(fasta=current, count=current, taxonomy=current, splitmethod=classify,
taxlevel=4, cutoff=0.03)
mothur > summary.seqs(fasta=current, count=current)
mothur > make.shared(list=current, count = current, label = 0.03)
mothur > remove.rare(shared=current, nseqs = 10, label = 0.03)
mothur > count.groups(shared=current)

APPENDIX C: SUPPLEMENTAL TABLES

Table C1. Synthetic Water Chemistry. Adapted from R. S. Brown, 2023. Minerals, dissolved organic carbon, and trace metals added to 135 L of DI water in sterile reservoirs for synthetic water used in flow-through incubation experiments. Mineral solutions were dissolved in DI water in individual ziplock bags prior to addition based on Order of Addition above, to yield the Final Reservoir Concentrations. Concentrated trace metal solutions were prepared in 1 L volumes and stored in sterile 1 L Nalgene bottles (Thermo Scientific, Waltham, MA) prior to addition.

Chemical Compound	Order of Addition	Final Reservoir Concentration (mg/L)	Mass of Compound Added (g)
Mineral Solution			**135 L Reservoir**
$NaHCO_3$	1	70.00	9.450
KCl	2	3.00	0.405
KH_2PO_4	3	4.40	0.594
$MgSO_4 \cdot 7H_2O$	4	27.00	3.645
$CaCl_2$	5	20.00	2.700
$NaNO_3$	6	60.00	8.100
$Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$	13	3.00	0.405
Dissolved Organic Carbon			
$C_6H_{12}O_6$	12	1.00	0.135
Trace Metal Solution			**1 L Nalgene Bottle**
$MnCl_2$	7	0.50	6.760
$CoCl_2 \cdot 6H_2O$	8	0.10	1.350
$ZnSO_4 \cdot 7H_2O$	9	0.05	0.675
$CuCl2 \cdot 2H_2O$	10	0.02	0.270
$NaMoO_4 \cdot 2H_2O$	11	0.03	0.405

Table C2. Richness (S) GAMM Results. SE; Standard Error, **e.d.f.**; expected degrees of freedom, **Ref d.f.**; reference degrees of freedom.

Component	**Term**	**Estimate**	***SE***	***T***	***P***	
A. parametric coefficients	(Intercept)	2,400.211	79.018	30.375	0.0000	***
Component	**Term**	***e.d.f.***	***Ref d.f.***	***F***	***P***	
B. smooth terms	s(Soil Nitrogen)	0.899	5	257.647	0.0006	***
	s(Soil Nitrogen, Partition)	7.014	30	143.299	0.0000	***
	s(Soil Nitrogen, Habitat)	2.900	19	102.699	0.0007	***
	s(Soil Nitrogen, Partition, Habitat)	0.001	60	0.001	0.4653	
	s(Latitude, Longitude)	0.017	96	0.108	0.3806	
	s(Site)	5.598	7	268.139	0.0002	***
	s(State)	0.606	1	154.361	0.1077	
	s(Year)	0.000	1	0.000	0.9225	
	s(Sample ID)	76.478	146	188.657	0.0000	***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.446, Deviance explained 0.505
-REML : 3971.817, Scale est: 1.000, N: 559
Family: Scaled t(7.102,220.356), Link function: Identity

Table C3. Relative Abundance by Partition. SD; Standard Deviation. Summary statistics depicting mean and standard deviation of relative abundance aggregated by partition (Figure 6).

	Field Top		Field Bottom		Lab Top		Lab Bottom	
Taxonomic Group	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Acidobacteria	13.314	5.486	17.726	7.774	12.746	5.361	19.169	8.143
Gammaproteobacteria	15.853	5.081	13.701	5.549	19.079	6.941	13.534	5.460
Actinobacteria	15.653	9.096	12.989	7.531	12.903	8.885	11.337	7.364
Alphaproteobacteria	11.260	5.188	10.255	4.337	10.535	4.865	9.686	4.314
Deltaproteobacteria	9.752	3.566	8.459	3.203	10.750	3.733	8.849	3.375
Chloroflexi	7.052	2.494	8.598	3.286	6.062	2.115	8.537	3.077
Verrucomicrobia	6.132	2.778	5.833	2.402	6.322	2.299	6.011	2.209
Bacteroidetes	5.244	4.729	3.096	2.662	6.408	4.542	3.191	2.888
Planctomycetes	4.024	1.229	4.323	1.651	3.819	1.013	4.018	1.258
Gemmatimonadetes	2.692	2.549	3.308	2.122	2.063	1.450	3.268	2.064
Unclassified Bacteria	1.153	0.640	1.639	1.091	1.143	0.661	1.698	1.024
Nitrospirae	0.949	0.970	1.672	1.567	0.959	0.962	1.956	1.280
Rokubacteria	0.834	0.961	1.908	2.171	0.640	0.741	2.111	2.388
Firmicutes	1.266	1.053	1.605	1.748	1.058	0.849	1.545	1.396
Latescibacteria	0.839	0.606	1.176	0.908	0.780	0.498	1.209	0.826
Cyanobacteria	0.910	2.009	0.384	1.013	1.490	4.278	0.400	0.666
Other	3.074	2.377	3.327	2.564	3.243	2.275	3.479	2.479

Table C4. Kruskal-Wallis Test (Partition) Results. ***d.f.***; degrees of freedom. Kruskal-Wallis results of relative abundance by partition (Figure 6). Significant *p*-values are depicted in bold.

Taxonomic Group	***d.f.***	***H-Value***	***P-Value***
Acidobacteria	3	93.709	**0.0000**
Gammaproteobacteria	3	63.588	**0.0000**
Actinobacteria	3	19.786	**0.0032**
Alphaproteobacteria	3	8.792	0.5474
Deltaproteobacteria	3	30.120	**0.0000**
Chloroflexi	3	87.447	**0.0000**
Verrucomicrobia	3	6.138	1.0000
Bacteroidetes	3	88.754	**0.0000**
Planctomycetes	3	5.144	1.0000
Gemmatimonadetes	3	57.243	**0.0000**
Unclassified Bacteria	3	54.004	**0.0000**
Nitrospirae	3	86.273	**0.0000**
Rokubacteria	3	108.367	**0.0000**
Firmicutes	3	19.845	**0.0031**
Latescibacteria	3	29.626	**0.0000**
Cyanobacteria	3	46.575	**0.0000**
Other	3	4.872	1.0000

Table C5. Relative Abundance by Habitat. ***SD***; Standard Deviation. Summary statistics depicting mean and standard deviation of relative abundance aggregated by habitat (Figure 7).

	Remnant Forest		**Shallow Water**		**Tree Planting**	
Taxonomic Group	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
Acidobacteria	15.780	6.602	14.384	7.217	16.691	7.878
Gammaproteobacteria	14.644	5.432	16.415	7.083	15.741	6.049
Actinobacteria	15.607	6.975	6.973	4.799	16.085	9.165
Alphaproteobacteria	13.514	4.645	7.762	4.561	9.762	3.037
Deltaproteobacteria	8.423	3.440	11.971	2.897	8.403	3.190
Chloroflexi	6.942	2.602	7.393	2.415	8.235	3.512
Verrucomicrobia	6.241	2.218	6.285	2.650	5.757	2.419
Bacteroidetes	3.408	2.763	7.546	5.324	3.062	2.207
Planctomycetes	4.290	1.370	4.038	1.515	3.825	1.026
Gemmatimonadetes	2.505	1.540	2.319	1.895	3.531	2.575
Unclassified Bacteria	1.020	0.616	1.996	1.154	1.283	0.648
Nitrospirae	1.400	1.490	1.866	1.366	0.965	0.814
Firmicutes	1.042	1.418	2.056	1.411	1.109	0.871
Rokubacteria	1.436	1.926	1.458	2.229	1.224	1.302
Latescibacteria	1.095	0.866	0.882	0.688	1.003	0.669
Cyanobacteria	0.338	0.709	1.154	1.820	0.946	3.678
Spirochaetes	0.271	0.436	1.490	1.306	0.176	0.265
Other	2.046	1.305	4.012	1.896	2.201	1.063

Table C6. Kruskal-Wallis Test (Habitat) Results. ***d.f.***; degrees of freedom. Kruskal-Wallis results of relative abundance by habitat (Figure 7). Significant *p* values are depicted in bold.

Taxonomic Group	*d.f.*	*H-Value*	*P-Value*
Acidobacteria	2	12.823	**0.0295**
Gammaproteobacteria	2	4.603	1.0000
Actinobacteria	2	175.795	**0.0000**
Alphaproteobacteria	2	130.935	**0.0000**
Deltaproteobacteria	2	127.982	**0.0000**
Chloroflexi	2	14.631	**0.0120**
Verrucomicrobia	2	5.424	1.0000
Bacteroidetes	2	96.161	**0.0000**
Planctomycetes	2	12.376	**0.0369**
Gemmatimonadetes	2	44.101	**0.0000**
Unclassified Bacteria	2	117.727	**0.0000**
Nitrospirae	2	45.354	**0.0000**
Firmicutes	2	113.384	**0.0000**
Rokubacteria	2	1.464	1.0000
Latescibacteria	2	6.332	0.7596
Cyanobacteria	2	101.996	**0.0000**
Spirochaetes	2	252.807	**0.0000**
Other	2	158.915	**0.0000**

Table C7. PERMANOVA Results of Bray Curtis Metric. ***d.f.***; degrees of freedom, ***SS***; Sum of squares

Main Effects	***d.f.***	***SS***	R^2	***pseudoF***	***P-Value***
Site	7	19.604	0.133	14.462	**0.001**
Habitat	2	12.870	0.087	33.230	**0.001**
Partition	3	5.793	0.039	9.972	**0.001**
Richness	1	3.269	0.022	16.882	**0.001**
N Concentration	1	0.659	0.004	3.405	**0.002**
Residual	544	105.345	0.714		
Total	558	147.541	1.000		

Significance values based on 999 permutations

Table C8. Factor level Pairwise PERMANOVA Results. ***d.f.***; degrees of freedom

		Bray - Curtis		**Raup - Crick**	
Pairwise Comparison	***d.f.***	R^2	*P*	R^2	*P*
Tree Planting vs Shallow Water	1, 368	0.091	**0.001**	0.550	**0.001**
Remnant Forest vs Tree Planting	1, 392	0.027	**0.001**	0.092	**0.001**
Remnant Forest vs Shallow Water	1, 352	0.088	**0.001**	0.480	**0.001**
Field Bottom vs Field Top	1, 278	0.025	**0.001**	0.094	**0.001**
Field Bottom vs Lab Bottom	1, 274	0.004	0.254	0.015	0.119
Field Bottom vs Lab Top	1, 279	0.043	**0.001**	0.136	**0.001**
Field Top vs Lab Bottom	1, 276	0.033	**0.001**	0.114	**0.001**
Field Top vs Lab Top	1, 281	0.008	**0.008**	0.004	0.385
Lab Bottom vs Lab Top	1, 278	0.047	**0.001**	0.135	**0.001**

Significance values based on 999 permutations

Table C9. PERMANOVA Results of Raup-Crick Metric. ***d.f.***; degrees of freedom, ***SS***; Sum of squares

Main Effects	***d.f.***	***SS***	R^2	***pseudoF***	***P-Value***
Site	7	15.923	0.298	1,007.224	**0.001**
Habitat	2	26.687	0.500	5,908.198	**0.001**
Partition	3	6.476	0.121	955.885	**0.001**
Richness	1	3.457	0.065	1,530.674	**0.001**
N Concentration	1	-0.410	-0.008	-181.344	1.000
Residual	544	1.229	0.023		
Total	558	53.363	1.000		

Significance values based on 999 permutations

Table C10. nosZA GAMM Results. ***SE***; Standard Error, ***e.d.f.***; expected degrees of freedom, ***Ref d.f.***; reference degrees of freedom.

Component	Term	Estimate	*SE*	*T*	*P*	
A. parametric coefficients	(Intercept)	7.686	0.472	16.274	0.0000	***
	Field Bottom Partition	0.762	0.283	2.690	0.0072	**
	Lab Top Partition	1.093	0	3.766	0.0002	***
	Lab Bottom Partition	1.477	0	5.097	0.0000	***
	Shallow Water Habitat	1.266	0	4.151	0.0000	***
	Tree Planting Habitat	0.677	0	2.271	0.0232	*
	Field Bottom:Shallow Water	-1.096	0	-2.623	0.0087	**
	Lab Top:Shallow Water	-0.383	0	-0.878	0.3799	
	Lab Bottom:Shallow Water	-1.453	0	-3.327	0.0009	***
	Field Bottom:Tree Planting	-0.380	0	-0.929	0.3527	
	Lab Top:Tree Planting	-1.508	0	-3.593	0.0003	***
	Lab Bottom:Tree Planting	-1.688	0	-4.068	0.0000	***

Table C10 nosZA GAMM Results continued.

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(N_2 Flux)	0.000	23	0.000	0.8190	
	s(N_2 Flux):Field Top.Remnant Forest	0.000	21	0.000	0.8090	
	s(N_2 Flux):Field Bottom.Remnant Forest	0.000	21	0.000	0.6750	
	s(N_2 Flux):Lab Top.Remnant Forest	0.803	19	5.403	0.0260	*
	s(N_2 Flux):Lab Bottom.Remnant Forest	0.000	22	0.000	0.4543	
	s(N_2 Flux):Field Top.Shallow Water	0.000	22	0.000	0.7890	
	s(N_2 Flux):Field Bottom.Shallow Water	0.000	23	0.000	0.4196	
	s(N_2 Flux):Lab Top.Shallow Water	0.000	22	0.000	0.3731	
	s(N_2 Flux):Lab Bottom.Shallow Water	0.169	22	0.249	0.2146	
	s(N_2 Flux):Field Top.Tree Planting	0.119	21	0.144	0.2933	

Table C10. nosZA GAMM Results continued.

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(N_2 Flux):Field Bottom.Tree Planting	0.000	20	0.000	0.5527	
	s(N_2 Flux):Lab Top.Tree Planting	0.000	22	0.000	0.4412	
	s(N_2 Flux):Lab Bottom.Tree Planting	3.152	23	12.920	0.0142	*
	s(Site)	5.470	7	210.420	0.0102	*
	s(State)	0.255	1	12.098	0.2440	
	s(Year)	0.726	1	231.883	0.0576	.

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.258, Deviance explained 0.251
-REML : 761.911, Scale est: 1.000, N: 416
Family: Scaled t(3,1.008), Link function: Identity

Table C11. nosZB GAMM Results. ***SE***; Standard Error, ***e.d.f.***; expected degrees of freedom, ***Ref d.f.***; reference degrees of freedom.

Component	Term	Estimate	*SE*	*T*	*P*	
A. parametric coefficients	(Intercept)	7.418	0.179	41.381	0.0000	***
	Field Bottom Partition	0.054	0.155	0.347	0.7288	
	Lab Top Partition	0.305	0	1.947	0.0515	.
	Lab Bottom Partition	0.298	0	1.885	0.0594	.
	Shallow Water Habitat	0.476	0	2.881	0.0040	**
	Tree Planting Habitat	0.172	0	1.069	0.2849	
	Field Bottom:Shallow Water	-0.226	0	-1.023	0.3063	
	Lab Top:Shallow Water	-0.300	0	-1.312	0.1896	
	Lab Bottom:Shallow Water	-0.797	0	-3.440	0.0006	***
	Field Bottom:Tree Planting	0.070	0	0.324	0.7463	
	Lab Top:Tree Planting	-0.544	0	-2.468	0.0136	*
	Lab Bottom:Tree Planting	-0.691	0	-3.138	0.0017	**

Table C11. nosZB GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(N_2 Flux)	0.000	23	0.000	0.8655	
	s(N_2 Flux):Field Top.Remnant Forest	1.593	22	7.153	0.0182	*
	s(N_2 Flux):Field Bottom.Remnant Forest	0.342	17	0.587	0.2165	
	s(N_2 Flux):Lab Top.Remnant Forest	0.000	21	0.000	0.4315	
	s(N_2 Flux):Lab Bottom.Remnant Forest	0.492	22	1.072	0.1310	
	s(N_2 Flux):Field Top.Shallow Water	0.000	23	0.000	0.8019	
	s(N_2 Flux):Field Bottom.Shallow Water	0.837	13	7.049	0.0097	**
	s(N_2 Flux):Lab Top.Shallow Water	0.062	22	0.107	0.1707	
	s(N_2 Flux):Lab Bottom.Shallow Water	0.598	22	1.199	0.1365	
	s(N_2 Flux):Field Top.Tree Planting	0.000	21	0.000	0.7841	

Table C11. nosZB GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(N_2 Flux)	0.000	23	0.000	0.8655	
	s(N_2 Flux):Field Bottom.Tree Planting	0.000	20	0.000	0.9884	
	s(N_2 Flux):Lab Top.Tree Planting	0.000	22	0.000	0.5454	
	s(N_2 Flux):Lab Bottom.Tree Planting	4.753	22	32.251	0.0001	***
	s(Site)	6.590	7	125.011	0.0000	***
	s(State)	0.000	1	0.000	0.8401	
	s(Year)	0.000	1	0.000	0.8116	

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.204, Deviance explained 0.215
-REML : 508.185, Scale est: 1.000, N: 416
Family: Scaled t(3,0.518) , Link function: Identity

Table C12. nxrB (NO_2) GAMM Results. ***SE***; Standard Error, ***e.d.f.***; expected degrees of freedom, ***Ref d.f.***; reference degrees of freedom.

Component	Term	Estimate	*SE*	*T*	*P*	
A. parametric coefficients	(Intercept)	6.865	0.226	30.390	0.0000	***
	Field Bottom Partition	0.044	0.094	0.468	0.6400	
	Lab Top Partition	-0.025	0	-0.269	0.7877	
	Lab Bottom Partition	0.031	0	0.332	0.7396	
	Shallow Water Habitat	-0.461	0	-4.704	0.0000	***
	Tree Planting Habitat	-0.329	0	-3.545	0.0004	***
	Field Bottom:Shallow Water	0.032	0	0.230	0.8182	
	Lab Top:Shallow Water	0.136	0	0.988	0.3233	
	Lab Bottom:Shallow Water	0.162	0	1.203	0.2291	
	Field Bottom:Tree Planting	0.098	0	0.752	0.4520	
	Lab Top:Tree Planting	-0.058	0	-0.448	0.6541	
	Lab Bottom:Tree Planting	-0.022	0	-0.167	0.8675	

Table C12. nxrB (NO_2) GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(NO_2 Flux)	0.431	9	1.759	0.0997	.
	s(NO_2 Flux):Field Top.Remnant Forest	0.352	22	0.762	0.1569	
	s(NO_2 Flux):Field Bottom.Remnant Forest	0.657	23	2.232	0.0842	.
	s(NO_2 Flux):Lab Top.Remnant Forest	0.000	23	0.000	0.9530	
	s(NO_2 Flux):Lab Bottom.Remnant Forest	0.310	21	0.662	0.1592	
	s(NO_2 Flux):Field Top.Shallow Water	0.856	10	6.861	0.0048	**
	s(NO_2 Flux):Field Bottom.Shallow Water	0.567	9	1.775	0.0741	.
	s(NO_2 Flux):Lab Top.Shallow Water	1.772	20	12.472	0.0009	***
	s(NO_2 Flux):Lab Bottom.Shallow Water	0.000	19	0.000	0.5959	
	s(NO_2 Flux):Field Top.Tree Planting	2.575	23	13.969	0.0037	**

Table C12. nxrB (NO_2) GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
	s(NO_2 Flux):Field Bottom.Tree Planting	0.000	23	0.000	0.2615	
	s(NO_2 Flux):Lab Top.Tree Planting	1.104	23	6.288	0.0139	*
	s(NO_2 Flux):Lab Bottom.Tree Planting	0.000	22	0.000	0.2165	
	s(Site)	5.890	7	383.890	0.0147	*
	s(State)	0.825	1	1,475.428	0.0168	*
	s(Year)	0.001	1	0.001	0.3248	

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.249, Deviance explained 0.330
-REML : 419.156, Scale est: 1.000, N: 516
Family: Scaled t(3,0.342), Link function: Identity

Table C13. nxrB (NO_3) GAMM Results. SE; Standard Error, **e.d.f.**; expected degrees of freedom, **Ref d.f.**; reference degrees of freedom.

Component	Term	Estimate	*SE*	*T*	*P*	
A. parametric coefficients	(Intercept)	6.840	0.226	30.294	0.0000	***
	Field Bottom Partition	0.057	0.093	0.612	0.5406	
	Lab Top Partition	-0.010	0	-0.108	0.9138	
	Lab Bottom Partition	0.042	0	0.450	0.6525	
	Shallow Water Habitat	-0.509	0	-5.034	0.0000	***
	Tree Planting Habitat	-0.310	0	-3.317	0.0009	***
	Field Bottom:Shallow Water	0.064	0	0.446	0.6558	
	Lab Top:Shallow Water	0.181	0	1.308	0.1910	
	Lab Bottom:Shallow Water	0.209	0	1.489	0.1365	
	Field Bottom:Tree Planting	0.094	0	0.719	0.4724	
	Lab Top:Tree Planting	-0.023	0	-0.178	0.8583	
	Lab Bottom:Tree Planting	-0.020	0	-0.152	0.8796	

Table C13. nxrB (NO_3) GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
B. smooth terms	s(NO_3 Flux)	1.369	23	10.636	0.0146	*
	s(NO_3 Flux):Field Top.Remnant Forest	0.095	16	0.078	0.3696	
	s(NO_3 Flux):Field Bottom.Remnant Forest	0.629	23	1.374	0.1445	
	s(NO_3 Flux):Lab Top.Remnant Forest	0.000	23	0.000	0.4588	
	s(NO_3 Flux):Lab Bottom.Remnant Forest	0.000	23	0.000	0.9471	
	s(NO_3 Flux):Field Top.Shallow Water	0.686	20	1.342	0.1462	
	s(NO_3 Flux):Field Bottom.Shallow Water	1.061	20	2.603	0.1098	
	s(NO_3 Flux):Lab Top.Shallow Water	0.000	20	0.000	0.6370	
	s(NO_3 Flux):Lab Bottom.Shallow Water	0.700	20	1.216	0.1877	
	s(NO_3 Flux):Field Top.Tree Planting	0.940	22	1.913	0.1374	

Table C13. nxrB (NO_3) GAMM Results continued

Component	Term	*e.d.f.*	*Ref d.f.*	*T*	*P*	
	s(NO_3 Flux):Field Bottom.Tree Planting	0.000	23	0.000	0.7460	
	s(NO_3 Flux):Lab Top.Tree Planting	0.000	23	0.000	0.7361	
	s(NO_3 Flux):Lab Bottom.Tree Planting	0.000	23	0.000	0.5724	
	s(Site)	5.850	7	674.546	0.0260	*
	s(State)	0.831	1	1,663.417	0.0156	*
	s(Year)	0.000	1	0.000	0.3282	

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.247, Deviance explained 0.303
-REML : 425.813, Scale est: 1.000, N: 516
Family: Scaled t(3,0.354), Link function: Identity

APPENDIX D: SUPPLEMENTAL FIGURES

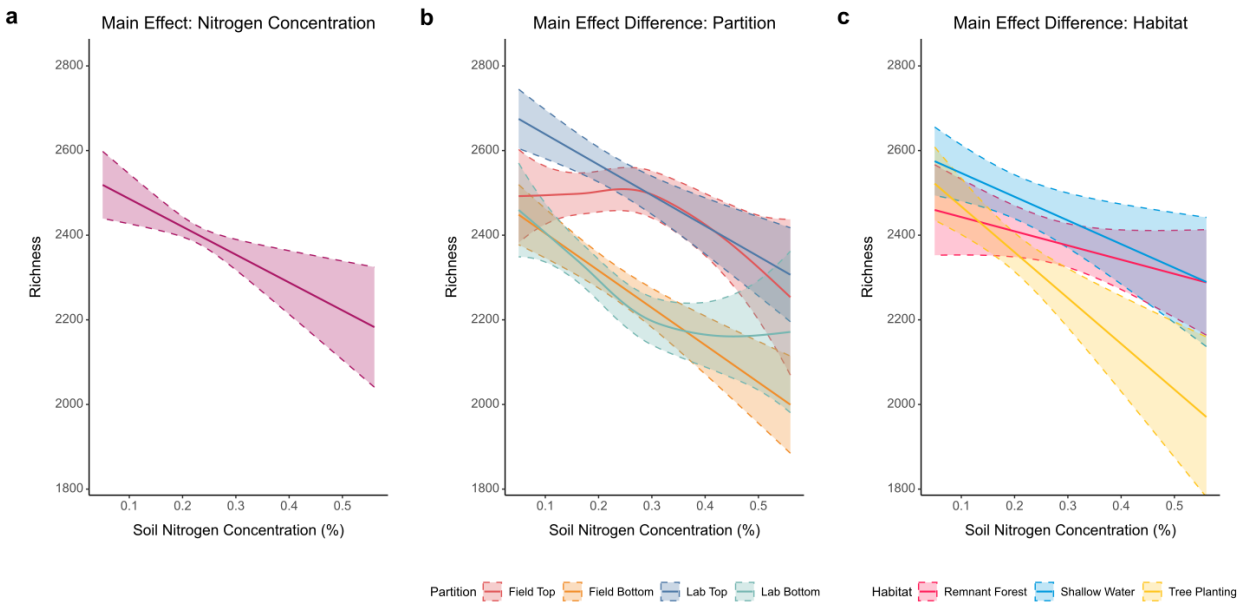


Figure D1. Richness Declines with Soil N concentration. Significant GAMM model smooths are shown with Baysian confidence intervals for (a) the Main Effect of Soil N concentration on Richness colored in purple. Partial Effects Difference Smooths depict the differences in estimated Richness between the main effect and smooth effect for each level colored by (b) partition and (c) habitat along the gradient of soil N

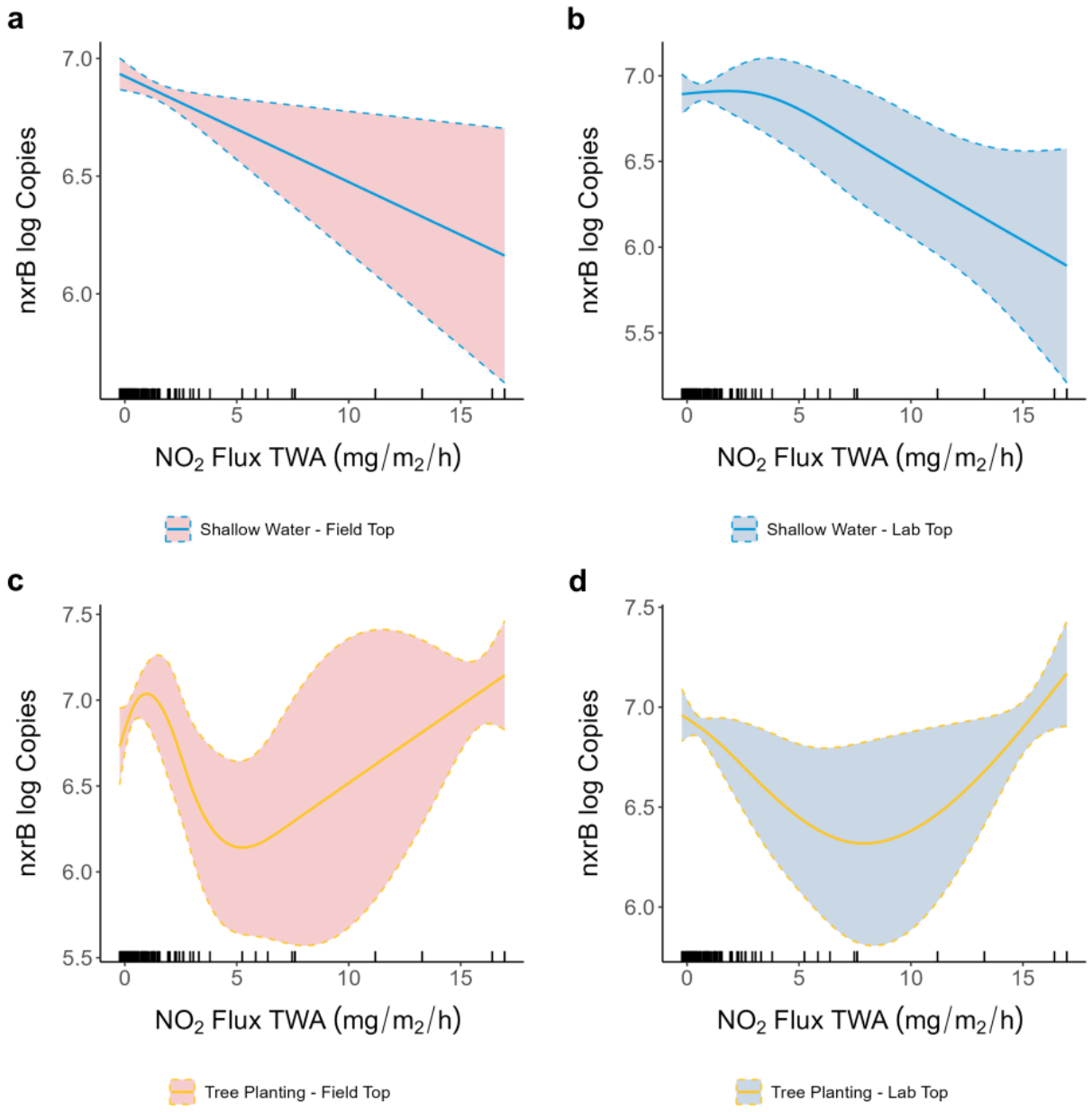


Figure D2. Trends in nxrB Gene Abundance Estimated with NO_2 Flux Rates. Significant GAMM model smooths depicted with Baysian confidence intervals and rug plots of sampling effort are shown for the three-way interactions of N_2 predicting nxrB log copies with shallow water field tops (a), shallow water lab tops tops (b), tree planting field tops (c), and tree planting lab tops (d).

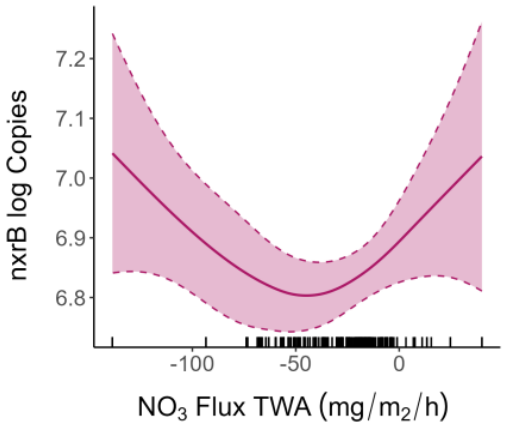
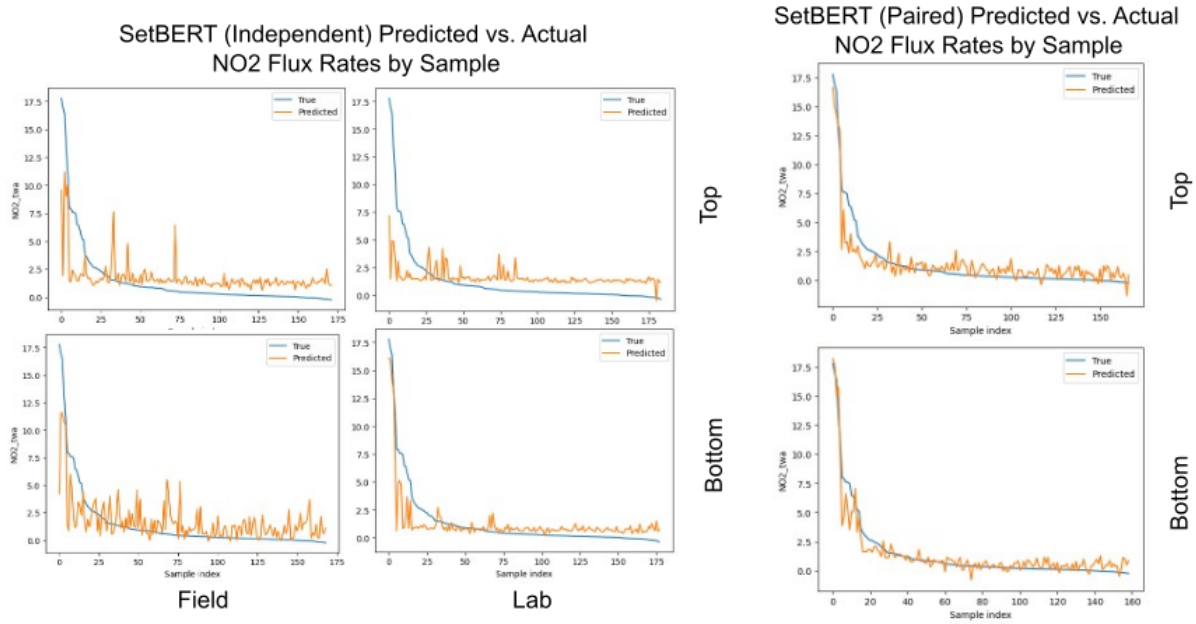


Figure D3. Trends in nxrB Gene Abundance Estimated with NO_3 Flux Rates. Significant GAMM model smooth is depicted with Baysian confidence intervals and rug plot of sampling effort for the main effect of NO_3 Flux on estimated nxrB gene abundance



SetBERT vs. GLMNET NRMSE

Partition	GLMNET	SetBERT
Field Top	**0.2702**	0.7662
Lab Top	**0.4489**	0.8709
Field Bottom	0.8728	**0.6886**
Lab Bottom	0.7855	**0.4732**

SetBERT vs. GLMNET NRMSE

Partition	GLMNET	SetBERT (Independent)	SetBERT (Paired)
Field Top	**0.2702**	0.7662	**0.3927**
Lab Top	**0.4489**	0.8709	
Field Bottom	0.8728	0.6886	**0.2940**
Lab Bottom	0.7855	0.4732	

Figure D4. SetBERT Predictive Accuracy of Independent and Paired models. SetBERT model predictions vs observed flux rates for NO_2 flux compared to GLMNET results for independent and paired soil partitions

APPENDIX E: SUPPLEMENTAL RESULTS

Richness GAMM

The richness GAMM (Appendix Table C2) accounted for just over half of the variation in the dataset, expressed as deviance explained as the models are non-gaussian estimates (GAMM, Deviance Explained = 50.5%, -REML = 3971.82). The intercept of richness was significantly greater than 0, and estimated to be 2,400 unique OTUs across samples (GAMM, est. = 2400.21, SE = 77.96, T = 31.83, $P < 0.0001$). Soil N concentration had a significant and negative main effect on estimated richness (GAMM, e.d.f. = 0.899, F = 257.651, $P < 0.001$). Estimated richness was greatest at the lowest measured N concentrations, and declined over the range of increasing N concentration (Appendix Fig. D1a). A significant partial effect of soil N by partition was detected (GAMM, e.d.f. = 7.015, F = 143.3, $P < 0.001$), where the smooth effects are decomposed and the difference is taken between the main effect smooth and the main effect smooth for each level of partition (Appendix Fig. D1b).

Smooth differences for each partition were visualized, by adding the intercept, and main effect smooth to the partial effect smooth for each partition (Appendix Fig D1b). Field top samples maintained an estimated OTU richness of approximately 2,400 OTUs across measured soil N concentrations of 0.1% to 0.3%, before experiencing a decline in richness to increasing soil N beyond the 0.3% threshold. In contrast, lab top samples experienced a decrease in estimated richness across the entire range of soil N, but had greater estimated richness than field top samples which maintained richness (0.1-0.3% soil N). These results indicate that top samples experience a positive shift in estimated OTU richness (50-150 OTUs) as a result of the incubation treatment (Appendix Fig. D1b), compared to field top samples over the lower range of soil N (0.1-0.2% soil N). Between the range of 0.2-0.5% soil N, estimated richness smooths for both field top and field bottom partitions are similar, and follow the main effect trend of declining soil richness with increasing soil N. Field bottom and lab bottom partitions are estimated to have approximately 200 fewer OTUs (Appendix Fig. D1b) than either field top or field bottom partitions across the range of soil N. Field bottom and lab bottom smooths respond very similarly to the main effect across increasing soil N concentration, however lab bottom samples appear to have an enriched richness estimate at the greatest values of soil N. The difference in partial smooth effects across levels of partition indicates that field and lab top samples have greater estimated richness compared to field and lab bottom samples, however the degree to which independent partitions show differences in estimated richness is dependent on soil N concentration.

A significant partial effect of soil N by habitat was found (GAMM, e.d.f. = 2.900, F = 102.699, $P < 0.001$) across independent levels of habitat difference smooths (Appendix Fig. D5c). Remnant forest and shallow water habitats show little difference from the main effect of N concentration, and show a significant decline in estimated richness across increasing N concentration. Shallow water habitats appear to have a marginally higher estimated richness than the remnant forest (Appendix Fig. D1c), and estimated richness declines more sharply in shallow water habitats as a response to increasing N concentration (Appendix Fig. D1c). Tree planting

habitats show the greatest difference among habitat types, and respond to increased N concentration with a rapid decline in estimated richness. Tree planting habitats are estimated to contain approximately 100-200 fewer unique OTUs (Appendix Fig. D1c), than either remnant forest or shallow water habitats, across the soil N range of 0.2 - 0.5% (Appendix Fig. D1c). Significant differences were not detected for the two factor interaction smooth of soil N by partition and habitat (GAMM, e.d.f. = 0.001, F = 0.001, P = 0.465). Significant random effects were detected for easement site (GAMM, e.d.f. = 5.598, F = 268.141, P < 0.001), and sample ID (GAMM, e.d.f. = 76.478, F = 188.657, P < 0.001), however, no significant effects were observed for sampling point locations (GAMM, e.d.f. = 0.017, F = 0.108, P = 0.381), state of collection (GAMM, e.d.f. = 0.606, F = 154.362, P = 0.101), or year of collection (GAMM, e.d.f. = 0.000, F = 0.000, P = 0.923).

QPCR

The nxrB GAMM predicted by NO_3 flux rate (Supplemental Table E13) was the only qPCR GAMM model that exhibited a significant main effect trend (GAMM, Deviance Explained = 30.3%, - REML = 425.813). The main effect of NO_3 on estimated nxrB gene abundance (GAMM, e.d.f. = 1.36, F = 10.636, P < 0.05) exhibits a concave smooth trend centered about average NO_3 flux measurements. No significant smooths were observed for the three-way interaction between NO_3 flux, habitat, and partition. Significant random effects were observed for easement sites (GAMM, e.d.f. = 5.850, F = 674.546, P < 0.05), and state (GAMM, e.d.f. = 0.831, F = 1,663.417, P < 0.05), but not year of collection (GAMM, e.d.f. = 0, F = 0, P = 0.328).