Microbial Ecology (2025) 88:22
https://doi.org/10.1007/s00248-025-02516-1

RESEARCH

# 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]**

Received: 13 December 2024 / Accepted: 13 March 2025


**Abstract**
Shifts in agricultural land use over the past 200 years have led to a loss of nearly 50% of existing wetlands in the USA, 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 the 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 predict nutrient flux using microbial assemblage composition. Soil microbial assemblages showed fine-scale turnover in soil cores collected across the topsoil horizon (0–5 cm; 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, the predictive accuracy of flux rates was improved when implementing deep learning models that paired connected samples across time.

**Keywords** Biogeochemical cycling · Deep learning · Nitrogen cycle · Nutrient cycling · Wetland restoration

## 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 [1]. The primary cause of this degradation is nutrient over-enrichment, attributed to agriculture and the excess discharge of nitrogen (N) and phosphorus (P) [2]. 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 [3]. Recent federal efforts in the USA 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 landowners to collaborate with non-profits and researchers to abate dissolved nutrient loads from agricultural runoff [4].

Numerous ecosystem services have been restored as part of this program including nutrient cycling and retention and the reestablishment of wildlife habitat [5–7]. Restored wetlands have shown patterns of biogeochemical cycling including increased rates of carbon (C) sequestration and reduction in greenhouse gas emission [8], increased P sorption through modification of site hydrology over the course of restoration [9], and localized shifts in functional N-cycling

✉ Donald M. Walker
Donald.Walker@mtsu.edu

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

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

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

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

Published online: 02 April 2025

gene abundances at the watershed-scale [10]. 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 [11].

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 [12] and soil formation [13] and collectively drives the biogeochemical cycling of N, P, and organic C [14, 15]. 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 [11, 16]. Shifts in soil microbial assemblage structure and their associated functions have been shown to respond to several factors including hydrologic connectivity among sites [17], wetting and drying regimes [18], land-use practices [17, 19, 20], soil depth [21], and historic landscape use such as fertilizer amendments [19, 22]. Ecological restoration practice has been shown to respond to and positively impact soil biodiversity, and measured metabolic function correlates with restoration age [19, 23–25], 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 [26].

Ecological studies have historically focused on determining patterns of community composition and their distribution across the landscape [27]. The observed patterns in these studies are defined and constrained by the *scale* at which the study is conducted [27–29]. *Ecological scale* is defined by three axes: *spatial*, *temporal*, and *phylogenetic* [30, 31], each of which is composed of two elements, *grain* and *extent* [29, 32]. *Spatial scale* is determined by the distance, area, or volume being studied; *temporal scale* refers to the time period encompassing the sampling effort, while *phylogenetic scale* is defined by the taxonomic, functional, or trophic group(s) being investigated [31, 33]. *Grain* refers to the breadth of the smallest sampling unit, while *extent* refers to the scope of all samples collected in the study [31, 32]. 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 the 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 [34–36]. Trends in microbiome composition and function across soil depth have been well documented [37]. Past studies have shown that soil microbiomes exhibit significant millimeter-scale differences by depth, in addition to quantified abiotic differences at the same scale [38]. Given that abiotic factors, taxonomic composition and abundance, and functional nutrient cycling genes have been shown to vary as a result of soil depth [21, 37, 39], soil strata must be considered to build predictive models of nutrient cycling using microbiome composition [40].

Soils represent a multi-scalar continuum of ecological gradients and conditions rather than a single homogeneous environment [41]. 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 [41, 42]. Bacterial assemblages have been shown to respond to environmental changes with shifts in both $\alpha$- and $\beta$-diversity as a result of both deterministic and stochastic processes such as environmental filtering, diversification, dispersal limitation, and drift [33]. 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, our 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 (0–5 cm); (2) use quantitative PCR (qPCR) to determine whether measured nutrient turnover is predictive of functional nitrogen cycling gene abundance; and (3) develop predictive models of nutrient turnover at the landscape scale using bacterial microbiome composition. We 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 (Supplemental Fig. F1). 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 1 to 4 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 h) aluminum sleeve (Supplemental Fig. F2a). At each sampling location, a sample rod was inserted approximately 5 cm into the ground (Supplemental Fig. F2b). Soils collected within the spatula (1 cm diameter, ~ 5 cm of sediment) were then separated into three even "partitions" (~ 1.6 cm each; Supplemental Fig. F2ef), 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), placed on ice within 30 min of collection, and transported to Middle Tennessee State University (MTSU; Murfreesboro, TN) within 6 h 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 × 15 cm depth; Supplemental Fig. F2c, 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 (Supplemental Methods File C), while the second was used to measure total C, total N, and extractable phosphorus (Supplemental Methods File C).

## Flow-Through Incubation Experiment

A parallel flow-through incubation design (Supplemental Figs. F3, F4a, 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 [43–45]. Synthetic water (Supplemental Table E1) [46] was developed and maintained for all experimental incubations [47]. This experiment was designed to measure maximal $NO_3$ and $PO_4$ flux rates within the inundated experimental cores (Supplemental Methods File C).

Nutrient flux rates were calculated for each dissolved nutrient or gas using Eq. C1 (Supplemental Methods File C). 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 (Supplemental Methods Eq. C2) were calculated for each core that represented the average flux rate of a given nutrient over the course of the experiment.

## Post-Incubation Soil Sampling

Post-incubation, surface water was siphoned off, and large soil cores were transported to a BSL2 hood (Supplemental Fig. F4c, 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 (Supplemental Fig. F4d, 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-h 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 2 h. Approximately 1.25 g of soil was added into a 2 ml microcentrifuge tube pre-filled with 900 µL of sterile millipore water (2-h 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 was 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 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) [48], 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 nM 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 [49] using mothur v1.45.3 [50, 51] to quality control, process, and cluster amplicon sequence data into operational taxonomic units (OTUs) at 97% sequence similarity [52]. The SILVA v132 reference database was used to assign taxonomy to OTUs [53]. Raw sequence reads were deposited in the National Center for Biotechnology Information: BioProject PRJNA1129551. All mothur code is included in Supplemental File A.

Contaminating sequence reads were identified with algorithmic detection and removed (in R package *decontam*) using a probability threshold of 0.45 [54]. Post-decontamination, rare OTUs were removed if they did not meet a minimum abundance threshold of 10 reads across all samples [55, 56]. 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.

## 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 degenerative 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 [57], 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. [58] (Supplemental Methods File C).

## Statistical Analysis

### Measures of Assemblage Diversity

Bacterial taxonomic diversity ($\alpha$-diversity) and composition ($\beta$-diversity) were 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 [59]. 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 Eq. C3 (Supplemental Methods File C) [60–63].

### Modeling Assemblage Diversity

To evaluate the relationship between soil bacterial assemblage alpha ($\alpha$) 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 [64–67]. We tested for the effect of soil N concentration on assemblage richness and Hill-Shannon diversity. The relationship between soil N and measures of $\alpha$-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 [68], resulting in the 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 the final fit was then validated by examining the plot of residuals [69].

### 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 ($\beta$) 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 [70]. 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 (Supplemental Methods File C).

### 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 [71], and columns with non-zero sums were retained. Machine learning models were constructed using the *caret* package in R [72] to implement the Generalized Linear Models with Elastic Net regularization (GLMNET) [73].

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 [74, 75]. Values of NRMSE standardized by SD (0–1) represent the ratio of explained vs unexplained variation by the regression against the response variable [76]. 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 [74, 75].

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. 1a, Table 1) . Significant partial effects were detected for the smooths of soil N by sample partition (Fig. 1b) and soil N by habitat

type (Fig. 1c), 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 and 0.3% (Fig. 1b). 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 to 0.3%, shallow water habitats exhibit an increase in detected *effective* OTUs (Fig. 1c). 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. 1c). Additional analysis of trends in richness can be found in the supplement content (Supplemental Results File D, Supplemental Table E2, Supplemental Fig. F5).

## Assemblage Relative Abundance

An assessment of taxonomic relative abundance by sample partition (Supplemental Table E3) showed several statistically significant differences among partitions, within taxonomic groups (Fig. 2, Supplemental Table E4). 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 (Supplemental Table E4). No significant differences were detected among soil partitions within Alphaproteobacteria, Verrucomicrobia, Planctomycetes, 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 the field bottom and lab bottom partitions was the relative abundance of Nitrospirae, which showed significant enrichment in lab bottom partitions (Table 2).

Taxonomic relative abundance assessed by habitat type (Supplemental Table E5) revealed statistically significant differences across multiple taxonomic groups (Fig. 3, Supplemental Table E6). 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" (Supplemental Table E6). More significant differences across taxonomic groups occurred between remnant forest

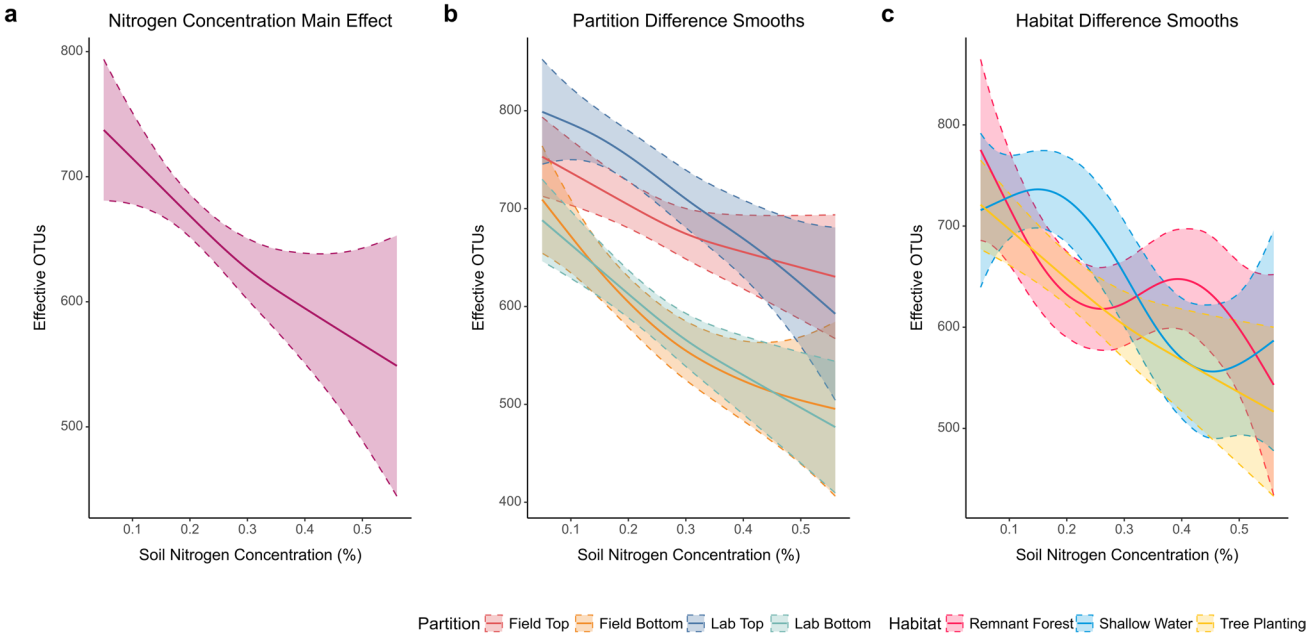


**Fig. 1** Effective OTUs decline with increased soil nitrogen concentration. Significant GAMM model smooths are shown with Bayesian confidence intervals for **a** the main effect of soil N concentration on effective OTUs colored in purple. Partial effects difference smooths depict the differences in 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

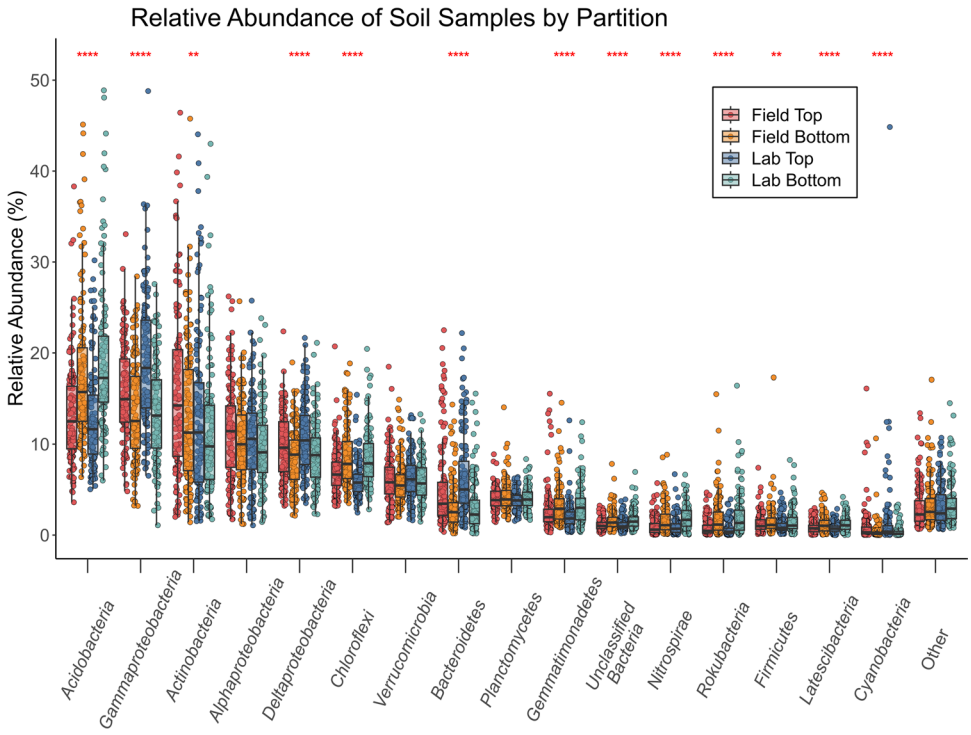


**Fig. 2** Assemblage relative abundance by partition. Box plots of relative abundance (%) by phylum and class (for phylum Proteobacteria) colored by soil sample partition. Red asterisks indicate a significant difference detected by the Kruskal–Wallis test between partitions by taxonomic group

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

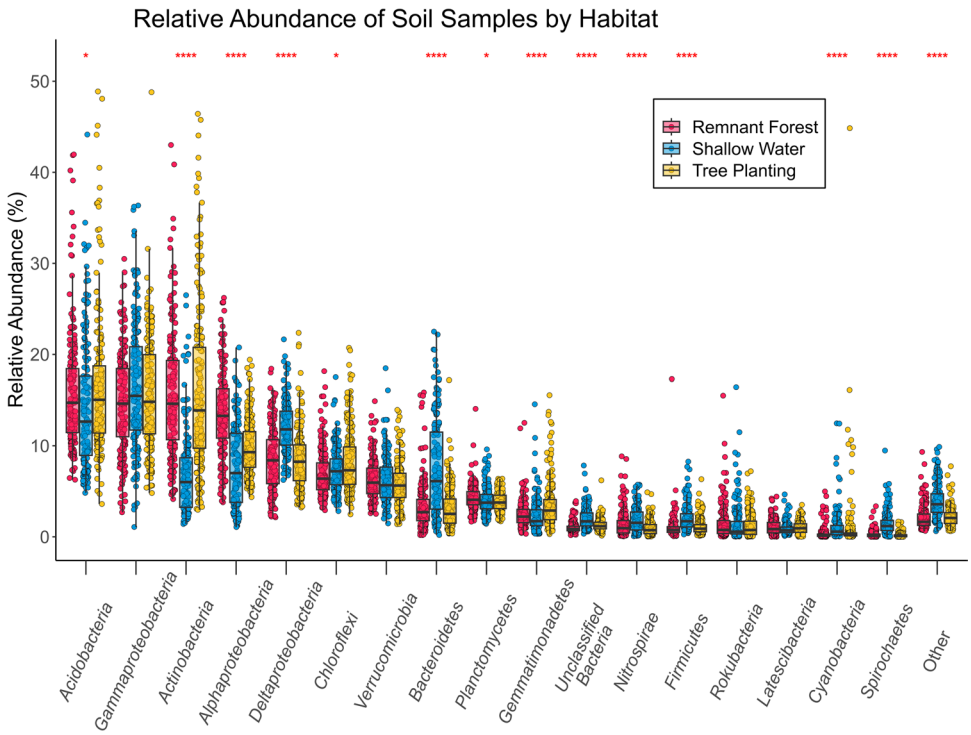


**Fig. 3** Assemblage relative abundance by habitat. Box plots of relative abundance (%) by phylum and class (for phylum Proteobacteria) colored by habitat type. Red asterisks indicate a significant difference as detected by the Kruskal–Wallis test between habitats by taxonomic group

bacterial assemblage taxonomic composition than either is to shallow water habitat (Table 3).

## Beta Diversity

Variation partitioning of assemblage composition assessed by the Bray–Curtis dissimilarity (Fig. 4a, b, Supplemental Table E7) showed significant effects of easement site, habitat (Fig. 4b), partition (Fig. 4a), richness, and soil N concentration, on the soil microbiome. Pairwise PERMANOVA of partitions showed significant differences for all but field bottom/lab bottom (Supplemental Table E8). Independent pairwise PERMANOVA analysis of habitat

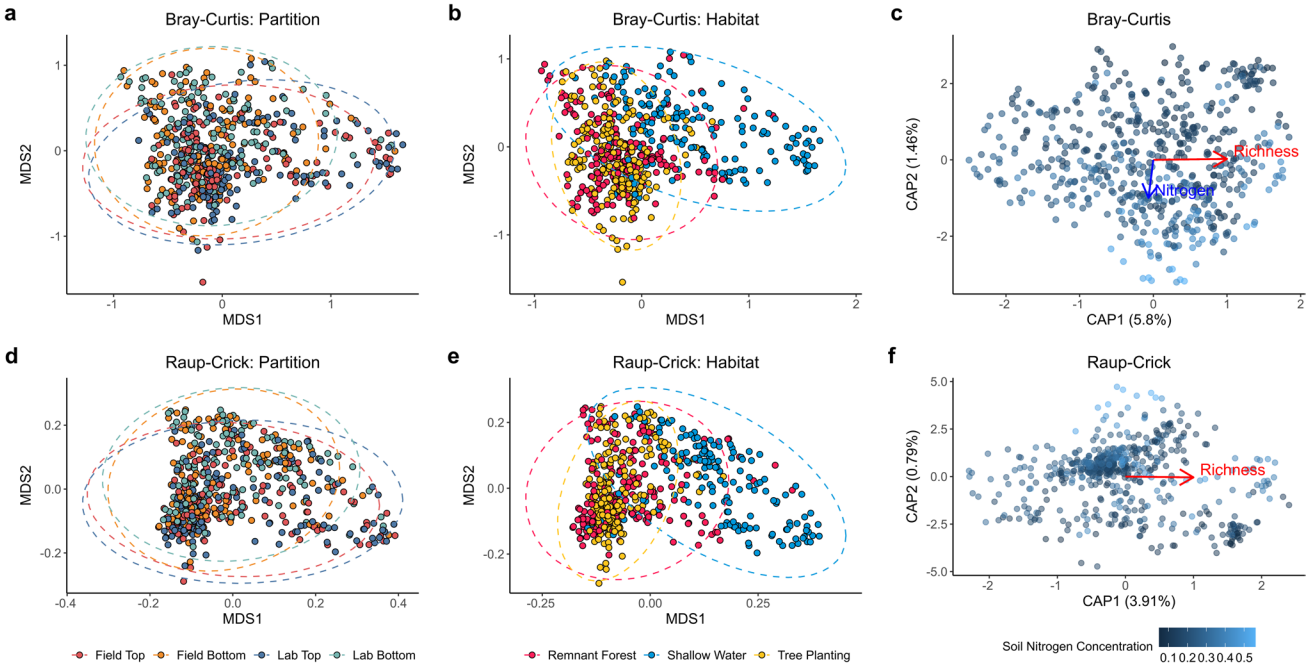


**Fig. 4** Ordinations of assemblage beta diversity. The first two dimensions of NMDS plots of the Bray–Curtis $\beta$-diversity index colored by partition (**a**) and habitat type (**b**), contrasted by plots of Raup-Crick $\beta$-diversity colored by partition (**d**) and habitat (**e**), are shown. 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)

**Table 1** GAMM results of the Hill-Shannon diversity model

| 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 $R^2$: 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

*SE* standard error, *e.d.f.* expected degrees of freedom, *Ref d.f.* reference degrees of freedom

**Table 2** Dunn's test results of relative abundance by pairwise partitions (Fig. 2)

| Taxonomic group | Field top—field bottom | | Field top—lab top | | Field top—lab bottom | | Field bottom—lab top | | Field bottom—lab bottom | | Lab top—lab bottom | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *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 |

Significant *p*-values are depicted in bold

**Table 3** Dunn's test results of relative abundance by pairwise habitats (Fig. 3)

| Taxonomic group | Remnant forest—shallow water | | Remnant forest—tree planting | | Shallow water—tree planting | |
|---|---|---|---|---|---|---|
| | *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** |

Significant *p*-values are depicted in bold

showed significant differences across all combinations of habitat type comparisons (Supplemental Table E8).

Variation partitioning of assemblage composition assessed using Raup-Crick dissimilarity (Fig. 4d, e, Supplemental Table E9) revealed significant effects of easement site, habitat (Fig. 4e), partition (Fig. 4d), 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 (Supplemental Table E8). Independent pairwise PERMANOVA of habitat showed significant differences across all combinations of habitat type (Supplemental Table E8). 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. 4c; 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 the increase in richness increased along the CAP1 axis. We 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 had a more negative CAP2 value. The Raup-Crick dbRDA (Fig. 4e; 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.

## 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 (Supplemental Table E10). 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. 5a), as was for tree planting in lab bottoms (Fig. 5c). While log copy numbers of nosZ-B did not differ significantly among partitions (Supplemental Table E11), significant smooth terms were detected for the three-way smooth interaction of $N_2$, partition, and habitat

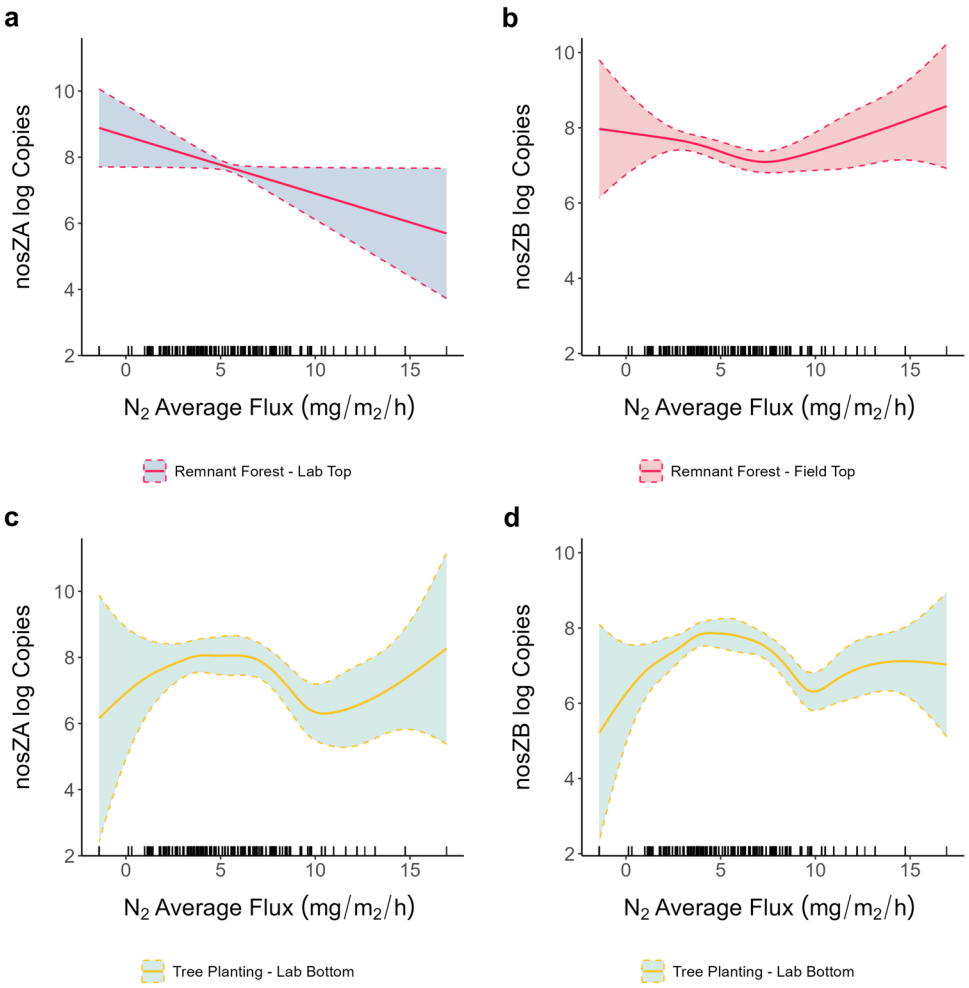


**Fig. 5** Trends in nosZ-A and nosZ-B gene abundance estimated with $N_2$ flux rates significant GAMM model smooths depicted with Bayesian confidence intervals and rug plots of sampling effort are shown for the three-way interactions of $N_2$ predicting noszA log copies with remnant forest lab tops (**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**)

factors that included remnant forest in field tops (Fig. 5b) and tree planting in lab bottoms (Fig. 5d). Additional significant parametric effects can be found in Supplementary Results, in addition to trends for other functional gene markers.

## Machine Learning and Deep Neural Networks

GLMNET models showed significant differential predictive accuracy of soil partitions across nutrient flux rates (Fig. 6a). 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. 6b) 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).

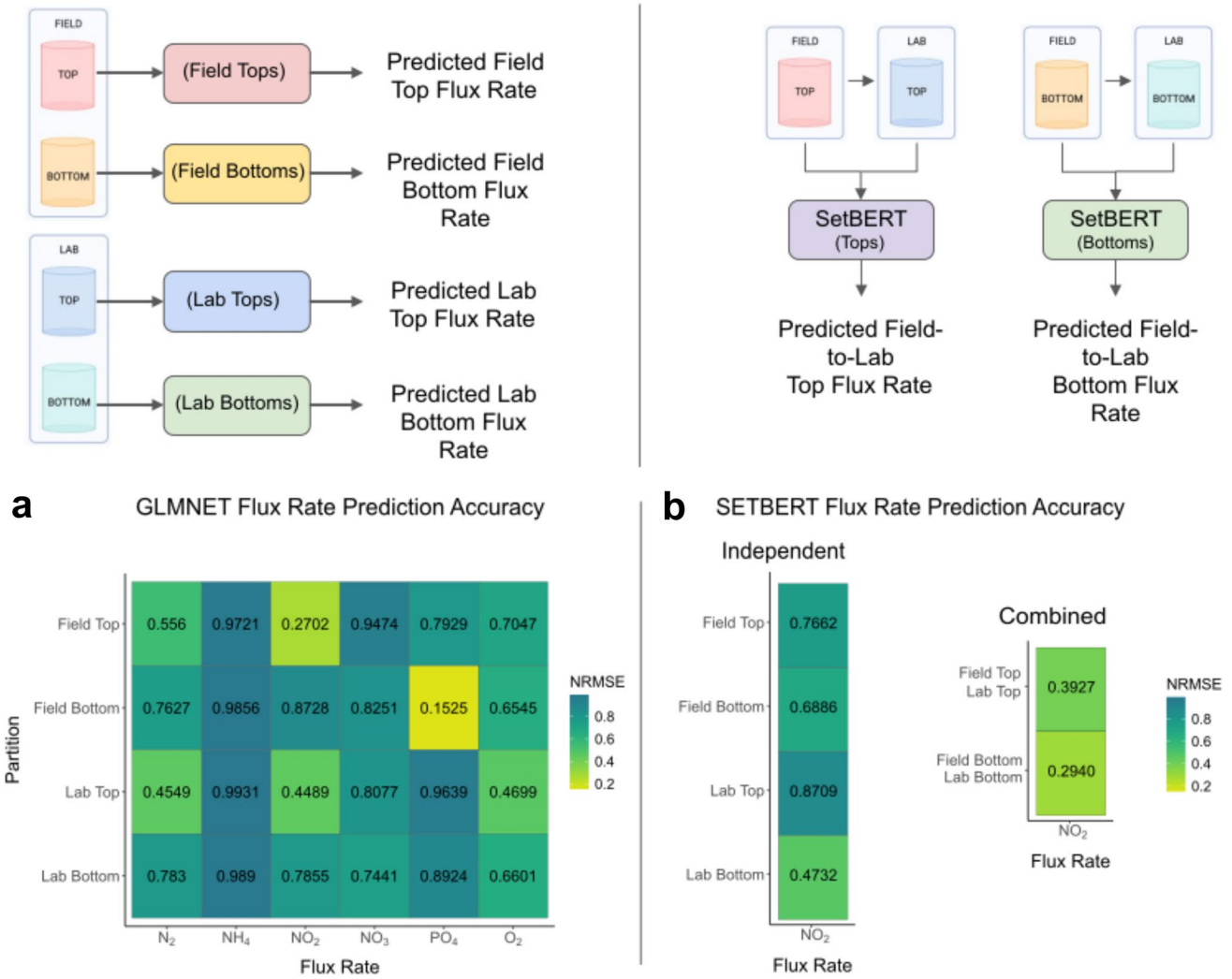


**Fig. 6** Machine learning and deep neural network prediction of nutrient flux rates. Accuracy of independent GLMNET models predicting nutrient flux rates from partition microbial assemblage OTU counts aggregated to genera (**a**). SetBERT model accuracy (**b**) of predicted $NO_2$ model accuracy for 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 (Supplemental Results File D) and Hill-Shannon diversity (Fig. 1a) indicate that bacterial assemblage diversity decreased with N concentration in sampled soils. These results are consistent with previous studies that have shown bacterial phylotype [77], relative abundance [78], and plant diversity [79] 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 [10, 42, 80] across microhabitats on the landscape and its interaction with N concentration. This differential response was strongly linked to moisture gradients [17, 18], soil formation processes [13], and soil structure [21]. 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 [81, 82] or may also be in response to historical agricultural practices [8, 17, 19, 20] and their lasting effect on microbial diversity [83].

Significant differences in the relative abundance of sampled microbial assemblages were observed when the phylogenetic scale was subdivided into phylum and class level (Figs. 2 and 3). These differences align with the potential environmental heterogeneity among soil partitions (Fig. 2), controlled lab incubations (Fig. 2), and restoration practices (Fig. 3). These findings are consistent with previous literature which highlights microbial compositional differences driven by soil depth [21, 37, 38, 41], moisture availability [17, 18], and microhabitats [25, 41, 84]. The importance of *spatial* and *temporal scales* 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. 2). Members of the phylum Nitrospirae represent a diverse range of aerobic chemolithotrophs that are involved in key nutrient cycling pathways, including nitrification [85].

Soil assemblage $\beta$-diversity, as assessed using the abundance-based Bray–Curtis metric, was primarily influenced by soil partition (Fig. 4a) and habitat type (Fig. 4b). A strata-specific response was observed for field and lab top samples, consistent with changes in $\alpha$-diversity described earlier. These results suggest that top soils may experience rapid shifts in abundance during the incubation experiments [26]. Raup-Crick dissimilarity revealed significant turnover between partitions (Fig. 4d) and habitat types (Fig. 4e). 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. 4c) indicated that the N concentration on the landscape significantly contributed to differences in abundance across the landscape (Fig. 4c), while Raup-Crick analysis revealed factors other than N concentration are key contributors to measured richness (Fig. 4f). Past studies have shown significant shifts in $\beta$-diversity associated with N amendment on the landscape [21, 84], and our results highlight the importance of using traditional metrics alongside null models to determine whether $\beta$-diversity shifts are significantly related to trends in measured $\alpha$-diversity. Patterns observed across $\alpha$- and $\beta$-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 can be estimated using $N_2$ gas flux, and whether also 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 [40]. 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 $\alpha$- and $\beta$-diversity metrics, may well be the driving factors 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 [84]. 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 the predictive accuracy of landscape functions utilizing soil microbiome composition, as they may incorporate data that spans the entirety of the ecological scale to produce better-fitting models and bring a picture of microbial dynamics into focus.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00248-025-02516-1.

**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 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 those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or Natural Resource Conservation Service.

**Author Contributions** Study conceptualization and design was conducted by N.R.A. Material preparation and data collection were performed by N.R.A., R.S.B., S.D., S.G.W. and K.C.M. Formal analysis was conducted by N.R.A. and D.W.L. and J.L.P. developed and contributed to setBERT analysis. Funding acquisition was conducted by D.M.W., J.N.M., and J.L.P. The main manuscript text was written by N.R.A. Figures in the manuscript were prepared by N.R.A. with contributions from D.W.L and J.L.P. for Fig. 6 and supplemental figure F8. All authors reviewed and commented on previous versions of the manuscript and approved the final manuscript.

**Funding** This work was supported by the Natural Resource Conservation Services sub-award 220858, the National Science Foundation grants EF-2125065 and CAREER 2236580 to D. Walker, the Molecular Biosciences Program at MTSU, and the National Science Foundation grant DEB 1933925 to J. Phillips.

**Data Availability** Raw sequence reads were deposited in the National Center for Biotechnology Information: BioProject PRJNA1129551. Analysis code is provided within the supplementary information files. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Competing Interests** The authors declare no competing interests.

**Disclaimer** Opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation or Natural Resource Conservation Service.



## References

1. Glibert PM (2020) Harmful algae at the complex nexus of eutrophication and climate change. Harmful Algae 91:101583. https://doi.org/10.1016/j.hal.2019.03.001
2. Glibert PM, Maranger R, Sobota DJ, Bouwman L (2014) The Haber Bosch–harmful algal bloom (HB–HAB) link. Environ Res Lett 9:105001. https://doi.org/10.1088/1748-9326/9/10/105001
3. Nazari-Sharabian M, Ahmad S, Karakouzian M (2018) Climate change and eutrophication: a short review. Eng Technol Appl Sci Res 8:3668–3672. https://doi.org/10.48084/etasr.2392
4. Comín FA, Romero JA, Astorga V, García C (1997) Nitrogen removal and cycling in restored wetlands used as filters of nutrients for agricultural runoff. Water Sci Technol 35:255–261. https://doi.org/10.1016/S0273-1223(97)00076-0
5. Mitsch WJ, Bernal B, Nahlik AM et al (2013) Wetlands, carbon, and climate change. Landsc Ecol 28:583–597
6. Mitsch WJ, Gosselink JG (2000) The value of wetlands: importance of scale and landscape setting. Ecol Econ 35:25–33. https://doi.org/10.1016/S0921-8009(00)00165-8
7. Zedler JB, Kercher S (2005) Wetland resources: status, trends, ecosystem services, and restorability. Annu Rev Environ Resour 30:39–74. https://doi.org/10.1146/annurev.energy.30.050504.144248
8. Hemes KS, Chamberlain SD, Eichelmann E et al (2019) Assessing the carbon and climate benefit of restoring degraded agricultural peat soils to managed wetlands. Agric For Meteorol 268:202–214. https://doi.org/10.1016/j.agrformet.2019.01.017
9. Xu G, Ren Y, Yue M et al (2022) Phosphorus sorption capacity in soils from freshwater restored coastal wetlands increased with restoration age. Geoderma 422:115926. https://doi.org/10.1016/j.geoderma.2022.115926
10. Kasak K, Espenberg M, Anthony TL et al (2021) Restoring wetlands on intensive agricultural lands modifies nitrogen cycling microbial communities and reduces N2O production potential. J Environ Manage 299:113562. https://doi.org/10.1016/j.jenvman.2021.113562
11. Correa-Garcia S, Constant P, Yergeau E (2023) The forecasting power of the microbiome. Trends Microbiol 31:444–452. https://doi.org/10.1016/j.tim.2022.11.013
12. Chaparro JM, Sheflin AM, Manter DK, Vivanco JM (2012) Manipulating the soil microbiome to increase soil health and plant fertility. Biol Fertil Soils 48:489–499. https://doi.org/10.1007/s00374-012-0691-4
13. Sokol NW, Slessarev E, Marschmann GL et al (2022) Life and death in the soil microbiome: how ecological processes influence biogeochemistry. Nat Rev Microbiol 20:415–430. https://doi.org/10.1038/s41579-022-00695-z
14. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034–1039. https://doi.org/10.1126/science.1153213
15. Lappalainen HK, Kerminen V-M, Petäjä T et al (2016) 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. Atmos Chem Phys 16:14421–14461. https://doi.org/10.5194/acp-16-14421-2016
16. O'Brien SL, Gibbons SM, Owens SM et al (2016) Spatial scale drives patterns in soil bacterial diversity: spatial scale drives soil diversity. Environ Microbiol 18:2039–2051. https://doi.org/10.1111/1462-2920.13231
17. Tomasek A, Staley C, Wang P et al (2017) Increased denitrification rates associated with shifts in prokaryotic community composition caused by varying hydrologic connectivity. Front Microbiol 8:2304. https://doi.org/10.3389/fmicb.2017.02304
18. Veach AM, Zeglin LH (2020) Historical drought affects microbial population dynamics and activity during soil drying and re-wet. Microb Ecol 79:662–674. https://doi.org/10.1007/s00248-019-01432-5
19. Dai Z, Liu G, Chen H et al (2020) Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. ISME J 14:757–770. https://doi.org/10.1038/s41396-019-0567-9
20. Rasche F, Cadisch G (2013) The molecular microbial perspective of organic matter turnover and nutrient cycling in tropical agroecosystems - What do we know? Biol Fertil Soils 49:251–262. https://doi.org/10.1007/s00374-013-0775-9
21. Tang Y, Yu G, Zhang X et al (2018) Changes in nitrogen-cycling microbial communities with depth in temperate and subtropical forest soils. Appl Soil Ecol 124:218–228. https://doi.org/10.1016/j.apsoil.2017.10.029
22. Creamer RE, Hannula SE, Leeuwen JPV et al (2016) Ecological network analysis reveals the inter-connection between soil biodiversity and ecosystem function as affected by land use across Europe. Appl Soil Ecol 97:112–124. https://doi.org/10.1016/j.apsoil.2015.08.006
23. Cookson WR, Osman M, Marschner P et al (2007) Controls on soil nitrogen cycling and microbial community composition across land use and incubation temperature. Soil Biol Biochem 39:744–756. https://doi.org/10.1016/j.soilbio.2006.09.022
24. Fernández-Martínez MA, Pointing SB, Pérez-Ortega S et al (2017) Functional ecology of soil microbial communities along a glacier forefield in Tierra del Fuego (Chile). Int Microbiol 161–173. https://doi.org/10.2436/20.1501.01.274
25. Wang C, Wang S, Wu B et al (2021) Ecological restoration treatments enhanced plant and soil microbial diversity in the degraded alpine steppe in Northern Tibet. Land Degrad Dev 32:723–737. https://doi.org/10.1002/ldr.3754
26. Urakawa H, Bernhard AE (2017) Wetland management using microbial indicators. Ecol Eng 108:456–476. https://doi.org/10.1016/j.ecoleng.2017.07.022
27. Schneider DC (2001) The rise of the concept of scale in ecology. Bioscience 51:545. https://doi.org/10.1641/0006-3568(2001)051[0545:TROTCO]2.0.CO;2

28. Levin SA (1992) The problem of pattern and scale in ecology: the Robert H MacArthur Award Lecture. Ecology 73:1943–1967. https://doi.org/10.2307/1941447
29. Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3:385. https://doi.org/10.2307/2389612
30. Graham CH, Storch D, Machac A, Isaac N (2018) Phylogenetic scale in ecology and evolution. Glob Ecol Biogeogr 27:175–187. https://doi.org/10.1111/geb.12686
31. Ladau J, Eloe-Fadrosh EA (2019) Spatial, temporal, and phylogenetic scales of microbial ecology. Trends Microbiol 27:662–669. https://doi.org/10.1016/j.tim.2019.03.003
32. O'Neill RV, Johnson AR, King AW (1989) A hierarchical framework for the analysis of scale. Landsc Ecol 3:193–205. https://doi.org/10.1007/BF00131538
33. Vellend M (2010) Conceptual synthesis in community ecology. Q Rev Biol 85:183–206. https://doi.org/10.1086/652373
34. Hartemink AE, Zhang Y, Bockheim JG et al (2020) Soil horizon variation: a review. In: Advances in Agronomy. Elsevier, pp 125–185
35. Ivanova EA, Pershina EV, Shapkin VM et al (2020) Shifting prokaryotic communities along a soil formation chronosequence and across soil horizons in a South Taiga ecosystem. Pedobiologia 81–82:150650. https://doi.org/10.1016/j.pedobi.2020.150650
36. Naumova NB, Belanov IP, Alikina TY, Kabilov MR (2021) Undisturbed soil pedon under birch forest: characterization of microbiome in genetic horizons. Soil Syst 5:14. https://doi.org/10.3390/soilsystems5010014
37. Naylor D, McClure R, Jansson J (2022) Trends in microbial community composition and function by soil depth. Microorganisms 10:540. https://doi.org/10.3390/microorganisms10030540
38. Cai Y-J, Liu Z-A, Zhang S et al (2022) Microbial community structure is stratified at the millimeter-scale across the soil–water interface. ISME Commun 2:53. https://doi.org/10.1038/s43705-022-00138-z
39. Qian L, Yu X, Gu H et al (2023) Vertically stratified methane, nitrogen and sulphur cycling and coupling mechanisms in mangrove sediment microbiomes. Microbiome 11:71. https://doi.org/10.1186/s40168-023-01501-5
40. Young JM, Skvortsov T, Kelleher BP et al (2019) Effect of soil horizon stratigraphy on the microbial ecology of alpine paleosols. Sci Total Environ 657:1183–1193. https://doi.org/10.1016/j.scitotenv.2018.11.442
41. Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15:579–590. https://doi.org/10.1038/nrmicro.2017.87
42. Furtak K, Gałązka A (2019) Edaphic factors and their influence on the microbiological biodiversity of the soil environment. Postępy Mikrobiol - Adv Microbiol 58:375–384. https://doi.org/10.21307/PM-2019.58.4.375
43. Groffman PM, Altabet MA, Böhlke JK et al (2006) Methods for measuring denitrification: diverse approaches to a difficult problem. Ecol Appl 16:2091–2122. https://doi.org/10.1890/1051-0761(2006)016[2091:MFMDDA]2.0.CO;2
44. Miller-Way T, Twilley R (1996) Theory and operation of continuous flow systems for the study of benthic-pelagic coupling. Mar Ecol Prog Ser 140:257–269. https://doi.org/10.3354/meps140257
45. Nifong RL, Taylor JM, DeVilbiss S (2022) Spatial and temporal patterns of benthic nutrient cycling define the extensive role of internal loading in an agriculturally influenced oxbow lake. Biogeochemistry 159:413–433. https://doi.org/10.1007/s10533-022-00935-7
46. Brown RS (2023) Assessing nutrient retention and removal measurements among restored floodplain wetlands. Dissertation, Tennessee Technological University
47. Speir SL, Taylor JM, Scott JT (2017) Seasonal differences in relationships between nitrate concentration and denitrification rates in ditch sediments vegetated with rice cutgrass. J Environ Qual 46:1500–1509. https://doi.org/10.2134/jeq2016.11.0450
48. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303
49. Kozich JJ, Westcott SL, Baxter NT et al (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120. https://doi.org/10.1128/AEM.01043-13
50. Schloss PD (2020) Reintroducing mothur: 10 years later. Appl Environ Microbiol 86:e02343-e2419. https://doi.org/10.1128/aem.02343-19
51. Schloss PD, Westcott SL, Ryabin T et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541
52. Grajal-Puche A, Murray CM, Kearley M et al (2020) Microbial assemblage dynamics within the American alligator nesting ecosystem: a comparative approach across ecological scales. Microb Ecol 80:603–613. https://doi.org/10.1007/s00248-020-01522-9
53. Quast C, Pruesse E, Yilmaz P et al (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219
54. Davis NM, Proctor DM, Holmes SP et al (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6:226. https://doi.org/10.1186/s40168-018-0605-2
55. Brown SP, Veach AM, Rigdon-Huss AR et al (2015) Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecol 13:221–225. https://doi.org/10.1016/j.funeco.2014.08.006
56. Cao Q, Sun X, Rajesh K et al (2021) Effects of rare microbiome taxa filtering on statistical analysis. Front Microbiol 11:607325. https://doi.org/10.3389/fmicb.2020.607325
57. Hallin S, Philippot L, Löffler FE et al (2018) Genomics and ecology of novel N2O-reducing microorganisms. Trends Microbiol 26:43–55. https://doi.org/10.1016/j.tim.2017.07.003
58. Keeley RF, Rodriguez-Gonzalez L, Class USFG et al (2020) Degenerate PCR primers for assays to track steps of nitrogen metabolism by taxonomically diverse microorganisms in a variety of environments. J Microbiol Methods 175:105990. https://doi.org/10.1016/j.mimet.2020.105990
59. R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
60. Bullen PS (2003) The power means. In: Handbook of means and their inequalities. Mathematics and its applications, vol 560. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-0399-4_3
61. Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432. https://doi.org/10.2307/1934352
62. Jost L (2006) Entropy and diversity. Oikos 113:363–375. https://doi.org/10.1111/j.2006.0030-1299.14714.x
63. Roswell M, Dushoff J, Winfree R (2021) A conceptual guide to measuring species diversity. Oikos 130:321–338. https://doi.org/10.1111/oik.07202
64. Wood SN (2003) Thin-plate regression splines. J R Stat Soc B 65:95–114. https://doi.org/10.1111/1467-9868.00374
65. Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc B 73:3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x
66. Wood SN (2017) Generalized additive models: an introduction with R, 2nd ed. Chapman and Hall/CRC

67. Wood SN, Pya N, Säfken B (2016) Smoothing parameter and model selection for general smooth models (with discussion). J Am Stat Assoc 111:1548–1575. https://doi.org/10.1080/01621459.2016.1180986
68. Pedersen EJ, Miller DL, Simpson GL, Ross N (2019) Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 7:e6876. https://doi.org/10.7717/peerj.6876
69. Simpson G (2018) Introducing gratia. Bottom Heap
70. Chase JM, Kraft NJB, Smith KG et al (2011) Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2:art24. https://doi.org/10.1890/ES10-00117.1
71. Armour CR, Topçuoğlu BD, Garretto A, Schloss PD (2022) A Goldilocks principle for the gut microbiome: taxonomic resolution matters for microbiome-based classification of colorectal cancer. 13:. https://doi.org/10.1128/mbio.03161-21
72. Kuhn M (2008) Building predictive models in *R* using the **caret** package. J Stat Softw 28:. https://doi.org/10.18637/jss.v028.i05
73. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33:. https://doi.org/10.18637/jss.v033.i01
74. Moriasi DN, Arnold JG, Van Liew MW et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50:885–900. https://doi.org/10.13031/2013.23153
75. Ranatunga T, Tong STY, Yang YJ (2016) An approach to measure parameter sensitivity in watershed hydrological modelling. Hydrol Sci J 1–17. https://doi.org/10.1080/02626667.2016.1174335
76. Otto SA (2019) How to normalize the RMSE. In: www.marinedatascience.co, https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/. Accessed 10/24/2023
77. Yuan X, Knelman JE, Wang D et al (2017) Patterns of soil bacterial richness and composition tied to plant richness, soil nitrogen, and soil acidity in alpine tundra. Arct Antarct Alp Res 49:441–453
78. Li J-G, Shen M-C, Hou J-F et al (2016) Effect of different levels of nitrogen on rhizosphere bacterial community structure in intensive monoculture of greenhouse lettuce. Sci Rep 6:25305
79. Midolo G, Alkemade R, Schipper AM et al (2019) Impacts of nitrogen addition on plant species richness and abundance: a global meta-analysis. Glob Ecol Biogeogr 28:398–413
80. Thakur MP, Wright AJ (2017) Environmental filtering, niche construction, and trait variability: the missing discussion. Trends Ecol Evol 32:884–886
81. Connell JH, Slatyer RO (1977) Mechanisms of succession in natural communities and their role in community stability and organization. Am Nat 111:1119–1144
82. McCook L (1994) Understanding ecological community succession: causal models and theories, a review. Vegetatio 110:115–147
83. Gupta A, Singh UB, Sahu PK et al (2022) Linking soil microbial diversity to modern agriculture practices: a review. Int J Environ Res Public Health 19:3141. https://doi.org/10.3390/ijerph19053141
84. Mueller RC, Belnap J, Kuske CR (2015) Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. Front Microbiol 6:157127. https://doi.org/10.3389/fmicb.2015.00891
85. Paul E, Frey S (2023) Soil microbiology, ecology and biochemistry. Elsevier