Developing Indicators of Nutrient Pollution in Streams Using 16S rRNA Gene Metabarcoding of Periphyton-Associated Bacteria

Erik M. Pilgrim¹,*, Nathan J. Smucker¹, Huiyun Wu², John Martinson¹, Christopher T. Nietch¹, Marirosa Molina³, John A. Darling³, Brent R. Johnson¹
¹United States Environmental Protection Agency, Office of Research and Development, Cincinnati, OH 45268, USA
²School of Public Health & Tropical Medicine, Tulane University, New Orleans, LA 70112, USA
³United States Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC 27711, USA

Abstract

Indicators based on nutrient-biota relationships in streams can inform water quality restoration and protection programs. Bacterial assemblages could be particularly useful indicators of nutrient effects because they are species-rich, important contributors to ecosystem processes in streams, and responsive to rapidly changing conditions. Here, we sampled 25 streams weekly (12–14 times each) and used 16S rRNA gene metabarcoding of periphyton-associated bacteria to quantify the effects of total phosphorus (TP) and total nitrogen (TN). Threshold indicator taxa analysis identified assemblage-level changes and amplicon sequence variants (ASVs) that increased or decreased with increasing TP and TN concentrations (i.e., low P, high P, low N, and high N ASVs). Boosted regression trees confirmed that relative abundances of gene sequence reads for these four indicator groups were associated with nutrient concentrations. Gradient forest analysis complemented these results by using multiple predictors and random forest models for each ASV to identify portions of TP and TN gradients at which the greatest changes in assemblage structure occurred. Synthesized statistical results showed bacterial assemblage structure began changing at 24 μg TP/L with the greatest changes occurring from 110 to 195 μg/L. Changes in the bacterial assemblages associated with TN gradually

---

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

*Correspondence: pilgrim.erik@epa.gov.

Author Contributions: E.M.P., N.J.S. and C.T.N. designed the project with input from M.M., J.A.D. and B.R.J., N.J.S., H.W., J.M. and E.M.P. provided the data analyses. All authors contributed to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/w14152361/s1. Table S1: Site Coordinates; Table S2: Total phosphorus TITAN results for amplicon sequence variants (ASV) with >95% purity and 95% reliability, change points (CP), sample frequencies (Freq), z-scores, and percentile distributions of bootstrapped change points. Indicator column denotes high P or low P ASVs that increased or decreased, respectively, with greater TP concentrations; Table S3: Total nitrogen TITAN results for amplicon sequence variants (ASV) with >95% purity and 95% reliability, change points (CP), sample frequencies (Freq), z-scores, and percentile distributions of bootstrapped change points. Indicator column denotes high N or low N ASVs that increased or decreased, respectively, with greater TN concentrations; Table S4: Amplicon sequence variants identified as having random forest models with R² values > 0 in gradient forest analysis; Figure S1: Comparison of TP and TN concentrations to watershed area.
occurred from 275 to 855 μg/L. Taxonomic and phylogenetic analyses showed that low nutrient ASVs were commonly Firmicutes, Verrucomicrobiota, Flavobacterales, and Caulobacteriales, Pseudomonadales, and Rhodobacterales of Proteobacteria, whereas other groups, such as Chitinophagales of Bacteroidota, and Burkholderiales, Rhizobiales, Sphingomonadales, and Steroidobacterales of Proteobacteria comprised the high nutrient ASVs. Overall, the responses of bacterial ASV indicators in this study highlight the utility of metabarcoding periphyton-associated bacteria for quantifying biotic responses to nutrient inputs in streams.

**Keywords**

phosphorus; nitrogen; agriculture; periphyton; biomonitoring; bioassessment; 16S; gradient forest; boosted regression trees; threshold indicator taxa analysis; TITAN

1. Introduction

Nutrient pollution of freshwater ecosystems is a persistent and growing concern around the world [1-3], often being the most common stressor affecting streams [4-7]. Inputs of nitrogen or phosphorus create environmental stress by causing changes in the primary producer communities, reductions in biodiversity, alterations to food webs, hypoxia, and harmful algal blooms [8-12]. These negative impacts are expected to continue to increase as human populations grow, requiring more food production and watershed development, leading to further stress and alteration of freshwater and coastal ecosystems [13-15].

The development of methods for measuring ecosystem responses to nutrient pollution is therefore crucial for understanding how to mitigate or prevent ecological damage. Decision makers, regulatory agencies, and other end users need the best tools possible to measure, address, and alleviate the negative effects of nutrient pollution.

In aquatic environments, the benthic bacterial assemblage is part of the periphyton (or stream biofilm community) that is critical to biogeochemical processing. Stream biogeochemistry is driven by water quality and abiotic-biotic interactions within the periphyton matrix [16,17]. Bacteria are biogeochemical workhorses, responsible for important processes such as nitrification, denitrification, metabolism of pollutants, and producing hydrolytic enzymes that promote the turnover of organic carbon and nutrients [18-21]. These bacteria-mediated processes, coupled with their abilities to quickly replicate and increase biomass in response to changes in environmental conditions [22,23], offer a basal response to changing nutrient conditions in streams [24]. As a result, the responses of the bacterial component of stream periphyton to nitrogen and phosphorus concentrations can be used as indicators that provide context for understanding nutrient effects on streams [25,26]. Tools that help elucidate periphytic microbial responses to nutrient pollution therefore can provide valuable information for understanding and managing the effects of nitrogen and phosphorus pollution.

Aquatic bacterial assemblages, however, can be challenging to characterize. Many microbes are impossible to culture and study in the laboratory, including the bulk of taxa that would be of interest in ecological and environmental studies [27,28]. For this reason, molecular genetic techniques like DNA metabarcoding have been employed to sample and analyze
a variety of microbial communities [17,29]. DNA metabarcoding has become a valuable tool for gathering extensive amounts of microbial community data that would have been very costly or nearly impossible a few years ago [30-34]. Recent studies in freshwater environments have shown bacterial assemblage responses to nutrients [26,35,36], flow and fine sediment [37], temperature and pH [36], and conductivity [35].

In this study of streams in southwestern Ohio, USA, we applied DNA metabarcoding to the bacterial assemblages associated with periphyton. Using samples collected weekly from sites representing a broad range of nitrogen and phosphorus conditions within a large watershed, we analyzed bacterial taxa, defined as amplicon sequence variants (ASVs), and the associated water chemistry data with threshold indicator taxa analysis (TITAN; [38]), boosted regression trees (BRT; [39,40]), and gradient forests (GF; [41]) to determine how bacterial assemblages and individual taxa respond to different nutrient concentrations. We also used phylogenetic analysis to show the distribution of high and low nutrient indicator taxa among the major bacterial groups.

2. Materials and Methods

2.1. Site Selection and Sampling

Twenty-five sites in second- to third-order wadeable streams within the East Fork of the Little Miami River watershed of southwestern Ohio (USA) (Figure 1, Supplementary Table S1) were chosen based on historical monitoring data and site characterizations to comprise a gradient of both nitrogen and phosphorus concentrations across sites while minimizing possible confounding effects of non-nutrient factors. The watershed experiences a temperate seasonal climate and the 25 stream sites covered a range of watershed sizes (15.8–82.3 km², median 37.6 km², mean 40.3 km² ± 16.4 km²), percent agriculture coverage (0–88%, median 53%, mean 48% ± 29%), percent forest coverage (9–59%, median 32%, mean 33% ± 14%), and percent urban coverage (0–69%, median 8%, mean 19% ± 20%).

Sites were sampled weekly from early July through September 2016 with another 1–2 samples taken in October. Each site had 13–14 periphyton samples (n = 342 among sites) and 10–12 nutrient samples collected (TP n = 281, TN n = 280 among sites). At each site, water was collected in 1-L acid-washed polypropylene bottles for nutrient analyses and periphyton was scraped and composited from five rocks from five equally spaced transects over a 75-m stream reach collected with a firm-bristled brush on a cordless drill using a 6.7 cm² plastic guide. Periphyton samples were stored frozen at −80 °C until thawed for DNA analysis. Water samples for total phosphorus (TP) and total nitrogen (TN) were stored in the dark at 4 °C until analyzed within 24 h or were frozen at −20 °C for analysis within 14 days using a QuikChem 8500 nutrient autoanalyzer system (Lachat Instruments, Milwaukee, WI, USA). Acid persulfate wet digestions were conducted prior to using the molybdate and antimony potassium tartrate reaction and ascorbic acid reduction method to measure TP [42,43]. Total nitrogen was measured using an alkaline wet oxidation persulfate method prior to cadmium reduction [44,45].
2.2. DNA Metabarcoding Workflow

Upon slow thawing in a refrigerator, samples were immediately filtered through sterile polycarbonate (0.8 μm) filters. These filters were chosen to capture the algal and cyanobacterial portion of the periphyton (see [46]) but also were expected to capture most of the bacteria present in the periphyton. We recognize, however, that some loss of the smallest members (<0.8 μm) of the bacterial community could have occurred. Material collected on the filter was scraped and subsampled (~50 mg) for DNA extraction. This periphyton subsample was then ground in a 1.5-mL microcentrifuge tube with a pestle after treatment with liquid nitrogen, and the ground sample was extracted with the Qiagen DNeasy PowerLyzer PowerSoil kit (Germantown, MD, USA) according to the manufacturer’s protocol, with the addition of initial tissue digestion with proteinase K for at least 2 h at 56°C. Extraction blanks were processed approximately every 100 samples. Isolated DNA was quantified with PicoGreen on a Bio-Tek Synergy HT1 Microplate Reader (Winooski, VT, USA) and then normalized to 10 ng/μL prior to amplification by polymerase chain reaction (PCR).

PCR was used on the normalized DNA samples to amplify a segment of the 16S ribosomal gene covering parts of variable regions V3 to V5, targeting bacteria. This genetic locus is commonly used in metabarcoding studies of microbial communities [33,47,48]. PCRs, including negative controls, were run at 20 μL final volumes that included 2 μL of Qiagen 10X PCR buffer (with MgCl₂), 0.6 μL of 25 mM MgCl₂, 1 μL each of the primers (forward, 515F: GTGCCAGCMGCCGCGGTAA; reverse, 806R: GGACTACHVGGGTWTCTAAT), 0.4 μL of 10 mM dNTPs, 4 μL of 1X BSA, 8.9 μL sterile water, 0.1 μL Taq polymerase (Qiagen), and 2 μL of normalized template DNA. The reaction program for the thermal cycler was 94 °C for 2 m 30 s, then 35 cycles of 94 °C for 30 s, 50 ° C for 60 s, and 72 °C for 60 s and a final 10m extension at 72 °C. Each template was amplified in triplicate and those replicates were then pooled and cleaned (Qiagen Qiaquick 96 PCR Purification Kit). Negative controls and extraction blanks were not included beyond this step as amplicon bands were not present.

Cleaned PCR amplicons went through a second round of PCR (dual indexing) to add adapters and indexes for runs on an Illumina MiSeq (San Diego, CA, USA). The original PCR primers had upstream (5’) adapter sequences (forward: ACACTGACGACATGGTTCTACA; reverse: TACGGTAGCAGAGACTTGGTCT) to provide priming sites for the dual index PCR primers. This index PCR had a reaction program of eight cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, with a final extension of 72 °C for 5 min. These index PCR amplicons were cleaned with the AMPure XP Kit (Beckman Coulter Life Science, Indianapolis, IN, USA). Cleaned amplicons were quantified with PicoGreen (as above), normalized in Qiagen EB buffer, and then all index PCR plates were pooled into a single solution using 3 μL of each amplicon. This pooled sample was then sequenced on an Illumina MiSeq using a 500-cycle (v2) MiSeq Sequencing Kit (paired-end, 250 base pairs) according to the manufacturer’s protocols.
2.3. Bioinformatics

QIIME2 was used for bioinformatic metagenomics analysis (version 2020; https://qiime2.org, accessed on 3 Jun 2020) [49]. Sample metadata were created by following the template for the project. Raw reads were imported into QIIME2 followed the “Casava 1.8 paired-end demultiplexed fastq” format. Quality filtration, denoising, merging of paired reads, and chimera removal were carried out by using the DADA2 pipeline [50] implemented in the QIIME2 environment. Filtration parameters were set as: 25 bases trimming at the left side of forward and reverse reads, and truncation of forward and reverse reads up to 250 bases. Raw DNA reads were not clustered into operational taxonomic units (OTUs), and instead kept as amplicon sequence variants (ASVs) to avoid reducing the diversity of potential indicator taxa and because OTUs from clustering are difficult to compare with other data sets. ASVs with only a single DNA sequence, however, were excluded. Although ASVs for this genetic locus could be considered analogous to species, we maintain a more conservative stance and only refer to them as ASVs. Taxonomic assignments were made with SILVA (Supplementary File) [51].

2.4. Statistical Analyses

Our objectives were to characterize bacterial assemblage–environment relationships and to identify indicator taxa responsive to TP and TN concentrations. Quantifying and summarizing these assemblage and taxa changes can help identify possible management targets for TP or TN concentrations. For each statistical analysis, we used data from all sampling events for each site because this incorporated the variability in nutrient concentrations over time and provides a fuller representation of nutrient conditions at sites, likely leading to more robust indicator development. All statistical analyses used relative abundances of gene sequence reads for each ASV in a sample.

We used nonmetric multidimensional scaling (NMDS) to visualize changes in bacterial assemblage structure among samples and Spearman correlations to examine relationships of axis scores with TN, TP, watershed land cover, and bacterial metrics (see below). For multivariate statistics, we originally included taxa with >1% relative abundance in > 1% of samples, which reduces the effects of noise introduced by including many rare taxa [52] or DNA sequencing errors such as tag jumping, and is a reasonable and conservative criterion [52-56]. This criterion, however, left us with using <60% of the total gene sequence reads among all samples. So, to compromise between excluding a large proportion of assemblages and including an excessive number of rare ASVs, we included ASVs in the NMDS that first were observed in >5 samples and then those that comprised 75% of the total gene sequence reads when ordered by decreasing average relative abundance (429 of 18,406 ASVs). The requirement for observations in >5 samples also meets a recommended criterion for inclusion in threshold indicator taxa analysis (TITAN). For the NMDS, relative abundances of ASV gene sequence reads were square-root transformed, the Bray–Curtis dissimilarity coefficient was used, and the ordination was rotated to maximize variation along the first axis. We used the vegan package and the metaMDS function [57] in R v. 4.0.3 [58] to conduct the NMDS.
We used TITAN to characterize assemblage level changes and to identify ASVs that either increased or decreased with increasing TP or TN concentrations [38]. TITAN uses bootstrapping and indicator species analysis to identify the point along an environmental gradient at which each ASV has the greatest change in its frequency and relative abundance. Response magnitudes are standardized to $z$-scores to facilitate cross-taxon comparisons. Bootstrapping also identifies ASVs with consistently strong responses that increase or decrease in >95% of bootstrapped replicates. For ease of communication, we subsequently refer to ASVs that decrease and ASVs that increase with greater TP or TN concentrations as low P or N taxa and high P or N taxa, respectively. The distribution of bootstrapped ASV change points and sum $z$-scores of all indicator taxa at each observed TP or TN concentration (i.e., partition) shows how taxa and overall assemblages change along the TP or TN gradient [59]. We used R v. 3.4.3 [56] and the ‘TITAN2’ package [60] to conduct TITAN for TP and TN gradients with 1000 bootstraps. We included all ASVs used in the NMDS for consistency among statistical analyses.

For each sample, we summed the relative abundances of ASVs in each indicator group—low P, high P, low N, and high N bacteria—which could be used as metrics interpreted in a manner like that of traditional biomonitoring based on other organisms in streams (e.g., [61-63]). Spearman correlations were used to describe relationships of bacterial metrics with nutrients and watershed land cover. We also used boosted regression trees to examine relationships between bacterial metrics and multiple predictor variables. Boosted regression tree analysis is a machine learning method that combines results from thousands of regression trees into a model with high predictive performance [39,40]. This analysis can handle correlated predictors, determine the relative importance of predictors, model a variety of responses, and generate partial dependence plots showing the effects of each predictor while controlling for the mean effects of other variables in the model. The partial dependence plots were used to describe nutrient-bacteria relationships. We used a 10-fold cross-validation procedure, which used all data for training and validation steps, to evaluate model performance and to determine mean correlations between fitted and raw values and mean predictive performance of models reported as the percentage of null deviance explained. We used TP, TN, and conductivity as predictors and each of the four bacterial metrics as response variables. Using R v. 3.4.3 [58] and the ‘dismo’ package [64], we set tree complexity = 2, learning rate = 0.001, and bag fraction = 0.5, which created more than the recommended minimum of 1000 trees for each model [65].

Lastly, we used gradient forest analysis to characterize relationships between multi-taxa assemblages, using ASV relative abundances, and multiple environmental predictor variables in one statistical analysis. This analysis provides complementary results to TITAN, which uses multi-taxa responses to a single predictor based on relative abundance and occurrence data, and to boosted regression trees, which model single variable responses to multiple predictors. Gradient forest analysis combines results from random forest models for each ASV with $R^2$ values > 0 into a measure of overall change in assemblage structure [41]. For each ASV, the split values and their importance (i.e., variation explained by the partitioning) are collated using the R package ‘extendedForest’ and then used to compute multi-species and assemblage change functions along each environmental gradient and to produce visualizations of results with the R package ‘gradientForest’ [41]. Split density
plots, which incorporate the weighted importance of ASV splits and are standardized as a ratio to the density of data, show where the greatest rates of assemblage change occur along environmental gradients (ratios > 1). We interpreted peaks as the predictor values at which greatest change occurred, and the increase and decrease in splits surrounding these peaks as denoting unique regions of change along the gradient when ratios were > 1. Random permutations are used to determine the conditional importance of each predictor variable by measuring subsequent degradation in model performance.

2.5. Phylogenetic Analysis

Sequence data for the 429 ASVs used in the indicator analyses (see above) were aligned with Clustal W using default settings in MEGA-X [66]. To help assess the potential relatedness of ASVs marked as nutrient indicators, a maximum likelihood (ML) tree was generated in MEGA-X using the generalized time-reversible model with gamma-distributed rates and invariant sites (GTR + I + G). This more complex model of evolution was chosen for the ML analysis given that the ASVs were spread across multiple bacterial phyla. The ML tree was drawn and annotated in iTOL v5 [67].

3. Results

3.1. Descriptive Summary of Data

TP values for water collected at the stream sites ranged from 18 to 889 μg/L (N = 280, median 171 μg/L, mean ± standard deviation = 220 ± 179 μg/L). TN values for water collected at the stream sites ranged from 76 to 6560 μg/L (N = 281, median 621 μg/L, mean ± standard deviation = 778 ± 615 μg/L). The gradients of TP and TN concentrations were continuous and well represented with no gaps or overrepresentation of certain values (see [46] for further details). Historical conductivity values for water collected at the stream sites ranged from 373 to 789 μS/cm (N = 25, median 544 μS/cm, mean ± standard deviation = 561 ± 112 μS/cm).

The MiSeq run of the 342 periphyton samples generated nearly 6 million sequence reads (mean: 17,538 reads/sample; median: 17,179 reads/sample; minimum: 3258 reads; maximum: 36,102 reads) of appropriate length. Analysis of those reads found 18,406 ASVs with 2 or more reads. The top 75% of ASVs (based on reads per ASV) used for the NMDS and indicator analyses comprised 429 ASVs with approximately 4.5 million sequence reads.

Nonmetric multidimensional scaling showed that changes in bacterial assemblage structure among samples were associated with percent forest and percent agriculture in watersheds and with TP and TN concentrations (Figure 2). More forested land cover, less agriculture, and lower TP and TN concentrations were associated with decreasing axis 1 scores and increasing axis 2 scores. Increasing relative abundances of low P bacteria and decreasing relative abundances of high P and high N bacteria (see TITAN results below) were associated with decreasing axis 1 scores and increasing axis 2 scores. Relative abundances of low N bacteria increased with decreasing axis 1 scores.
3.2. Bacterial Indicators and Relationships with TP and TN

Of the 429 ASVs used in TITAN, 105 were associated with lower TP concentrations (low P bacteria) and 138 were associated with higher TP concentrations (high P bacteria) among sites (Supplementary Table S2). Bootstrapped assemblage change points indicated that the greatest changes in low P bacteria occurred between 110 and 152 μg TP/L, with an overall change point at 136 μg TP/L (Figure 3a, Table 1). Of the 105 low P ASVs, most had change points within a narrow range of TP concentrations when compared to those of high P ASVs (Figure 3b, Supplementary Table S2). Bootstrapped assemblage change points indicated that the greatest changes in high P bacteria occurred between 137 and 243 μg TP/L, with a change point at 183 μg TP/L (Figure 3a, Supplementary Table S1). Of the 138 high P ASVs, most had broad distributions of bootstrapped change points when compared to those of low P ASVs, indicating that these high P ASVs had gradual changes in their relative abundances and occurrences across a wide range of TP concentrations (Figure 3b, Supplementary Table S2).

Of the same 429 ASVs used in TITAN, 63 were associated with lower TN concentrations (low N bacteria) and 74 were associated with higher TN concentrations (high N bacteria) among sites (Supplementary Table S3). Bootstrapped assemblage change points indicated that the greatest changes in low N bacteria occurred between 462 and 695 μg TN/L, with an overall change point at 467 μg TN/L (Figure 3c, Table 2). Bootstrapped assemblage change points indicated that the greatest changes in high N bacteria occurred between 665 and 812 μg TN/L, with an overall change point at 772 μg TN/L (Figure 3c, Table 2). In general, most low N and high N ASV change points occurred within narrow ranges of the total TN gradient, but a greater proportion of high N ASVs than low N ASVs had broad bootstrapped distributions of change points, indicating that these high N ASVs had more gradual changes in their relative abundances and occurrences than did low N ASVs as TN concentrations increased (Figure 3d, Supplementary Table S3).

Relative abundances of low P and low N ASVs decreased and those of high P and high N ASVs increased as watershed percent agriculture and nutrient concentrations increased (Table 3). Boosted regression tree models explained 54–67% of the deviance in bacterial metrics, and the cross-validated deviance explained ranged from 45 to 56% (Table 4), showing that the models performed similarly when predicting withheld data. TP was the most important predictor for low P ASVs, high P ASVs, but also high N ASVs, while TN was the most important predictor of only low N ASVs. Partial dependence plots identified ranges of TP and TN gradients within which large changes in bacterial metrics occurred (Figure 4). Relative abundances of low P bacteria had a small decrease from 42 to 61 μg TP/L followed by their largest decrease from 108 to 184 μg TP/L and smaller decreases from 241 to 311, 352 to 378, and 494 to 509 μg TP /L. High P bacteria had a small increase in relative abundance from 36 to 82 μg TP/L followed by their largest increase from 94 to 184 μg TP/L and another small increase from 287 to 356 μg TP/L. Relative abundances of low N bacteria had a large decrease from 271 to 459 μg TN/L followed by a smaller decrease from 459 to 859 μg TN/L. High N bacteria gradually increased in relative abundance from 279 to 756 μg TN/L.
Gradient forest analysis included 278 ASVs with $R^2$ values > 0 (maximum = 0.57, mean = 0.13, median = 0.10), including 61 with $R^2$ > 0.20 (Supplementary Table S4). Peaks in the standardized split density plot showed that the greatest changes in bacterial assemblage structure occurred between 24 and 75, 75 and 152, and 250 and 325 μg TP/L, with smaller changes occurring between 174 and 217 and 325 and 386 μg TP/L (Figure 5). For TN, a gradual change in bacterial assemblage structure occurred between 193 and 928 μg/L, with a small second response around 1458 μg/L (Figure 5).

3.3. Summary of Changes in Bacterial Assemblages

Complementary results from these statistical analyses showed that bacterial assemblages were strongly associated with TP and TN gradients. Results were collated to identify TP and TN concentrations at which notable changes in bacterial assemblages occurred (Figure 6, Tables 1 and 2). In general, changes along TP and TN gradients were marked by shifts from dominance by low nutrient ASVs to high nutrient ASVs with major points of overall community change being highlighted by TITAN change points, steep portions of partial dependence plots, and peaks in gradient forest split density plots. Bacterial assemblages began having small changes from 24 to 82 μg TP/L with midpoints of multiple responses occurring around 50 μg TP/L. This initial change was followed by larger changes, including continued decreases in relative abundances of low P taxa, increases in those of high P taxa, TITAN change points, and additional large changes in gradient forest analysis between 110 and 195 μg TP/L. The final major changes in bacterial assemblages occurred between 275 and 365 μg TP/L, after which high P taxa dominated. The body of statistical evidence suggests that the bacterial assemblage structure experienced large, gradual changes with initial large decreases in relative abundances of low N ASVs from 275 to 565 μg TN/L followed by the greatest increase in relative abundances of high N ASVs and bootstrapped change points for high N taxa from 565 to 855 μg TN/L.

3.4. Taxonomy and Phylogenetic Analysis

Phylogenetic analysis of the 429 ASVs determined how the different types of indicators are spread amongst the various bacterial orders (Figure 7). Most bacterial phyla and classes were recovered as monophyletic with a few exceptions (e.g., Caulobacterales nested within Rhizobiales, and Rhodobacterales and Pseudomonadales are not monophyletic), though investigating the evolutionary history of bacteria was not the goal of the analysis. All clades with more than four ASVs had at least one ASV that was found to be an indicator of nutrient condition. Many ASVs responded to both TP and TN concentrations with 67 ASVs as indicators of both high P and high N and 45 ASVs as indicators of both low P and low N.

The majority of ASVs (N = 285) were members of the phylum Proteobacteria (Figure 7, Table 5). Other phyla in the top 75% with more than 10 ASVs included Actinobacteria (48 ASVs), Cyanobacteria (26 ASVs), Bacteroidota (16 ASVs), Firmicutes (14 ASVs), Planctomycetota (14 ASVs), and Verrucomicrobiota (11 ASVs) (Figures 7 and 8). Bacterial phyla with few ASVs (N < 5) included Chloroflexi, Acidobacteriota, Gemmatimonadota, Desulfovacterota, Deinococcales, Fusobacteriota, Myxococcota and Nitrospirota.
Within the Proteobacteria, the order Rhizobiales had the most ASVs (N = 79) and also the most ASVs found to be indicators in TITAN (N = 49) with 18 ASVs that were both indicators for high P and high N conditions, 10 ASVs that were indicators of high P, 1 ASV that was a high N indicator, 4 ASVs that were both indicators for low P and low N, 12 ASVs that were indicators of low P, and 4 ASVs that were indicators of low N (Table 5). Other seemingly important orders found within Proteobacteria included Sphingomonadales (59 total ASVs, 39 ASVs as indicators with 11 ASVs as indicators of both high P and high N and 19 ASVs as indicators of only high P), Burkholderiales (47 total ASVs, 27 ASVs as indicators with 9 ASVs as indicators of both high P and high N and 9 ASVs as indicators of only high P), Rhodobacterales (34 total ASVs, 21 ASVs as indicators with 4 ASVs as indicators of only high P, 11 ASVs as indicators of both low P and low N, and 4 ASVs as indicators of only low P), Pseudomonadales (16 total ASVs with 13 ASVs as indicators with 3 ASVs as indicators of just high P, 7 ASVs as indicators of only low P, and 3 ASVs as indicators only low N), and Xanthomonadales (16 total ASVs, 9 ASVs as indicators with 2 ASVs as indicators of only high P, 2 ASVs as indicators of both low P and low N, and 3 ASVs as indicators of only low P). Outside the Proteobacteria, other potentially important phyla/orders include Micrococcales (13 total ASVs, 10 ASVs as indicators) of the Actinobacteria, Chitinophagales (10 total ASVs, 6 ASVs as indicators) of the Bacteroidota, Cyanobacteria (26 total ASVs, 18 ASVs as indicators), and Verrucomicrobiota (11 total ASVs, 9 ASVs as indicators).

Indicators of high and low nutrient conditions were dispersed across all bacterial groups (Figures 7 and 8). ASVs classified under Firmicutes, Verrucomicrobiota, Flavobacteriales, and Pseudomonadales, Rhodobacterales, and Xanthomonadales of Proteobacteria were more; often correlated with low nutrient concentrations. ASVs classified under Chitinophagales of Bacteroidota, and Burkholderiales, Rhizobiales, Sphingomonadales, and Steroidobacterales of Proteobacteria, were correlated with high nutrient conditions.

4. **Discussion**

Our analyses showed distinct changes in bacterial assemblages associated with gradients of TP and TN concentrations, and considerable numbers of ASVs were significant responders to the phosphorus and nitrogen conditions (from TITAN, 243 of 429 for TP and 137 of 429 ASVs for TN, and from gradient forest analysis, 278 of 429 ASVs). Metabarcoding resulted in more than 18,000 bacterial ASVs, of which the 429 most frequent, comprising 75% of the total relative abundance represented over 15 different bacterial phyla. This work adds to the growing field of genetic characterization of bacterial communities for indicators of ecological or environmental conditions in streams and can inform the future applications of bacterial indicators in monitoring programs [26,36].

Our study identified many bacterial taxa that have significant relationships with either or both TP and TN. Most of these relationships show responses to TP over a concentration range of 25–195 μg/L and to TN over a range of 275–855 μg/L (Figure 5), which is similar to the ranges found for diatom responses in this watershed (20–150 μg TP/L and 150–850 μg “TN/L), though diatom assemblages tended to respond at much lower TP concentrations [41]. In general, TITAN results suggested that tow nutrient bacterial ASVs exhibited...
sharp declines in their relative abundances and occurrences as nutrient concentrations increased. This pattern could be due to greater numbers of high nutrient ASVs that, despite mostly having gradual increases across wider ranges of nutrient conditions, quickly and cumulatively outcompeted low nutrient ASVs as nutrient concentrations increased. While the bacterial indicator ASVs described herein likely respond to nutrient concentrations, they also could be responding to nutrient-related changes in other benthic community members [68-71].

Non-photosynthesizing bacterial groups fill biochemical roles that are very different from those of cyanobacterial or algal members of periphyton communities. For example, Sphingomonadales are known for secreting extracellular matrices as they proliferate, Actinobacteria often function as decomposers, and some Rhizobiales (e.g., *Rhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Sinorhizobium*, *Azorhizobium*, and *Allorhizobium*) are major players in symbiotic N2-fixation in nature [72]. Bacteria also are important contributors to nutrient cycling within periphyton and produce extracellular enzymes that are important to phosphorus, nitrogen, and carbon uptake from organic sources [19], and their enzyme activity can be stimulated by algal productivity associated with greater amounts of nutrients [73-76]. More studies are necessary to uncover whether the bacterial indicators found herein are responding directly or indirectly to TP and TN conditions, but that does not change their potential importance as indicators for biomonitoring. Our results are informative for developing nutrient indicators in other regions, and although the data and analyses from this work are based on sites within a single watershed, the breadth of temporal coverage and the common watershed scale of monitoring programs increase the likelihood that these approaches are applicable to other, similar temperate watersheds.

DNA metabarcoding applied to biological monitoring is not without its limitations or biases. Complicating factors for the further development of these molecular genetic tools include concerns related to sample preservation [77], DNA extraction method [78,79], PCR bias [80,81], number of gene copies [82], uniformity across bioinformatic pipelines [83], and status of genetic databases [84]. However, research advances continue to address these concerns and minimize their impacts in molecular studies. For most taxa studied with DNA metabarcoding for environmental applications, critics often discuss comparisons of genetic data to traditional taxonomic methods, but such a concern does not apply to bacteria. Isolation and taxonomic identification of these periphyton-associated bacteria would require a myriad of culture methods. It is widely accepted that the percentage of bacteria that can be cultured is minimal relative to the estimated number of total bacterial species richness (10^7–10^9) [85,86]. DNA-based methods for the identification of bacterial communities still present the best opportunity to generate useful data in a timely and cost-effective manner.

Recent studies have demonstrated the potential utility of DNA metabarcoding for assessing responses of bacterial communities to anthropogenic impacts in multiple environmental contexts, and increasingly these responses are being captured in biotic indices informative to managers and decision-makers [87-89]. For instance, benthic bacterial communities have been shown to mirror responses of established macrofaunal indicators to nutrient enrichment associated with salmon farming [90], and 16S rRNA gene metabarcoding has been incorporated into a multi-trophic Metabarcoding Biotic Index found to be strongly correlated

Water (Basel). Author manuscript; available in PMC 2023 July 30.
with those disturbances [91]. Biotic indices based solely on bacterial community profiles are similarly strongly correlated with impacts of salmon aquaculture, and those indices are sufficiently robust to be replicable across multiple independent laboratories [92], suggesting potential utility for routine biomonitoring. Of particular interest with respect to the current study is the fact that “de novo” methods for identifying bacterial indicators—i.e., taxonomy-free approaches based on a statistical analysis of ASVs—performed better in these contexts than a taxonomy-based approach in which 16S DNA sequence data was used to calculate a previously developed bacterial biotic index [93]. In addition, complementary analyses associated with our work also found that bacteria-nutrient relationships remained strong over time, except immediately following hydrologic disturbance, and that bacterial metrics best represent long-term nutrient concentrations, though they were still correlated with short-term concentrations as well [94]. In addition, comprehensive and complementary analyses of temporal variability among sites in our study found that bacteria-nutrient relationships remained strong over time, except immediately following hydrologic disturbance, and that bacterial metrics best represent long-term nutrient concentrations, though they also were still correlated with short-term concentrations [94]). These additional analyses also showed that indicators developed using multiple samples over time performed well despite temporal variability in both bacterial assemblages and nutrient concentrations.

Bacterial DNA metabarcoding has also proven useful to assess a variety of other anthropogenic environmental stressors. Recent studies demonstrate that indices based on bacterial community structure respond strongly to environmental impacts of oil and gas drilling and production [95], urban pollution in coastal mangrove forests [96], and metal pollution associated with cement plants [97]. Although many of these studies focus on benthic bacterial assemblages in estuarine and marine environments, similar approaches have been developed for freshwater habitats, including a recent study developing biological indicators of river health based on changes to bacterial communities in periphyton [98]. Our research thus adds to a growing literature illustrating the potential value of 16S metabarcoding for routine biomonitoring and provides further evidence of the utility of taxonomy-free approaches for developing informative indicators based on bacterial responses to environmental stress.

While our work highlights the potential utility of DNA metabarcoding applied to periphyton bacteria as nutrient indicators, these results suggest that other areas of metagenomic research would likely prove useful. Future studies could investigate not only the diversity of bacterial taxa present but also the genetic capabilities of those taxa related to processing nutrients. Studies of RNA from periphyton bacteria coupled with nutrient chemistry would likely show which nutrient processing genes are activated in response to nitrogen or phosphorus availability [99,100] (Graves et al. 2016; Dai et al. 2019) while also providing data on which bacterial taxa are functioning in these conditions [101,102] (Leff et al. 2015; Li et al. 2018). Analysis of these bacterial biochemical pathways could also be informative to how these bacteria interact with other members of the periphyton community. To date, microbial metagenomic nutrient processing studies have not been coupled with nutrient indicator analyses, but such work would be a valuable addition to environmental management tools applied to nutrient pollution.
5. Conclusions

Periphyton and other algal-based communities continue to be high-priority organisms collected as part of monitoring programs for their use as indicators of nutrient pollution in aquatic ecosystems [17,103]. Our analyses demonstrated that DNA metabarcoding applied to periphyton-associated bacteria provides similar results to DNA metabarcoding applied to benthic diatoms for quantifying nutrient responses in a watershed affected by agricultural and urbanization stresses. Multiple statistical methods identified assemblage changes along TP and TN gradients that could serve as possible targets for decision-makers tasked with mitigating nutrient inputs. While the results indicated that potential TP concentration targets based on bacterial indicators were higher than those found for diatom indicators in the same watershed, these responses still provide useful ecological insights into how nutrient pollution affects aquatic ecosystems. These targets may be informative for the development of nutrient criteria related to land use, best management practices for reducing nutrients in watersheds, and discharge permits. Further development of technologies, methods, databases, and analytical methods will continue to improve the capabilities of applying genetic/genomic approaches to biological monitoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank Michael McManus and Roy Martin for their help with developing site selection within the watershed, Jeni Jones for field sampling, and Barry Wiechman and Sara Okum for their work in the laboratory. Roy Martin and Richard Devereux provided useful comments to a previous draft of the manuscript. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA. This manuscript has EPA tracking number ORD-047389. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. All research was funded by the United States Environmental Protection Agency.

Data Availability Statement:

Bacterial 16S (and also diatom rbcL) and nutrient data are available at https://www.ncbi.nlm.nih.gov/bioproject/592969 and https://doi.org/10.23719/1504034.

References

1. Beusen AHW; Bouwman AF; Van Beek LPH; Mogollón JM; Middelburg JJ Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. Biogeosciences 2016, 13, 2441–2451.
2. Sabo RD; Clark CM; Bash J; Sobot D; Cooter E; Dobrowski JP; Houlton BZ; Rea A; Schwede D; Morford SL; et al. Decadal shift in nitrogen inputs and fluxes across the contiguous United States: 2002–2012. J. Geophys. Res.-Biogeosci 2019, 124, 3104–3124.
3. Sabo RD; Clark CM; Gibbs DA; Metson GS; Todd MJ; LeDue SD; Greiner D; Fry JA; Polinsky R; Yang Q; et al. Phosphorus inventory for the conterminous United States (2002–2012). J. Geophys. Res.-Biogeosci 2021, 126, e2020JG005684.
4. United States Environmental Protection Agency. National Rivers and Streams Assessment 2008–2009: A Collaborative Survey; EPA/841/R-16/007; Office of Water and Office of Research and Development: Washington, DC, USA, 2016.
5. Dodds WK; Smith VH Nitrogen, phosphorus, and eutrophication in streams. Inland Waters 2016, 6, 155–164.
6. Carvalho L; Mackay EB; Cardoso AC; Baatrup-Pedersen A; Birk S; Blackstock KL; Borics G; Borja A; Feld CK; Ferreira MT; et al. Protecting and restoring Europe’s waters: An analysis of the future development needs of the Water Framework Directive. Sci. Total Environ 2019, 658, 1228–1238. [PubMed: 30677985]
7. Poikane S; Várbiró G; Kelly M; Birk S; Phillips G Estimating river nutrient concentrations consistent with good ecological condition: More stringent nutrient thresholds needed. Ecol. Indic 2021, 121, 107017.
8. Wang L; Robertson DM; Garrison PL Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: Implication to nutrient criteria development. Environ. Manag. 2007, 39, 194–212.
9. Weijters MJ; Janse JH; Alkemade R; Verhoeven JTA Quantifying the effect of catchment land use and water nutrient concentrations on freshwater river and stream biodiversity. Aquat. Conserv 2009, 19, 104–112.
10. Stevenson RJ; Bennett BJ; Jordan DN; French RD Phosphorus regulates stream injury by filamentous green algae, DO, and pH with thresholds in responses. Hydrobiologia 2012, 695, 25–42.
11. Wurtsbaugh WA; Paerl HW; Dodds WK Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. WIREs Water 2019, 6, e1373.
12. Smucker NJ; Beaulieu JJ; Nietch CT; Young JL Increasingly severe cyanobacterial blooms and deep-water hypoxia coincide with warming temperatures in reservoirs. Glob. Chang. Biol 2021, 27, 2507–2519. [PubMed: 33774887]
13. Howarth RW; Sharples A; Walker D Sources of nutrient pollution to coastal waters in the United States: Implications for achieving coastal water quality goals. Estuaries 2002, 25, 656–676.
14. Vörösmarty CJ; McIntyre PB; Gessner MO; Dudgeon D; Prusevich A; Green P; Glidden S; Bunn SE; Sullivan CA; Reidy Liebermann C; et al. Global threats to human water security and river biodiversity. Nature 2010, 467, 555–561. [PubMed: 20882010]
15. Gilbert PM; Beusen AHW; Harrison JA; Durr HH; Bouwman AF; Lanuette GA Changing land-, sea-, and airscapes: Sources of nutrient pollution affecting habitat suitability for harmful algae. In Global Ecology and Oceanography of Harmful Algal Blooms; Ecological Studies (Analysis and Synthesis); Gilbert P, Berdalet E, Burford M, Pitcher G, Zhou M, Eds.; Springer: Cham, Switzerland, 2018; Volume 232.
16. Caruso G; La Ferla R; Azzaro M; Zoppini A; Marino G; Petochi T; Corinaldesi C; Leonardi M; Zaccoone R; Fonda Umami S; et al. Microbial assemblages for environmental quality assessment: Knowledge, gaps and usefulness in the European marine strategy framework directive. Crit. Rev. Microbiol. 2016, 42, 883–904. [PubMed: 26585708]
17. Sagova-Mareckova M; Boenigk J; Bouchez A; Cermakova K; Chonova T; Cordier T; Eisenheld U; Elserk T; Fazi S; Fleituch T; et al. Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. Water Res. 2021, 191, 116767. [PubMed: 33418487]
18. Mulholland PJ; Helton AM; Poole GC; Hall RO; Hamilton SK; Peterson BJ; Tank JL; Ashkenas LR; Cooper LW; Dahm CN; et al. Stream denitrification across biomes and its response to anthropogenic nitrate loading. Nature 2008, 452, 202–205. [PubMed: 18337819]
19. Sinsabaugh RL; Hill BH; Follstad Shah JJ Eoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. Nature 2009, 462, 795–798. [PubMed: 20010687]
20. Besemer K Biodiversity, community structure and function of biofilms in stream ecosystems. Res. Microbiol. 2015, 166, 774–781. [PubMed: 26027773]
21. Battin TJ; Besemer K; Bengtsson MM; Romani AM; Packmann A The ecology and biogeochemistry of stream biofilms. Nat. Rev. Microbiol. 2016, 14, 251–263. [PubMed: 26972916]
22. Antwis RE; Griffiths SM; Harrison XA; Aranega-Bou P; Arce A; Bettridge AS; Brailsford FL; de Menezes A; Devaynes A; Forbes KM; et al. Fifty important research questions in microbial ecology. FEMS Microbiol. Ecol. 2017, 93, fix044.
23. Abreu CI; Friedman J; Andersen Woltz VL; Gore J Mortality causes universal changes in microbial community composition. Nat. Commun 2019, 10, 2120. [PubMed: 31073166]

24. Romero F; Acuña V; Font C; Freixa A; Sabater S Effects of multiple stressors on river biofilms depend on the time scale. Sci. Rep 2019, 9, 15810. [PubMed: 31676856]

25. Seviour RJ; Mino T; Onuki M The microbiology of biological phosphorus removal in activated sludge systems. FEMS Microbiol. Rev 2003, 27, 99–127. [PubMed: 12697344]

26. Simonin M; Voss KA; Hassett BA; Rocca JD; Wang SY; Bier RL; Violin CR; Wright JP; Bernhardt ES In search of microbial indicator taxa: Shifts in stream bacterial communities along an urbanization gradient. Environ. Microbiol 2019, 21, 3652–3668.

27. Steen AD; Crits-Christoph A; Carini P; DeAngelis KM; Fierer N; Lloyd KG; Thrash JC High proportions of bacteria and archaea across most biomes remain uncultured. ISME J. 2019, 13, 3126–3130. [PubMed: 31388130]

28. Bodor A; Boundedjoun N; Vincze GE; Kis AE; Laczi K; Bende G; Szilagyi A; Kovacs A; Perei K; Rakhely G Challenges of unculturable bacteria: Environmental perspectives. Rev. Environ. Sci. Biotechnol 2020, 19, 1–22.

29. Lau KEM; Washington VJ; Fan V; Neale MW; Lear G; Curran J; Lewis GD A novel bacterial community index to assess stream ecological health. Freshwater Biol. 2015, 60, 1988–2002.

30. Ji Y; Ashton L; Pedley SM; Edwards DP; Tang Y; Nakamura A; Kitching R; Dolman PM; Woodcock P; Edwards FA; et al. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol. Lett 2013, 16, 1245–1257. [PubMed: 23910579]

31. Stein ED; Martinez MC; Stiles S; Miller PE; Zakharov EV Is DNA barcoding actually cheaper and faster than traditional morphological methods: Results from a survey of freshwater bioassessment efforts in the United States? PLoS ONE 2014, 9, e95525. [PubMed: 24755838]

32. Keck F; Vasselton V; Tapolczai K; Rimet F; Bouchez A Freshwater biomonitoring in the information age. Front. Ecol. Environ 2017, 15, 266–274.

33. Pawlowski J; Kelly-Quinn M; Altermatt F; Apotheloz-Perret-Gentil L; Begja P; Boggero A; Borja A; Bouchez A; Cordier T; Domaizon I; et al. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ 2018, 637–638, 1295–1310. [PubMed: 29801222]

34. Wetterstrand KA DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available online: www.genome.gov/sequencingcostsdata (accessed on 15 November 2021).

35. Bock C; Jensen M; Forster D; Marks S; Nuy J; Psenner R; Beisser D; Boenigk J Factors shaping community patterns of protists and bacteria on a European scale. Environ. Microbiol 2020, 22, 2243–2260. [PubMed: 32203262]

36. Wood SA; Biessy L; Pingram MA; Hamer MP; Wagenhoff A; Clapcott J; Pearman JK Temporal and spatial variation in bacterial communities on uniform substrates in non-wadeable rivers. Environ. DNA 2021, 3, 1023–1034.

37. Juvigny-Khenafou NPD; Piggott JJ; Atkinson D; Zhang Y; Wu N; Matthaei CD Fine sediment and flow velocity impact bacteria community and functional profile more than nutrient enrichment. Ecol. Appl 2021, 31, e02212. [PubMed: 32754996]

38. Baker ME; King RS A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods Ecol. Evol 2010, 1, 25–37.

39. Friedman JH Greedy function approximation: A gradient boosting machine. Ann. Stat 2001, 29, 1189–1232.

40. Hastie T; Tibshirani R; Friedman J Elements of Statistical Learning: Data Mining, Inference and Prediction; Springer: New York, NY, USA, 2009.

41. Ellis N; Smith SJ; Pitcher CR Gradient forests: Calculating importance gradients on physical predictors. Ecology 2012, 93, 156–168. [PubMed: 22486096]

42. United States Environmental Protection Agency. Method 365.1, Revision 2.0: Determination of Phosphorus by Semi-Automated Colorimetry; EPA-600/R-93/100; United States Environmental Protection Agency: Washington, DC, USA, 1993.

43. Tucker S Determination of Orthophosphate in Waters by Flow Injection Analysis Colorimetry (High Throughput): QuikChem® Method 10-115-01-1-V; Lachat Instruments: Loveland, CO, USA, 2008.
44. Patton CJ; Kryskalla JR Methods of Analysis by the U.S. Geological Survey National Water Quality Laboratory: Evaluation of Alkaline Persulfate Digestion as an Alternative to Kjeldahl Digestion for Determination of Total and Dissolved Nitrogen and Phosphorus in Water; 03-4174:33; U.S. Department of the Interior, U.S. Geological Survey: Reston, VA, USA, 2003.

45. Smith P; Bogren K Determination of Nitrate/Nitrite in Manual Persulfate Digestions: QuikChem® Method 10-107-04-4-A; Lachat Instruments: Loveland, CO, USA, 2003.

46. Smucker NJ; Pilgrim EM; Nietch CT; Darling JA; Johnson BR DNA metabarcoding effectively quantifies diatom responses to nutrients in streams. Ecol. Appl 2020, 30, e02205. [PubMed: 32602216]

47. Werner JJ; Loren O; Hugenholtz P; DeSantis TZ; Walters WA; Caporaso JG; Angenstein LT; Knight R; Ley RE Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. ISME J. 2012, 6, 94–103. [PubMed: 21716311]

48. Hall EK; Bernhardt ES; Bier RL; Bradford MA; Boot CM; Cotner JB; del Giorgio PA; Evans SE; Graham EB; Jones SE; et al. Understanding how microbiomes influence the systems they inhabit. Nat. Microbiol 2018, 3, 977–982. [PubMed: 30143799]

49. Bolyen E; Rideout JR; Dillon MJ; Abnet CC; Al-Ghalith GA; Alexander H; Alm EJ; Arumugam M; Asnicar F; et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol 2019, 37, 852–857. [PubMed: 31341288]

50. Callahan BJ; McMurdie PJ; Rosen MJ; Han AW; Johnson AJA; Holmes SP DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 2016, 13, 581–583. [PubMed: 27214047]

51. Glöckner FO; Yilmaz P; Quast C; Gerken J; Bacto; Ciuprina A; Bruns G; Yarza P; Peplies J; Westram R; et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol 2017, 261, 169–176. [PubMed: 28648396]

52. McCune B; Grace JB Analysis of Ecological Communities; MjM Software Design: Glenden Beach, OR, USA, 2002.

53. Lavoie I; Dillon PJ; Campeau S The effect of excluding diatom taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment. Ecol. Indic 2009, 9, 213–225.

54. Rimet F; Bouchez A Biomonitoring river diatoms: Implications of taxonomic resolution. Ecol. Indic 2012, 15, 92–99.

55. Poos MS; Jackson DA Addressing the removal of rare species in multivariate bioassessments: The impact of methodological choices. Ecol. Indic 2012, 18, 82–90.

56. Rothenberger MB; Swaffield T; Calomeni AJ; Cabrey CD Multivariate analysis of water quality and plankton assemblages in an urban estuary. Estuaries Coasts 2014, 37, 695–711.

57. Hill BH; Herlihy AT; Kaufmann PR; Stevenson RJ; McCormick FH; Johnson CB Use of periphyton assemblage data as an index of biotic integrity. J. N. Am. Benthol. Soc 2000, 19, 50–67.
64. Hijmans RJ; Phillips S; Leathwick J; Elith J. Dismo: Species Distribution Modeling. 2017. Available online: https://cran.r-project.org/web/packages/dismo/dismo.pdf (accessed on 17 July 2020).

65. Elith J; Leathwick JR; Hastie T. A working guide to boosted regression trees. J. Anim. Ecol 2008, 77, 802–813. [PubMed: 18397250]

66. Kumar S; Stecher G; Li M; Knyaz C; Tamura K. MEGA X: Molecular evolutionary genetic analysis across computing platforms. Mol. Biol. Evol 2018, 35, 1547–1549. [PubMed: 29722887]

67. Letunic I; Bork P. Interactive tree of life (iTOl) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021, 49, W293–W296. [PubMed: 33885785]

68. Amin SA; Parker MS; Armbrust EV. Interactions between diatoms and bacteria. Microbiol. Mol. Biol. Rev 2012, 76, 667–684. [PubMed: 22933565]

69. Stanish LF; O’Neill SP; Gonzalez A; Legg TM; Knelman J; McKnight DM; Spaulding S; Nemergut DR. Bacteria and diatom co-occurrence patterns in microbial mats from polar desert streams. Environ. Microbiol 2012, 15, 1115–1131. [PubMed: 22998505]

70. Wyatt KH; Turetsky MR. Algae alleviate carbon limitation of heterotrophic bacteria in a boreal peatland. J. Ecol 2015, 103, 1165–1171.

71. Koedooder C; Stock W; Willems A; Mangelinckx S; de Troch M; Vyverman W; Sabbe K. Diatom-bacteria interactions modulate the composition and productivity of benthic diatom biofilms. Front. Microbiol 2019, 10, 1255. [PubMed: 31231340]

72. Beeckmans S; Xie JP. Reference Module in Biomedical Sciences; Elsevier: Amsterdam, The Netherlands, 2015.

73. Espeland EM; Francoeur SN; Wetzel RG. Influence of algal photosynthesis on biofilm bacterial production and associated glucosidase and xyllosidase activities. Microb. Ecol 2001, 42, 524–530. [PubMed: 12024235]

74. Rier ST; Kuehn KA; Francoeur SN. Algal regulation of extracellular enzyme activity in stream microbial communities associated with inert substrata and detritus. J. N. Am. Benthol. Soc 2007, 26, 439–449.

75. Smucker NJ; Vis ML. Acid mine drainage affects the development and function of epilithic biofilms in streams. J. N. Am. Benthol. Soc 2011, 30, 728–738.

76. Pope CA; Halvorson HM; Findlay RH; Francoeur SN; Kuehn KA. Light and temperature mediate algal stimulation of heterotrophic activity on decomposing leaf litter. Freshwater Biol. 2020, 65, 1210–1222.

77. Stein ED; White BP; Mazor RD; Miller PE; Pilgrim EM. Evaluating ethanol-based sample preservation to facilitate the use of DNA barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. PLoS ONE 2013, 8, e51273. [PubMed: 23308979]

78. Gaget V; Keulen A; Lau M; Monis P; Brookes JD. DNA extraction from benthic cyanobacteria: Comparative assessment and optimization. J. Appl. Microbiol 2017, 122, 294–304. [PubMed: 27779795]

79. Majaneva M; Diserud OH; Eagle SHC; Hajibabaei M; Ekrem T. Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. Metabarcoding Metagenom. 2018, 2, e26664.

80. Gibson JF; Shokralla S; Curry C; Baird DJ; Monk WA; King I; Hajibabaei M. Large-scale biomonitring of remote and threatened ecosystems via high-throughput sequencing. PLoS ONE 2015, 10, e0138432. [PubMed: 26488407]

81. Elbrecht V; Vamos EE; Meissner K; Aroviita J; Leese F. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods Ecol. Evol. 2017, 8, 1265–1275.

82. Vasselon V; Bouchez A; Rimet F; Jacquet S; Trobajo R; Corniquel M; Tapolczai K; Domaizon I. Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods Ecol. Evol. 2018, 9, 1060–1069.

83. Mathon L; Valentini A; Guerin P-E; Normandieu E; Noel C; Lionnet C; Boulanger E; Thuiller W; Bernatchez L; Mouillot D; et al. Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. Mol. Ecol. Resour 2021, 21, 2565–2579. [PubMed: 34002951]
84. Trebitz AS; Hoffman JC; Grant GW; Billehaus TM; Pilgrim EM Potential for DNA-based identification of Great Lakes fauna: Match and mismatch between taxa inventories and DNA barcode libraries. Sci. Rep 2015, 5, 12162. [PubMed: 26199185]
85. Stewart EJ Growing unculturtable bacteria. J. Bacteriol 2012, 194, 4151–4160. [PubMed: 22661685]
86. Schloss PD; Handelsman J Status of the microbial census. Microbiol. Mol. Biol. Rev 2004, 68, 686–691. [PubMed: 15590780]
87. Cordier T; Lanzen A; Apotheloz-Perret-Gentil L; Stoeck T; Pawlowski J Embracing environmental genomics and machine learning for routine biomonitoring. Trends Microbiol. 2019, 27, 387–397. [PubMed: 30554770]
88. Zolkifli N; Sharuddin SS; Yusoff MZM; Hassan MA; Maeda T; Ramli N A review of current and emerging approaches for water pollution monitoring. Water 2020, 12, 3417.
89. Michan C; Blasco J; Alhama J High-throughput molecular analyses of microbiomes as a tool to monitor the wellbeing of aquatic environments. Microb. Biotechnol 2021, 14, 870–885. [PubMed: 33559398]
90. Stoeck T; Frühe L; Forster D; Cordier T; Martins CIM; Pawlowski J Environmental DNA metabarcoding of benthic bacterial communities indicates the benthic footprint of salmon aquaculture. Mar. Pollut. Bull 2018, 127, 139–149. [PubMed: 29475645]
91. Keeley N; Wood SA; Pochon X Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. Ecol. Indic 2018, 85, 1044–1057.
92. Dully V; Balliet H; Frühe L; Däumer M; Thielen A; Gallie S; Berrill I; Stoeck T Robustness, sensitivity and reproducibility of eDNA metabarcoding as an environmental biomonitoring tool in coastal salmon aquaculture—An inter-laboratory study. Ecol. Indic 2021, 121, 107049.
93. Aylagas E; Atalah J; Sanchez-Jerez P; Pearman JK; Casado N; Asensi J; Toledo-Guedes K; Carvalho S A step towards the validation of bacteria biotic indices using DNA metabarcoding for benthic monitoring. Mol. Ecol. Resour 2021, 21, 1889–1903. [PubMed: 33825307]
94. Smucker NJ; Pilgrim EM; Wu H; Nietch CT; Darling JA; Molina M; Johnson BR; Yuan LL Characterizing temporal variability in streams supports nutrient indicator development using diatom and bacterial DNA metabarcoding. Sci. Total Environ 2022, 154960. [PubMed: 35378187]
95. Laroche O; Wood SA; Tremblay LA; Ellis JI; Lear G; Pochon X A cross-taxa study using environmental DNA/RNA metabarcoding to measure biological impacts of offshore oil and gas drilling and production operations. Mar. Pollut. Bull 2018, 127, 97–107. [PubMed: 29475721]
96. Torres GG; Figueroa-Galvis I; Munoz-Garcia A; Polania J; Vanegas J Potential bacterial bioindicators of urban pollution in mangroves. Environ. Pollut 2019, 255 Pt 2,113293. [PubMed: 31563776]
97. da Costa Silva TA; de Paula M Jr.; Silva WS; Lacorte GA Can moderate heavy metal soil contaminations due to cement production influence the surrounding soil bacterial communities? Ecotoxicology 2022, 31, 134–148. [PubMed: 34748159]
98. Pin L; Eiler A; Fazi S; Friberg N Two different approaches of microbial community structure characterization in riverine epilithic biofilms under multiple stressors conditions: Developing molecular indicators. Mol. Ecol. Resour 2021, 21, 1200–1215. [PubMed: 33529477]
99. Graves CJ; Makrides EJ; Schmidt VT; Giblin AE; Cardon ZG; Rand DM Functional response of salt marsh microbial communities to long-term nutrient enrichment. Appl. Environ. Microbiol 2016, 82, 2862–2871. [PubMed: 26944843]
100. Dai Z; Liu G; Chen H; Chen C; Wang J; Ai S; Wei D; Li D; Ma B; Tang C; et al. Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. ISME J. 2020, 14, 757–770. [PubMed: 31827246]
101. Leff JW; Jones SE; Prober SM; Barberan A; Borer ET; Firn JL; Harpole WS; Hobbie SE; Hofmockel KS; Knops JMH; et al. Consistent response of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proc. Natl. Acad. Sci. USA 2015, 112, 10967–10972. [PubMed: 26283343]
102. Li Y; Jing H; Xia X; Cheung S; Suzuki K; Liu H Metagenomic insights into the microbial community and nutrient cycling in the western subarctic Pacific Ocean. Front. Microbiol 2018, 9, 623. [PubMed: 29670596]

103. Bailet B; Bouchez A; Franc A; Frigerio JM; Keck F; Karjalainen SM; Rimet F; Schneider S; Kahler M Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding Metagenom. 2019, 3, e34002.
Figure 1.
Map of collection sites in the East Fork of the Little Miami River Watershed.
Figure 2.
Nonmetric multidimensional scaling ordination showing Spearman correlations of axis scores with total phosphorus (TP), total nitrogen (TN), watershed percent forest, watershed percent agriculture, and relative abundances of low P, low N, high P, and high N bacterial ASVs. Vectors are scaled to span the range of possible correlation coefficients (−1 to 1) along NMDS axes.
Figure 3.
Threshold indicator taxa analysis showing sumZ scores and change points of ASVs for TP (a,b) and TN (c,d). Filtered refers to results using only ASVs identified as being pure (consistent response direction in >95% of bootstrap replicates) and reliable (consistently significant responses in >95% of bootstrap replicates). Red shows increaser sumZ scores and ASVs; blue shows decreaser sumZ scores and ASVs. For (a,c), circles; show the sumZ scores of decreaser tend increaser ASVs at each observed nutrient concentration, and cumulative frequencies show the distribution of assemblage change points (max sumZ) based on 100 bootstraps. For (b,d), open circles show change points for each ASV (y-axis tick mares) scaled according to magnitude; of $z$ scores, and dotted liners show the 5th and 95th quantiles based on 1000 bootstraps. Narrow peaks in sumZ scores, steep increases
in the cumulative frequency curves, and multiple ASV change points occurring within a narrow range of TP or TN concentrations suggest nutrient thresholds at these concentrations. Broad peaks in sumZ scores, gradual increases in cumulative frequency curves, and gradual addition of ASV change points indicate more gradual responses to nutrient concentration and a longer gradient of community change.
Figure 4.
Partial dependence plots from boosted regression trees showing responses of bacterial metrics to TP (a,b) or TN (c,d) while controlling for the average effect of other variables. FF = fitted functions. Rug plots show deciles of predictor values.
Figure 5.
Gradient Forest analysis results showing bacterial ASV assemblage responses to TP (a) and TN (b). Compositional change based on aggregating ASV responses were determined by split importance and values along TP and TN gradients (bars) and the ratios (blue lines) of split density (black lines) to data density (red lines). Peaks and regions of standardized split density plots with ratios above 1 (horizontal dashed blue lines) mark portions of the TP or TN gradient within which ASV compositional change is relatively greater other points along the nutrient gradient.
Figure 6.
Summary of change points for TP and TN across the indicator analyses. Circles denote the mid-response value with horizontal bars show 5th to 95th quantiles for TITAN results and the beginning and end of responses for boosted regression trees and gradient forest results. Light blue circles are for low nutrient taxa, red circles are for high nutrient taxa, and gray circles are assemblage changes from gradient forest analytes. Bacterial responses occurred at multiple values along the TP or TN gradients and those are noted as sumZ1, sumZ2, BRT1, BRT2, etc. Vertical dashed lines and gray bars mark concentrations on the TP or TN gradients with substantial changes in bacterial assemblage. (LP = low phosphorus, LN = low nitrogen, HP = high phosphorus, HN = high nitrogen, CP = change point, BRT = boosted...
regression tree analysis, GF = gradient forest analysis). GF2 has no horizontal bars because it only briefly exceeded the density-of-splits to density-of-data ratio of 1.
Figure 7.
Phylogenetic tree of the 429 ASVs comprising 75% of the total relative abundance among all samples. Indicator ASVs from TITAN are marked. Major bacterial taxonomic groups are color coded in the legend in the order they appear in the tree, clockwise from the cyanobacteria near the top. Groups with only 1 or 2 members are unmarked.
Figure 8.
Breakdown of the indicator ASVs within the top 15 bacterial groups by count (a) or by proportion (b).
Table 1.
Summary of TP responses (Sorted by mid-response). CP is change point (mid-response) for TITAN results with start and end being 5th and 95th percentiles of bootstrapped change points. BRT = boosted regression trees, GF = gradient forest analysis, LP = low phosphorus ASVs, HP = high phosphorus ASVs. The sumZ responses are from TITAN distributions of bootstrapped community change points.

| Response  | Start of Response (μg/L) | Mid-Response (μg/L) | End of Response (μg/L) |
|-----------|--------------------------|---------------------|------------------------|
| GF TP1    | 24                       | 50                  | 75                     |
| BRT LP1   | 44                       | 52                  | 61                     |
| BRT HP1   | 36                       | 59                  | 82                     |
| LP sumz1  | 97                       | 113                 | 125                    |
| GF TP2    | 75                       | 114                 | 152                    |
| LP CP     | 110                      | 136                 | 152                    |
| LP sumz2  | 125                      | 138                 | 152                    |
| BRT HP2   | 94                       | 139                 | 184                    |
| BRT LP2   | 108                      | 146                 | 184                    |
| HP sumz1  | 113                      | 155                 | 170                    |
| HP sumz2  | 170                      | 180                 | 200                    |
| HP CP     | 137                      | 183                 | 243                    |
| GF TP3    | 174                      | 196                 | 217                    |
| BRT LP3   | 241                      | 276                 | 311                    |
| GF TP4    | 250                      | 287                 | 325                    |
| BRT HP3   | 287                      | 321                 | 356                    |
| GF TP5    | 325                      | 356                 | 386                    |
| BRT LP4   | 352                      | 365                 | 378                    |
| BRT LP5   | 494                      | 502                 | 509                    |
Table 2.
Summary of TN responses (Sorted by mid-response). CP is change point (mid-response) for TITAN results with start and end being 5th and 95th percentiles of bootstrapped change points. BRT = boosted regression trees, GF = gradient forest analysis, LN = low nitrogen ASVs, HN = high nitrogen ASVs. The sumZ responses are from TITAN distributions of bootstrapped community change points. GF2 had no discernable start or end of response because it only briefly exceeded the density-of-splits to density-of-data ratio of 1.

| Response  | Start of Response (µg/L) | Mid-Response (µg/L) | End of Response (µg/L) |
|-----------|--------------------------|---------------------|------------------------|
| BRT LN1   | 271                      | 365                 | 459                    |
| LN CP     | 462                      | 467                 | 695                    |
| LN sumz1  | 398                      | 471                 | 534                    |
| BRT HN    | 279                      | 517                 | 756                    |
| GF1       | 193                      | 560                 | 928                    |
| LN sumz2  | 534                      | 592                 | 641                    |
| BRT LN2   | 459                      | 659                 | 859                    |
| LN sumz3  | 641                      | 694                 | 748                    |
| HN sumz1  | 617                      | 699                 | 748                    |
| HN CP     | 665                      | 772                 | 812                    |
| HN sumz2  | 748                      | 786                 | 850                    |
| GF2       | -                        | 1458                | -                      |
Table 3.
Correlations among bacteria metrics, % watershed agriculture (% ag), total nitrogen (TN), and total phosphorus (TP). Bacteria metric correlations with % agriculture used means of all samples within each site (n = 25), whereas TP n = 280, TN n = 281.

|          | % ag  | TN   | TP   |
|----------|-------|------|------|
| Low P    | −0.84 | −0.59| −0.68|
| High P   | 0.86  | 0.62 | 0.71 |
| Low N    | −0.77 | −0.64| −0.47|
| High N   | 0.88  | 0.64 | 0.71 |
| TP       | 0.85  | 0.66 |      |
| TN       | 0.65  |      |      |
Table 4.
Results from boosted regression tree models for each bacterial metric showing deviance explained as a percentage of the null deviance for observed data and of deviance explained using cross-validated (CV) data. EC = electrical conductivity.

| Metric       | Relative Importance (%) | Observed Deviance Explained (%) | CV Deviance Explained (%) | CV Correlation | TP  | TN  | EC  |
|--------------|-------------------------|---------------------------------|---------------------------|----------------|-----|-----|-----|
| Low P ASVs   | 0.59                    | 0.45 ± 0.05                     | 0.69 ± 0.02               | 55.0           | 25.1| 19.9|
| High P ASVs  | 0.67                    | 0.56 ± 0.06                     | 0.77 ± 0.02               | 46.2           | 28.8| 25  |
| Low N ASVs   | 0.54                    | 0.46 ± 0.08                     | 0.70 ± 0.04               | 13.7           | 76.2| 10.1|
| High N ASVs  | 0.60                    | 0.50 ± 0.10                     | 0.79 ± 0.03               | 40.5           | 24.6| 34.9|
Table 5.

Nutrient indicator ASVs by bacterial phylum/order.

| Phylum         | Order          | Total | No. Indicators | High P and N | High P Only | High N Only | Low P and N | Low P Only | Low N Only |
|----------------|----------------|-------|----------------|--------------|-------------|-------------|--------------|-------------|-------------|
| Proteobacteria | Rhizobiales    | 79    | 49             | 18           | 10          | 1           | 4            | 12          | 4           |
|                | Sphingomonadales| 59    | 39             | 11           | 19          | 6           | 2            | 1           |             |
|                | Burkholderiales| 47    | 27             | 9            | 1           | 4           | 3            | 1           |             |
|                | Rhodobacterales| 34    | 21             | 4            | 11          | 4           | 2            |             |             |
|                | Pseudomonadales| 16    | 13             | 3 **         | 1           | 7           | 3 **         |             |             |
|                | Xanthomonadales| 16    | 9              | 1            | 2           | 2           | 3            | 1           |             |
|                | Caulobacterales| 12    | 4              |              |             | 1           |              |             |             |
|                | Steroidobacterales| 7 | 5             | 4            | 1           |             |             |             |             |
|                | Acetobacterales| 3     | 2              | 1            | 1           |             |             |             |             |
|                | Enterobacterales| 3    | 1              |              |             |             |             |             |             |
| Gammaproteobacteria | * | 2    | 1              | 1            |             |             |             |             |             |
|                | Azospirillales | 1     | 1              |              |             |             |             |             |             |
|                | Reyranellales  | 1     | 1              | 1            |             |             |             |             |             |
|                | Rickettsiales  | 1     | 1              | 1            |             |             |             |             |             |
| Alphaproteobacteria | * | 1    | 1              |              |             |             |             |             |             |
|                | “PLTA13”       | 1     | 1              |              |             |             |             |             |             |
|                | “PHOS-HD29”    | 1     | 1              | 1            |             |             |             |             |             |
|                | Tistrellales   | 1     | 0              |              |             |             |             |             |             |
| Actinobacteriota | Micrococcales | 13    | 10             | 3            | 1           | 1           | 3            | 1           | 1           |
|                | Propionibacterales| 8 | 5             | 4            |             |             |             |             |             |
|                | Microtrichales | 8     | 4              | 2            | 1           | 1           |             |             |             |
|                | Corynebacteriales| 5    | 3              | 1            | 1           |             |             |             |             |
|                | “PeM15”        | 3     | 1              |              |             |             |             |             |             |
|                | Gaiellales     | 3     | 1              | 1            |             |             |             |             |             |
|                | Solirubrobacterales| 3 | 1             |              |             |             |             |             |             |
|                | Actinobacteriota | * | 3    | 0              |              |             |             |             |             |
|                | Frankiales     | 1     | 1              |              |             |             |             |             |             |
| Actinobacteriota | Kineosporiales | 1     | 1              |              |             |             |             |             |             |
| Phylum                  | Order                  | Total | No. Indicators | High P and N | High P Only | High N Only | Low P and N | Low P Only | Low N Only |
|------------------------|------------------------|-------|----------------|--------------|-------------|-------------|-------------|------------|------------|
| Cyanobacteria          |                        | 14    | 10             | 2            | 1           | 1           | 6           |
|                        | Cyanobacteria          | 7     | 5              | 2            | 1           | 1           |             |            |            |
|                        | “SepB-3”               | 4     | 2              | 1            | 1           |             |             |            |            |
| Bacteroidota           | Chitinophagales        | 10    | 6              | 4            | 1           |             |             |            |            |
|                        | Flavobacteriales       | 5     | 5              | 1            |             |             |             |            |            |
|                        | Cytophagales           | 1     | 0              |              |             |             |             |            |            |
| Firmicutes             | Bacillales             | 8     | 5              | 1            | 1           | 2           | 1           |
|                        | Exiguobacterales       | 4     | 3              |              |             |             |             |            |            |
|                        | Alicyclobacillales     | 1     | 1              |              |             |             |             |            |            |
|                        | Lactobacillales        | 1     | 0              |              |             |             |             |            |            |
| Planctomycetota        | Prelilulales           | 7     | 4              | 1            | 2           | 1           |
|                        | Gemmatales             | 5     | 3              | 1            |             |             |             |            |            |
|                        | Isophaerales           | 1     | 1              |              |             |             |             |            |            |
|                        | Planctomycetales       | 1     | 1              |              |             |             |             |            |            |
| Verrucomicrobiota      | Verrucomicrobales      | 11    | 9              | 1            |             |             |             | 5          | 2          |
| Chloroflexi            | Chloroflexi            | 4     | 2              | 1            | 1           |             |             |            |            |
| Acidobacteriota        | Vicinamibacterales     | 2     | 2              |              |             |             |             | 2          |            |
|                        | Blastocatellae         | 2     | 0              |              |             |             |             |            |            |
| Gemmatimonadota        | Gemmatimonadales       | 2     | 1              |              |             |             |             |            | 1          |
| Desulfobacterota       | Desulfobulbales        | 1     | 1              |              |             |             |             |            |            |
| Thermi/Deinococci      | Deinococcales          | 1     | 1              |              |             |             |             |            |            |
| Fuso-bacteriota        | Fusoibacteria          | 1     | 0              |              |             |             |             |            |            |
| Myxococcota            | Myxococcota            | 1     | 0              |              |             |             |             |            |            |
| Nitrospirota           | Nitrospireales         | 1     | 0              |              |             |             |             |            |            |

* Some ASVs could only be reliably identified to phylum/class.

** Within this order, one ASV was an indicator of high P and an indicator of low N.