

Meter-scale variation within a single transect demands attention to taxon accumulation curves in riverine microbiome studies

Bingdi Liu^{1,#}, Lin Zhang^{1,#}, Jason H. Knouft^{2,3}, Fangqiong Ling (✉)^{1,4,5,6}

¹ Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

² Department of Biology, Saint Louis University, St. Louis, MO 63103, USA

³ National Great Rivers Research and Education Center, East Alton, MO 62035, USA

⁴ Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

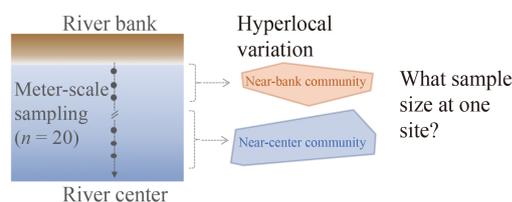
⁵ Division of Biological and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO 63110, USA

⁶ Division of Computational and Data Science, Washington University in St. Louis, St. Louis, MO 63130, USA

HIGHLIGHTS

- Riverine microbiomes exhibited hyperlocal variation within a single transect.
- Certain family-level taxa directionally associated with river center and bank.
- Taxon accumulation curves within a transect urges more nuanced sampling design.

GRAPHIC ABSTRACT



ARTICLE INFO

Article history:

Received 17 January 2022

Revised 23 March 2022

Accepted 30 March 2022

Available online 6 May 2022

Keywords:

Microbiome

Freshwater

Taxon accumulation curve

Community assembly

ABSTRACT

Microbial communities inhabiting river ecosystems play crucial roles in global biogeochemical cycling and pollution attenuation. Spatial variations in local microbial assemblages are important for detailed understanding of community assembly and developing robust biodiversity sampling strategies. Here, we intensely analyzed twenty water samples collected from a one-meter spaced transect from the near-shore to the near-center in the Meramec River in eastern Missouri, USA and examined the microbial community composition with 16S rRNA gene amplicon sequencing. Riverine microbiomes across the transect exhibited extremely high similarity, with Pearson's correlation coefficients above 0.9 for all pairwise community composition comparisons. However, despite the high similarity, PERMANOVA revealed significant spatial differences between near-shore and near-center communities ($p = 0.001$). Sloan's neutral model simulations revealed that within-transect community composition variation was largely explained by demographic stochasticity ($R^2 = 0.89$). Despite being primarily explained by neutral processes, LefSe analyses also revealed taxa from ten families of which relative abundances differed directionally from the bank to the river center, indicating an additional role of environmental filtering. Notably, the local variations within a river transect can have profound impacts on the documentation of alpha diversity. Taxon-accumulation curves indicated that even twenty samples did not fully saturate the sampling effort at the genus level, yet four, six and seven samples were able to capture 80% of the phylum-level, family-level, and genus-level diversity, respectively. This study for the first time reveals hyperlocal variations in riverine microbiomes and their assembly mechanisms, demanding attention to more robust sampling strategies for documenting microbial diversity in riverine systems.

© The Author(s) 2022. This article is published with open access at link.springer.com and journal.hep.com.cn

✉ Corresponding author

E-mail: fangqiong@wustl.edu

These authors contributed equally to this work.

Special Issue—Frontier Progresses from Chinese-American Professors of Environmental Engineering and Science (Responsible Editors: Xing Xie, Jinkai Xue & Hongliang Zhang)

1 Introduction

Streams and rivers provide numerous essential ecosystem services to human societies, including provisioning of drinking water, irrigation, and water for industrial needs, while also supporting biodiversity and ecological systems (Gilvear et al., 2016). Fundamentally, rivers and streams

are now recognized for their roles in global biogeochemical cycling, which involves active transformation of organic compounds rather than passively channeling them to the sea (Ensign and Doyle, 2006; Cole et al., 2007; Withers and Jarvie, 2008; Battin et al., 2009). The major drivers of nutrient cycling in rivers are microorganisms (Falkowski, Fenchel, and Delong, 2008; Battin et al., 2009; Fasching et al., 2020). Equally important, streams experiencing human impacts can harbor taxa that have been selected through water treatment systems and runoff from the landscape, including microbes that have evolved antibiotic resistance (McLellan et al., 2015). In order to understand the impact of rivers on biogeochemical cycles and examine anthropogenic impacts on natural environments, it is pertinent to develop robust understanding of microbial diversity, distribution, and functions in riverine systems.

Recent studies have revealed riverine microbial communities to be diverse, dynamic, and spatially heterogeneous (Read et al., 2015; Savio et al., 2015; Cruaud et al., 2020). In particular, spatial variables have emerged as the dominant factors driving alpha and beta diversity, more so than any physical and chemical parameters (Read et al., 2015; Savio et al., 2015). In terms of alpha diversity, while the River Continuum Concept predicts alpha diversity to increase as a river moves from the headwaters to the river mouth (Vannote et al., 1980), observations in riverine microbial systems have shown various different phenomena, where alpha diversity could positively or negatively associate with the mean dendritic distances/lengths from headwaters (Read et al., 2015; Savio et al., 2015), highlighting the need for understanding spatial scales at which community assembly processes operate in river systems. In addition, the intriguing spatial variations in riverine microbial communities have led to a refined employment of community assembly theories to unify the observations about the riverine microbial systems at a catchment scale. The community assembly processes in riverine microbiomes can be conceptualized as dispersals, births/deaths, and selections. Riverine microbial communities are part of a complex aquatic network where microorganisms can disperse from soil and groundwater and undergo species sorting by the local environment and ecological interactions, while the balance between dispersal and species sorting is regulated by hydraulic factors, relative locations in a network, and local water residence time (Read et al., 2015; RuizGonzalez2015; Savio et al., 2015).

Despite our increased understanding about community assembly processes on the catchment scale, community dynamics, spatial variation in community structure, and underlying community assembly mechanisms at the local scale are less understood, although with notable exceptions (Crump et al., 2012; Gweon et al., 2021). In terms of local variations, habitat has emerged as an

important driver of variation in community structure. While the type of habitat studied is context dependent, in riverine systems, “habitat” can refer to the river substrate, water column, or at a larger scale, the average characteristics of the river channel, or potentially other aspects of the physical environment. Crump et al. (2012) showed that soil water, hyporheic water, stream water, and lake water contain shared and unique taxa (Crump et al., 2012). Gweon et al. (2021) showed that free-living, particle-associated biofilms on benthic stones and rocks, and sediment type differed significantly in alpha and beta diversity across multiple sites at different locations in the river network (Gweon et al., 2021). While habitat type has been addressed, the free-living communities in the water column are usually assumed to be well-mixed, and oftentimes, one single water sample is taken as a representative sample of the local community (Read et al., 2015; Ruiz-González et al., 2015; Gweon et al., 2021). Critical questions remain to be addressed as to whether there are local variations in the stream microbial communities, whether hyperlocal variations would play a role in observed diversity and taxon abundance, and eventually affect our ability to infer community assembly processes from field data.

In this study, we address these challenges through an intense hyperlocal sampling in a transect of the Meramec River, which is located in eastern Missouri, USA. and serves as the primary source of drinking water for more than 200,000 people (The Missouri Department of Natural Resources, 2015). Here, we define “hyperlocal sampling” as sampling intensively within the geographical area that is previously represented as one site. We investigated the composition of riverine microbial communities along a river transect in order to 1) reveal potential within-transect spatial variation in riverine microbial diversity and composition, 2) identify key microbial taxa that drive within-transect variations, and 3) examine the role of within-transect variation in sampling strategies to estimate microbial diversity. Our approach was intended to reveal the small-scale variability in riverine communities by intensely sampling and analyzing microbiomes in 1-meter intervals from near the river bank to near the river center within a single transect. Upon spatial structuring being detected, linear discriminant analysis was performed to identify the microbial taxa driving the within-transect variations. Taxon-accumulation curves were generated to examine the effect of transect sample size on observed diversity. Finally, in light of recent studies exploring community assembly processes in riverine assemblages, we discuss the importance of considering taxon sampling efforts when investigating community assembly patterns and processes from field data. This study provided a unique opportunity to better understand microbial community assembly in riverine systems and facilitate the development of sampling strategies to enable more robust understandings of community assembly in rivers.

2 Methods

2.1 Sampling and DNA extraction

Water samples were collected on December 13, 2019 from the Meramec River in eastern Missouri (38.50194° N, 90.59150° W, Fig. 1A). The river discharge on that day was 1,560 cubic feet per second (44.2 cubic meters per second) recorded from a USGS stream gage located approximately 250-meter downstream from the sample collection site (07019000 Meramec River near Eureka, MO). We collected twenty grab samples (250-mL each sample) across a 20-meter transect that started 10 meters from the river bank (Fig. 1B). Each sample was collected at a depth of 0.1 m upstream of the person collecting the sample to avoid contamination. Samples were returned to shore, secured with a cap, and then filtered to collect biomass within 1 hour of collection. Water samples were pre-filtered using a 40- μ m screen (Fisherbrand, Fisher Scientific, USA), and then biomass was retrieved by a 0.22- μ m cellulose esters membrane using a Millipore Sigma Microfil V Filtration Device (Millipore Sigma, St. Louis, MO, USA). The samples were transferred to the laboratory over ice. Each filter was dissected using sterile scalpels and transferred into a Lysing Matrix E tube of Fast DNA SPIN Kit (MP Biomedicals, USA) and preserved in -80°C until DNA extraction.

2.2 DNA extraction

DNA was extracted from the filter in the Lysing Matrix E

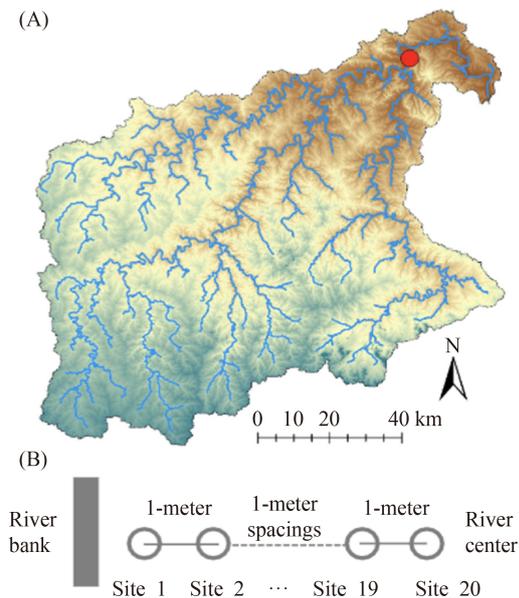


Fig. 1 Sampling location and sampling scheme. (A) Location of the sampling transect in the Meramec River watershed. (38.502075°N, 90.591522°W). (B) Schematics of intensive transect sampling. Twenty water microbiome samples were taken in 1-meter intervals from near-shore to near-center in the transect.

tube using the FastDNA SPIN Kit for Soil (MP Biomedicals, USA) within a week of sample collection. The DNA concentrations were quantified using AccuBlue® High Sensitivity dsDNA Quantitation Kit with DNA Standards (Biotium, San Francisco, CA, USA).

2.3 16S rRNA gene amplicon sequencing

Two steps of PCR were performed to prepare the 16S rRNA library (Preheim et al., 2013). In the first step, the extracted DNA template was amplified using a universal primer set 515F (GTGCCAGCMGCCGCGGTAA)/806R (GGACTACHVGGGTWTCTAAT) targeting the V4 region of the 16S rRNA gene. In the second step, unique indexing sequences were added to each library. Every sample was characterized in two technical replicates. The final products were pooled at equal molarity and then purified with Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). Sequencing was performed on an Illumina Miseq platform (2×250) at the Edison Family Center for Genome Sciences and Systems Biology at Washington University in St. Louis.

2.4 16S rRNA gene amplicon sequence denoising

Raw sequencing reads were demultiplexed using Illumina's bcltofastq2 at the sequencing facility. Upon receiving demultiplexed raw fastq paired-end reads, we completed quality filtering, denoising, and generated the amplicon sequence variant (ASV) table using the QIIME 2 platform (Bolyen et al., 2019). To achieve a quality score above 30, the forward reads were trimmed at 10 bp and truncated at 180 bp; reverse reads were trimmed at 10 bp and truncated at 150 bp. The trimmed reads were denoised and merged using the DADA2 algorithm as implemented in the q2-dada2 plugin (Callahan et al., 2016). DADA2 implements a quality-aware model of Illumina amplicon errors and resolves differences as little as one nucleotide. Reads with a number of expected errors more than 2 and detected chimeras were discarded. ASV table was thus generated. Taxonomy classification of the resulting ASVs was performed using a multinomial Naive Bayes classifier which was trained on a SILVA 138 reference database (Quast et al., 2013). The confidence threshold was set at 0.7 as default. To construct the taxonomy classifier, QIIME 2 compatible reference files were downloaded from GitHub, then reads from the reference database were extracted based on matches to our primer set using QIIME plugin feature-classifier extract-reads. The classifier was then trained using feature-classifier fit-classifier-naive-bayes. The generated QIIME artifacts were imported into R using qiime2R package for further analysis.

The ASV table was filtered to exclude ASVs with less than three reads across the entire study and those that

failed to be assigned to a domain (these sequences added up to 0.14% across all samples in the study). ASVs assigned to the Eukarya domain were excluded (these sequences added up to 0.16% across samples in the study). A Pearson correlation coefficient was calculated between technical replicates of the same sample based on relative abundance using the “cor” function in R base functions to examine technical robustness. All technical replicates resulted in values above 0.99 (Table S1), indicating high technical reproducibility. Then, technical replicates were merged by the “merge_samples” command in Phyloseq (McMurdie and Holmes, 2013). Rarefaction curves were examined using the “rarefy” function in Vegan (Oksanen et al., 2020). All twenty samples reached a plateau on rarefaction curves at ASV level (Figure S1). Following that, a single rarefaction was performed (set.seed=1234) to subsample all samples to the same sequencing depth, 108,899 seqs/sample, which was the minimum read depth across all samples.

2.5 Taxonomic profiles

Pearson correlation of 190 sample pairs based on relative abundance was computed using the R base “cor” function to examine the correlation in taxonomic compositions of riverine microbial communities. The agglomeration at class level and calculation of the relative abundances for each class in the individual sample were performed using phyloseq functions “tax_glom” and “transform_sample_counts”. All plots were created in ggplot (Wickham et al., 2019).

2.6 Diversity analyses

The alpha diversity of ASV profiles was characterized using the Shannon index and observed richness computed by the “diversity” and “specnumber” functions in vegan (Oksanen et al., 2020). Corresponding mean values of alpha diversity were calculated. First, read counts were converted to relative abundance after rarefying using the rarefy_even_depth() function. Next, Bray-Curtis dissimilarities were computed between samples based on the ASV relative abundances as the equation below, where x_{ij} and x_{ik} are relative abundances of ASV i in samples j and k (Eq. 1). Then, Bray-Curtis dissimilarities were examined with the permutational multivariate analysis of variance (PERMANOVA) using the function “adonis2” and the analysis of multivariate homogeneity of group dispersions (PERMDISP) using the function “betadisper” from vegan (Oksanen et al., 2020) with set.seed=1234. Besides Bray-Curtis distances, Jaccard distances were computed between samples based on presence and absence of ASVs as in Eq. (2), where A and B are numbers of species present in two compared samples and J are the number of shared species. After that, principal coordinates analysis (PCoA) was performed on the Bray-Curtis dissimilarity

matrix and Jaccard dissimilarity matrix using the “ordinate” function in Phyloseq based on the “vegdist” function in vegan and the “pcoa” function in ape (Paradis et al., 2004; Oksanen et al., 2020). Last, visualizations of the PCoA results were generated by the “plot_ordination” function.

$$\text{Bray-Curtis distance} = \frac{\sum_{i=1}^n |x_{ij} - x_{ik}|}{\sum_{i=1}^n |x_{ij} + x_{ik}|}, \quad (1)$$

$$\text{Jaccard distance} = \frac{A + B - 2J}{A + B - J}. \quad (2)$$

2.7 Linear discriminant analysis (LDA)

LDA effect size measurement was performed using the package LefseR (Segata et al., 2011; Khleborodova, 2020) with set.seed=1234. Briefly, Lefse examined if features from two sample groups were significantly different in abundances by first screening for those taxa that rejected a null hypothesis of equal average abundance, and then screening by effect sizes. As a first screen, differentially abundant features were screened for by passing a Kruskal-Wallis test with an adjusted alpha setting at 0.05 as a coarse screen. Next, a linear discriminant analysis was performed to screen for differentially abundant taxa by primary LDA scores. The primary LDA score of a taxon was calculated by averaging the differences between actual group means with the differences between group means along the first linear discriminant axis. LefseR would output a list of features with log10 transformed LDA score larger than 2. Then, a plot with cumulative LDA scores from ASVs of the same family was generated. Last, to examine the abundance differences at a family level, we combined ASVs identified as differentially abundant taxa by their family-level classification for visualization.

2.8 Taxon-accumulation curves

Taxon-accumulation curves at taxonomic units of phylum, family, and genus were generated using the function “specaccum” from vegan (Oksanen et al., 2020). The confidence intervals were estimated from the mean value and standard deviation of 10000 random permutations with ci (the Z value) = 1.96, ci.type=“bar”.

3 Results

3.1 Taxonomic compositions of the Meramec River transect near Eureka, Missouri

Twenty water samples were taken in one-meter intervals along a river transect in the Meramec River in eastern Missouri, and the samples were analyzed for 16S rRNA gene amplicons (Fig. 1). These samples resulted in a total

of 2,700,925 reads and 2,975 ASVs. A total of 2,177,980 reads and 2,953 ASVs remained after rarefaction to 108,899 seqs/sample. Four bacterial phyla were abundant along the transect (average relative abundance > 3%, Fig. 2A). These were *Bacteroidota* ($43 \pm 4.5\%$), *Proteobacteria* ($39 \pm 4.3\%$), *Actinobacteriota* ($11 \pm 0.74\%$), and *Cyanobacteria* ($3.1 \pm 0.42\%$). Several phyla were detected at average relative abundances above or close to 0.1% across river water samples, which were *Verrucomicrobiota* ($1.1 \pm 0.12\%$), *Firmicutes* ($0.56 \pm 0.12\%$), *Campilobacterota* ($0.48 \pm 0.065\%$), *Patescibacteria* ($0.34 \pm 0.060\%$), *Desulfobacterota* ($0.18 \pm 0.036\%$), *Bdellovibrionota* ($0.14 \pm 0.034\%$), *Acidobacteriota* ($0.11 \pm 0.025\%$), *Chloroflexi* ($0.11 \pm 0.023\%$) and *Planctomycetota* ($0.094 \pm 0.029\%$).

Within *Bacteroidota*, the *Bacteroidia* class was the most abundant ($43 \pm 4.5\%$ of the entire community), and within *Proteobacteria*, the *Gammaproteobacteria* class was the most abundant ($35 \pm 3.7\%$), followed by *Alphaproteobacteria* ($4.1 \pm 0.76\%$). Nine classes were detected at average relative abundances above or close to 0.1% across river water samples, which were *Verrucomi-*

crobiae ($0.94 \pm 0.12\%$), *Campylobacteria* ($0.48 \pm 0.065\%$), *Acidimicrobiia* ($0.45 \pm 0.082\%$), *Clostridia* ($0.35 \pm 0.077\%$), *Parcubacteria* ($0.25 \pm 0.062\%$), *Bacilli* ($0.17 \pm 0.052\%$), *Desulfuromonadia* ($0.15 \pm 0.033\%$), *Chlamydiae* ($0.14 \pm 0.063\%$) and *Bdellovibrionia* ($0.088 \pm 0.029\%$). Several other bacterial classes were detected albeit at smaller fractions (average relative abundance < 0.1%), including *Planctomycetes*, *Gracilibacteria*, *Vicinamibacteria*, *Omnitrophia* and *Oligoflexia* (Fig. 2A). Among the *Bacteroidota* phylum, the detected taxa primarily related to *Flavobacteriales*, followed by *Cytophagales* and *Chitinophagales*. Among all bacterial families detected, *Flavobacteriaceae* and *Comamonadaceae* are the two most abundant, taking up a total of over 50% sequencing reads across the transect. The twenty most abundant bacterial families are provided in Table S2.

3.2 Riverine microbiome composition across the transect were highly similar

Strikingly, the microbial communities across the transect were highly similar. When compared against each other

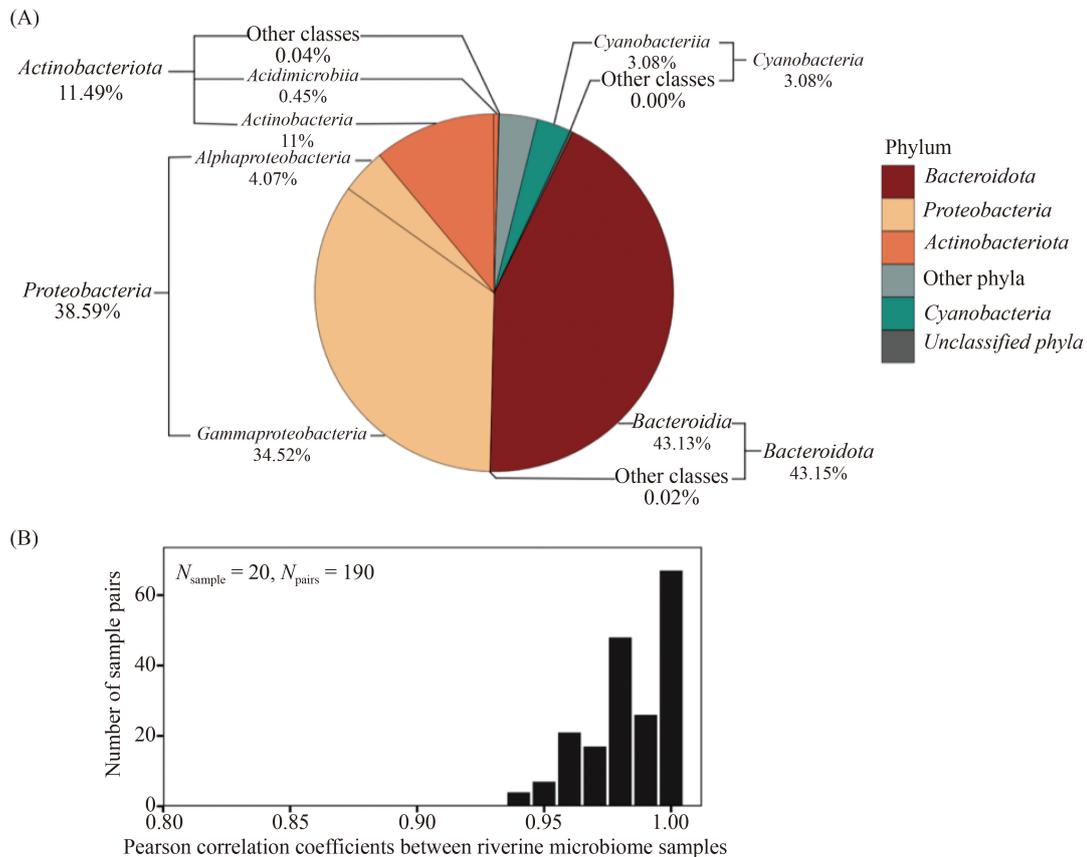


Fig. 2 High reproducibility and microbiome profiles of sites within a transect. (A) The average relative abundance of abundant phyla (> 3% in relative abundance) and classes (> 0.05% in relative abundance) in all river water microbiome samples. The taxa that were assigned as “NA” at phylum level were grouped into “unclassified Phyla” and the phyla with mean relative abundance less than 3% were grouped into “other phyla”. The classes from “other phyla” were not labeled and classes assigned with “NA” or with mean relative abundance less than 0.05% were grouped into “other Class” under their own phyla. (B) The distribution of Pearson correlation coefficients of 190 pairs from twenty sites.

using a Pearson correlation coefficient, the 190 unique sample pairs resulting from the 20 samples ranged from 0.94 to 1.00 and exhibited a unimodal distribution peaking close to one (Fig. 2B). The extremely high similarity across a transect supports the suggestion that rivers are open ecosystems where mixing is pervasive.

3.3 Spatial differences in microbial community composition emerged between near-shore and near-center groups despite overall similarities

Observing the strong correlation between samples, we further asked: are there any fine differences between community compositions across a transect? Using principal coordinate analysis (PCoA) as an unsupervised learning tool with no a priori assumptions about any grouping, we visualized Bray-Curtis dissimilarities between the samples. In the visualization from the first two principal coordinates, the riverine microbiome samples that were collected near the riverbank (samples 1–10) grouped separately from those collected near the river center (samples 11–20, Fig. 3A). It should be noted that the group naming was made to reflect the pattern in Fig. 3A. The PCoA from Jaccard distances showed a similar grouping, indicating that community composition differences were caused by both presence and absence of microbial taxa and their differing relative abundances (Figure S2).

It should be noted that the within-transect community composition differentiation emerged despite an overall high similarity among samples. When we examined the pairwise Bray-Curtis dissimilarities of samples within a near-shore or near-center cluster in the PCoA space (Fig. 3A, PERMANOVA $p = 0.001 < 0.05$, PERMDISP $p = 0.3 > 0.05$, mean distance to centroid for riverbank group = 0.055, mean distance to centroid for river center group = 0.06), the average within-group Bray-Curtis dissimilarity was as low as 0.08 (Fig. 3B, 90 unique samples pairs). When examined between these two groups, the average Bray-Curtis dissimilarity was as low as 0.14 (Fig. 3C, 100 unique sample pairs). Even between groups, community composition was extremely similar. These results indicate fine spatial differences within an overall very similar transect continuity. In terms of fluid dynamics, mass and heat transfer, the riverbank can present slower flow, higher temperature, and chemicals from the soil, which could all potentially lead to varied environments across a river transect.

3.4 Informative taxa were identified across a river transect

Since we detected the differentiation of microbial communities across the river transect, we explored the taxa driving this difference. Thus, an LDA analysis was performed for river water samples from the two clusters in Fig. 3A. Seven ASVs were determined to be associated with the near-center cluster (Table S3), and these ASVs were classified into five families: *Comamonadaceae*,

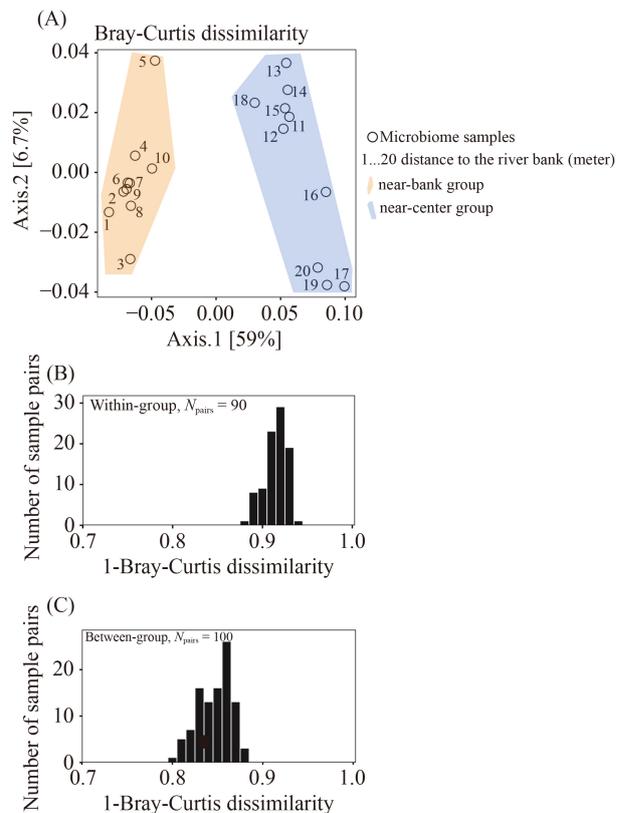


Fig. 3 Near-shore and near-center groups emerged in PCoA analysis of Bray-Curtis dissimilarities. (A) The principal coordinate analysis resulting from Bray-Curtis dissimilarities revealed a dramatic difference in composition between near-shore and near-center samples. Circles indicate samples; numbers indicate the distance between the sampling site and the shore. Orange and blue polygons indicate near-shore and near-center groups. (B) Histogram of 1-Bray-Curtis dissimilarity values from 90 pairs of within-group comparisons in near-shore or near-center groups. (C) Histogram of 1-Bray-Curtis dissimilarity values from 100 pairs between-group comparisons across near-shore or near-center groups.

Microbacteriaceae, *Burkholderiaceae*, *Rhodobacteraceae*, and *Sphingomonadaceae*. Twelve ASVs were associated with the river bank water microbiome and were classified to five families (Table S3): *Flavobacteriaceae*, *Crocinitomicaceae*, *Spirosomaceae*, *Methylophilaceae*, and *Sporichthyaceae* (Fig. 4A).

It should be noted that while statistically significant variation was detected, such variations occurred in a gradual manner (Fig. 4B). The most enriched ASVs near shore were *Flavobacteriaceae*, *Crocinitomicaceae*, and *Spirosomaceae*; all of these were *Bacteroidetes*. *Bacteroidetes* are dominant phyla in soil environments due to their traits including the ability to secrete diverse carbohydrate-active enzymes (Larsbrink and McKee, 2020). Potentially, the taxa that were more abundant near the bank were seeded from the soil or plant debris. While the differentiative ASVs presented a small fraction of the diversity, their relative abundance cannot be ignored. The *Flavobacteriaceae* taxa that are also highly variable took

up above 8% of all samples. Differentiative ASVs under *Comamonadaceae* constituted 5%–15% relative abundance.

3.5 Impact of hyperlocal variation on microbiome sampling

The within-transect variation of riverine microbiomes demands us to consider: how will the hyperlocal variation affect our ability to document the diversity at the riverine microbiome at one location? Therefore, we employed taxon accumulation curves to examine the impact of sample size on representation of diversity at a local site. We found that even at the phylum level, the taxon accumulation curve did not become saturated after 20 samples were collected (Fig. 5A). When within-transect sample size reaches twenty, the estimated means of richness at the phylum, family, and genus level were 43 phyla, 413 families, and 716 genera, respectively. The sample size to achieve 80% of the highest richness are 4 samples, 6 samples, and 7 samples for phylum, family, and genus levels, respectively (Fig. 5). These results support that hyperlocal variation has a profound impact on taxon diversity sampling. The pervasive approach of one sample per site can severely underestimate local diversity in riverine microbial assemblages.

4 Discussion

Our results demonstrate a clear structuring of the bacterial communities within a single transect of a stream even though the overall community composition and alpha diversity were quite similar among samples (Figs. 2 and 3). The community composition differences were mainly driven by differentially abundant taxa from near-shore to the near-center of the river (Fig. 4). Taxa of which the abundance varied directionally were detected by LefSe analyses. We are not certain what factor is driving differences in community composition between near-shore and near-center samples, which appear to occur between samples 10 and 11. Such a distinct break was unexpected, and we were not able to capture the source of this variation in our sample design. However, we hypothesize that substrate and/or hydraulic variation is causing differentiated representation of microbes in the water column. The taxon abundance curves were non-saturated even with twenty samples collected across the transect (Fig. 5), despite the ASV-level rarefaction curve reaching a plateau in every sample in terms of sequencing efforts (Fig. S1).

While taxon accumulation curves are classic tools in ecology to document biodiversity and examine sampling efforts, to our best knowledge, our study is the first to examine microbial taxon accumulation curves along a river transect. The finding that even at the phylum level,

twenty samples did not fully capture the riverine microbial community diversity was surprising. Nevertheless, with an extensive sampling scheme and deep sequencing depth, our taxon accumulation curves likely captured many of the rare taxa in the system. Despite the importance of taxon abundance curves in ecological experimental design, the results here should not be taken as prescriptive. Instead, we hope this study could motivate future site-specific taxon accumulation curve studies, the data from which will eventually drive a consensus in sample size determination for riverine microbial community studies.

Our findings could hold significance for many applications comparing alpha diversity across sites in riverine microbial assemblages, yet perhaps the most pertinent immediate next question would be to encourage a more deliberate sampling approach when exploring the recruitment of taxa into aquatic systems as a river develops. Recent studies have considered the dispersal of microorganisms from soil, soil water, and sediment to river water as important mechanisms in riverine community assembly (Crump et al., 2012; Ruiz-González et al., 2015; Gweon et al., 2021). In particular, Ruiz-González et al. (2015) found that across a catchment of a La Cote-Nord region of Quebec, there was a group of hypothetical “seed taxa” that rose in their relative abundance as the river Strahler order increased (Ruiz-González et al., 2015). The fact that some later-dominant seed microorganisms could appear at low relative abundance at their entry suggests that sampling efforts would be important in capturing these taxa and infer *the order of* their dispersals, which are essential to infer the role of priority effects in shaping community assemblages (Fukami, 2015; Svoboda et al., 2018). Thus, a more deliberate approach in sampling microbial assemblages in riverine systems that considers taxon abundance curves may help us understand the role of priority effects in shaping community assemblages estimated from field-collected data.

We identified informative taxa characterized by large relative abundance differences in the near-shore and near-center groups. The taxa that were associated with the river center, namely, *Comamonadaceae*, *Microbacteriaceae*, *Burkholderiaceae*, *Rhodobacteraceae* and *Sphingomonadaceae*, are taxa that are frequently associated with aquatic environments and even prevalent in potable water (Newton et al., 2011; Ling et al., 2018). In contrast, multiple ASVs associated with the near-shore group were *Flavobacteriaceae* and *Crocinitomicaceae*. These groups are chemoorganotrophs and previously detected in various aquatic and soil environments. In addition, *Spirocomaceae* was associated with near-shore groups. Microorganisms under the *Spirochetes* phylum are known to have a twisting mobility (Charon et al., 2012), perhaps, the mobility of this lineage played a role in their dispersals. The list of organisms revealed in our analysis as near-shore associations can help form more detailed

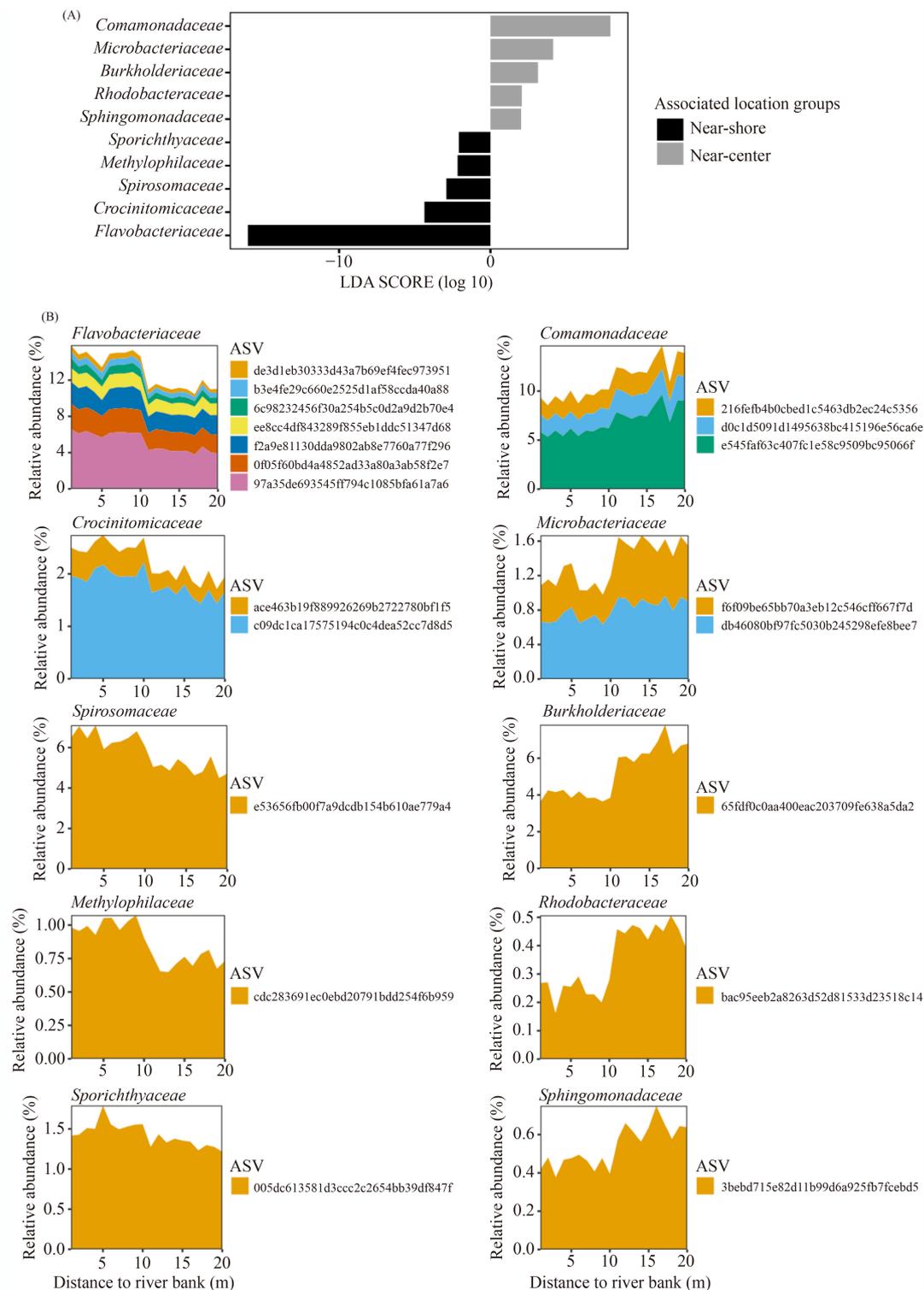


Fig. 4 The informative ASVs explained the community composition difference in near-shore and near-center groups. (A) Linear discriminant analysis (LDA) effect size measurement at the ASV level and accumulated with the family names. (B) Area plots of relative abundance of informative ASVs accumulating at the family level.

hypotheses about soil-to-water microbial dispersals.

It should be noted that multiple taxa exhibited directional variation, as shown in Fig. 4. However, not all taxa exhibited this variation, suggesting that stochasticity

played a role in community composition. The prevalence of taxa is better explained by the Sloan's version of the neutral model than statistical null models (Supplementary Text, Table S4), suggesting that the community

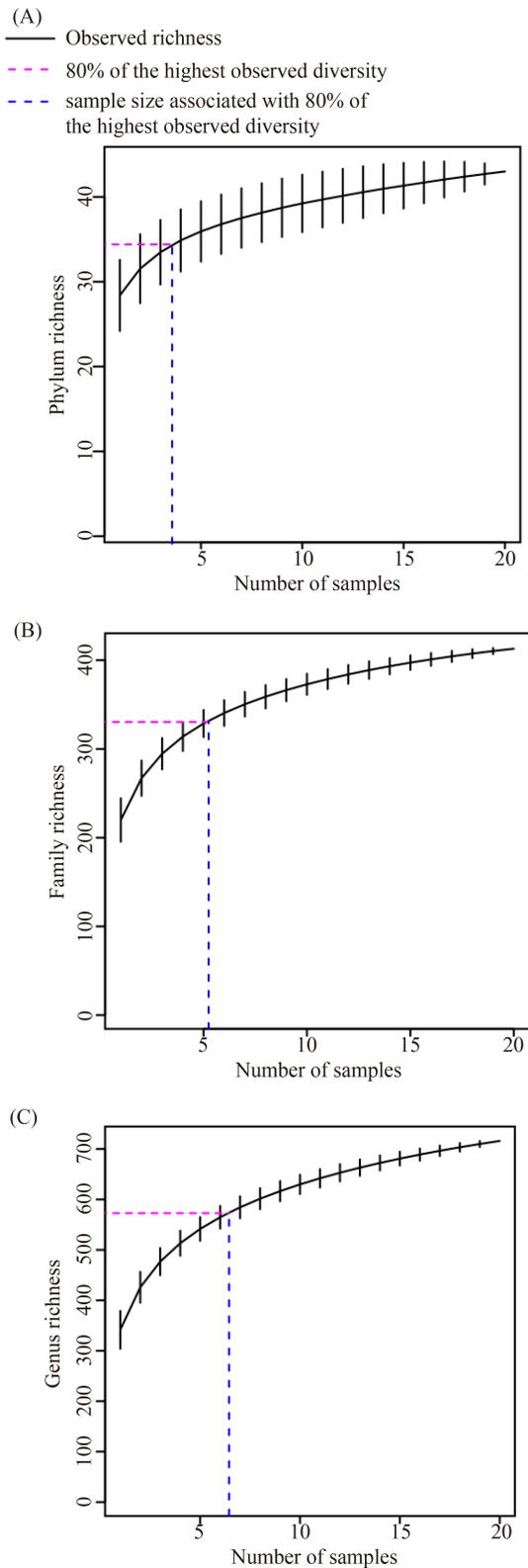


Fig. 5 Taxon accumulative curves within a transect exhibiting the detection efficiency for the richness of phylum (A), family (B), and genus (C). The vertical bars indicate the confidence intervals, and the lines indicate the mean number of taxa accumulated.

composition variations cannot be explained solely by variation due to sampling artefacts.

We acknowledge that, in this study, the community composition was examined using 16S rRNA gene amplicon sequencing, and hence limited our ability to examine sub-species diversity and variability. Studies characterizing microbial diversity from Vibrionaceae in coastal environments have suggested that multiple genotypes can co-exist within a natural aquatic environment and exhibit different functions (Thompson et al., 2005). While strain diversity has not been directly characterized in lotic systems, we believe this is an important area to explore in future research because strain diversity has helped explain functional redundancy in many microbial assemblages (Louca et al., 2018). Currently, strain diversity is only starting to be integrated into the study of community assembly processes, yet enlightening results have been suggested to explain alternative community types in biofilm communities (Leventhal et al., 2018). Nevertheless, our results, which encourage a more deliberate approach in sampling microbial diversity from local communities, could aid the discovery of genetic diversity in riverine microbial communities from the same samples, and eventually facilitate the integration of strain diversity in disentangling community assembly processes in riverine systems.

5 Conclusions

In this study, we examined the hyperlocal variation of microbial communities by intensively sampling a single transect in the Meramec River watershed. Our PERMANOVA analysis detected significant near-bank and near-center differences in the microbial communities despite overall high similarity. Sloan's neutral model simulations revealed that within-transect community composition variation was largely explained by demographic stochasticity ($R^2 = 0.89$). Taxa indicative of near-bank or near-center differences were identified through LefSE analysis. After observing the within-transect variation, we constructed taxon-accumulation curves at the phylum, family, and genus levels to examine the effect of within-transect variation on documenting alpha diversity. Sampling effort at as many as twenty samples did not fully saturate the diversity at the genus level, yet four, six and seven samples were able to capture 80% of the phylum-level, family-level, and genus-level diversity, respectively. Our results are useful for designing future field studies on community assembly, as research comparing across sites and locations can incorporate taxon-accumulation curves to draw more nuanced conclusions about the dispersals of taxa. This study for the first time reveals hyperlocal variations in riverine microbiomes and their assembly mechanisms, demanding attention to more robust sampling strategies for documenting microbial diversity in riverine systems.

Acknowledgements This work is partially supported by a Living Earth Collaborative Seed Grant to FL and JHK and a Ralph Powe Junior Faculty Enhancement Award to FL.

Electronic Supplementary Material Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s11783-022-1543-6> and is accessible for authorized users.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Battin T J, Luysaert S, Kaplan L A, Aufdenkampe A K, Richter A, Tranvik L J (2009). The boundless carbon cycle. *Nature Geoscience*, 2(9): 598–600
- Bolyen E, Rideout J R, Dillon M R, Bokulich N A, Abnet C C, Al-Ghalith G A, Alexander H, Alm E J, Arumugam M, Asnicar F, et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8): 852–857
- Callahan B J, McMurdie P J, Rosen M J, Han A W, Johnson A J A, Holmes S P (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7): 581–583
- Charon N W, Cockburn A, Li C, Liu J, Miller K A, Miller M R, Motaleb M A, Wolgemuth C W (2012). The unique paradigm of spirochete motility and chemotaxis. *Annual Review of Microbiology*, 66(1): 349–370
- Cole J J, Prairie Y T, Caraco N F, McDowell W H, Tranvik L J, Striegl R G, Duarte C M P, Kortelainen J A, Downing J A, Middelburg J J, Melack J (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* (New York, N.Y.), 10(1): 172–185
- Cruaud P, Vigneron A, Fradette M S, Dorea C C, Culley A I, Rodriguez M J, Charette S J (2020). Annual bacterial community cycle in a seasonally ice-covered river reflects environmental and climatic conditions. *Limnology and Oceanography*, 65(S1): S21–S37
- Crump B C, Amaral-Zettler L A, Kling G W (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME journal*, 6(9): 1629–1639
- Ensign S H, Doyle M W (2006). Nutrient spiraling in streams and river networks. *Journal of Geophysical Research. Biogeosciences*, 111(G4)
- Falkowski P G, Fenchel T, Delong E F (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879): 1034–1039
- Fasching C, Akotoye C, Bižić M, Fonvielle J, Ionescu D, Mathavarajah S, Zoccarato L, Walsh D A, Grossart H, Xenopoulos M A (2020). Linking stream microbial community functional genes to dissolved organic matter and inorganic nutrients. *Limnology and Oceanography*, 65(S1): S71–S87
- Fukami T (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1): 1–23
- Gilvear D J, Greenwood M T, Thoms M C, Wood P J (2016). *River Science: Research and Management for the 21st Century*. Hoboken: John Wiley & Sons
- Gweon H S, Bowes M J, Moorhouse H L, Oliver A E, Bailey M J, Acreman M C, Read D S (2021). Contrasting community assembly processes structure lotic bacteria metacommunities along the river continuum. *Environmental Microbiology*, 23(1): 484–498
- Khleborodova A (2020). Lefser: R implementation of the LEfSE method for microbiome biomarker discovery (Version R package version 140)
- Larsbrink J, McKee L S (2020). Bacteroidetes bacteria in the soil: Glycan acquisition, enzyme secretion, and gliding motility. *Advances in Applied Microbiology*, 110: 63–98
- Leventhal G E, Boix C, Kuechler U, Enke T N, Sliwerska E, Holliger C, Cordero O X (2018). Strain-level diversity drives alternative community types in millimetre-scale granular biofilms. *Nature Microbiology*, 3(11): 1295–1303
- Ling F, Whitaker R, LeChevallier M W, Liu W T (2018). Drinking water microbiome assembly induced by water stagnation. *The ISME journal*, 12(6): 1520–1531
- Louca S, Polz M F, Mazel F, Albright M B N, Huber J A, O'Connor M I, Ackermann M, Hahn A S, Srivastava D S, Crowe S A, Doebeli M, Parfrey L W (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, 2(6): 936–943
- McLellan S L, Fisher J C, Newton R J (2015). The microbiome of urban waters. *International microbiology: The official journal of the Spanish Society for Microbiology*, 18(3): 141–149
- McMurdie P J, Holmes S (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4): e61217
- Newton R J, Jones S E, Eiler A, McMahan K D, Bertilsson S (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and molecular biology reviews*, 75(1): 14–49
- Oksanen J, Blanchet F G, Kindt R, Legendre P, Minchin P R, O'Hara R B, Simpson G L, Solymos P, Stevens M H H, Wagner H (2020). *Vegan: Community Ecology Package* (Version R package version 25–7)
- Paradis E, Claude J, Strimmer K (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* (Oxford, England), 20(2): 289–290
- Preheim S P, Perrotta A R, Friedman J, Smilie C, Brito I, Smith M B, Alm E (2013). Computational methods for high-throughput comparative analyses of natural microbial communities. *Methods in Enzymology*, 531: 353–370
- Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F O (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue): D590–D596
- Read D S, Gweon H S, Bowes M J, Newbold L K, Field D, Bailey M J,

- Griffiths R I (2015). Catchment-scale biogeography of riverine bacterioplankton. *ISME journal*, 9(2): 516–526
- Ruiz-González C, Niño-García J P, Del Giorgio P A (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters*, 18(11): 1198–1206
- Savio D, Sinclair L, Ijaz U Z, Parajka J, Reischer G H, Stadler P, Blaschke A P, Blöschl G, Mach R L, Kirschner A K T, Farnleitner A H, Eiler A (2015). Bacterial diversity along a 2600 km river continuum. *Environmental Microbiology*, 17(12): 4994–5007
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett W S, Huttenhower C (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6): R60
- Svoboda P, Lindström E S, Ahmed Osman O, Langenheder S (2018). Dispersal timing determines the importance of priority effects in bacterial communities. *ISME journal*, 12(2): 644–646
- The Missouri Department of Natural Resources (2015). The state of our Missouri waters-Meramec river watershed. Retrieved from www.ewgateway.org/wp-content/uploads/2017.08/MRP-MeramecRiverWatershed.pdf
- Thompson J R, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt D E, Benoit J, Sarma-Rupavtarm R, Distel D L, Polz M F (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science*, 307(5713): 1311–1313
- Vannote R L, Minshall G W, Cummins K W, Sedell J R, Cushing C E (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(1): 130–137
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemond G, Hayes A, Henry L, Hester J, et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43): 1686
- Withers P J A, Jarvie H P (2008). Delivery and cycling of phosphorus in rivers: A review. *Science of the total environment*, 400(1–3): 379–395