Nutrient dynamics and stream order influence microbial community patterns along a 2914 km transect of the Mississippi River

Michael W. Henson[1], Jordan Hanssen[2], Greg Spooner[2], Patrick Fleming[2], Markus Pukonen[2], Frederick Stahr[3], and J. Cameron Thrash[1*]

[1] Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, U.S.A.
[2] O.A.R. Northwest, Seattle, WA 98103, U.S.A.
[3] School of Oceanography, University of Washington, Seattle, WA 98195, U.S.A.

*Correspondence:

J. Cameron Thrash
Louisiana State University
Department of Biological Sciences
202 Life Sciences Bldg.
Baton Rouge, LA 70803
Phone: 225-578-8210
Fax: 225-578-2597
thrashc@lsu.edu

Running title: Microbial regime changes on the Mississippi River

Keywords: Mississippi River, WGCNA, microbial ecology, eutrophication

Abstract

Draining 31 states and roughly 3 million km^2, the Mississippi River (MSR) and its tributaries constitute an essential resource to millions of people for clean drinking water, transportation, agriculture, and industry. Since the turn of the 20^{th} century, MSR water quality has continually rated poorly due to human activity. Acting as first responders, microorganisms can mitigate, exacerbate, and/or serve as predictors for water quality, yet we know little about their community structure or ecology at the whole river scale for large rivers. We collected both biological (16S and 18S rRNA gene amplicons) and physicochemical data from 38 MSR sites over nearly 3000 km from Minnesota to the Gulf of Mexico. Our results revealed a microbial community composed of similar taxa to other rivers but with unique trends in the relative abundance patterns among phyla, OTUs, and the core microbiome. Furthermore, we observed a separation in microbial communities that mirrored the transition from an 8^{th} to 10^{th} Strahler order river at the Missouri River confluence, marking a different start to the lower MSR than the historical distinction at the Ohio River confluence in Cairo, IL. Within MSR microbial assemblages we identified subgroups of OTUs from the phyla Acidobacteria, Bacteroidetes, Oomycetes, and Heterokonts that were associated with, and predictive of, the important eutrophication nutrients nitrate and phosphate. This study offers the most comprehensive view of MSR microbiota to date, provides important groundwork for higher resolution microbial studies of river perturbation, and identifies potential microbial indicators of river health related to eutrophication.

Introduction

By connecting terrestrial, lotic, and marine systems, rivers perform vital roles in both the transport and processing of compounds in all major global biogeochemical cycles (Richey *et al.* 2002; Ensign and Doyle 2006; Withers and Jarvie 2008; Battin *et al.* 2009; Savio *et al.* 2015). Within the carbon cycle alone, rivers collectively discharge organic carbon to the oceans at over 0.4 Pg C yr^{-1} (Cauwet *et al.* 2002). Perhaps more importantly, rivers are generally net heterotrophic (Cole *et al.* 2007), indicating that they not only transport organic matter but host active metabolic processing of it as well. Conservative estimates place heterotrophic output of the world's fluvial networks (streams, rivers, and estuaries) at 0.32 Pg C yr^{-1} (Cole and Caraco 2001; Battin *et al.* 2009). Although rivers contain a small minority of global fresh water at any given moment, the considerable volumes that pass through these systems make them relevant to models attempting to quantify global elemental transformations. However, despite the fact that microbial functions likely play a vital role in ecosystem health for both rivers themselves and their places of discharge, microbial functions in rivers remain understudied.

At 3734 km, the Mississippi River (MSR) is the fourth longest on earth, draining 31 U.S. states and two Canadian provinces- a watershed consisting of 41% of the continental U.S (Turner and Rabalais 2003; Dagg *et al.* 2004). The MSR is a major source of drinking water for many U.S. cities; a critical thoroughfare for transportation, commerce, industry, agriculture, and recreation; and conveys the vestiges of human activity to the Gulf of Mexico (GOM). In New Orleans, the average flow rate is over 16,990 cubic meters s^{-1} (cms) (Rabalais *et al.* 1996), but can exceed 84,000 cms during flood stages (Singh 2012), and carries over 150 x 10^9 kg of suspended sediment into the northern GOM annually (Dagg *et al.* 2004, 2005). The MSR also transports considerable amounts of carbon, nitrogen and phosphorus, with average annual fluxes

of 2.1 TgC yr^{-1}, > 1.4 TgN $year^{-1}$, and > 0.14 TgP yr^{-1}, respectively (Goolsby and Battaglin 2001; Cauwet *et al.* 2002; Aulenbach *et al.* 2007). Globally, the MSR represents 0.8% of the dissolved organic carbon flux to the worlds oceans (Cauwet *et al.* 2002). When considering total nitrogen, up to 62% can occur as nitrate (NO_3^-) (Goolsby and Battaglin 2001). This massive discharge of excess eutrophic nutrients, primarily from corn and soybean (nitrogen) and animal manure (phosphorus) runoff (McIsaac *et al.* 2001; Turner and Rabalais 2004; Alexander *et al.* 2008; Schilling *et al.* 2010; Duan *et al.* 2014; Staley *et al.* 2014a), fuels one of the largest marine zones of seasonal coastal hypoxia in the world (Rabalais *et al.* 2002, 2007; Bianchi *et al.* 2010; Bristow *et al.* 2015). Studying microbial relationships to river eutrophication will improve our understanding of their contributions to either mitigating or exacerbating nutrient input.

Far from a homogenous jumble of organisms ferried downriver, microbial community composition changes with distance from the river mouth and/or from the influence of tributaries (Kolmakova *et al.* 2014; Read *et al.* 2015; Savio *et al.* 2015), resulting from altered nutrient concentrations (Staley *et al.* 2014a; Van Rossum *et al.* 2015; Meziti *et al.* 2016), differing dissolved organic matter (DOM) sources (Ruiz-González *et al.* 2013; Zeglin 2015; Blanchet *et al.* 2016), and land use changes (Staley *et al.* 2014b; Van Rossum *et al.* 2015; Zeglin 2015). Past studies of the Thames, Danube, Yenisei, and Columbia Rivers have found that planktonic river microbial assemblages were dominated by the phyla Actinobacteria, Proteobacteria, and Bacteriodetes; and taxa such as acI Actinobacteria, *Polynucleobacter* spp., GKS9 and LD28 *Betaproteobacteria,* CL500-29 Actinobacteria, LD12 SAR11 *Alphaproteobacteria,* and *Novosphingobium* spp. (Crump *et al.* 1999; Kolmakova *et al.* 2014; Read *et al.* 2015; Savio *et al.* 2015).

Specific to the MSR, previous 16S rRNA gene amplicon and metagenomic studies have

109 demonstrated that microbial assemblages in the Minnesota portion of the river (Lake Itasca to La
110 Crescent) correlated with organic carbon concentration, total dissolved solids, and land use
111 changes (Staley *et al.* 2013, 2014a; b). Species richness increased near developed land with
112 greater concentrations of nitrate and nitrite, as opposed to sites near pastureland that had greater
113 organic carbon (Staley *et al.* 2014a). Similar to past studies of other lotic systems (Crump *et al.*
114 2012; Read *et al.* 2015; Savio *et al.* 2015), Actinobacteria and Proteobacteria increased in
115 relative abundance along the upper MSR, while OTUs belonging to Bacteroidetes decreased
116 (Staley *et al.* 2013). A follow up time-series study over two summers showed a distinct seasonal
117 signal: samples from late summer of both years were more similar to each other than those from
118 early summer (Staley *et al.* 2015). A more recent study of another portion of the MSR (above the
119 Missouri River confluence to Natchez, La.) found bacterial communities formed distinct
120 regimes, likely due to the strong influence of their respective tributaries (Jackson *et al.* 2014;
121 Payne *et al.* 2017).

122 Researchers have suggested that some of these patterns supported the application of the
123 River Continuum Concept (RCC) (Vannote *et al.*, 1980) to river microbiota. The RCC postulates
124 that as a river increases in size, the influences of riparian and other inputs will decrease as the
125 river establishes a dominant core community (Vannote *et al.* 1980). Richness will increase with
126 stream order complexity before decreasing in higher order rivers (Vannote *et al.*, 1980).
127 Therefore, as continuous systems with increasing volumes and residence times, river microbiota
128 should transition from experiencing strong influences of mass effects from terrestrial, riverbed,
129 and tributary sources to systems where species sorting plays a more important role (Crump *et al.*
130 2012; Besemer *et al.* 2013; Savio *et al.* 2015; Niño-García *et al.* 2016).

Accordingly, studies examining the linkages between RCC and microbial structuring mechanisms (e.g. species sorting, mass effects) have found that some of the results indeed supported the RCC. Specifically, numerous lotic studies found that upstream and headwaters amassed high species richness, before richness decreased further downstream (Crump *et al.* 2012; Kolmakova *et al.* 2014; Savio *et al.* 2015; Niño-García *et al.* 2016). This follows the expected pattern outlined by the RCC where early stream and river microbial communities are heavily influenced by a large input of allocthonous organisms from sediment and groundwater into the river network. As the rivers progressed, such mass effects likely heeded to species sorting, a process requiring an increased water residence time to allow for the selection of species based on local environmental conditions (Crump *et al.* 2012; Savio *et al.* 2015; Niño-García *et al.* 2016). These types of river modulations, influenced by hydrology and environmental conditions, have correlated with decreased species richness and increased core community membership (the taxa found consistently throughout the entire river) in multiple systems (Crump *et al.* 2012; Besemer *et al.* 2013; Savio *et al.* 2015; Niño-García *et al.* 2016). An analyses of an upper portion of the MSR provided evidence for the importance of species sorting (Staley *et al.* 2015). However, in a study of the Thames River, researchers found no association between microbial assemblages and any measured variables other than residence time, leading them to suggest that ecological succession was the primary driver of microbial community composition (Read *et al.* 2015). Furthermore, a 1300 km transect of the lower MSR indicated that richness actually increased instead of decreasing in particle-associated, and to a lesser extent, free-living, communities, suggesting that the MSR may have fundamentally different mechanisms driving microbial diversity than other rivers (Payne *et al.* 2017).

153 Complicating matters, particle-associated communities in rivers (frequently defined as
154 those found on filters of > ~3 µm remain distinct from their free-living counterparts (Crump *et*
155 *al.* 1999; Riemann and Winding 2001; Jackson *et al.* 2014), potentially due to increased
156 production rates from access to readily obtainable carbon (Crump and Baross 1996; Crump *et al.*
157 1998, 1999). Typical river microorganisms associated with particles include OTUs related to the
158 Bacteroidetes clades *Flavobacteria* and *Cytophaga*, Planctomycetes, *Rhizobium* spp., and
159 *Methylophilaceae* spp. (Crump and Baross 1996; Allgaier and Grossart 2006; D'Ambrosio *et al.*
160 2014; Jackson *et al.* 2014). However, denoting consistent trends in particle community
161 composition must be tempered by recent evidence suggesting substrate availability and chemical
162 queues may trigger organisms to switch between free-living and particle-associated lifestyles
163 (Grossart 2010; D'Ambrosio *et al.* 2014). Rivers constitute complex and highly dynamic
164 ecosystems from a metacommunity perspective.
165 Although important insights have been gained from recent research on portions of the
166 MSR (Staley *et al.* 2013, 2015, 2016; Payne *et al.* 2017), microbiological transects at the whole-
167 river or catchment scale have yet to be completed. This study aimed to *i)* compare the structure
168 and abundance of size-fractionated MSR microbial community populations to those in other
169 rivers, *ii)* examine within-river heterogeneity of microbial communities, and *iii)* identify MSR
170 microorganisms most strongly associated with eutrophication- all at a near whole-river scale, for
171 both prokaryotes and microscopic eukaryotes. During the fall of 2014, we completed the most
172 extensive microbiological survey of the Mississippi River to date with a continual rowed transect
173 over 70 days. Rowers from the adventure education non-profit OAR Northwest collected
174 samples from Minneapolis, MN to the Gulf of Mexico (2918 km) (Fig. 1A). Our findings expand
175 the current information available on microbial assemblages in major lotic ecosystems; further

delineate the relationships between microbial structure and stream order, nutrients, and volume; and identify MSR taxa predictive of the eutrophication nutrients nitrate and phosphate.

Materials and Methods

Sampling and Cell Counts

We used rowboats and a simple filtration protocol (Supplementary Information) to collect water from 39 sites along a continual-rowed transect of the MSR, starting in Lake Itasca and ending in the GOM, over 70 days from September 18th to November 26th, 2014. Sites were chosen to be near major cities and above and below large tributaries. After some samples were removed due to insufficient sequence data, contamination, or incomplete metadata (see below), the final usable set of samples included 38 sites starting at Minneapolis (Fig. 1A, Table S1, General Information). Most sampling occurred within the body of the river, although due to safety issues, three samples were collected from shore (Table S1, General Information). We collected duplicate samples at each site, but because separate rowboat teams frequently collected these sometimes several dozen meters apart, they cannot be considered true biological replicates and we have treated them as independent samples.

At each site, we filtered 120 mL of water sequentially through a 2.7 µm GF/D filter (Whatman GE, New Jersey, USA) housed in a 25 mm polycarbonate holder (TISCH, Ohio, USA) followed by a 0.2 µm Sterivex filter (EMD Millipore, Darmstadt, Germany) with a sterile 60 mL syringe (BD, New Jersey, USA). We refer to fractions collected on the 2.7 µm and 0.22 µm filters as > 2.7 µm and 0.2-2.7 µm, respectively. Flow-through water from the first 60 mL was collected in autoclaved acid-washed 60 mL polycarbonate bottles. Both filters were wrapped

in parafilm, and together with the filtrate, placed on ice in Yeti Roadie 20 coolers (Yeti, Austin, TX) until shipment to LSU. Further, 9 mL of whole water for cell counts was added to sterile 15 mL Falcon tubes containing 1 mL of formaldehyde and placed into the cooler. The final cooler containing samples from sites P-Al had substantial ice-melt. Though our filters were wrapped in parafilm, we processed melted cooler water alongside our other samples to control for potential contamination in these filters. Given that some of our samples were expected to contain low biomass, we also included duplicate process controls for kit contamination (Salter *et al.* 2014; Weiss *et al.* 2014) with unused sterile filters.

Flow-through 0.2 µm filtered water from each collection was analyzed for SiO_4, PO_4^{3-}, NH_4^+, NO_3^-, and NO_2^- (µg/L) at the University of Washington Marine Chemistry Laboratory (http://www.ocean.washington.edu/story/Marine+Chemistry+Laboratory). Aboard-rowboat measurements were taken for temperature and turbidity. We determined turbidity by deploying a secchi disk (Wildco, Yulee, FL), while drifting with the current so the line hung vertically. It was lowered until no longer visible, then raised until just visible, and measured for its distance below the waterline. We then calculated secchi depth from the average of two measurements. Temperature was measured with probes from US Water Systems (Indianapolis, IN), rinsed with distilled water between samples. Samples for cell counts were filtered through a 2.7 µm GF/D filter, stained with 1x Sybr Green (Lonza), and planktonic cells were enumerated using the Guava EasyCyte (Millipore) flow cytometer as previously described (Thrash *et al.* 2015).

Throughout the transect, seven cooler exchanges occurred to ensure samples were not exposed to extended durations at *in situ* temperatures. Cooler temperatures were monitored with HOBO loggers (Onset, Bourne, MA) to ensure that samples stayed at $\leq 4^{o}C$. On average, a sample spent 7.6 days (median = 6 days, range=1 day minimum, 16 days maximum) at $\leq 4^{o}C$

222 before delivery to LSU. Though samples were not frozen immediately, previous research has
223 provided evidence that soil and water samples stored at 4ºC, and for different durations (≥ 14
224 days), did not show significant alterations to most community relative abundances, structure, or
225 composition (Lauber *et al.* 2010; Rubin *et al.* 2013; Tatangelo *et al.* 2014).
226
227 *DNA extraction and Sequencing*
228 DNA was extracted from both filter fractions and controls using a MoBio PowerWater DNA kit
229 (MoBio Laboratories, Carlsbad, CA) following the manufacturer's protocol with one minor
230 modification: in a biosafety cabinet (The Baker Company, Stanford, ME), Sterivex filter
231 housings were cracked open using sterilized pliers and filters were then removed by cutting
232 along the edge of the plastic holder with a sterile razor blade before being placed into bead-
233 beating tubes. DNA was eluted with sterile MilliQ water, quantified using the Qubit2.0
234 Fluorometer (Life Technologies, Carlsbad, CA), and stored at -20° C. Bacterial and archaeal
235 sequences were amplified at the V4 region of the 16S rRNA gene using the 515f and 806r primer
236 set (Caporaso *et al.* 2012), and eukaryotic sequences from the V9 region of the 18S rRNA gene
237 using the 1391r and EukBR primer set (Amaral-Zettler *et al.* 2009). Amplicons were sequenced
238 on an Illumina MiSeq as paired-end 250 bp reads at Argonne National Laboratory. Sequencing
239 of the 16S and 18S rRNA gene amplicons resulted in 13,253,140 and 13,240,531 raw sequences,
240 respectively.
241
242 *Sequence Analysis*
243 We analyzed amplicon data with Mothur v.1.33.3 (Schloss *et al.* 2009) using the Silva v.119
244 database (Pruesse *et al.* 2007; Quast *et al.* 2013). Briefly, 16S and 18S rRNA gene sequences

were assembled into contigs and discarded if the contig had any ambiguous base pairs, possessed repeats greater than 8 bp, or were greater than 253 or 184 bp in length, respectively. Contigs were aligned using the Silva rRNA v.119 database, checked for chimeras using UCHIME (Edgar *et al.* 2011), and classified also using the Silva rRNA v.119 database. Contigs classified as chloroplast, eukaryotes, mitochondria, or "unknown;" or as chloroplast, bacteria, archaea, mitochondria, or "unknown;" were removed from 16S or 18S rRNA gene data, respectively. The remaining contigs were clustered using the cluster.split() command into operational taxonomic units (OTUs) using a 0.03 dissimilarity threshold ($OTU_{0.03}$). After these steps, 146,725 and 131,352 OTUs remained for the 16S and 18S rRNA gene communities, respectively.

Sample quality control

To evaluate the potential for contamination from extraction kits, cooler water in the last set of samples, or leaking/bursting of pre-filters, all samples were evaluated with hierarchical clustering and NMDS analysis. Hierarchical clustering was performed in R using the *hclust* function with methods set to "average", from the vegan package (Oksanen *et al.* 2015). Samples were removed from our analysis if they were observed to be outliers in both the NMDS and hierarchical clustering such that they grouped with our process controls. The process and cooler water controls were extreme outliers in both, as was sample L2 (Fig. S1, S2). Sterivex and prefilter samples generally showed strong separation with the exception of three 16S rRNA gene samples- STER X2, W2, S2 (Fig. S1, S2). The only other samples that were removed were due to missing chemical data (Lake Itasca1-2, A1-2) or failed sequencing (16S STER Af1; 16S PRE S2, X2; 18S PRE O1). Not including process or cooler water controls, 152 samples were sequenced each for prokaryotic and eukaryotic communities. After these QC measures, 144 and

149 samples remained in the analyses from the 16S and 18S rRNA gene amplicons, respectively. Further, to control for potential contaminants, any OTU with greater 20 reads in the process or cooler controls was removed from the data set. 146,725 and 131,327 OTUs remained after these steps for 16S and 18S rRNA gene communities, respectively.

Alpha and Beta Diversity

$OTU_{0.03}$ analyses were completed with the R statistical environment v.3.2.1(R Developement Core Team 2015). Using the package PhyloSeq (McMurdie and Holmes 2013), alpha-diversity was first calculated on the non-normalized, unfiltered OTUs using the "estimate richness" command within PhyloSeq, which calculates six alpha diversity metrics (McMurdie and Holmes 2013). After estimating alpha diversity, potentially erroneous rare OTUs, defined here as those without at least two sequences in 20% of the sites, were discarded. After this filter, the dataset contained 950 and 724 16S and 18S rRNA gene OTUs, respectively. For site-specific community comparisons, OTU counts were normalized using the package DESeq2 (Love *et al.* 2014) with a variance stabilizing transformation (Learman *et al.* 2016). Beta-diversity between samples was examined using Bray-Curtis distances via ordination with non-metric multidimensional scaling (NMDS). Analysis of similarity (ANOSIM) was used to test for significant differences between groups of samples (e.g lower versus upper MSR) using the *anosim* function in the vegan package (Oksanen *et al.* 2015). The influence of Hellinger transformed environmental parameters on beta-diversity was calculated in R with the *envfit* function. Relative abundances plots were made using the R package ggplot2. Linear and non-linear regressions were made using the command *stat_smooth()* with the flag "level=0.95" and method = "lm" or "loess" for linear and non-linear regressions, respectively.

291

292 *Network analyses and modeling*

293 To identify specific OTUs with strong relationships to environmental parameters (e.g. turbidity,
294 NO_3^-), we employed weighted gene co-expression network analysis (WGCNA) (Langfelder and
295 Horvath 2008) as previously described for OTU relative abundances (Guidi *et al.* 2016). First, a
296 similarity matrix of nodes (OTUs) was created based on pairwise Pearson correlations across
297 samples. This was transformed into an adjacency matrix, which represents the strength at which
298 two OTUs within the matrix are connected (adjacency). To do this, the similarity matrix was
299 raised to a soft threshold power (p; p = 6 for 16S and 18S > 2.7 µm, p = 4 for 16S 0.2-2.7 µm, p
300 = 9 for 18S 0.2-2.7 µm) that maximizes OTU connections with others and ensures scale-free
301 topology fit. Submodules of highly co-correlating OTUs were assigned a color (e.g. yellow,
302 green, turquoise) as defined with a topological overlap measure and hierarchical clustering,
303 which identified submodules based on an OTU's weighted correlations to other OTUs within and
304 outside the network. Each submodule, represented by an eigenvalue, was pairwise Pearson
305 correlated to individual Hellinger-transformed environmental measurements (Figs. S7-10A). To
306 explore the relationship of submodule structure to these parameters, OTUs within the submodule
307 were plotted using their individual correlation to the parameter of interest (y-axis, here nitrate or
308 phosphate) and their submodule membership (x-axis), defined as the strength of an OTU
309 (number of connections) to other OTUs within the submodule (Figs. S7-10B, D). Strong
310 correlations between submodule structure and an environmental parameter facilitate
311 identification of OTUs that are highly correlated to that parameter. To evaluate the predictive
312 relationship between a submodule and a parameter, we employed partial least square regression
313 (PLS) analysis. PLS maximizes the covariance between two parameters (e.g., OTU abundances

and nitrate concentration) to define the degree to which the measured value (OTU abundance) can predict the response variable (nutrient concentration). The PLS model was permutated 1000 times and Pearson correlations were calculated between the response variable and leave-one-out cross-validation (LOOCV) predicted values. Modeled values were then compared with measured values to determine the explanatory power of the relationships (Figs. S7-10C, E). Relative contributions of individual OTUs to the PLS regression were calculated using value of importance in the projection (VIP) (Chong and Jun 2005) determination. PLS was run using the R package *pls* (Mevik and Wehrens 2007), while VIP was run using additional code found here: http://mevik.net/work/software/VIP.R. Though other environmental parameters had strong Pearson correlations to submodules, only submodules with strong, positive correlations to measures related to eutrophication were selected, in accordance with our study objectives, as these were most likely to provide candidate taxa with remediation potential.

Accession numbers

Community 16S and 18S rRNA gene sequence fastq files are available at the NCBI Sequence Read Archive under the accession numbers: SRR3485674 - SRR3485971 and SRR3488881 - SRR3489315, respectively.

Code Availability

All code used for Mothur, SeqENV, PhyloSeq, WGCNA, and PLS regression analyses can be found on the Thrash lab website (http://thethrashlab.com/publications) with the reference to this manuscript linked to “Supplemental Information”.

Results

Using rowboats and a simple syringe-based filtration protocol, we measured 12 biological, chemical, and physical parameters (e.g. 16S and 18S rRNA gene communities, NH_4^+, river speed, etc.) from 38 sites along a 2918 km transect of the MSR (Fig. 1A). River order increases dramatically at the Missouri confluence [eighth to tenth Strahler order (Pierson *et al.* 2008)], which corresponded to overall discharge (Fig. 1A) and beta diversity changes discussed below. Therefore, we refer to this juncture as the separator between the upper (0 km – 1042 km, Sites A-S) and lower MSR (1075-2914 km, Sites T-Al). Within the upper MSR, NO_3^-, PO_4^{3-}, and NO_2^- were variable but generally increased downriver until peak concentrations near the confluences of the Illinois and Missouri Rivers. This gave way to lower and more consistent concentrations along the lower MSR (Fig. 1B). Turbidity (inversely related to secchi disk visibility) increased steadily downriver to a maximum near the Illinois and Missouri River confluences (1042 km, Site S) (Fig. 1B, Secchi Disk), then trended downwards for the rest of the transect. Planktonic (< 2.7 µm) cell counts varied between 1 and $3x10^6$ cells$\cdot mL^{-1}$ in the upper MSR, and decreased to high 10^5 cells$\cdot mL^{-1}$ in the lower MSR (Fig. 1B, Planktonic Cell Counts). Water temperature ranged from 19°C (133km, Site E) to 11.7°C (2552 km, Site Ag), and river speed, excluding three sites sampled from shore, was between 5.5 mph at Site Y and 0.4 mph (597 km, Site L) (Table S1, General Information). Spearman rank correlations of the measured environmental parameters showed strong positive correlations between nitrate, phosphate, distance, and increased turbidity; while nitrate and phosphate both strongly correlated negatively to water temperature and river speed (Table S1, Spearman Rank).

Bacterial and archaeal communities

We observed a clear distinction between the 0.2-2.7 µm and > 2.7 µm 16S rRNA gene communities (ANOSIM R = 0.65, P = 0.001) (Fig. S1A). Both size fractions also showed significant community separation between sites above and below the Missouri River confluence (ANOSIM, > 2.7 µm: R = 0.44, P=0.001; 0.2-2.7 µm: R = 0.48, P = 0.001) (Fig. 2A, B), that the Eukaryotic fractions mirrored (below)(Fig. 2C, D). Overall richness was significantly higher in the lower vs. upper rivers (Table S1, Richness comparisons). Taken separately, most measurements indicated that richness increased with river distance in the upper and lower rivers, and this trend was common for both the > 2.7 µm and 0.2-2.7 µm fractions (Fig. S3A, B). Evenness measurements had less consistent trends.

Phosphate and turbidity had the highest correlation with the separation between the upper and lower > 2.7 µm communities (R^2 = 0.58, R^2 = 0.54, respectively), with water temperature (R^2 = 0.52) also contributing (Table 1). At an OTU level, taxa related to the acI clade (Actinobacteria) and unclassified *Bacillaceae* (R^2 > 0.77, P = 0.001) contributed most to the separation between the upper and lower > 2.7 communities, with OTUs related to the *Bacillales, Gemmatimonadaceae, Peptococcaceae*, and *Micromonosporaceae* clades also a factor (R^2 > 0.70, P = 0.001) (Table 2). For the 0.2-2.7 µm fraction, nitrate were the strongest correlating environmental factors with the distinction between upper and lower communities (R^2 = 0.443), although phosphate, turbidity, and water temperature (R^2 > 0.385 for each) also contributed (Table 1). OTUs related to *Flavobacterium* and an unclassified Bacterium (closest NCBI BLAST hit *Acidovorax* sp., KM047473), most strongly contributed to the separation between the 0.2-2.7 µm communities (R^2 > 0.503, P = 0.0001) (Table 2). Other important OTUs driving the separation between upper and lower communities belonged to the clades Bacteroidetes, *Microbacteriaceae*, *Clostridiales*, and *Holophagaceae* (R^2 > 0.49, P = 0.0001) (Table 2).

383 At the phylum level, Proteobacteria, Actinobacteria, and Bacteroidetes dominated
384 bacterial communities in both fractions (Figs. 3A and B) (Table S1, 16S OTU table norm.).
385 Proteobacteria abundance in the > 2.7 µm fraction fluctuated over the course of the transect (Fig.
386 3A), whereas their 0.2-2.7 µm counterparts generally increased in relative abundance downriver
387 (Fig. 3B). 0.2-2.7 µm Bacteroidetes and Actinobacteria generally decreased in the upper river
388 and stabilized in the lower river. These phyla showed considerable abundance variation in the >
389 2.7 µm fraction. Cyanobacteria in the > 2.7 µm fraction negatively correlated with increased
390 turbidity (Spearman rank = 0.67), consistent with lower irradiance (Fig. 3A). Both > 2.7 µm and
391 0.2-2.7 µm Acidobacteria increased in abundance downriver and positively correlated with river
392 distance (Wilcoxon single ranked test, P = < 0.01) (Fig. S4A, B). Within the 0.2-2.7 µm fraction,
393 the five most abundant OTUs were classified as a LD12 (OTU11; Proteobacteria), two acI clade
394 OTUs (OTU4, OTU7; Actionbacteria), a *Limnohabitans* sp. (OTU2; Proteobacteria), and a LD28
395 (OTU8; Proteobacteria) (Table S1, 16S OTU table norm.). Comparatively, an unclassified
396 *Methylophilaceae* (OTU1; Proteobacteria), a *Planktothrix* sp. (OTU12; Cyanobacteria), a NS11-
397 12 marine group (OTU21; Bacteroidetes), an Candidatus Aquirestis sp. (OTU17; Bacteroidetes),
398 and an unclassified *Sphingobacteriales* (OTU25; Bacteroidetes) were the most abundant OTUs
399 in the > 2.7 µm fraction (Table S1, 16S OTU table norm.). Archaea occurred at much lower
400 relative abundances: we found only eight OTUs belonging to the phyla Euryarchaeota and
401 Thaumarchaeota, collectively. These OTUs were classified as *Methanobacterium* sp. (OTU714,
402 1968) and soil Crenarchaeotic Group (OTU370, 389, 983, and 1253), *Candidatus*
403 Nitrosoarchaeum sp. (OTU1093) from the phylum Euryarcheota, and an unclassified
404 Thaumarchaeota (OTU2951) from the phylum Thaumarchaeota (Table S1, 16S OTU table
405 norm.). Overall, Thaumarchaeota increased in abundance along the transect more in the > 2.7 µm

406 fraction compared to the 0.2-2.7 µm fraction (Fig. S4A, B). In both fractions, we only detected
407 Euryarcheota at very low abundances. Importantly, the primers used in this study may miss some
408 archaeal taxa (Parada *et al.* 2015).
409
410 *Microbial eukaryotic communities*
411 Eukaryotic communities, observed via the 18S rRNA gene, also showed a significant separation
412 between > 2.7 µm and 0.2-2.7 µm fractions (ANOSIM R = 0.689, P = 0.001) (Fig. S1B). As
413 expected due to generally larger cell sizes in microbial eukaryotes compared to prokaryotes,
414 species richness remained higher in the > 2.7 µm vs. 0.2-2.7 µm fractions (Fig. S3C, D). Both
415 the > 2.7 µm and 0.2-2.7 fractions also showed a significant separation between the upper and
416 lower MSR communities (ANOSIM, > 2.7 µm: R = 0.696, P = 0.001; 0.2-2.7 µm: R = 0.576, P =
417 0.001) (Fig. 2C, D). Overall richness in the > 2.7 µm fraction was also higher in the lower vs.
418 upper river, but this was not true for the 0.2-2.7 µm fraction (Table S1, Richness comparisons).
419 Richness in the > 2.7 µm fraction increased along both the upper and lower river, similarly to
420 prokaryotic communities, but remained relatively stable within the 0.2-2.7 µm fraction.
421 Phosphate was the top environmental factor correlating to the distinction between
422 eukaryotic communities in both filter fractions (> 2.7 µm, R^2 = 0.49; 0.2-2.7µm R^2 = 0.56)
423 (Table 1). No other factors had correlations > 0.38 (Table 1). At the OTU level, taxa related to an
424 unclassified Ochrophyta (OTU63) and an unclassified Eukaryote (OTU1) separated the 0.2-2.7
425 µm communities (R^2 > 0.645, P = 0.0001), while the same unclassified Eukaryote OTU (OTU1)
426 and a second unclassified Eukaryote (OTU222) contributed most to separating the > 2.7 µm
427 communities (R^2 > 0.80, P = 0.0001) (Table 2).

Stramenopiles (or Heterokonts), encompassing diatoms and many other forms of algae, and OTUs that could not be classified at the phylum level dominated both the > 2.7 µm and 0.2-2.7 µm communities (Fig. 3C, D). Stramenopiles accounted for over 25% of both communities, with higher abundances in the upper vs. lower river. We observed a similar trend of disparate abundances between the upper and lower river for > 2.7 µm Cryptomonadales and 0.2-2.7 µm Nucletmycea, the latter of which include fungi (Fig. 3C; Table S1, 18S OTU table norm.). Within the 0.2-2.7 µm fraction, we identified the five most abundant OTUs as three unclassified *Bacillariophytina* (OTU7, OTU14, OTU9), a *Pythium* sp. (OTU170), and an unclassified *Cryptomonas* (OTU11) (Table S1, 18S OTU table norm.). Comparatively, two unclassified *Eukaryotes* (OTU2 and OTU1), an unclassified *Stramenopiles* (OTU3), an unclassified *Perkinsidae* (OTU13), and an unclassified *Chrysophyceae* (OTU6) had the highest abundance in the > 2.7 µm fraction (Table S1, 18S OTU table norm).

The Mississippi River Core Microbiome

We defined the core microbiome as those OTUs detectable after normalization in greater than 90% of the sites. The 16S rRNA gene > 2.7 µm and 0.2-2.7 µm core microbiomes consisted of 82 and 98 OTUs, respectively, classified into eight different phyla- Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Verrucomicrobia, Chloroflexi, Chlorobi, Gemmatimonadetes- and composed of taxa such as freshwater SAR11 (LD12, Proteobacteria), *Limnohabitans* sp. (Proteobacteria), *Polynucleobacter* sp. (Proteobacteria), acI clade (Actinobacteria), LD28 clade (Proteobacteria), and *Planktothrix* sp. (Cyanobacteria) (Table S1, 16S 0.2-2.7 µm Core Comm. and 16S > 2.7 µm Core Comm.). Core microbome relative abundance in both fractions decreased along the upper river but stabilized in the lower river (Fig.

451 4A). We confirmed this effect by analyzing the upper and lower core microbiomes separately.
452 Although the total OTU numbers changed (81 and 116 OTUs in the upper MSR and 160 and 144
453 OTUs in the lower MSR for the > 2.7 µm and 0.2-2.7 µm fractions, respectively), the trends
454 remained the same (Fig. S5).

455 Eighty OTUs comprised the > 2.7 µm 18S rRNA gene core microbiome (Fig. 4B). We
456 classified these as *Alveolata*, *Cryptophceae*, *Nucletmycea*, *Stramenopiles*, or unclassified
457 Eurkaryota (Table S1, 18S > 2.7 µm Core Comm.). Again, consistent with larger organism sizes,
458 and thus fewer OTUs overall, the 0.2-2.7 µm Eukaryotic core microbiome comprised only 21
459 OTUs (Table S1, 18S 0.2-2.7 µm Core Comm.). These OTUs consisted of *Alveolata*,
460 *Nucletmycea*, *Stramenopiles*, or unclassified Eurkaryota (Table S1, 18S 0.2-2.7 µm Core
461 Comm.). While the 0.2-2.7 µm core microbiome remained relatively stable along the river, the >
462 2.7 µm core decreased along the upper MSR before stabilizing in the lower river, similarly to
463 that of the prokaryotes (Fig. 4B).
464
465 *Network analyses identify taxa associated with and predictive of eutrophication*
466 We applied Weighted Gene Correlation Network Analysis (WGCNA) to identify co-occurring
467 groups of OTUs (submodules) that also had significant associations with the eutrophication
468 nutrients phosphate and nitrate. Of the submodules identified through WGCNA as being most
469 strongly correlated to phosphate and nitrate, we restrict our discussion to those modeled via PLS
470 analysis to predict > 50% of the measured nutrient concentration. Our additional PLS analyses
471 for those submodules predicting < 50% of nitrate and/or phosphate concentrations are included
472 in Figure S6 and Figure S7-S10.

Three submodules in the prokaryotic and eukaryotic fractions were strongly associated with phosphate. The 0.2-2.7 µm prokaryotic submodule most associated with phosphate was composed of 51 OTUs (Table 3), and had moderate correlation between the submodule structure and phosphate (Fig. S7D). Strong correlations between submodule structure and the measured nutrient suggests that individual submodule OTUs that also have strong correlations to the nutrient are the most important organisms associated with that nutrient (Langfelder and Horvath 2008). PLS modeling determined that this prokaryotic submodule predicted 80% of measured phosphate concentrations (Figure S7E, Table 3). Variable importance in the projection (VIP) analysis found OTUs corresponding to an unclassified *Holophagaceae*, an unclassified *Gemmatimonadaceae*, and an unclassified *Burkholderiaceae* were the three most important in the PLS model for phosphate (Fig. 5A, Table S1, 16S 0.2-2.7 µm PO4). OTU322 (Acidobacteria subgroup 6), had moderate correlation to phosphate but high node centrality (n = 51; Fig. 5A, Table S1, 16S 0.2-2.7 µm PO4), corroborating evidence that freshwater sediment Acidobacteria subgroup 6 often occur in co-culture with Alphaproteobacteria and may be metabolically connected (Kielak *et al.* 2016 and refs. within). A submodule from the 0.2-2.7 µm eukaryotic fraction was also highly predictive of phosphate (Table 3, Fig. S8D). We identified the top four VIP taxa as an unclassified *Peronosporomycetes*, an unclassified *Ochrophyta*, an unclassified Eukaryote, and an unclassified *Stramenopiles* (Fig. 6A, Table S1, 18S 0.2-2.7 µm PO4). All four OTUs had Pearson correlations with phosphate greater than 0.60, among which two were negative (Fig. 6A, Table S1, 18S 0.2-2.7 µm PO4). Unfortunately, the most highly interconnected OTU within the submodule remained unclassified even at the Phylum level (Fig. 6A, Table S1, 18S 0.2-2.7 µm PO4). A > 2.7 µm eukaryotic submodule could predict 62% of

495 measured phosphate variance (Table 3). An unclassified Eukaryote and an unclassified
496 *Peronosporomycetes* occupied top two VIP positions (Fig. 6B, Table S1, 16S > 2.7 µm PO4).
497 Only two submodules had strong associations with nitrate. The > 2.7 µm prokaryotic
498 submodule most strongly correlated with nitrate contained 133 OTUs (Table 3, Fig. S9A).
499 However, despite a low structural correlation to nitrate, the submodule strongly predicted
500 measured nitrate in the PLS model (Figure S9B, Table 3). The four highest VIP scoring OTUs
501 were an *Anabaena* sp., a *Flavobacterium* sp., an unclassified bacterium, and a member of the
502 *Sphingobacteriales* NS11-12 marine group (Fig. 5B, Table S1, 16S > 2.7 µm NO3). OTUs with
503 the highest node centrality (> 20) belonged to the *Sphingomonadales* (*Alphaproteobacteria*) and
504 *Sphingobacteriales* (Bacteroidetes), and each positively correlated with nitrate (R > 0.48, Fig.
505 5B, Table S1, 16S > 2.7 µm NO3). The > 2.7 µm eukaryotic submodule most associated with
506 nitrate predicted 57% of observed variation in nitrate concentrations (Table 3). OTUs with the
507 top VIP scores were two unclassified *Chrysophyceae*, an unclassified *Ochrophyta*, and an
508 unclassified Diatom (Fig. 6C, S10; Table S1, 18S > 2.7 µm NO3).
509

510 **Discussion**

511 Understanding microbial communities on a whole-river scale is essential for linking the complex
512 relationships between microorganisms, metabolism, and water quality. At 2914 km, this study is
513 the largest transect of the MSR to date, and the first to include both prokaryotic and eukaryotic
514 data by size fractions. This is also the largest river yet surveyed at this level of geographic
515 completion, which may account for some of its distinctive microbiological features. Although the
516 MSR comprised similar aquatic microbial taxa found within other rivers (Zwart *et al.* 2002;
517 Cottrell *et al.* 2005; Ghai *et al.* 2011; Fortunato *et al.* 2013; Jackson *et al.* 2014; Read *et al.* 2015;

518 Savio *et al.* 2015; Meziti *et al.* 2016), it hosted unique relative abundance and core microbiome
519 trends, indicating that rivers have distinguishing microbiome characteristics. Furthermore, we
520 observed two different community regimes separated at the Missouri River confluence and
521 mirrored by the measured physio-chemical properties of the MSR (Fig. 1B). This is an important
522 hydrological feature of the MSR that strongly influences its microbiology, and further
523 distinguishes it from other rivers. Our study also identified taxa associated with, and predictive
524 of, the eutrophication nutrients nitrate and phosphate that provide important targets for future
525 study and may assist in detecting and quantifying imminent changes in river water quality.
526 MSR microbial communities separated into two distinct fractions of microbial
527 assemblages, 0.2-2.7 µm and > 2.7 µm (ANOSIM, R = 0.65, P = 0.001), throughout the river
528 transect, a common feature of aquatic ecosystems (DeLong *et al.* 1993; Crump *et al.* 1999;
529 Allgaier and Grossart 2006; D'Ambrosio *et al.* 2014; Jackson *et al.* 2014). However, despite
530 sharing similar taxa with other rivers (Ghai *et al.* 2011; Fortunato *et al.* 2012; Crump *et al.* 2012;
531 Read *et al.* 2015; Savio *et al.* 2015; Meziti *et al.* 2016; Niño-García *et al.* 2016), MSR microbial
532 community abundances were distinct at a phylum level from previous work on Thames River
533 (Read *et al.* 2015), Danube River (Savio *et al.* 2015), Canadian Boral Rivers (Niño-García *et al.*
534 2016), Toolik catchment (Crump *et al.* 2012), and Yenisei River (Kolmakova *et al.* 2014).
535 Specifically, within the 0.2-2.7µm MSR prokaryotic fraction, Proteobacteria remained the most
536 abundant phylum throughout the river, while Bacteroidetes and Actinobacteria decreased
537 downriver (Fig. 3). In contrast, other river studies found headwaters dominated by Bacteroidetes
538 that gave way to more abundant Actinobacteria and Proteobacteria further downriver (Read *et al.*
539 2015; Savio *et al.* 2015). Similarly, the trends observed in the upper portion of the MSR (Staley
540 *et al.* 2013), where Bacteroidetes and Actinobacteria decreased while Proteobacteria increased,

diverged from the trends found in this study (Fig. 3). Further, while no clear increasing or decreasing trends were observed in Payne *et al.* (2017), Actinobacteria occurred with the highest relative abundances in the planktonic fraction, whereas in this study, Proteobacteria were the most abundant phylum. This may be attributable to the difference in sampling year or possibly methodology, such as an amplification step that was not undertaken in this work (Payne *et al.* 2017). Furthermore, although many of the dominant taxa found in other rivers also occupied the MSR (LD12 (freshwater SAR11, *Alphaproteobacteria*), *Limnohabitans* sp. (*Betaproteobacteria*), LD28 (*Betaproteobacteria*), acI clade (Actinobacteria), and *Algoriphagu*s sp. (Bacteriodetes) (Ghai *et al.* 2011; Fortunato *et al.* 2013; Kolmakova *et al.* 2014; Read *et al.* 2015; Savio *et al.* 2015; Meziti *et al.* 2016), we observed OTU-level abundance differences at a whole-river scale. For example, OTUs belonging to the acI and LD12 freshwater clades continually increased in abundance towards the river mouth in the Thames and Danube Rivers (Read *et al.* 2015; Savio *et al.* 2015), whereas MSR OTUs classified as acI (OTUs 3, 4 and 7) and LD12 (OTU 11) did not show any distinct change in abundance throughout the transect. Specific to studies of portions of the MSR,

The generally greater richness (Table S1, Richness comparisons) and decreased core microbial community abundances (Fig. 4) in the lower vs. upper MSR contrasted predictions by the RCC (Vannote *et al.*, 1980) and observations in previous studies from both large and small rivers (Read *et al.* 2015; Savio *et al.* 2015, Niño-García *et al.* 2016) where these trends were reversed. Importantly, we did not sample the true headwaters of the MSR (Lake Itasca to above St. Cloud), and at the point of first sampling, the MSR already constituted an eighth order river. Therefore, some of the trends predicted by the RCC may have been missed, although increasing richness was also observed in the lower MSR during a different year (Payne *et al.* 2017).

564 Ultimately, the variant observations between this MSR survey and other river studies may result
565 from different sampling methodologies or the particular timing of sampling. Previous studies
566 conducted on six artic rivers (Crump *et al.* 2009), two smaller temperate rivers (Parker and
567 Ipswich Rivers) (Crump and Hobbie 2005), and the upper portion of the MSR (Staley *et al.*
568 2015) showed that seasonality was important in structuring bacterial communities over three year
569 and two year sampling periods.

570 However, our observed differences may also stem from biological signal related to
571 unique environmental conditions, human impacts (Meziti *et al.* 2016; Payne *et al.* 2017), and
572 changes in hydrology and the level of river engineering (Ruiz-González *et al.* 2013, 2015;
573 Freimann *et al.* 2015; Niño-García *et al.* 2016) with distance. For instance, the most striking
574 difference between the MSR and other rivers was the distinct separation of prokaryotic and
575 eukaryotic community regimes at the Missouri River confluence. This separation matched
576 physio-chemical changes observed for the MSR such as increased Strahler river order (8 to 10)
577 and more consistent nutrient concentrations in the lower river compared to the upper (Fig. 1B)
578 (Pierson *et al.* 2008). These results are bolstered by another observation depicting a major
579 separation of communities above and below the Missouri confluence, observed during a different
580 year (2012) and roughly three months prior (late summer) to the current study (Payne *et al.*
581 2017). This similarity to our results suggests the major influence of the Missouri River
582 confluence is not seasonally dependent.

583 Although the overall microbial community patterns in the MSR differed from other
584 rivers, we hypothesize that similar ecological processes drove the separation between the upper
585 and lower MSR communities, namely changes in the importance of immigration, emigration, and
586 resource gradient dynamics (Crump *et al.* 2012; Savio *et al.* 2015; Niño-García *et al.* 2016).

587 Specifically, our data suggest that mass effects play a role in structuring microbial communities
588 in the upper MSR, although instead of only in the headwaters, this process continues for almost a
589 third of the length of the river, possibly due to contributions from large tributaries and abundant
590 dams. This contrasts with the conclusion made by Staley *et al.* (2015) for a smaller transect in
591 the upper MSR, where they suggested species sorting was the primary influence structuring
592 upper MSR microbial communities. Increased turbidity correlated with decreased core
593 microbiome relative abundance (Spearman rank correlation, > 2.7 µm R = 0.53; 0.2-2.7 µm R =
594 0.63) in the upper MSR, and nutrients like phosphate and nitrate continually increased along the
595 upper river in the current study (Fig. 1B). Similarly, richness also increased along the upper river
596 (except in the 0.2-2.7 µm 18S communities) (Fig. S3). These patterns are consistent with
597 communities under the influence of strong mass effects in the upper MSR, whereby inputs from
598 allochthonous sources continually contributed nutrients, particulate matter, and additional
599 microbial taxa.

600 Conversely, nitrate, phosphate, turbidity, and core microbiome relative abundance all
601 stabilized in the lower MSR (Figs. 1B, 4), suggesting a change in the relative influence of
602 external sources for these variables. We speculate that once the MSR grew to a tenth order river,
603 its large volume and size buffered it from allochthonous inputs. Nevertheless, overall richness
604 was greater in the lower river compared to the upper (except in the 0.2-2.7 µm 18S communities)
605 (Table S1, Richness comparisons), and, like the upper river, richness generally increased in the
606 lower river (Fig. S3), matching a previous observation (Payne *et al.* 2017). However, since the
607 relative abundance of core microbiome taxa remained comparatively stable in the lower river
608 (Fig. 4), increased richness likely occurred primarily from the emergence of low abundance
609 OTUs. The extent to which a local or regional event impacts downriver populations is dependent

610 on the success of allochthonous taxa associated with such an event to become established within
611 the autochthonous population (Crump *et al.* 2012; Niño-García *et al.* 2016). The larger the river,
612 the less likely any newly immigrated taxa will establish dominance. Thus, a plausible scenario
613 for the regime change observed at the Missouri River confluence is that the influence of mass
614 effects was drastically reduced by the increased size of the lower river, and that allochtonous
615 taxa immigrating there were relegated to lower abundances.

616 Alternatively, some of the increase in richness may have occurred through native
617 community differentiation within the lower river. An overall community shift from a mixture of
618 allochthonous members to a "native" population requires growth rates greater than residence
619 time over a given distance (Crump *et al.* 2004). Though river speed increased along the MSR,
620 the effective residence time also increased since taxa no longer experienced rapidly changing
621 environmental variables (Fig. 1B). As a result, the lower river may have provided opportunities
622 for microbial community differentiation based on microniches within the river, which is possible
623 considering average prokaryotic growth rates (Savio *et al.* 2015), especially among particle-
624 associated (> 2.7 µm) taxa (Crump *et al.* 1998). Thus, although the aggregate patterns in
625 particular phyla and the overall taxonomic richness within the MSR differ from other systems,
626 similar ecological processes may still drive these patterns, but the relative proportion of the river
627 whereby mass effects vs. species sorting dominates fosters unique community assemblages
628 (Niño-García *et al.* 2016). To resolve the relative importance of mass effects vs. species sorting
629 in the upper and lower rivers, future work will need to incorporate explicit measurements of
630 immigrating and emigrating taxa, as well as growth rates for different organisms at different
631 geographic positions.

632 Within this dynamic river system, we also sought to define the microorganisms most
633 associated with eutrophication nutrients to provide specific taxa for future study of nitrate and
634 phosphate uptake and/or metabolism, and also as plausible biological indicators of river trophic
635 state. By focusing on OTUs with the highest VIP scores in the PLS models, we identified several
636 bacterial and eukaryotic taxa that fit these criteria (Figs. 5, 6). Organisms in the
637 *Sphingomonadaceae* (e.g., *Novosphingobium* spp.) contributed strongly to the PLS models
638 predicting nitrate with both > 2.7 µm (Fig. 5B) and 0.2-2.7 µm (Fig. S6A) size fractions (Table
639 S1, 16S 0.2-2.7 µm VIP NO3 and 16S > 2.7 µm VIP NO3). *Novosphingobium* spp. isolates have
640 previously been associated with eutrophic environments (Trusova and Gladyshev 2002; Zwart *et*
641 *al.* 2002; Addison *et al.* 2007; Li *et al.* 2012) and some can reduce nitrate (Addison *et al.* 2007;
642 Li *et al.* 2012), making these specific OTUs candidates for nitrate metabolism in the river water
643 column. An *Anabaena* sp. (OTU40) from the core microbiome (Table S1, 16S > 2.7 µm Core
644 Comm.) had the top VIP score within the > 2.7 µm submodule that could predict 69% of nitrate
645 concentrations, and correlated negatively with nitrate (Table 3, Fig. 5B). The nitrogen-fixing
646 *Anabaena* spp. (Allen and Arnon 1955) typically bloom in low dissolved inorganic nitrogen
647 (DIN) conditions (Wood *et al.* 2010), making the absence of these consistent with high DIN. An
648 unclassified *Holophagaceae* OTU (OTU33) had the highest VIP score within the prokaryotic
649 0.2-2.7 µm submodule that predicted 80% of the variance in phosphate concentrations (Table 3,
650 Fig. 5A). This same OTU was also a key driver of the beta diversity separation between upper
651 and lower river communities (Table 2), and a member of the core microbiome (Table S1, 16S
652 0.2-2.7 µm Core Comm.). The *Holophagaceae* belong to the Acidobacteria phylum, and the
653 Ohio River contained a much higher abundance of Acidobacteria relative to other tributaries in a
654 previous study (Jackson *et al.* 2014). Notably, Acidobacteria increased with river distance in our

study as well (Fig. S4), with a peak in the 0.2-2.7 µm fraction near the Arkansas River confluence. This increase in "free-living" Acidobacteria downriver is distinct from other whole river studies, making these organisms, and the *Holophagaceae* OTUs in particular, potentially important organisms for the MSR river basin specifically.

Within Eukaryotes, multiple different algae, diatom, and Oomycetes OTUs occupied submodules highly predictive of nitrate and phosphate (Fig. 6A-C), and specifically *Chrysophyceae* taxa from both size fractions correlated strongly with nitrate (Fig. 6C, Fig. S6E). *Chrysophyceae* (golden algae) commonly occupy river systems (Necchi Jr 2016) including the MSR (Korajkic *et al.* 2015), can be autotrophic and mixotrophic (Jansson *et al.* 1996), and may serve as predators of prokaryotes (Caron *et al.* 1990). Further, multiple OTUs classified as *Peronosporomycetes* were important in predicting phosphate in both size fractions (Fig 6A, B, Table S1, 18S 0.2-2.7 µm VIP PO4 and 18S > 2.7 µm VIP PO4). *Peronosporomycetes* are fungus-like eukaryotic organisms known to be pathogenic in fish, plants, and mammals (Dick 2003; Islam and von Tiedemann 2011). While we also identified many other eukaryotic OTUs as important predictors of nutrients, poor taxonomic resolution hindered our ability to discuss them further. Improved cultivation and systematics of key microbial eukaryotes will be vital to understanding river nutrient dynamics.

While the most geographically comprehensive analysis to date for the MSR, this study only encompasses a snapshot in time. As noted previously, seasonal changes that have been observed in other rivers (Crump and Hobbie 2005; Crump *et al.* 2009; Smith *et al.* 2010; Fortunato *et al.* 2013; Staley *et al.* 2015) undoubtedly influence this dynamic system, though the permanence of the regime change at the Missouri River confluence remains in question. Future studies should incorporate microbial responses, at a whole-river scale, to seasonal pulse events

(e.g. rain, snow melt), and how river size and volume may buffer local microbial communities from allochthonous inputs (Zeglin 2015). Our current research highlights the distinctiveness of MSR microbial communities and the complexities influencing their structure within the MSR ecosystem. The observed association between changes in Strahler's river order, nutrient dynamics, and community composition indicates the importance of hydrology on the spatial dynamics structuring microbial communities (Freimann *et al.* 2015; Zeglin 2015; Niño-García *et al.* 2016) and provides baseline information for future MSR studies that incorporate greater temporal and spatial resolution. With river water quality of growing local and global importance (Vorosmarty *et al.* 2010; Russell and Weller 2013), the candidate taxa predictive of eutrophic nutrients determined herein also provide important targets for further research into their roles for indicating, and potentially improving, river health.

Acknowledgements

This work was supported by the Louisiana State University Department of Biological Sciences, College of Science, and Office of Research and Economic Development; and the College of the Environment at the University of Washington. Portions of this research were conducted with high performance computing resources provided by Louisiana State University (http://www.hpc.lsu.edu). The authors thank Dr. Caroline Fortunato and Dr. Gary King for thoughtful comments on the manuscript. The authors also thank the countless volunteers, schools, and organizations that facilitated the research. We specifically thank Pete Weess, Jessica Zimmerman, Katy Welch, Brian Moffitt, and David Cheney for helping organize the shipment of coolers between sites, and the OAR Northwest sponsors- Seattle Yacht Club Foundation, Yeti Coolers, and the National Mississippi River Museum and Aquarium. A full list of sponsors and

volunteers can be found on the OAR Northwest website (oarnorthwest.org). We would also like to thank Dr. Matthew Sullivan and Dr. Simon Roux for their help with coding the WGCNA and sPLS analyses. Lastly, we would like to thank Mrs. Ginger Thrash, who connected the Thrash lab to OAR Northwest.

Conflict of Interest

The authors declare no competing financial interests.

References

Addison, S. L., S. M. Foote, N. M. Reid, and G. Lloyd-Jones. 2007. Novosphingobium nitrogenifigens sp. nov., a polyhydroxyalkanoate-accumulating diazotroph isolated from a New Zealand pulp and paper wastewater. Int. J. Syst. Evol. Microbiol. **57**: 2467–2471.

Alexander, R. B., R. A. Smith, G. E. Schwarz, E. W. Boyer, J. V. Nolan, and J. W. Brakebill. 2008. Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin. Environ. Sci. Technol. **42**: 822–830.

Allen, M. B., and D. I. Arnon. 1955. Studies on nitrogen-fixing blue-green algae. I. Growth and nitrogen fixation by Anabaena cylindrica Lemm. Plant Physiol. **30**: 366.

Allgaier, M., and H.P. Grossart. 2006. Seasonal dynamics and phylogenetic diversity of free-living and particle-associated bacterial communities in four lakes in northeastern Germany. Aquat. Microb. Ecol. **45**: 115–128.

Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS One **4**: e6372.

Aulenbach, B. T., H. T. Buxton, W. A. Battaglin, and R. H. Coupe. 2007. Streamflow and nutrient fluxes of the Mississippi-Atchafalaya River Basin and subbasins for the period of record through 2005.

Battin, T. J., L. a. Kaplan, S. Findlay, C. S. Hopkinson, E. Marti, A. I. Packman, J. D. Newbold, and F. Sabater. 2009. Biophysical controls on organic carbon fluxes in fluvial networks. Nat. Geosci. **2**: 595–595.

Besemer, K., G. Singer, C. Quince, E. Bertuzzo, W. Sloan, and T. J. Battin. 2013. Headwaters are critical reservoirs of microbial diversity for fluvial networks. Proc. Biol. Sci. **280**: 20131760.

Bianchi, T. S., S. F. DiMarco, J. H. Cowan, R. D. Hetland, P. Chapman, J. W. Day, and M. A. Allison. 2010. The science of hypoxia in the northern Gulf of Mexico: A review. Sci. Total Environ. **408**: 1471–1484.

Blanchet, M., O. Pringault, C. Panagiotopoulos, D. Lefèvre, B. Charrière, J.F. Ghiglione, C. Fernandez, F. L. Aparicio, C. Marrasé, P. Catala, L. Oriol, J. Caparros, and F. Joux. 2016. When riverine dissolved organic matter (DOM) meets labile DOM in coastal waters: changes in bacterial community activity and composition. Aquat. Sci. , doi:10.1007/s00027-

016-0477-0

Bristow, L. A., N. Sarode, J. Cartee, A. Caro-Quintero, B. Thamdrup, and F. J. Stewart. 2015. Biogeochemical and metagenomic analysis of nitrite accumulation in the Gulf of Mexico hypoxic zone. Limnol. Oceanogr. **60**: 1733–1750.

Caporaso, J. G., C. L. Lauber, W. a Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. a Gilbert, G. Smith, and R. Knight. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. **6**: 1621–1624.

Caron, D. A., K. G. Porter, and R. W. Sanders. 1990. Carbon, nitrogen, and phosphorus budgets for the mixotrophic phytoflagellate Poterioochromonas malhamensis (Chrysophyceae) during bacterial ingestion. Limnol. Oceanogr. **35**: 433–443.

Cauwet, G., D. A. Hansell, and C. A. Carlson. 2002. DOM in the coastal zone, p. 579–609. *In* Biogeochemistry of marine dissolved organic matter. Academic Press, San Diego, CA.

Chong, I. G., and C. H. Jun. 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. **78**: 103–112.

Cole, J. J., and N. F. Caraco. 2001. Carbon in catchments: Connecting terrestrial carbon losses with aquatic metabolism. *Marine and Freshwater Research*. 101–110.

Cole, J. J., Y. T. Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, C. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg, and J. M. Melack. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems **10**: 171–184.

Cottrell, M. T., L. A. Waidner, L. Yu, and D. L. Kirchman. 2005. Bacterial diversity of metagenomic and PCR libraries from the Delaware River. Environ. Microbiol. **7**: 1883–1895.

Crump, B. C., L. A. Amaral-Zettler, and G. W. Kling. 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME J **6**: 1629–1639.

Crump, B. C., E. V. Armbrust, and J. A. Baross. 1999. Phylogenetic Analysis of Particle-Attached and Free-Living Bacterial Communities in the Columbia River, Its Estuary, and the Adjacent Coastal Ocean. Appl. Environ. Microbiol. **65**: 3192–3204.

Crump, B. C., and J. A. Baross. 1996. Particle-attached bacteria and heterotrophic plankton associated with the Columbia River estuarine turbidity maxima. Part. Bact. heterotrophic Plankt. Assoc. with Columbia River Estuar. Turbid. maxima

Crump, B. C., J. A. Baross, and C. A. Simenstad. 1998. Dominance of particle-attached bacteria in the Columbia River estuary, USA. Aquat. Microb. Ecol. **14**: 7–18.

Crump, B. C., and J. E. Hobbie. 2005. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol. Oceanogr. **50**: 1718–1729.

Crump, B. C., C. S. Hopkinson, M. L. Sogin, and J. E. Hobbie. 2004. Microbial Biogeography along an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time. Appl. Environ. Microbiol. **70**: 1494–1505.

Crump, B. C., B. J. Peterson, P. A. Raymond, R. M. W. Amon, A. Rinehart, J. W. McClelland, and R. M. Holmes. 2009. Circumpolar synchrony in big river bacterioplankton. Proc. Natl. Acad. Sci. **106**: 21208–21212.

D'Ambrosio, L., K. Ziervogel, B. Macgregor, A. Teske, and C. Arnosti. 2014. Composition and enzymatic function of particle-associated and free-living bacteria: a coastal/offshore comparison. ISME J. 1–13.

Dagg, M., R. Benner, S. Lohrenz, and D. Lawrence. 2004. Transformation of dissolved and

785 particulate materials on continental shelves influenced by large rivers: Plume processes.
786 Cont. Shelf Res. **24**: 833–858.
787 Dagg, M. J., T. S. Bianchi, G. A. Breed, W. J. Cai, S. Duan, H. Liu, B. A. McKee, R. T. Powell,
788 and C. M. Stewart. 2005. Biogeochemical characteristics of the lower Mississippi River,
789 USA, during June 2003. Estuaries **28**: 664–674.
790 DeLong, E. F., D. G. Franks, and A. Alldredge. 1993. Phylogenetic diversity of aggregate-
791 attached vs. free-living marine bacteria assemblages. Limnol. Ocean. **38**: 924–934.
792 Dick, M. . 2003. Straminipilous Fungi: Systematics of the Peronosporomycetes Including
793 Accounts of the Marine Straminipilous Protists, the Plasmodiophorids and Similar
794 Organisms. Mycopathologia **156**: 385–386.
795 Duan, S., R. T. Powell, and T. S. Bianchi. 2014. High frequency measurement of nitrate
796 concentration in the Lower Mississippi River, USA. J. Hydrol. **519**: 376–386.
797 Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves
798 sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.
799 Ensign, S. H., and M. W. Doyle. 2006. Nutrient spiraling in streams and river networks. J.
800 Geophys. Res. Biogeosciences **111**, doi:10.1029/2005JG000114
801 Fortunato, C. S., A. Eiler, L. Herfort, J. A. Needoba, T. D. Peterson, B. C. Crump. 2013.
802 Determining indicator taxa across spatial and seasonal gradients in the Columbia River
803 coastal margin. ISME J **7**: 1899–1911.
804 Fortunato, C. S., L. Herfort, P. Zuber, A. M. Baptista, and B. C. Crump. 2012. Spatial variability
805 overwhelms seasonal patterns in bacterioplankton communities across a river to ocean
806 gradient. ISME J **6**: 554–563.
807 Freimann, R., H. Bürgmann, S. E. G. Findlay, and C. T. Robinson. 2015. Hydrologic linkages
808 drive spatial structuring of bacterial assemblages and functioning in alpine floodplains TL -
809 6. Front. Microbiol. **6 VN-re**, doi:10.3389/fmicb.2015.01221
810 Van Geel, B., L. R. Mur, M. Ralska-Jasiewiczowa, and T. Goslar. 1994. Fossil akinetes of
811 Aphanizomenon and Anabaena as indicators for medieval phosphate-eutrophication of Lake
812 Gosciaz (Central Poland). Rev. Palaeobot. Palynol. **83**: 97–105.
813 Ghai, R., F. Rodŕíguez-Valera, K. D. McMahon, D. Toyama, R. Rinke, T. C. S. de Oliveira, J.
814 W. Garcia, F. P. de Miranda, and F. Henrique-Silva. 2011. Metagenomics of the water
815 column in the pristine upper course of the Amazon river. PLoS One **6**,
816 doi:10.1371/journal.pone.0023785
817 Goolsby, D. A., and W. A. Battaglin. 2001. Long-term changes in concentrations and flux of
818 nitrogen in the Mississippi River basin, USA. Hydrol. Process. **15**: 1209–1226.
819 Grossart, H. P. 2010. Ecological consequences of bacterioplankton lifestyles: Changes in
820 concepts are needed. Environ. Microbiol. Rep. **2**: 706–714.
821 Guidi, L., S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L.
822 Berline, and J. R. Brum. 2016. Plankton networks driving carbon export in the oligotrophic
823 ocean. Nature
824 Islam, M. T., and A. von Tiedemann. 2011. 2,4-Diacetylphloroglucinol suppresses
825 zoosporogenesis and impairs motility of Peronosporomycete zoospores. World J. Microbiol.
826 Biotechnol. **27**: 2071–2079.
827 Jackson, C. R., J. J. Millar, J. T. Payne, and C. A. Ochs. 2014. Free-living and particle-associated
828 bacterioplankton in large rivers of the Mississippi River Basin demonstrate biogeographic
829 patterns. Appl. Environ. Microbiol. **80**: 7186–7195.
830 Jansson, M., P. Blomqvist, A. Jonsson, and A. BergstrÖm. 1996. Nutrient limitation of

bacterioplankton, autotrophic and mixotrophic phytoplankton, and heterotrophic nanoflagellates in Lake Örträsket. Limnol. Oceanogr. **41**: 1552–1559.

Kielak, A. M., C. C. Barreto, G. A. Kowalchuk, J. A. van Veen, and E. E. Kuramae. 2016. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front. Microbiol. **7**: 744.

Kolmakova, O. V., M. I. Gladyshev, A. S. Rozanov, S. E. Peltek, and M. Y. Trusova. 2014. Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. FEMS Microbiol. Ecol. **89**: 442–450.

Korajkic, A., L. W. Parfrey, B. R. McMinn, Y. V. Baeza, W. VanTeuren, R. Knight, and O. C. Shanks. 2015. Changes in bacterial and eukaryotic communities during sewage decomposition in Mississippi river water TL - 69. Water Res. **69 VN-r**: 30–39.

Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted gene co-expression network analysis. BMC Bioinformatics **9**: 559.

Lauber, C. L., N. Zhou, J. I. Gordon, R. Knight, and N. Fierer. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol. Lett. **307**: 80–86.

Learman, D. R., M. W. Henson, J. C. Thrash, B. Temperton, P. M. Brannock, S. R. Santos, A. R. Mahon, and K. M. Halanych. 2016. Biogeochemical and microbial variation across 5500 km of Antarctic surface sediment implicates organic matter as a driver of benthic community structure. Front. Microbiol. **7**, doi:10.3389/fmicb.2016.00284

Li, H.-F., Z. T. Zou, B. Z. Li, X. Wang, J. S. Yang, and H. L. Yuan. 2012. Novosphingobium sediminis sp. nov., isolated from the sediment of a eutrophic lake. J. Gen. Appl. Microbiol. **58**: 357–362.

Love, M. I., S. Anders, and W. Huber. 2014. Differential analysis of count data - the DESeq2 package. Genome Biol. **15**: 550.

McIsaac, G. F., M. B. David, G. Z. Gertner, and D. a Goolsby. 2001. Nitrate flux in the Mississippi River. Nature **414**: 166–167.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One **8**: e61217.

Mevik, B. H., and R. Wehrens. 2007. The pls Package: Principle Component and Partial Least Squares Regression in R. J. Stat. Softw. **18**: 1–24.

Meziti, A., D. Tsementzi, K. Ar. Kormas, H. Karayanni, and K. T. Konstantinidis. 2016. Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. Environ. Microbiol.

Necchi Jr, O. 2016. Heterokonts (Xanthophyceae and Chrysophyceae) in Rivers, p. 153–158. *In* River Algae. Springer.

Niño-García, J. P., C. Ruiz-González, and P. A. del Giorgio. 2016. Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. ISME J. , doi:10.1038/ismej.2015.226

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2015. vegan: Community Ecology Package. R package version 2.2-1. R Packag. version **1**: R package version 2.2-1.

Parada, A. E., D. M. Needham, and J. A. Fuhrman. 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ. Microbiol.

Payne, J. T., J. J. Millar, C. R. Jackson, and C. A. Ochs. 2017. Patterns of variation in diversity of the Mississippi river microbiome over 1,300 kilometers. PLoS One **12**: e0174890.

877 Pierson, S. M., B. J. Rosenbaum, L. D. McKay, and T. G. Dewald. 2008. Strahler Stream Order
878 and Strahler Calculator Values in NHDPlus.
879 Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, F. O. Glockner, and F. O.
880 Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned
881 ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**: 7188–7196.
882 Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner.
883 2013. The SILVA ribosomal RNA gene database project: Improved data processing and
884 web-based tools. Nucleic Acids Res. **41**: 590–596.
885 Rabalais, N. N., R. E. Turner, B. K. Sen Gupta, D. F. Boesch, P. Chapman, and M. C. Murrell.
886 2007. Hypoxia in the northern Gulf of Mexico: Does the science support the Plan to
887 Reduce, Mitigate, and Control Hypoxia? Estuaries and Coasts **30**: 753–772.
888 Rabalais, N. N., R. E. Turner, D. Justić, Q. Dortch, W. J. Wiseman, B. K. Sen Gupta, and D.
889 Justic. 1996. Nutrient Changes in the Mississippi River and System Responses on the
890 Adjacent Continental Shelf. Estuaries **19**: 386.
891 Rabalais, N. N., R. E. Turner, and W. J. Wiseman. 2002. Gulf of Mexico Hypoxia, a.K.a. "the
892 Dead Zone." Annu. Rev. Ecol. Syst. **33**: 235–263.
893 R Developement Core Team. 2015. R: A Language and Environment for Statistical Computing.
894 R Found. Stat. Comput. **1**: 409.
895 Read, D. S., H. S. Gweon, M. J. Bowes, L. K. Newbold, D. Field, M. J. Bailey, and R. I.
896 Griffiths. 2015. Catchment-scale biogeography of riverine bacterioplankton. ISME J **9**:
897 516–526.
898 Richey, J. E., J. M. Melack, A. K. Aufdenkampe, V. M. Ballester, and L. L. Hess. 2002.
899 Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric
900 CO2. Nature **416**: 617–620.
901 Riemann, L., and A. Winding. 2001. Community dynamics of free-living and particle-associated
902 bacterial assemblages during a freshwater phytoplankton bloom. Microb. Ecol. **42**: 274–
903 285.
904 Van Rossum, T., M. A. Peabody, M. I. Uyaguari-Diaz, K. I. Cronin, M. Chan, J. R. Slobodan,
905 M. J. Nesbitt, C. A. Suttle, W. W. L. Hsiao, P. K. C. Tang, N. A. Prystajecky, and F. S. L.
906 Brinkman. 2015. Year-long metagenomic study of river microbiomes across land use and
907 water quality . Front. Microbiol. **6**: 1–15.
908 Rubin, B. E. R., S. M. Gibbons, S. Kennedy, J. Hampton-Marcell, S. Owens, and J. A. Gilbert.
909 2013. Investigating the Impact of Storage Conditions on Microbial Community
910 Composition in Soil Samples. PLoS One **8**: 1–9.
911 Ruiz-González, C., L. Proia, I. Ferrera, J. M. Gasol, and S. Sabater. 2013. Effects of large river
912 dam regulation on bacterioplankton community structure. FEMS Microbiol. Ecol. **84**: 316–
913 331.
914 Ruiz-González, C., G. Salazar, R. Logares, L. Proia, J. M. Gasol, and S. Sabater. 2015. Weak
915 Coherence in Abundance Patterns Between Bacterial Classes and Their Constituent OTUs
916 Along a Regulated River. Front. Microbiol. **6**: 1293.
917 Russell, T. A., and L. Weller. 2013. State of the River Report: Water Quality and River Health in
918 the Metro Mississippi River. Friends of the Mississippi River.
919 Salter, S. J., M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J.
920 Parkhill, N. J. Loman, and A. W. Walker. 2014. Reagent and laboratory contamination can
921 critically impact sequence-based microbiome analyses. BMC Biol. **12**: 87.
922 Savio, D., L. Sinclair, U. Z. Ijaz, J. Parajka, G. H. Reischer, P. Stadler, A. P. Blaschke, G.

Blöschl, R. L. Mach, A. K. T. Kirschner, A. H. Farnleitner, and A. Eiler. 2015. Bacterial diversity along a 2600 km river continuum. Environ. Microbiol. **17**: n/a-n/a.

Schilling, K. E., K.-S. Chan, H. Liu, and Y.-K. Zhang. 2010. Quantifying the effect of land use land cover change on increasing discharge in the Upper Mississippi River. J. Hydrol. **387**: 343–345.

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. a. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75**: 7537–7541.

Singh, V. 2012. Application of Frequency and Risk in Water Resources: Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14–17 May 1986, Louisiana State University, Baton Rouge, USA, Springer Science & Business Media.

Smith, M. W., L. Herfort, K. Tyrol, D. Suciu, V. Campbell, B. C. Crump, T. D. Peterson, P. Zuber, A. M. Baptista, and H. M. Simon. 2010. Seasonal Changes in Bacterial and Archaeal Gene Expression Patterns across Salinity Gradients in the Columbia River Coastal Margin. PLoS One **5**: e13312.

Staley, C., T. J. Gould, P. Wang, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2014a. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. Front Microbiol **5**: 524.

Staley, C., T. J. Gould, P. Wang, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2014b. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Front. Microbiol. **5**: 414.

Staley, C., T. J. Gould, P. Wang, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2015. Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River. Sci. Total Environ. **505**: 435–45.

Staley, C., T. J. Gould, P. Wang, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2016. Sediments and Soils Act as Reservoirs for Taxonomic and Functional Bacterial Diversity in the Upper Mississippi River. Microb. Ecol. **71**: 814–824.

Staley, C., T. Unno, G. TJ, B. Jarvis, J. Phillips, C. JB, T. J. Gould, B. Jarvis, J. Phillips, J. B. Cotner, and M. J. Sadowsky. 2013. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J. Appl. Microbiol. **115**: 1147–1158.

Tatangelo, V., A. Franzetti, I. Gandolfi, G. Bestetti, and R. Ambrosini. 2014. Effect of preservation method on the assessment of bacterial community structure in soil and water samples. FEMS Microbiol. Lett. **356**: 32–38.

Thrash, J. C., J. L. Weckhorst, and D. M. Pitre. 2015. Cultivating Fastidious Microbes, p. 1–22. *In* Protocols for Metagenomic Library Generation and Analysis in Petroleum Hydrocarbon Microbe Systems. Humana Press.

Trusova, M. Y., and M. I. Gladyshev. 2002. Phylogenetic diversity of winter bacterioplankton of eutrophic siberian reservoirs as revealed by 16S rRNA gene sequence. Microb. Ecol. **44**: 252–259.

Turner, R. E., and N. N. Rabalais. 2003. Linking landscape and water quality in the Mississippi river basin for 200 years. Bioscience **53**: 563–572.

Turner, R. E., and N. N. Rabalais. 2004. Suspended sediment, C, N, P, and Si yields from the Mississippi River Basin. Hydrobiologia **511**: 79–89.

969 Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. River
970 continuum concept. Can J Fish Aquat Sci **37**: 130–137.
971 Vorosmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S.
972 Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global
973 threats to human water security and river biodiversity. Nature **467**: 555–561.
974 Weiss, S., A. Amir, E. R. Hyde, J. L. Metcalf, S. J. Song, and R. Knight. 2014. Tracking down
975 the sources of experimental contamination in microbiome studies. Genome Biol. **15**: 564.
976 Withers, P. J. A., and H. P. Jarvie. 2008. Delivery and cycling of phosphorus in rivers: A review.
977 Sci. Total Environ. **400**: 379–395.
978 Wood, S. A., M. J. Prentice, K. Smith, and D. P. Hamilton. 2010. Low dissolved inorganic
979 nitrogen and increased heterocyte frequency: precursors to Anabaena planktonica blooms in
980 a temperate, eutrophic reservoir. J. Plankton Res. **32**: 1315–1325.
981 Zeglin, L. H. 2015. Stream microbial diversity in response to environmental changes: review and
982 synthesis of existing research. Front. Microbiol. **6**: 454.
983 Zwart, G., B. C. Crump, M. P. Kamst-van Agterveld, F. Hagen, and S. K. Han. 2002. Typical
984 freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of
985 lakes and rivers. Aquat. Microb. Ecol. **28**: 141–155.
986
987
988
989
990
991
992
993
994

995
996 **Tables**
997

Table 1. Summary of the correlations between environmental variables and microbial community beta diversity ordination on NMDS plots in Figure 2. +/- designates the direction of the association between the environmental variable and NMDS axis

Variable		NMDS1	NMDS2	P-value	R^2		NMDS1	NMDS2	P-value	R^2
PO_4^{3-}	16S rRNA Gene 0.2-2.7 µm	+	+	0.001	0.404	16S rRNA Gene > 2.7 µm	+	+	0.001	0.575
$SiSO_4$		-	+	0.003	0.183		-	-	0.015	0.115
NO_3^-		+	+	0.001	0.443		+	+	0.001	0.33
NH_4^+		-	-	0.002	0.178		-	-	0.004	0.189
NO_2^-		Not significant					Not significant			
River Speed		-	-	0.001	0.261		-	-	0.001	0.39
Water Temperature		-	-	0.001	0.385		-	-	0.001	0.524
Turbidity		-	-	0.001	0.409		-	-	0.001	0.536
PO_4^{3-}	18S rRNA Gene 0.2-2.7 µm	+	+	0.001	0.555	18S rRNA Gene > 2.7 µm	+	-	0.001	0.486
$SiSO_4$		-	+	0.001	0.264		-	-	0.001	0.265
NO_3^-		Not significant					+	+	0.001	0.2
NH_4^+		-	+	0.019	0.106		Not significant			
NO_2^-		Not significant					Not significant			
River Speed		-	-	0.001	0.282		-	-	0.001	0.345
Water Temperature		-	-	0.001	0.29		-	-	0.001	0.372
Turbidity		-	-	0.001	0.272		-	-	0.001	0.353

Table 2. Summary of the correlations between OTUs and microbial community beta diversity ordination on NMDS plots in Figure 2. +/- designates the direction of the association between the OTU and NMDS axis.

	OTU	Classification	NMDS1	NMDS2	P-value	R^2		OTU	Classification	NMDS1	NMDS2	P-value	R^2
16S rRNA Gene 0.2-2.7 µm	***OTU372***	Unc. Bacteria	+	-	0.0001	0.604	**16S rRNA Gene > 2.7 µm**	***OTU3***	acI clade	+	+	0.0001	0.796
	OTU135	*Flavobacterium* sp.	-	-	0.0001	0.522		***OTU181***	Unc. Bacillaceae	-	+	0.0001	0.773
	OTU29	*Fluviicola* sp.	-	-	0.0001	0.509		***OTU145***	*Desulfosporosinus* sp.	-	+	0.0001	0.742
	OTU48	Unc. Microbacteriaceae	-	+	0.0001	0.508		***OTU13***	Unc. Bacillales	-	+	0.0001	0.740
	OTU115	*Flavobacterium* sp.	-	+	0.0001	0.503		***OTU68***	Unc. Micromonosporaceae	-	+	0.0001	0.726
	OTU33	Unc. Holophagaceae	+	+	0.0001	0.501		***OTU202***	Unc. Bacillales	-	+	0.0001	0.725
	OTU161	*Flavobacterium* sp.	-	-	0.0001	0.501		***OTU60***	Unc. Gemmatimonadaceae	+	+	0.0001	0.716
	OTU100	*Clostridium* sp.	+	+	0.0001	0.497		***OTU89***	Unc. Bacillales	-	+	0.0001	0.702
	OTU85	Unc. Peptostreptococcaceae	+	+	0.0001	0.496		***OTU169***	Unc. Micromonosporaceae	-	+	0.0001	0.698
	OTU62	Unc. Comamonadaceae	-	+	0.0001	0.493		***OTU28***	Unc. Planococcaceae	-	+	0.0001	0.694
18S rRNA Gene 0.2-2.7 µm	***OTU1***	Unc. Eukaryota	+	+	0.0001	0.693	**18S rRNA Gene > 2.7 µm**	***OTU1***	Unc. Eukaryote	+	+	0.0001	0.870
	OTU63	Unc. Ochrophyta	-	-	0.0001	0.645		***OTU222***	Unc. Ochrophyta	+	+	0.0001	0.806
	OTU241	Unc. Ochrophyta	-	+	0.0001	0.642		***OTU3***	Unc. Stramenopiles	+	-	0.0001	0.796
	OTU133	Unc. Ochrophyta	-	+	0.0001	0.632		***OTU1984***	Unc. Eukaryote	+	-	0.0001	0.794
	OTU219	Unc. Eukaryota	+	+	0.0001	0.630		***OTU345***	Unc. Ochrophyta	-	-	0.0001	0.792
	OTU113	Unc. Alveolata	+	+	0.0001	0.616		***OTU112***	Unc. Cryptomonadales	+	+	0.0001	0.779
	OTU154	Unc. Ochrophyta	-	-	0.0001	0.598		***OTU63***	Unc. Eukaryote	+	+	0.0001	0.773
	OTU481	Unc. Eukaryota	-	-	0.0001	0.598		***OTU154***	Chrysophyceae clone P34.45	-	-	0.0001	0.773
	OTU457	Unc. Eukaryota	-	-	0.0001	0.596		***OTU155***	Bacillariophytina	+	+	0.0001	0.765
	OTU3	Unc. Stramenopile	+	+	0.0001	0.594		***OTU1146***	Unc. Eukaryote	+	+	0.0001	0.765

Table 3. Correlations between prokaryotic and eukaryotic submodules to nitrate and phosphate, as well as PLS model results

Fraction	*Gene*	*Nutrient*	*Submodule*	*Eigen Correlations*	*Total OTUs*	*PLS model*
0.2-2.7 µm	16S	NO_3^{2-}	Blue	R = 0.6, P = 7e-08	77	R^2=0.35; R = 0.65, P =1.e-09
> 2.7 µm	16S	NO_3^{2-}	Brown	R = 0.56, P = 3e-07	133	R^2 = 0.69; R = 0.83, P= < 2.2e-16
0.2-2.7 µm	16S	PO_4^{3-}	Turquoise	R = 0.53, P = 2e-06	51	R^2=0.80; R = 0.89, P = < 2.3e-16
> 2.7 µm	16S	PO_4^{3-}	Yellow	R = 0.53, P = 1e-06	80	R^2= 0.48; R = 0.77, P = 4.88e-15
0.2-2.7 µm	18S	NO_3^{2-}	Green	R = 0.39, P = 6e-04	39	R^2=0.38; R = 0.62, P =2.97e-9
> 2.7 µm	18S	NO_3^{2-}	Brown	R = 0.52, P = 2e-06	59	R^2=0.572; R = 0.759, P= 6.7e-15
0.2-2.7 µm	18S	PO_4^{3-}	Turquoise	R = 0.56, P = 2e-07	56	R^2=0.80; R = 0.89, P = < 2e-16
> 2.7 µm	18S	PO_4^{3-}	Red	R = 0.60, P = 2e-08	39	R^2 = 0.618; R = 0.799, P = < 2e-16

1003

Figure Legends

Figure 1. Sampling map (A) with graph inserts that represent the measured discharge rate (cubic meter second^{-1} [cms]) as recorded on the USGS gauges, and six environmental parameters measured along the transect, according to concentration or visible depth of secchi disk by distance (B). Throughout the figure and text, blue and red dots represent sampling locations above and below the Missouri River confluence, respectively, and are designated throughout as "upper" and "lower." Cell Counts only represent the Planktonic (< 2.7 µm) fraction.

Figure 2. Non-metric multidimensional scaling (NMDS) results for whole community. The four plots represent > 2.7 µm (A, C) and 0.2-2.7 µm (B, D) fractions for the 16S (A, B) and 18S (C, D) rRNA gene communities.

Figure 3. Relative abundance, according to transect distance, for the top 6 and 5 phyla in the 16S rRNA gene > 2.7 µm (A) and 0.2-2.7 µm (B) communities and 18S rRNA gene > 2.7 µm (C) and 0.2-2.7 µm (D) communities, respectively. Non-linear regressions with 95% CI (gray shading) are provided for reference.

Figure 4. Core microbiome aggregate abundance for the 16S (A) and 18S (B) rRNA gene. In each, triangles and circles points represent 0.2-2.7 µm and > 2.7 µm fractions, respectively. Non-linear regressions with 95% confidence intervals (CI) (gray shading) are provided for reference.

Figure 5. PLS results for the 16S rRNA gene community submodules most associated with phosphate (A) and nitrate (B). OTU correlation with a given nutrient is indicated on the y-axis according to the number of co-correlations (node centrality) on the x-axis. Community fractions: 0.2-2.7 µm (A) and > 2.7 µm (B). Circle size is proportional to VIP

scores, with top 10 VIP scoring and top node centrality OTUs labeled with their highest-resolution taxonomic classification and OTU number. Colors represent the taxonomic classification the phylum level. For all OTU taxonomic designations resulting from this analysis, see Table S1 “16S 0.2-2.7µm PO4” and “16S > 2.7µm NO3”.

Figure 6. PLS results for the 18S rRNA gene community submodules most associated with phosphate and nitrate. OTU correlation with a given nutrient is indicated on the y-axis according to the number of co-correlations (node centrality) on the x-axis. Community fractions: 0.2-2.7 µm (A) and > 2.7 µm (B, C). Circle size is proportional to VIP scores, with top 10 VIP scoring and top node centrality OTUs labeled with their highest-resolution taxonomic classification and OTU number. Colors represent the taxonomic classification the phylum level. For all OTU taxonomic designations resulting from this analysis, see Table S1 “18S 0.2-2.7µm NO3”, “16S > 2.7µm PO4”, and “16S >2.7µm NO3”

Supplemental Figure Legends

Figure S1. NMDS results for the 16S (A) and 18S (B) rRNA gene communities. In each, circles and triangles represent the > 2.7 µm and 0.2-2.7 µm fractions, respectively.

Figure S2. Hclust results for the 16S (A) and 18S (B) rRNA gene communities.

Figure S3. Separated upper and lower MSR Richness and Evenness indexes for the 16S > 2.7 µm (A) and 0.2-2.7 µm (B) and 18S > 2.7 (C) and 0.2-2.7 µm (D) communities. Linear regressions with 95% CI (gray shading) are provided for reference.

Figure S4. Relative abundance, by phylum, according to transect distance, for phyla accounting

1052 for > 0.1% of the total reads for the 16S rRNA (A, B) and 18S rRNA genes (C, D) > 2.7
1053 µm (A, C) and 0.2-2.7 µm (B, D) communities. Non-linear regressions with 95% CI (gray
1054 shading) are provided for reference.

1055 **Figure S5.** Separated upper and lower MSR core microbiome aggregate abundance for the 16S
1056 rRNA gene communities. For the upper and lower river, the core microbiome was defined
1057 separately requiring OTUs to have greater than one read in 90% of the samples. > 2.7 µm
1058 (A) and 0.2-2.7 µm (B) 16S rRNA gene communities in the upper and lower MSR.

1059 **Figure S6.** PLS results for the 0.2-2.7 µm and > 2.7 µm 16S (A-D) and 18S (E-H) rRNA gene
1060 community for selected submodules with nitrate and phosphate and a VIP score > 1.
1061 Correlation of submodule OTUs to nitrate and phosphate according to the number of co-
1062 correlations (node centrality) for 0.2-2.7 µm (A, B, E, F) and > 2.7 µm (C, D, G, H) 16S
1063 and 18S rRNA gene communities. Circle size is proportional to VIP scores, with top 10
1064 VIP scoring and top node centrality OTUs labeled with their highest-resolution taxonomic
1065 classification and OTU number. Colors represent the taxonomic classification the phylum
1066 level.

1067 **Figure S7.** WCGNA results for 0.2-2.7 µm 16S rRNA gene community submodules of interest
1068 based on Pearson correlations to environmental measurements (A). Color shading depicts
1069 the strength of the Pearson correlation with individual submodule's eigenvalue. Boxes
1070 indicate the selected submodule with its Pearson correlation and P-value. Graphs of the
1071 relationship between the selected submodule OTUs and the strength of the individual
1072 OTUs to nitrate (B) and phosphate (D). PLS regression of the predicted nutrient
1073 concentrations versus measured nutrient concentrations (C, E) using the corresponding
1074 selected submodules. Linear regressions with 95% CI (gray shading) are provided for

reference.

Figure S8. WCGNA results for 0.2-2.7 µm 18S rRNA gene community submodules of interest based on Pearson correlations to environmental measurements (A). Color shading depicts the strength of the Pearson correlation with individual submodule's eigenvalue. Boxes indicate the selected submodule with its Pearson correlation and P-value. Graphs of the relationship between the selected submodule OTUs and the strength of the individual OTUs to nitrate (B) and phosphate (D). PLS regression of the predicted nutrient concentrations versus measured nutrient concentrations (C, E) using the corresponding selected submodules. Linear regressions with 95% CI (gray shading) are provided for reference.

Figure S9. WCGNA results for > 2.7 µm 16S rRNA gene community submodules of interest based on Pearson correlations to environmental measurements (A). Color shading depicts the strength of the Pearson correlation with individual submodule's eigenvalue. Boxes indicate the selected submodule with its Pearson correlation and P-value. Graphs of the relationship between the selected submodule OTUs and the strength of the individual OTUs to nitrate (B) and phosphate (D). PLS regression of the predicted nutrient concentrations versus measured nutrient concentrations (C, E) using the corresponding selected submodules. Linear regressions with 95% CI (gray shading) are provided for reference.

Figure S10. WCGNA results for > 2.7 µm 18S rRNA gene community submodules of interest based on Pearson correlations to environmental measurements (A). Color shading depicts the strength of the Pearson correlation with individual submodule's eigenvalue. Boxes indicate the selected submodule with its Pearson correlation and P-value. Graphs of the

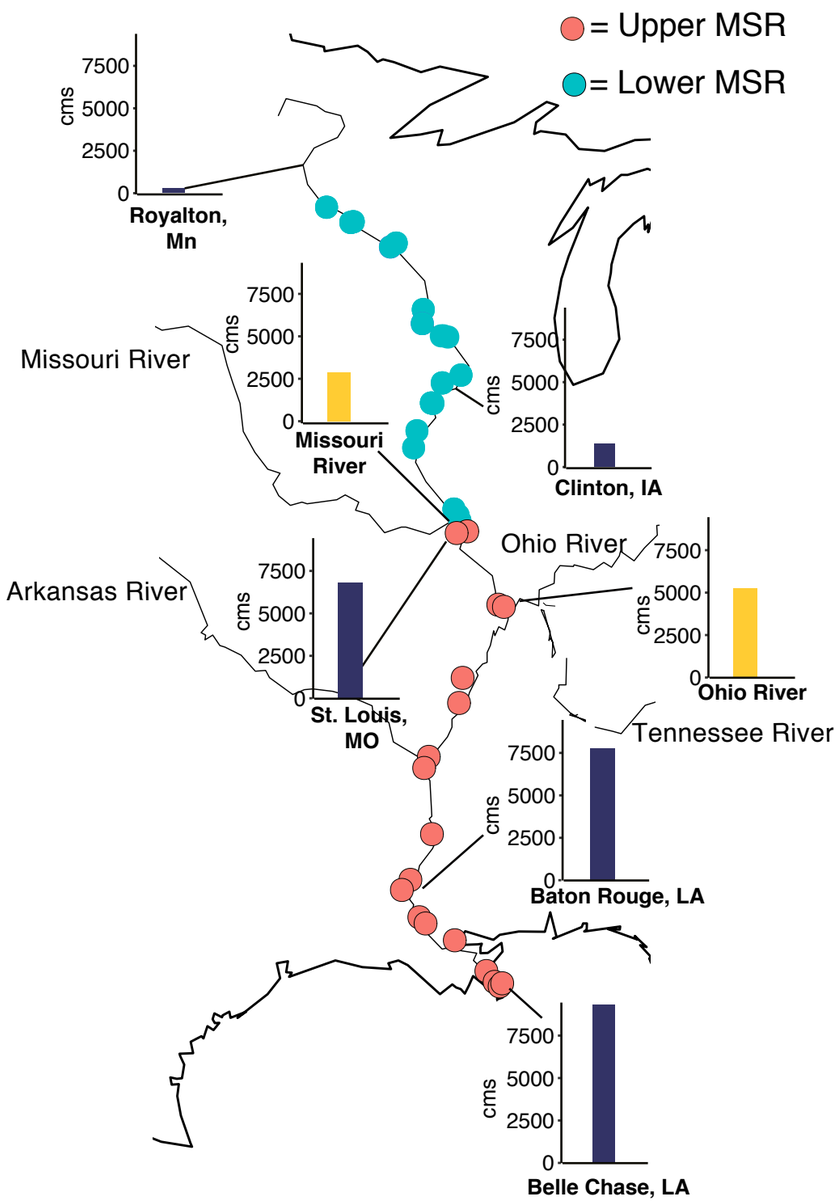
relationship between the selected submodule OTUs and the strength of the individual OTUs to nitrate (B) and phosphate (D). PLS regression of the predicted nutrient concentrations versus measured nutrient concentrations (C, E) using the corresponding selected submodules. Linear regressions with 95% CI (gray shading) are provided for reference.

Supporting Tables

Supplemental Table S1 is a spreadsheet, TableS1.xlsx. Tabs include General Information, Spearman Rank, 0.2-2.7 µm and > 2.7 µm 16S and 18S rRNA gene Core communities, 0.2-2.7 µm and > 2.7 µm 16S and 18S rRNA WGCNA, and 16S and 18S rRNA gene OTU tables (trimmed and normalized).

Additional Supplemental Information, including R scripts, Table S1, and our Mothur workflow are hosted on the Thrash Lab website at: thethrashlab.com/publications.

Figure 1A


= Upper MSR
= Lower MSR
Royalton, Mn
Missouri River
Missouri River
Clinton, IA
Ohio River
Arkansas River
St. Louis, MO
Ohio River
Tennessee River
Baton Rouge, LA
Belle Chase, LA
cms
7500
5000
2500
0

B

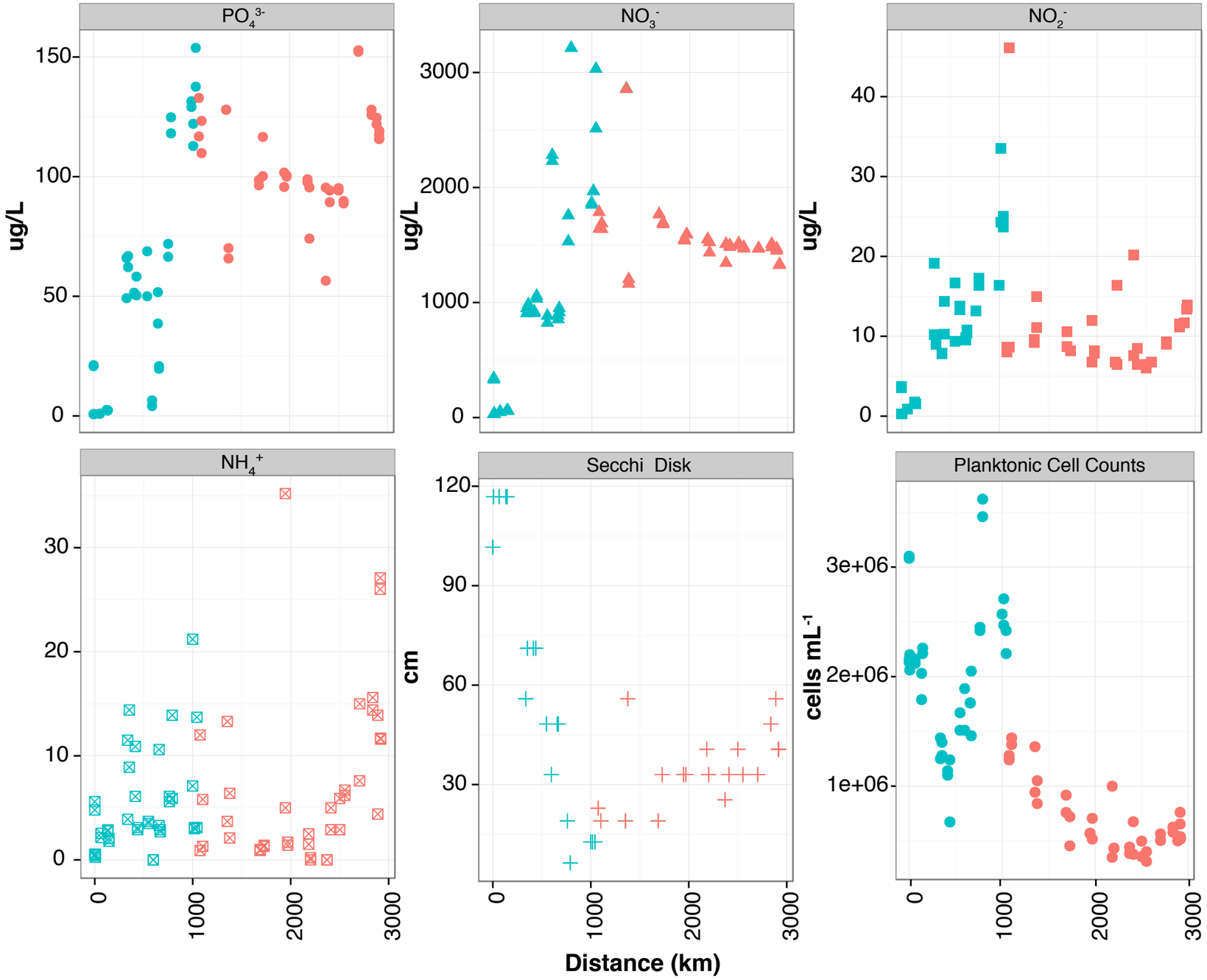
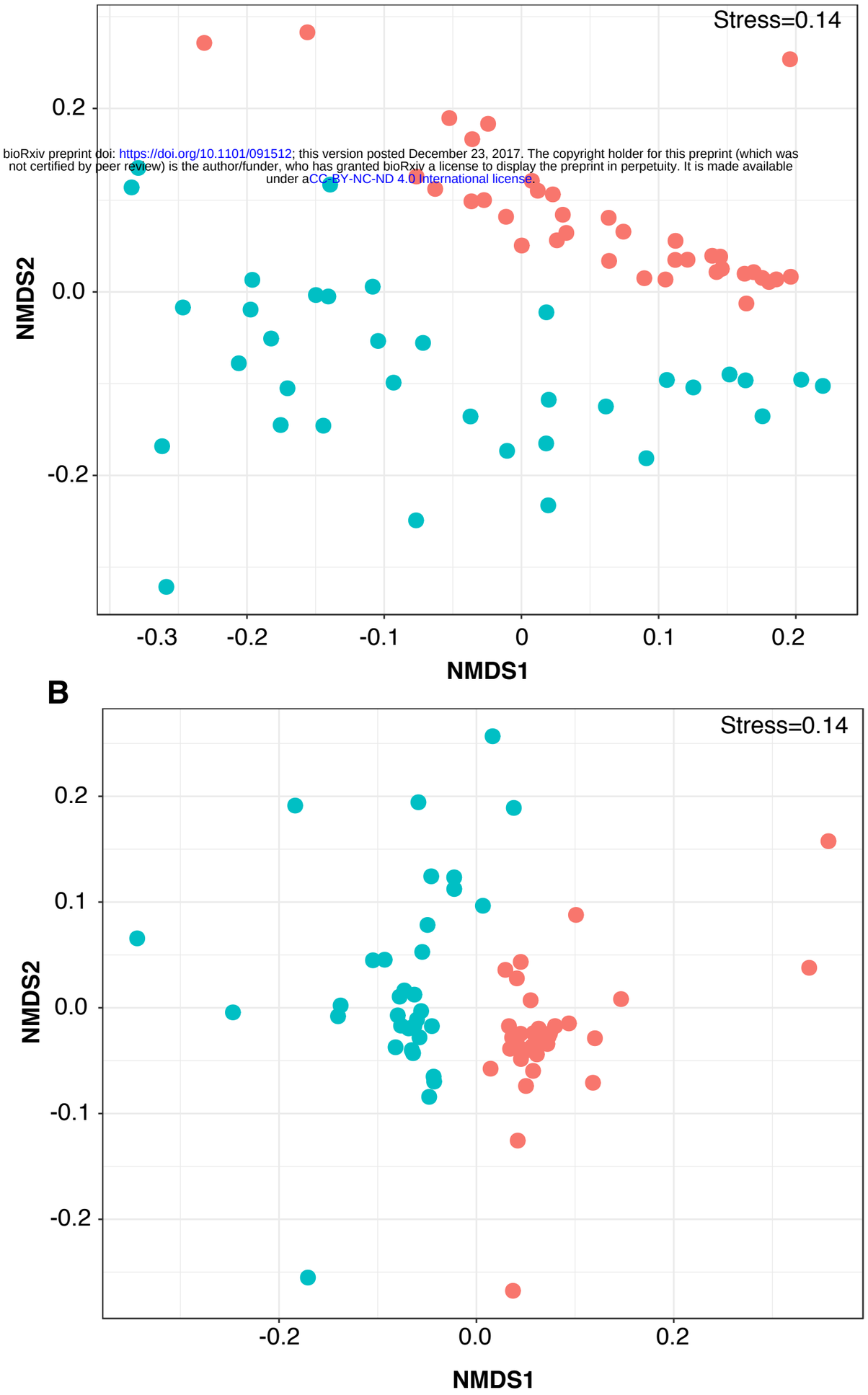
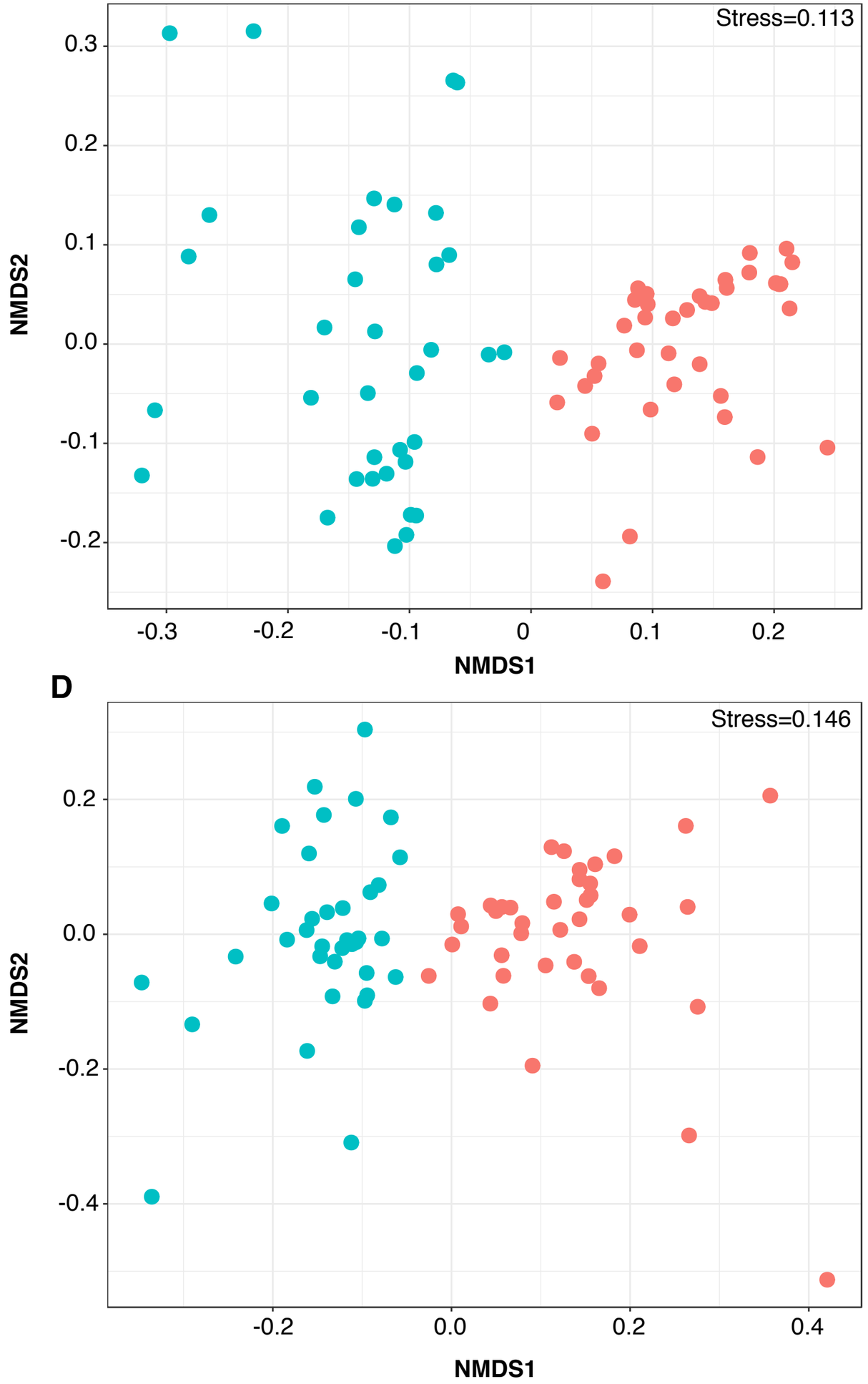

PO4 3-
NO3-
NO2-
NH4+
Secchi Disk
Planktonic Cell Counts
ug/L
cm
cells mL-1
Distance (km)

Figure 2A



C



B

D

Figure 3A

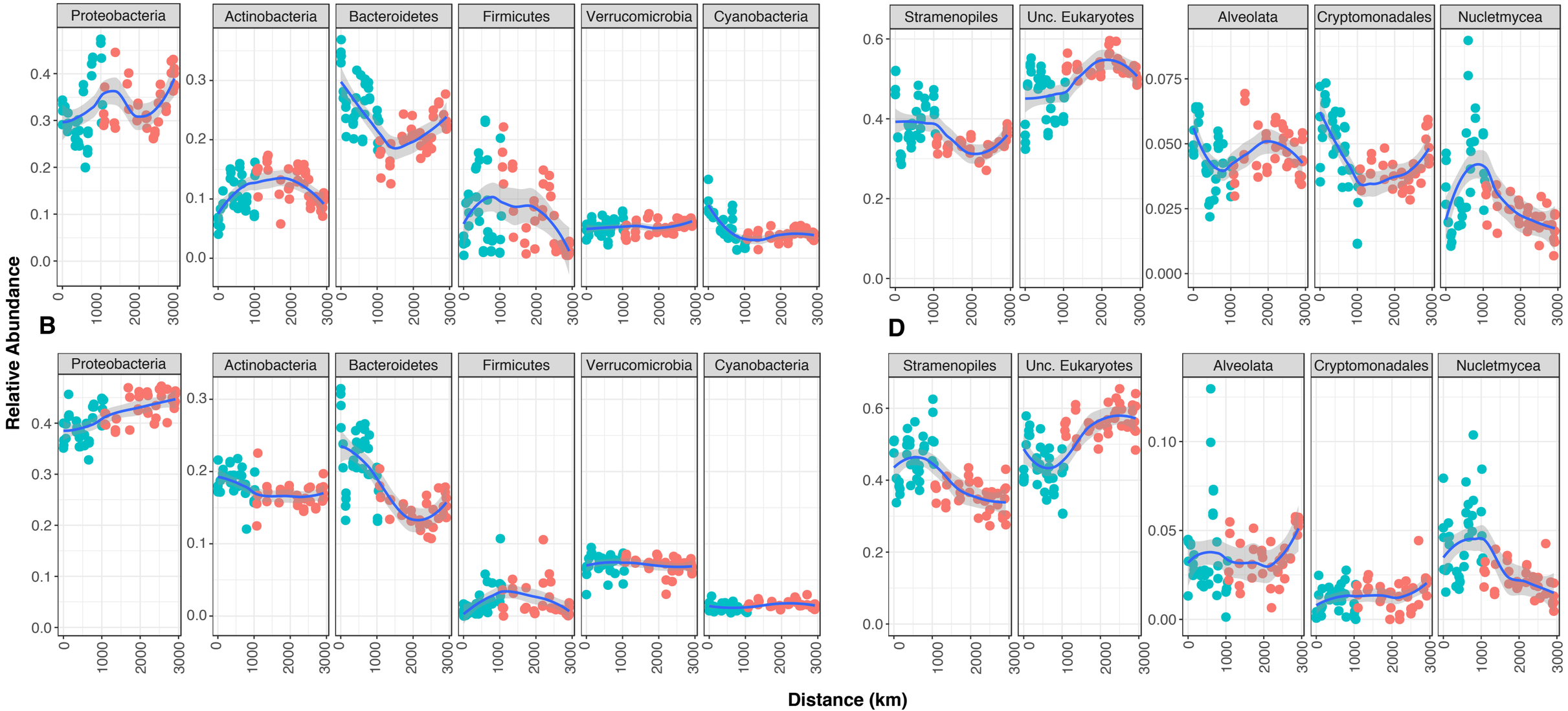

Proteobacteria
Actinobacteria
Bacteroidetes
Firmicutes
Verrucomicrobia
Cyanobacteria
B
C
Stramenopiles
Unc. Eukaryotes
Alveolata
Cryptomonadales
Nucletmycea
D
Relative Abundance
Distance (km)

Figure 4A

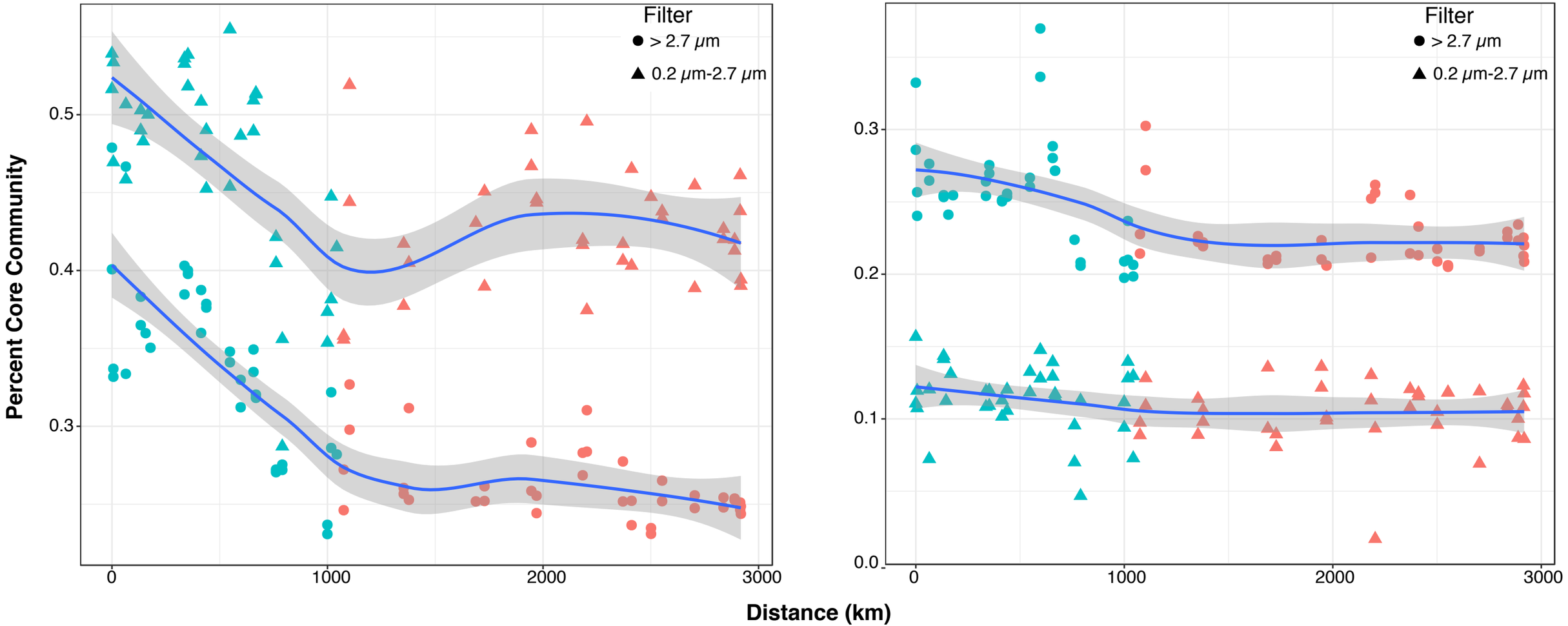

B
Filter
> 2.7 μm
0.2 μm-2.7 μm
Percent Core Community
0.5
0.4
0.3
0.2
0.1
0.0
0
1000
2000
3000
Distance (km)

Figure 5A

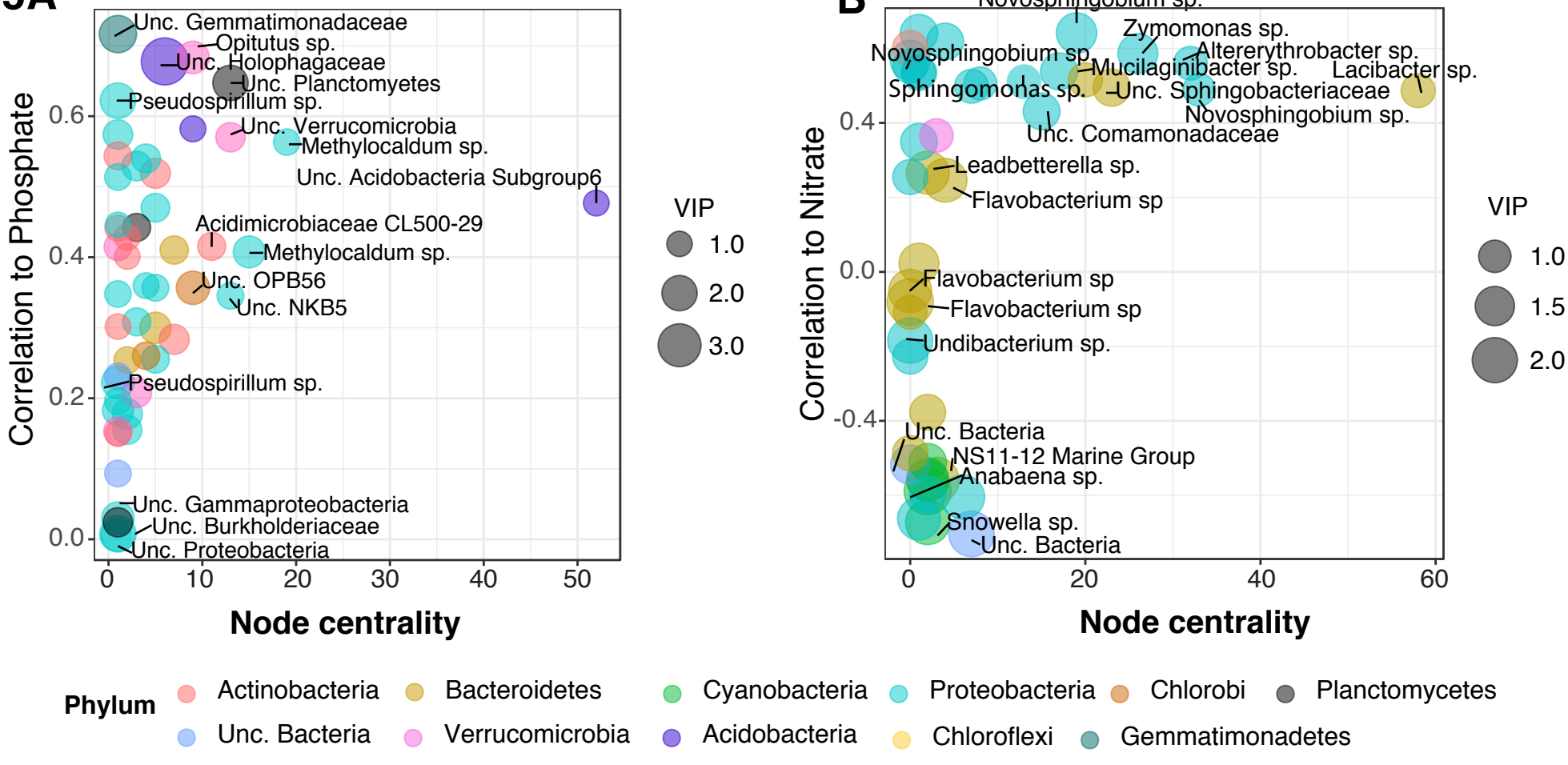
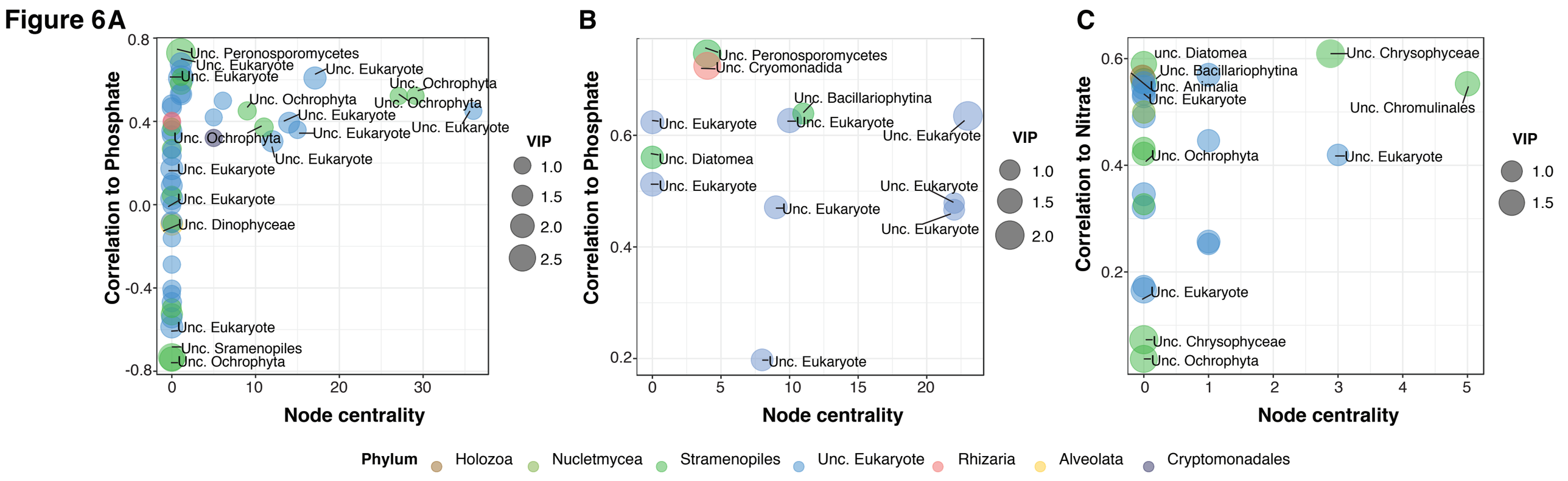

B
Unc. Gemmatimonadaceae
Opitutus sp.
Unc. Holophagaceae
Unc. Planctomyetes
Pseudospirillum sp.
Unc. Verrucomicrobia
Methylocaldum sp.
Unc. Acidobacteria Subgroup6
Acidimicrobiaceae CL500-29
Methylocaldum sp.
Unc. OPB56
Unc. NKB5
Pseudospirillum sp.
Unc. Gammaproteobacteria
Unc. Burkholderiaceae
Unc. Proteobacteria
Correlation to Phosphate
Node centrality
VIP
1.0
2.0
3.0
Novosphingobium sp.
Zymomonas sp.
Novosphingobium sp.
Altererythrobacter sp.
Mucilaginibacter sp.
Lacibacter sp.
Sphingomonas sp.
Unc. Sphingobacteriaceae
Novosphingobium sp.
Unc. Comamonadaceae
Leadbetterella sp.
Flavobacterium sp
Flavobacterium sp
Flavobacterium sp
Undibacterium sp.
Unc. Bacteria
NS11-12 Marine Group
Anabaena sp.
Snowella sp.
Unc. Bacteria
Correlation to Nitrate
Node centrality
VIP
1.0
1.5
2.0
Phylum
Actinobacteria
Bacteroidetes
Cyanobacteria
Proteobacteria
Chlorobi
Planctomycetes
Unc. Bacteria
Verrucomicrobia
Acidobacteria
Chloroflexi
Gemmatimonadetes

Figure 6


A
B
C
Correlation to Phosphate
Correlation to Nitrate
Node centrality
VIP
1.0
1.5
2.0
2.5
Unc. Peronosporomycetes
Unc. Eukaryote
Unc. Ochrophyta
Unc. Dinophyceae
Unc. Sramenopiles
Unc. Cryomonadida
Unc. Bacillariophytina
Unc. Diatomea
Unc. Chrysophyceae
Unc. Chromulinales
Unc. Animalia
Phylum
Holozoa
Nucletmycea
Stramenopiles
Unc. Eukaryote
Rhizaria
Alveolata
Cryptomonadales