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 20th 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 8th to 10th 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 600,000 cubic feet s^{-1} (cfs) (Rabalais et al. 1996), but can exceed 3 million cfs 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). This massive discharge includes excess nutrients (nitrogen and phosphorus), primarily from agricultural runoff

86 (McIsaac et al. 2001; Turner and Rabalais 2004; Schilling et al. 2010; Duan et al. 2014; Staley et
87 al. 2014a), that fuels one of the largest marine zones of seasonal coastal hypoxia in the world
88 (Rabalais et al. 2002, 2007; Bianchi et al. 2010; Bristow et al. 2015). Studying microbial
89 relationships to river eutrophication will improve our understanding of their contributions to
90 either mitigating or exacerbating nutrient input.

91 Far from a homogenous jumble of organisms ferried downriver, microbial community
92 composition changes with distance from the river mouth and/or from the influence of tributaries
93 (Lemke *et al.*, 2009; Winter *et a*l., 2007; Kolmakova *et al.*, 2014) (Read et al. 2015; Savio et al.
94 2015), resulting from altered nutrient concentrations (Staley et al. 2014b; Van Rossum et al.
95 2015; Meziti et al. 2016), differing dissolved organic matter (DOM) sources (Ruiz-González et
96 al. 2013; Blanchet et al. 2016), and land use changes (Staley et al. 2014a; Van Rossum et al.
97 2015; Zeglin 2015; Meziti et al. 2016). Past studies of the Thames, Danube, Yenisei, and
98 Columbia Rivers have found that planktonic river microbial assemblages were dominated by the
99 phyla Actinobacteria, Proteobacteria, and Bacteriodetes, and taxa such as acI Actinobacteria,
100 *Polynucleobacter* spp., GKS9 and LD28 *Betaproteobacteria*, CL500-29 Bacteriodetes, LD12
101 SAR11 *Alphaproteobacteria*, and *Novosphingobium* spp.(Zwart et al. 2002; Savio et al. 2015).
102 More recent 16S rRNA gene amplicon and metagenomic studies of the Minnesota portion of the
103 MSR corroborated previous research in other rivers that identified an increased proportion (Savio
104 et al. 2015) or richness (Read et al. 2015) of freshwater taxa with river distance, and an increased
105 abundance of "core" river microbes (Staley et al. 2013) with cumulative residence time (Savio *et*
106 *al.*, 2015, Winter *et al*., 2007; Read *et al.*, 2015; Staley *et al.*, 2013; Crump *et al*., 1996).

107 Researchers have suggested that these patterns supported the application of the River
108 Continuum Concept (RCC) (Vannote *et al*., 1980) to river microbiota. The RCC postulates that

as a river increases in size, the influences of riparian and other inputs will decrease as the river establishes a dominant core community. Richness will increase with stream order complexity before decreasing in higher order rivers (Vannote *et al.*, 1980). Therefore, as continuous systems with increasing volumes and residence times, river microbiota should transition from experiencing strong influences of mass effects from terrestrial and tributary sources to systems where species sorting plays a more important role (Fortunato et al. 2012; Besemer et al. 2013; Savio et al. 2015). Analysis of MSR microbial communities suggests the importance of point source inputs on local communities while regional scale differences are more influenced by mass immigration from sediment and tributaries (Staley et al. 2013, 2016; Payne et al. 2017).

Complicating matters, particle-associated communities in rivers (frequently defined as those found on filters of > ~3 μm) remain distinct from their free-living counterparts (Jackson *et al.*, 2014; Crump *et al.*, 1999; Riemann and Winding, 2001), potentially due to increased production rates from access to readily obtainable carbon (Crump *et al.*, 1999; Crump and Baross, 1996). Typical taxa associated with particles include OTUs related to the Bacteroidetes clades *Flavobacteria* and *Cytophaga*, Planctomycetes, *Rhizobium* spp., and *Methylophilaceae* spp. (Crump and Baross 1996; Allgaier and Grossart 2006; D'Ambrosio et al. 2014; Jackson et al. 2014). However, consistent trends in particle community composition are murky, with recent evidence suggesting organisms may switch between free-living and particle-associated lifestyles depending on substrate availability and chemical queues (Grossart 2010; D'Ambrosio et al. 2014). Thus, rivers constitute complex and highly dynamic ecosystems from a metacommunity perspective.

Although important insights have been gained from recent research on portions of the MSR (Staley et al. 2013, 2015, 2016; Payne et al. 2017), microbiological transects at the whole-

river or catchment scale have yet to be completed. This study aimed to *i)* compare size-fractionated MSR microbial community populations to those in other rivers, *ii)* examine within-river heterogeneity of microbial communities, and *iii)* identify MSR microorganisms most strongly associated with eutrophication- all at a near whole-river scale. During the fall of 2014, we completed the most extensive microbiological survey of the Mississippi River to date with a continual rowed transect over two months (70 days). Rowers from the adventure education non-profit OAR Northwest collected samples from Minneapolis, MN to the Gulf of Mexico (2918 km) (Fig. 1A). Our findings expand 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 continually 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). Most sampling occurred within the body of the river, although due to safety issues, three samples were collected from shore (Table S1). We collected duplicate samples at each site, but because

155 separate rowboat teams frequently collected these sometimes several dozen meters apart, they
156 cannot be considered true biological replicates and we have treated them as independent samples.
157 At each site, we filtered 120 mL of water sequentially through a 2.7 µm GF/D filter (Whatman
158 GE, New Jersey, USA) housed in a 25 mm polycarbonate holder (TISCH, Ohio, USA) followed
159 by a 0.2 µm Sterivex filter (EMD Millipore, Darmstadt, Germany) with a sterile 60 mL syringe
160 (BD, New Jersey, USA). We refer to fractions collected on the 2.7 µm and 0.22 µm filters as >
161 2.7 µm and 0.2-2.7 µm, respectively. Flow-through water from the first 60 mL was collected in
162 autoclaved acid-washed 60 mL polycarbonate bottles. Both filters were wrapped in parafilm, and
163 together with the filtrate, placed on ice in Yeti Roadie 20 coolers (Yeti, Austin, TX) until
164 shipment to LSU. Further, 9 mL of whole water for cell counts was added to sterile 15 mL
165 Falcon tubes containing 1 mL of formaldehyde and placed into the cooler. We monitored cooler
166 temperature with HOBO loggers (Onset, Bourne, MA) to ensure samples stayed at ≤ 4ºC. The
167 final cooler containing samples from sites P-Al had substantial ice-melt. Though our filters were
168 wrapped in parafilm, we processed melted cooler water alongside our other samples to control
169 for potential contamination in these filters. Given that some of our samples were expected to
170 contain low biomass, we also included duplicate process controls for kit contamination (Salter et
171 al. 2014; Weiss et al. 2014) with unused sterile filters. Flow-through 0.2 µm filtered water from
172 each collection was analyzed for SiO_4, PO_4^{3-}, NH_4^+, NO_3^-, and NO_2^- (µg/L) at the University of
173 Washington Marine Chemistry Laboratory
174 (http://www.ocean.washington.edu/story/Marine+Chemistry+Laboratory). Aboard-rowboat
175 measurements were taken for temperature and turbidity. We determined turbidity by deploying a
176 secchi disk (Wildco, Yulee, FL), while drifting with the current so the line hung vertically. It was
177 lowered until no longer visible, then raised until just visible, and measured for its distance below

178 the waterline. We then calculated secchi depth from the average of two measurements.
179 Temperature was measured with probes from US Water Systems (Indianapolis, IN), rinsed with
180 distilled water between samples. Samples for cell counts were filtered through a 2.7 µm GF/D
181 filter, stained with 1x Sybr Green (Lonza), and enumerated using the Guava EasyCyte
182 (Millipore) flow cytometer as previously described (Thrash et al. 2015).
183
184 *DNA extraction and Sequencing*
185 DNA was extracted from both filter fractions and controls using a MoBio PowerWater DNA kit
186 (MoBio Laboratories, Carlsbad, CA) following the manufacturer's protocol with one minor
187 modification: in a biosafety cabinet (The Baker Company, Stanford, ME), Sterivex filter
188 housings were cracked open using sterilized pliers and filters were then removed by cutting
189 along the edge of the plastic holder with a sterile razor blade before being placed into bead-
190 beating tubes. DNA was eluted with sterile MilliQ water, quantified using the Qubit2.0
191 Fluorometer (Life Technologies, Carlsbad, CA), and stored at -20° C. Bacterial and archaeal
192 sequences were amplified at the V4 region of the 16S rRNA gene using the 515f and 806r primer
193 set (Caporaso et al. 2012), and eukaryotic sequences from the V9 region of the 18S rRNA gene
194 using the 1391r and EukBR primer set (Amaral-Zettler et al. 2009). Amplicons were sequenced
195 on an Illumina MiSeq as paired-end 250 bp reads at Argonne National Laboratory. Sequencing
196 of the 16S and 18S rRNA gene amplicons resulted in 13,253,140 and 13,240,531 sequences,
197 respectively.
198
199 *Sequence Analysis*

200 We analyzed amplicon data with Mothur v.1.33.3 (Schloss et al. 2009) using the Silva v.119
201 database (Pruesse et al. 2007; Quast et al. 2013). Briefly, 16S and 18S rRNA gene sequences
202 were assembled into contigs and discarded if the contig had any ambiguous base pairs, possessed
203 repeats greater than 8 bp, or were greater than 253 or 184 bp in length, respectively. Contigs
204 were aligned using the Silva rRNA v.119 database, checked for chimeras using UCHIME (Edgar
205 et al. 2011), and classified also using the Silva rRNA v.119 database. Contigs classified as
206 chloroplast, eukaryotes, mitochondria, or "unknown;" or as chloroplast, bacteria, archaea,
207 mitochondria, or "unknown;" were removed from 16S or 18S rRNA gene data, respectively. The
208 remaining contigs were clustered using the cluster.split() command into operational taxonomic
209 units (OTUs) using a 0.03 dissimilarity threshold ($OTU_{0.03}$). After these steps, 146,725 and
210 131,352 OTUs remained for the 16S and 18S rRNA gene communities, respectively.
211
212 *Sample quality control*
213 To evaluate the potential for contamination from extraction kits, cooler water in the last set of
214 samples, or leaking/bursting of pre-filters, all samples were evaluated with hierarchical
215 clustering and NMDS analysis. Hierarchical clustering was preformed in R using the *hclust*
216 function with methods set to "average", from the vegan package (Oksanen et al. 2015). Samples
217 were removed from our analysis if they were observed to be outliers in both the NMDS and
218 hierarchical clustering such that they grouped with our process controls. The process and cooler
219 water controls were extreme outliers in both, as was sample L2 (Fig. S1, S2). Sterivex and
220 prefilter samples generally showed strong separation with the exception of three 16S rRNA gene
221 samples- STER X2, W2, S2 (Fig. S1, S2). The only other samples that were removed were due
222 to missing chemical data (Lake Itasca1-2, A1-2) or failed sequencing (16S STER Af1; 16S PRE

223 S2, X2; 18S PRE O1). Not including process or cooler water controls, 152 samples were
224 sequenced each for prokaryotic and eukaryotic communities. After these QC measures, 144 and
225 149 samples remained in the analyses from the 16S and 18S rRNA gene amplicons, respectively.
226 Further, to control for potential contaminants, any OTU with greater 20 reads in the process or
227 cooler controls was removed from the data set. 146,725 and 131,327 OTUs remained after these
228 steps for 16S and 18S rRNA gene communities, respectively.
229
230 *Alpha and Beta Diversity*
231 $OTU_{0.03}$ analyses were completed with the R statistical environment v.3.2.1(R Developement
232 Core Team 2015). Using the package PhyloSeq (McMurdie and Holmes 2013), alpha-diversity
233 was first calculated on the unfiltered OTUs using the "estimate richness" command within
234 PhyloSeq, which calculates Chao1 (McMurdie and Holmes 2013). After estimating chao1,
235 potentially erroneous rare OTUs, defined here as those without at least two sequences in 20% of
236 the data, were discarded. After this filter, the dataset contained 950 and 724 16S and 18S rRNA
237 gene OTUs, respectively. For site-specific community comparisons, OTU counts were
238 normalized using the package DESeq2 (Love et al. 2014) with a variance stabilizing
239 transformation (Learman et al. 2016). Beta-diversity between samples was examined using Bray-
240 Curtis distances via ordination with non-metric multidimensional scaling (NMDS). Analysis of
241 similarity (ANOSIM) was used to test for significant differences between groups of samples (e.g
242 lower versus upper MSR) using the *anosim* function in the vegan package (Oksanen et al. 2015).
243 The influence of environmental parameters on beta-diversity was calculated in R with the *envfit*
244 function.
245

246 *Network analyses and modeling*

247 To identify specific OTUs with strong relationships to environmental parameters (e.g. turbidity,
248 NO_3^-), we employed weighted gene co-expression network analysis (WGCNA) (Langfelder and
249 Horvath 2008) as previously described for OTU relative abundances (Guidi et al. 2016). First, a
250 similarity matrix of nodes (OTUs) was created based on pairwise Pearson correlations across
251 samples. This was transformed into an adjacency matrix by raising the similarity matrix to a soft
252 threshold power (p; p = 6 for 16S and 18S > 2.7 µm, p = 4 for 16S 0.2-2.7 µm, p = 9 for 18S 0.2-
253 2.7 µm) that ensured scale-free topology. Submodules of highly co-correlating OTUs were
254 defined with a topological overlap matrix and hierarchical clustering. Each submodule,
255 represented by an eigenvalue, was pairwise Pearson correlated to individual environmental
256 parameters (Figs. S7-10A). To explore the relationship of submodule structure to these
257 parameters, OTUs within the submodule were plotted using their individual correlation to the
258 parameter of interest (here nitrate or phosphate) and their submodule membership, defined as the
259 number of connections within the module (Figs. S7-10B, D). Strong correlations between
260 submodule structure and an environmental parameter facilitate identification of OTUs that are
261 highly correlated to that parameter. To evaluate the predictive relationship between a submodule
262 and a parameter, we employed partial least square regression (PLS) analysis. PLS maximizes the
263 covariance between two parameters (e.g., OTU abundance and nitrate concentration) to define
264 the degree to which the measured value (OTU abundance) can predict the response variable
265 (nutrient concentration). The PLS model was permutated 1000 times and Pearson correlations
266 were calculated between the response variable and leave-one-out cross-validation (LOOCV)
267 predicted values. Modeled values were then compared with measured values to determine the
268 explanatory power of the relationships (Figs. S7-10C, E). Relative contributions of individual

269 OTUs to the PLS regression were calculated using value of importance in the projection (VIP)
270 (Chong and Jun 2005) determination. PLS was run using the R package *pls* (Mevik and Wehrens
271 2007), while VIP was run using additional code found here:
272 http://mevik.net/work/software/VIP.R.
273
274 *Environmental Ontology*
275 Environmental ontology of individual 16S rRNA gene OTUs was determined using the SEQenv
276 (v1.2.4) pipeline (https://github.com/xapple/seqenv; Sinclair et al. 2016) as previously described
277 (Savio et al. 2015). Briefly, representative sequences of our OTUs were searched against the
278 NCBI nt database (updated on 07/01/2016) using BLAST and filtered for hits with a minimum of
279 99% identity. From each hit, a text query of the metadata was performed to identify terms
280 representing the sequence's environmental source. The text was mined for EnvO terms
281 (http://environmentontology.org/) and the frequency in which the terms appeared for each OTU
282 was recorded. Using the seq_to_names output provided by SEQenv, EnvO terms were formed
283 into six groups: Freshwater, Salt Water, Anthropogenic, Terrestrial, Sediment, and Unclassified
284 (Table S1). To be assigned a group, an OTU had to have the majority (> 50%) of its hits
285 classified to that term, while OTUs with equal distribution between two or more groups were
286 termed as Unclassified. OTUs that returned no significant hits to an EnvO term were assigned to
287 a seventh category, NA. OTUs and their corresponding relative abundances were merged based
288 on the assigned group and plotted with R.
289
290 *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), then trended downwards for the rest of the transect. Planktonic (< 2.7 μm) cell counts varied between 1 and $3x10^6$ cells·mL^{-1} in the upper MSR, and decreased to high 10^5 cells

314 ·mL^{-1} in the lower MSR (Fig. 1B). Water temperature ranged from 19°C (133km, Site E) to
315 11.7°C (2552 km, Site Ag), and river speed, excluding three sites sampled from shore, was
316 between 5.5 mph at Site Y and 0.4 mph (597 km, Site L) (Table S1). Spearman rank correlations
317 of the measured environmental parameters showed strong positive correlations between nitrate,
318 phosphate, distance, and increased turbidity; while nitrate and phosphate both strongly correlated
319 negatively to water temperature and river speed (Table S1).
320
321 *Bacterial and archaeal communities*
322 We observed a clear distinction between the 0.2-2.7 µm and > 2.7 µm 16S rRNA gene
323 communities (ANOSIM R = 0.65, P = 0.001) (Fig. S1A). Both size fractions had comparable
324 species richness and evenness that trended upwards downriver (Figs. S3A, B), although a peak
325 occurred for both fractions at sites O-Q (761-999 km) below the Des Moines River and above the
326 Illinois River (Fig. S3A, B). Both size fractions (NMDS stress = 0.14 for each) also showed a
327 significant separation between sites above and below the Missouri River confluence (ANOSIM,
328 > 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
329 fractions mirrored (below). Phosphate and turbidity had the highest correlation with the
330 separation between the upper and lower > 2.7 µm communities (r = 0.57, r= 0.54, respectively),
331 with water temperature and distance (r > 0.40) also contributing (Fig. 2A). At an OTU level, taxa
332 related to the acI clade (Actinobacteria) and unclassified *Bacillaceae* (r > 0.77, P = 0.001)
333 contributed most to the separation between the upper and lower > 2.7 communities, with OTUs
334 related to the *Bacillales, Gemmatimonadaceae, Peptococcaceae*, and *Micromonosporaceae*
335 clades also a factor (r > 0.70, P = 0.001) (Figure 2A, Table S1). For the 0.2-2.7 µm fraction,
336 distance and nitrate were the strongest correlating environmental factors with the distinction

337 between upper and lower communities (r = 0.59 and r = 0.47, respectively), although phosphate,
338 turbidity, and water temperature (r > 0.40 for each) also contributed (Fig. 2B). OTUs related to
339 *Flavobacterium* and an unclassified Bacterium (closest NCBI BLAST hit *Acidovorax* sp.,
340 KM047473), most strongly contributed to the separation between the 0.2-2.7 µm communities (r
341 > 0.52, P = 0.001). Other important OTUs driving the separation between upper and lower
342 communities belonged to the clades Bacteroidetes, *Microbacteriaceae*, *Clostridiales*, and
343 *Holophagaceae* (r > 0.49, P = 0.001) (Figure 2B, Table S1).
344 At the phylum level, Proteobacteria, Actinobacteria, and Bacteroidetes dominated
345 bacterial communities in both fractions (Figs. 3A and B) (Table S1). Proteobacteria in the > 2.7
346 µm fraction fluctuated widely in abundance (Fig. 3A), whereas their 0.2-2.7 µm counterparts
347 generally increased in relative abundance downriver (Fig. 3B). 0.2-2.7 µm Bacteroidetes and
348 Actinobacteria generally decreased in the upper river and stabilized in the lower river. These
349 phyla showed considerable abundance variation in the > 2.7 µm fraction. Cyanobacteria in the >
350 2.7 µm fraction negatively correlated with increased turbidity (Spearman rank = 0.67), consistent
351 with lower irradiance. Both > 2.7 µm and 0.2-2.7 µm Acidobacteria increased in abundance
352 downriver and positively correlated with river distance (Wilcoxon single ranked test, P = < 0.01)
353 (Fig. 3A, B). Within the 0.2-2.7 µm fraction, the five most abundant OTUs were classified as a
354 LD12 (OTU11), two acI clade OTUs (OTU4, OTU7), a *Limnohabitans* sp. (OTU2), and a LD28
355 (OTU8) (Table S1). Comparatively, an unclassified *Methylophilaceae* (OTU1), a *Planktothrix*
356 sp. (OTU12), a NS11-12 marine group (OTU21), an *Aquirestis* sp. (OTU17), and an unclassified
357 *Sphingobacteriales* (OTU25) were the most abundant OTUs in the > 2.7 µm fraction (Table S1).
358 Archaea occurred at much lower relative abundances: we found only eight OTUs belonging to
359 the phyla Euryarchaeota and Thaumarchaeota, collectively. Thaumarchaeota increased in

360 abundance along the transect more in the > 2.7 µm fraction compared to the 0.2-2.7 µm (Fig. 3A,
361 B). In both fractions, we only detected Euryarcheota at very low abundances. Importantly, the
362 primers used in this study may miss some archaeal taxa (Parada et al. 2015).
363
364 *Microbial eukaryotic communities*
365 Eukaryotic communities, observed via the 18S rRNA gene, also showed a significant separation
366 between > 2.7 µm and 0.2-2.7 µm fractions (ANOSIM R = 0.689, P = 0.001) (Fig. S1B). As
367 expected due to generally larger cell sizes in microbial eukaryotes compared to prokaryotes,
368 species richness remained higher in the > 2.7 µm vs. 0.2-2.7 µm fractions (Fig. S3C, D).
369 Richness in the > 2.7 µm fraction gradually increased downriver, similarly to prokaryotic
370 communities, but remained relatively stable within the 0.2-2.7 µm fraction. Both the > 2.7 µm
371 (stress = 0.113) and 0.2-2.7 µm (stress = 0.146) fractions also showed a significant separation
372 between the upper and lower MSR (ANOSIM, > 2.7 µm: R = 0.696, P = 0.001; 0.2-2.7 µm: R =
373 0.576, P = 0.001) (Fig. 2C, D). Distance and phosphate were the top two environmental factors
374 influencing this distinction (r = 0.75, r = 0.48; r = 0.70, r = 0.54, respectively) (Fig. 2C, D; Table
375 S1). No other factors had correlations > 0.40 (Table S1). At the OTU level, taxa related to an
376 unclassified Ochrophyta (OTU63) and an unclassified Eukaryote (OTU1) separated the 0.2-2.7
377 µm communities (r > 0.63, P= 0.001), while the same unclassified Eukaryote OTU (OTU1) and
378 a second unclassified Eukaryote (OTU222) contributed most to separating the > 2.7 µm
379 communities (r > 0.80, P = 0.001) (Figure 2C and D, Table S1).
380 Stramenopiles (or Heterokonts), encompassing diatoms and many other forms of algae,
381 and OTUs that could not be classified at the phylum level dominated both the > 2.7 µm and 0.2-
382 2.7 µm communities (Fig. 4). Stramenopiles accounted for over 25% of both communities, with

383 higher abundances in the upper vs. lower river. We observed a similar trend of disparate
384 abundances between the upper and lower river for > 2.7 µm Cryptomonadales and 0.2-2.7 µm
385 Nucletmycea, the latter of which include fungi (Fig. 4A; Table S1). Within the 0.2-2.7 µm
386 fraction, we identified the five most abundant OTUs as three unclassified *Bacillariophytina*
387 (OTU7, OTU14, OTU9), a *Pythium* sp. (OTU170), and an unclassified *Cryptomonas* (OTU11)
388 (Table S1). Comparatively, two unclassified *Eukaryotes* (OTU2 and OTU1), an unclassified
389 *Stramenopiles* (OTU3), an unclassified *Perkinsidae* (OTU13), and an unclassified
390 *Chrysophyceae* (OTU6) had the highest abundance in the > 2.7 µm fraction (Table S1).
391
392 *The Mississippi River Core Microbiome*
393 We defined the core microbiome as those OTUs detectable after normalization in greater than
394 90% of the samples. The 16S rRNA gene > 2.7 µm and 0.2-2.7 µm core microbiomes consisted
395 of 82 and 98 OTUs, respectively, classified into eight different phyla- Proteobacteria,
396 Actinobacteria, Bacteroidetes, Cyanobacteria, Verrucomicrobia, Chloroflexi, Chlorobi,
397 Gemmatimonadetes- and composed of taxa such as freshwater SAR11 (LD12), *Limnohabitans*
398 sp., *Polynucleobacter* sp., acI clade, LD28 clade, and *Planktothrix* sp. (Table S1). Core
399 microbome relative abundance in both fractions decreased along the upper river but stabilized in
400 the lower river (Fig. 5A). We confirmed this effect by analyzing the upper and lower core
401 microbiomes separately. Although the total OTU numbers changed (81 and 116 OTUs in the
402 upper MSR and 160 and 144 OTUs in the lower MSR for the > 2.7 µm and 0.2-2.7 µm fractions,
403 respectively), the trends remained the same (Fig. S4).
404 Eighty OTUs comprised the > 2.7 µm 18S rRNA gene core microbiome (Fig. 5B). We
405 classified these as *Alveolata*, *Cryptophceae*, *Nucletmycea*, *Stramenopiles*, or unclassified

406 Eurkaryota (Table S1). Again, consistent with larger organism sizes, and thus fewer OTUs
407 overall, the 0.2-2.7 µm Eukaryotic core microbiome comprised only 21 OTUs (Fig. 5B). These
408 OTUs consisted of *Alveolata*, *Nucletmycea*, *Stramenopiles*, or unclassified Eurkaryota (Table
409 S1). While the 0.2-2.7 µm core microbiome remained relatively stable along the river, the > 2.7
410 µm core decreased along the upper MSR before stabilizing in the lower river, similarly to that of
411 the prokaryotes (Fig. 5B).
412
413 *16S rRNA gene environmental ontology*
414 We successfully classified 136 of the 950 16S rRNA gene OTUs with EnvO terminology (Table
415 S1). Freshwater organisms dominated the OTU designations that we could assign, although their
416 relative abundance in both the > 2.7 µm and 0.2-2.7 µm fractions decreased along the upper river
417 before stabilizing in the lower MSR (Fig. S5A, B). However, LD12 (OTU11), the most abundant
418 OTU in our dataset and a well-established freshwater organism (Zwart et al. 2002; Newton et al.
419 2011; Salcher et al. 2011), did not receive an EnvO classification, demonstrating the limitations
420 of this technique with current database annotations. Terrestrial organisms from the > 2.7 µm
421 fraction decreased along the transect (Fig. S5A), while those in the 0.2-2.7 µm fraction remained
422 stable (Fig. S5B). Although representing a minor fraction of total OTUs, sediment-associated >
423 2.7 µm microorganisms fluctuated widely in abundance along the upper river before stabilizing
424 in the lower, whereas those in the 0.2-2.7 µm fraction occurred in similar abundances throughout
425 the river. Taxa associated with anthropogenic sources in both fractions increased along the river
426 (Fig. S5A-B).
427
428 *Network analyses identify taxa associated with and predictive of eutrophication*

429 We applied Weighted Gene Correlation Network Analysis (WGCNA) to identify co-occurring
430 groups of OTUs (submodules) that also had significant associations with the eutrophication
431 nutrients phosphate and nitrate. Of the submodules identified through WGCNA as being most
432 strongly correlated to phosphate and nitrate, we restrict our discussion to those modeled via PLS
433 analysis to predict > 50% of the measured nutrient concentration. Our additional PLS analyses
434 for those submodules predicting < 50% of nitrate and/or phosphate concentrations are included
435 in figure S6.

436 Three submodules in the prokaryotic and eukaryotic fractions were strongly associated
437 with phosphate. The 0.2-2.7 µm prokaryotic submodule most associated with phosphate was
438 composed of 51 OTUs (Table 1), and had moderate correlation between the submodule structure
439 and phosphate (Fig. S7D). Strong correlations between submodule structure and the measured
440 nutrient suggests that individual submodule OTUs that also have strong correlations to the
441 nutrient are the most important organisms associated with that nutrient (Langfelder and Horvath
442 2008). PLS modeling determined that this prokaryotic submodule predicted 80% of measured
443 phosphate concentrations (Table 1). Variable importance in the projection (VIP) analysis found
444 OTUs corresponding to an unclassified *Holophagaceae*, an unclassified *Gemmatimonadaceae*,
445 and an unclassified *Burkholderiaceae* were the three most important in the PLS model for
446 phosphate (Fig. 6A, Table S1). OTU322 (Acidobacteria subgroup 6), had moderate correlation to
447 phosphate but high node centrality (n = 51; Fig. 6A, Table S1), corroborating evidence that
448 freshwater sediment Acidobacteria subgroup 6 often occur in co-culture with
449 Alphaproteobacteria and may be metabolically connected (Kielak et al. 2016 and refs. within). A
450 submodule from the 0.2-2.7 µm eukaryotic fraction was also highly predictive of phosphate
451 (Table 1, Fig. S8D). We identified the top four VIP taxa as an unclassified *Peronosporomycetes*,

452 an unclassified *Ochrophyta*, an unclassified Eukaryote, and an unclassified *Stramenopiles* (Fig.
453 6F, Table S1). All four OTUs had Pearson correlations with phosphate greater than 0.60, among
454 which two were negative (Fig. 7A, Table S1). Unfortunately, the most highly interconnected
455 OTU within the submodule remained unclassified even at the Phylum level (Fig. 7A Table S1).
456 A > 2.7 µm eukaryotic submodule could predict 62% of measured phosphate variance (Table 1).
457 An unclassified Eukaryote and an unclassified *Peronosporomycetes* occupied top two VIP
458 positions (Fig. 7B, Table S1).
459 Only two submodules had strong associations with nitrate. The > 2.7 µm prokaryotic
460 submodule most strongly correlated with nitrate contained 133 OTUs (Table 1, Fig. S9A).
461 However, despite a low structural correlation to nitrate, the submodule strongly predicted
462 measured nitrate in the PLS model (Figure S9B, Table 1). The four highest VIP scoring OTUs
463 were an *Anabaena* sp., a *Flavobacterium* sp., an unclassified bacterium, and a member of the
464 *Sphingobacteriales* NS11-12 marine group (Fig. 6B, Table S1). OTUs with the highest node
465 centrality (> 20) belonged to the *Sphingomonadales* (*Alphaproteobacteria*) and
466 *Sphingobacteriales* (Bacteroidetes), and each positively correlated with nitrate (r > 0.48, Fig. 6B,
467 Table S1). The > 2.7 µm eukaryotic submodule most associated with nitrate predicted 57% of
468 observed variation in nitrate concentrations (Table 1). OTUs with the top VIP scores were two
469 unclassified *Chrysophyceae*, an unclassified *Ochrophyta*, and an unclassified Diatom (Fig. 7C,
470 S10; Table S1).
471
472 **Discussion**
473 Understanding microbial communities on a basin scale is essential for unveiling the complex
474 linkage between organisms, metabolism, and water quality. Our near whole river 2914 km

475 transect revealed that while the MSR comprised similar aquatic microbial taxa found within
476 other rivers (Ghai et al. 2011; Fortunato et al. 2013; Jackson et al. 2014; Read et al. 2015; Savio
477 et al. 2015; Meziti et al. 2016), the river hosted unique relative abundance and core microbiome
478 trends. Physio-chemical measurements demonstrated continual increases in nutrients and
479 sediment along the upper river, whereas in the lower river, nitrate, phosphate, nitrite, and
480 turbidity stayed more consistent (Fig. 1B). These distinct geographic patterns mirrored the
481 observed separation of two microbial regimes at the Missouri River confluence (Fig. 2). This
482 differed from the historical distinction of the upper and lower MSR at the Ohio River confluence
483 in Cairo, IL, but matched the changes in Strahler order from eight to ten (Pierson et al. 2008). In
484 general, both prokaryotic fractions and the > 2.7 µm fraction of eukaryotic communities
485 increased in richness downriver (Fig. S3). The relative abundance of core community taxa
486 decreased in the upper MSR before settling in the lower river (Fig. 5). Co-occurrence network
487 analyses identified taxa associated with, and predictive of, the eutrophication nutrients nitrate
488 and phosphate that provide important targets for future study and may assist in detecting and
489 quantifying imminent changes in river water quality.

490 Some qualities of the MSR matched other rivers. MSR microbial communities separated
491 into two distinct fractions of microbial assemblages, 0.2-2.7 µm and > 2.7 µm (ANOSIM, R =
492 0.65, P = 0.001), throughout the river transect (DeLong et al. 1993; Crump and Baross 1996;
493 Allgaier and Grossart 2006; D'Ambrosio et al. 2014; Jackson et al. 2014). These fractions were
494 composed of OTUs classified as LD12 (freshwater SAR11), *Limnohabitans* sp.
495 (Betaproteobacteria), LD28 (Betaproteobacteria), acI clade (Actinobacteria), and *Algoriphagu*s
496 sp. (Bacteriodetes). These taxa also dominated in other rivers (Ghai et al. 2011; Fortunato et al.
497 2013; Kolmakova et al. 2014; Read et al. 2015; Savio et al. 2015; Meziti et al. 2016).

498 However, despite the MSR generally sharing similar taxa with other rivers, abundance
499 patterns indicated a unique system. Phylum level abundances contrasted a previous study limited
500 to the Minnesota portion of the MSR (Staley et al. 2013), as well as studies of the Thames (Read
501 et al. 2015), Danube (Savio et al. 2015), and Columbia Rivers (Crump et al. 1999). Specifically,
502 within the 0.2-2.7μm MSR prokaryotic fraction, Proteobacteria remained the most abundant
503 phylum throughout the river, while Bacteroidetes and Actinobacteria decreased downriver (Fig.
504 3). In contrast, the other river studies identified Bacteroidetes-dominated headwaters that gave
505 way to more abundant Actinobacteria and Proteobacteria further downriver (Crump et al. 1999;
506 Staley et al. 2014a; Read et al. 2015; Savio et al. 2015). Although comparison beyond phylum
507 level is important for understanding community changes, most studies report only OTU
508 taxonomy making comparisons of OTU relative abundances difficult. OTUs belonging to the acI
509 and LD12 freshwater clades continually increased in abundance towards the river mouth in the
510 Thames and Danube Rivers (Read et al. 2015; Savio et al. 2015), whereas MSR OTUs classified
511 as acI (OTUs 3, 4 and 7) and LD12 (OTU 11) did not change in abundance throughout the
512 transect. Further, MSR OTU richness generally increased with distance, and core community
513 abundance decreased along the upper river in both fractions, contrasting predictions by the RCC
514 (Vannote *et al.*, 1980) and observations in previous studies from other rivers (Read et al. 2015;
515 Savio et al. 2015) where these trends were reversed. Importantly, we did not sample the true
516 headwaters of the MSR (Lake Itasca to above St. Cloud), and at the point of first sampling, the
517 MSR already constituted an eighth order river. Therefore, some of the trends predicted by the
518 RCC may have been missed. Ultimately, the variant observations between this MSR survey and
519 other river studies may result from different sampling methodologies or the particular timing of
520 sampling. However, these differences may also stem from biological signal related to unique

environmental conditions, human impacts, and changes in the level of river engineering with distance.

Within the MSR, we observed distinct prokaryotic and eukaryotic microbial community regimes separated at the Missouri River confluence. This separation matched physio-chemical changes observed for the MSR such as increased Strahler river order (8 to 10) and more consistent nutrient concentrations (Fig. 1). We hypothesize that the separation between the upper and lower MSR communities was driven by changes in the importance of immigration, emigration, and resource gradient dynamics. Specifically, our data suggests that mass effects play a role in structuring microbial communities in the upper MSR, although instead of only in the headwaters, this process continues for almost a third of the length of the river. Increased turbidity correlated with decreases in freshwater associated bacteria (Spearman rank correlation, > 2.7 μm $r = 0.55$; 0.2-2.7 μm $r = 0.51$) and the core microbiome relative abundance (Spearman rank correlation, > 2.7 μm $r = 0.53$; 0.2-2.7 μm $r = 0.63$) in the upper MSR. Similarly, nutrients like phosphate and nitrate continually increased along the upper river (Fig. 1). In contrast, these nutrients, turbidity, and core microbiome relative abundance all statbilized in the lower MSR. These patterns are consistent with communities under the influence of mass effects from tributaries in the upper MSR. However, once the MSR grew to a tenth order river, the large volume and size potentially buffered it from allochthonous influences, allowing species sorting effects to have greater influence over mass immigration. The lower river represents a more stable environment (e.g. nutrient concentrations) with its increased size and volume, contrasting the more variable upper MSR. Though river speed increased, the effective residence time also increased since taxa no longer experience rapidly changing environmental variables.

543 That we still observed increasing richness and variation between microbial communities
544 along the lower MSR might indicate the increased influence of environmental filtering
545 attributable to unmeasured bottom-up factors, such as the quality and quantity of DOM, or top-
546 down influences such as predation or viral lysis. An overall community shift from a mixture of
547 allochthonous members to a "native" population requires growth rates greater than residence
548 time over a given distance (Crump et al. 2004). The extent to which a local or regional event
549 impacts downriver populations would be dependent on the success of allochthonous taxa
550 associated with those events to establish within the autochthonous population. The lower river
551 provides ample opportunities for microbial community differentiation based on average
552 prokaryotic growth rates (Savio et al. 2015), especially among particle-associated (> 2.7 μm)
553 taxa (Crump et al. 1999). Thus, although the aggregate patterns in particular phyla and
554 taxonomic richness within the MSR differ from other systems, similar ecological processes may
555 still occur, but the relative proportion of the river whereby mass effects vs. species sorting
556 dominate fosters unique community dynamics.
557 Microorganisms most associated with eutrophication nutrients are important target taxa
558 for future study of nitrate and phosphate uptake and/or metabolizism, and also represent
559 plausible biological indicators of river trophic state. By focusing on OTUs with the highest VIP
560 scores in the PLS models, we identified several bacterial and eukaryotic taxa that fit these criteria
561 (Figs. 6, 7). Organisms in the *Sphingomonadaceae* (e.g., *Novosphingobium* spp.) contributed
562 strongly to the PLS models predicting nitrate with both > 2.7 μm (Fig. 6B) and 0.2-2.7 μm (Fig.
563 S6A) size fractions (Table S1). *Novosphingobium* spp. isolates have previously been associated
564 with eutrophic environments (Trusova and Gladyshev 2002; Zwart et al. 2002; Addison et al.
565 2007; Li et al. 2012) and some can reduce nitrate (Addison et al. 2007; Li et al. 2012), making

566 these specific OTUs candidates for nitrate metabolism in the river water column. An *Anabaena*
567 sp. (OTU40) from the core microbiome had the top VIP score within the > 2.7 µm submodule
568 that could predict 69% of nitrate concentrations, and correlated negatively with nitrate (Table 1,
569 Fig. 6B). The nitrogen-fixing *Anabaena* (Allen and Arnon 1955) typically bloom in low
570 dissolved inorganic nitrogen (DIN) conditions, making the absence of these consistent with high
571 DIN. An unclassified *Holophagaceae* OTU (OTU33) had the highest VIP score within the
572 prokaryotic 0.2-2.7 µm submodule that predicted 80% of the variance in phosphate
573 concentrations (Table 1, Fig. 6A). This same OTU was also a key driver of the beta diversity
574 separation between upper and lower river communities (Fig. 2B), and a member of the core
575 microbiome (Table S1). The *Holophagaceae* belong to the Acidobacteria phylum, and the Ohio
576 River has been reported to contain a much higher abundance of Acidobacteria relative to other
577 tributaries (Jackson et al. 2014). Notably, Acidobacteria increased with river distance in our
578 study as well (Fig. 3) with a peak in the 0.2-2.7 µm fraction around the Arkansas river. This
579 increase in "free-living" Acidobacteria downriver is unique among other whole river studies,
580 making these organisms, and the *Holophagaceae* OTUs in particular, potentially important
581 organisms for the MSR river basin specifically.
582 Within Eukaryotes, multiple different algae, diatom, and Oomycetes OTUs occupied
583 submodules highly predictive of nitrate and phosphate (Fig. 6A-C, Table S1), and specifically
584 *Chrysophyceae* taxa from both size fractions correlated strongly with nitrate (Fig. 7C, Fig. S6E).
585 *Chrysophyceae* (golden algae) commonly occupy river systems (Necchi Jr 2016) including the
586 MSR (Korajkic et al. 2015), can be autotrophic and mixotrophic (Jansson et al. 1996), and may
587 serve as predators of prokaryotes (Caron et al. 1990). Further, multiple OTUs classified as
588 *Peronosporomycetes* were important in predicting phosphate in both size fractions (Fig 7A and

589 B, Table S1). *Peronosporomycetes* are fungus-like eukaryotic organisms known to be pathogenic
590 in fish, plants, and mammals (Dick 2003; Islam and von Tiedemann 2011). While we also
591 identified many other eukaryotic OTUs as important predictors of nutrients, poor taxonomic
592 resolution hindered our ability to discuss them further. Improved cultivation and systematics of
593 key microbial eukaryotes will be vital to understanding river nutrient dynamics.

594 While the most geographically comprehensive analysis to date for the MSR, this study
595 only encompasses a snapshot in time. Seasonal changes that have been observed in the
596 Minnesota portion of the upper MSR and the Columbia River (Smith et al. 2010; Fortunato et al.
597 2013; Staley et al. 2015) undoubtedly influence this dynamic system. Future studies should
598 incorporate microbial responses, at a whole-river scale, to seasonal pulse events (e.g. rain, snow
599 melt) and how river size and volume may buffer local microbial communities from
600 allochthonous inputs (Zeglin 2015). Our current research highlights the distinctiveness of MSR
601 microbial communities and the complexities influencing their structure within the MSR
602 ecosystem. While similar to other riverine studies in some respects, the 2014 MSR microbes
603 differed in alpha diversity and core community membership compared to the other rivers. The
604 observed association between changes in Strahler's river order, nutrient dynamics, and
605 community composition indicates the importance of hydrology (Freimann et al. 2015; Zeglin
606 2015; Niño-García et al. 2016) on the spatial dynamics structuring microbial communities and
607 provides baseline information for future MSR studies that incorporate greater temporal and
608 spatial resolution. With water quality and river health of growing local and global importance
609 (Vorosmarty et al. 2010; Russell and Weller 2013), the determination of candidate taxa
610 predictive of eutrophic nutrients provides impetus for targeted research on their larger potential
611 in predicting, and potentially ameliorating, river quality.

Acknowledgements

This work was supported by the Department of Biological Sciences, College of Science, and the Office of Research and Economic Development at Louisiana State University; and the College of the Environment at 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. Gary King and Dr. Caroline Fortunato 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

635 nitrogenifigens sp. nov., a polyhydroxyalkanoate-accumulating diazotroph isolated from a
636 New Zealand pulp and paper wastewater. Int. J. Syst. Evol. Microbiol. **57**: 2467–2471.
637 Allen, M. B., and D. I. Arnon. 1955. Studies on nitrogen-fixing blue-green algae. I. Growth and
638 nitrogen fixation by Anabaena cylindrica Lemm. Plant Physiol. **30**: 366.
639 Allgaier, M., and H.-P. Grossart. 2006. Seasonal dynamics and phylogenetic diversity of free-
640 living and particle-associated bacterial communities in four lakes in northeastern Germany.
641 Aquat. Microb. Ecol. **45**: 115–128.
642 Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. A Method for
643 Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable
644 Regions of Small-Subunit Ribosomal RNA Genes. PLoS One **4**: e6372.
645 Battin, T. J., L. a. Kaplan, S. Findlay, C. S. Hopkinson, E. Marti, A. I. Packman, J. D. Newbold,
646 and F. Sabater. 2009. Biophysical controls on organic carbon fluxes in fluvial networks.
647 Nat. Geosci. **2**: 595–595.
648 Besemer, K., G. Singer, C. Quince, E. Bertuzzo, W. Sloan, and T. J. Battin. 2013. Headwaters
649 are critical reservoirs of microbial diversity for fluvial networks. Proc. Biol. Sci. **280**:
650 20131760.
651 Bianchi, T. S., S. F. DiMarco, J. H. Cowan, R. D. Hetland, P. Chapman, J. W. Day, and M. A.
652 Allison. 2010. The science of hypoxia in the northern Gulf of Mexico: A review. Sci. Total
653 Environ. **408**: 1471–1484.
654 Blanchet, M., O. Pringault, C. Panagiotopoulos, D. Lefèvre, B. Charrière, J.-F. Ghiglione, C.
655 Fernandez, F. L. Aparicio, C. Marrasé, P. Catala, L. Oriol, J. Caparros, and F. Joux. 2016.
656 When riverine dissolved organic matter (DOM) meets labile DOM in coastal waters:
657 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.

Christian Winter Gerhard Kavka, Robert L. Mach, and, T. H., and A. H. Farnleitner. 2007. Longitudinal Changes in the Bacterial Community Composition of the Danube River: a Whole-River Approach. AEM

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.

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., 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.

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 particulate materials on continental shelves influenced by large rivers: Plume processes. Cont. Shelf Res. **24**: 833–858.

Dagg, M. J., T. S. Bianchi, G. A. Breed, W.-J. Cai, S. Duan, H. Liu, B. A. McKee, R. T. Powell, and C. M. Stewart. 2005. Biogeochemical characteristics of the lower Mississippi River, USA, during June 2003. Estuaries **28**: 664–674.

DeLong, E. F., D. G. Franks, and A. Alldredge. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacteria assemblages. Limnol. Ocean. **38**: 924–934.

Dick, M. . 2003. Straminipilous Fungi: Systematics of the Peronosporomycetes Including Accounts of the Marine Straminipilous Protists, the Plasmodiophorids and Similar

Organisms. Mycopathologia **156**: 385–386.

Duan, S., R. T. Powell, and T. S. Bianchi. 2014. High frequency measurement of nitrate concentration in the Lower Mississippi River, USA. J. Hydrol. **519**: 376–386.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.

Ensign, S. H., and M. W. Doyle. 2006. Nutrient spiraling in streams and river networks. J. Geophys. Res. Biogeosciences **111**, doi:10.1029/2005JG000114

Fortunato, C. S., A. Eiler, L. Herfort, J. A. Needoba, T. D. Peterson, B. C. Crump, F. CS, A. Eiler, L. Herfort, N. JA, P. TD, and C. BC. 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. ISME J **7**: 1899–1911.

Fortunato, C. S., L. Herfort, P. Zuber, A. M. Baptista, and B. C. Crump. 2012. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. ISME J **6**: 554–563.

Freimann, R., H. Bürgmann, S. E. G. Findlay, and C. T. Robinson. 2015. Hydrologic linkages drive spatial structuring of bacterial assemblages and functioning in alpine floodplains TL - 6. Front. Microbiol. **6 VN-re**, doi:10.3389/fmicb.2015.01221

Van Geel, B., L. R. Mur, M. Ralska-Jasiewiczowa, and T. Goslar. 1994. Fossil akinetes of Aphanizomenon and Anabaena as indicators for medieval phosphate-eutrophication of Lake Gosciaz (Central Poland). Rev. Palaeobot. Palynol. **83**: 97–105.

Ghai, R., F. Rodŕíguez-Valera, K. D. McMahon, D. Toyama, R. Rinke, T. C. S. de Oliveira, J. W. Garcia, F. P. de Miranda, and F. Henrique-Silva. 2011. Metagenomics of the water column in the pristine upper course of the Amazon river. PLoS One **6**, doi:10.1371/journal.pone.0023785

727 Grossart, H. P. 2010. Ecological consequences of bacterioplankton lifestyles: Changes in
728 concepts are needed. Environ. Microbiol. Rep. **2**: 706–714.
729 Guidi, L., S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L.
730 Berline, and J. R. Brum. 2016. Plankton networks driving carbon export in the oligotrophic
731 ocean. Nature
732 Islam, M. T., and A. von Tiedemann. 2011. 2,4-Diacetylphloroglucinol suppresses
733 zoosporogenesis and impairs motility of Peronosporomycete zoospores. World J. Microbiol.
734 Biotechnol. **27**: 2071–2079.
735 Jackson, C. R., J. J. Millar, J. T. Payne, and C. a Ochs. 2014. Free-living and particle-associated
736 bacterioplankton in large rivers of the Mississippi River Basin demonstrate biogeographic
737 patterns. Appl. Environ. Microbiol. **80**: 7186–7195.
738 Jansson, M., P. Blomqvist, A. Jonsson, and A. BergstrÖm. 1996. Nutrient limitation of
739 bacterioplankton, autotrophic and mixotrophic phytoplankton, and heterotrophic
740 nanoflagellates in Lake Örträsket. Limnol. Oceanogr. **41**: 1552–1559.
741 Kielak, A. M., C. C. Barreto, G. A. Kowalchuk, J. A. van Veen, and E. E. Kuramae. 2016. The
742 Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front. Microbiol. **7**: 744.
743 Kolmakova, O. V., M. I. Gladyshev, A. S. Rozanov, S. E. Peltek, and M. Y. Trusova. 2014.
744 Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation
745 sequencing. FEMS Microbiol. Ecol. **89**: 442–450.
746 Korajkic, A., L. W. Parfrey, B. R. McMinn, Y. V. Baeza, W. VanTeuren, R. Knight, and O. C.
747 Shanks. 2015. Changes in bacterial and eukaryotic communities during sewage
748 decomposition in Mississippi river water TL - 69. Water Res. **69 VN-r**: 30–39.
749 Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted gene co-expression

750 network analysis. BMC Bioinformatics **9**: 559.

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

755 Lemke, M. J., E. K. Lienau, J. Rothe, T. A. Pagioro, J. Rosenfeld, and R. Desalle. 2009.
756 Description of freshwater bacterial assemblages from the upper Paran?? river floodpulse
757 system, Brazil. Microb. Ecol. **57**: 94–103.

758 Li, H.-F., Z. T. Zou, B. Z. Li, X. Wang, J. S. Yang, and H. L. Yuan. 2012. Novosphingobium
759 sediminis sp. nov., isolated from the sediment of a eutrophic lake. J. Gen. Appl. Microbiol.
760 **58**: 357–362.

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

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

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

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

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

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

River Algae. Springer.

Newton, R. J., and S. L. McLellan. 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front. Microbiol. **6**: 1–13.

Newton, R., S. Jones, A. Eiler, K. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol R **75**: 14–49.

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.

Pierson, S. M., B. J. Rosenbaum, L. D. McKay, and T. G. Dewald. 2008. Strahler Stream Order and Strahler Calculator Values in NHDPlus.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, F. O. Glockner, and F. O. Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**: 7188–7196.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner.

2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. **41**: 590–596.

Rabalais, N. N., R. E. Turner, B. K. Sen Gupta, D. F. Boesch, P. Chapman, and M. C. Murrell. 2007. Hypoxia in the northern Gulf of Mexico: Does the science support the Plan to Reduce, Mitigate, and Control Hypoxia? Estuaries and Coasts **30**: 753–772.

Rabalais, N. N., R. E. Turner, D. Justić, Q. Dortch, W. J. Wiseman, B. K. Sen Gupta, and D. Justic. 1996. Nutrient Changes in the Mississippi River and System Responses on the Adjacent Continental Shelf. Estuaries **19**: 386.

Rabalais, N. N., R. E. Turner, and W. J. Wiseman. 2002. Gulf of Mexico Hypoxia, a.K.a. "the Dead Zone." Annu. Rev. Ecol. Syst. **33**: 235–263.

R Developement Core Team. 2015. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. **1**: 409.

Read, D. S., H. S. Gweon, M. J. Bowes, L. K. Newbold, D. Field, M. J. Bailey, and R. I. Griffiths. 2015. Catchment-scale biogeography of riverine bacterioplankton. ISME J **9**: 516–526.

Richey, J. E., J. M. Melack, A. K. Aufdenkampe, V. M. Ballester, and L. L. Hess. 2002. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2. Nature **416**: 617–620.

Van Rossum, T., M. A. Peabody, M. I. Uyaguari-Diaz, K. I. Cronin, M. Chan, J. R. Slobodan, M. J. Nesbitt, C. A. Suttle, W. W. L. Hsiao, P. K. C. Tang, N. A. Prystajecky, and F. S. L. Brinkman. 2015. Year-long metagenomic study of river microbiomes across land use and water quality . Front. Microbiol. **6**: 1–15.

Ruiz-González, C., L. Proia, I. Ferrera, J. M. Gasol, and S. Sabater. 2013. Effects of large river

dam regulation on bacterioplankton community structure. FEMS Microbiol. Ecol. **84**: 316–331.

Russell, T. A., and L. Weller. 2013. State of the River Report: Water Quality and River Health in the Metro Mississippi River. Friends of the Mississippi River.

Salcher, M. M., J. Pernthaler, and T. Posch. 2011. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria "that rule the waves"(LD12). ISME J. **5**: 1242–1252.

Salter, S. J., M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J. Parkhill, N. J. Loman, and A. W. Walker. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. **12**: 87.

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.

Sinclair, L., U. Z. Ijaz, L. J. Jensen, M. J. L. Coolen, C. Gubry-Rangin, A. Chroňáková, A. Oulas, C. Pavloudi, J. Schnetzer, A. Weimann, A. Ijaz, A. Eiler, C. Quince, and E. Pafilis.

2016. <tt>Seqenv</tt> : linking sequences to environments through text mining. PeerJ **4**: e2690.

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

865 characterize the bacterial community of the Upper Mississippi River. J. Appl. Microbiol.
866 **115**: 1147–1158.
867 Thrash, J. C., J. L. Weckhorst, and D. M. Pitre. 2015. Cultivating Fastidious Microbes, p. 1–22.
868 *In* Protocols for Metagenomic Library Generation and Analysis in Petroleum Hydrocarbon
869 Microbe Systems. Humana Press.
870 Trusova, M. Y., and M. I. Gladyshev. 2002. Phylogenetic diversity of winter bacterioplankton of
871 eutrophic siberian reservoirs as revealed by 16S rRNA gene sequence. Microb. Ecol. **44**:
872 252–259.
873 Turner, R. E., and N. N. Rabalais. 2003. Linking landscape and water quality in the Mississippi
874 river basin for 200 years. Bioscience **53**: 563–572.
875 Turner, R. E., and N. N. Rabalais. 2004. Suspended sediment, C, N, P, and Si yields from the
876 Mississippi River Basin. Hydrobiologia **511**: 79–89.
877 Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. River
878 continuum concept. Can J Fish Aquat Sci **37**: 130–137.
879 Vorosmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S.
880 Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global
881 threats to human water security and river biodiversity. Nature **467**: 555–561.
882 Weiss, S., A. Amir, E. R. Hyde, J. L. Metcalf, S. J. Song, and R. Knight. 2014. Tracking down
883 the sources of experimental contamination in microbiome studies. Genome Biol. **15**: 564.
884 Withers, P. J. A., and H. P. Jarvie. 2008. Delivery and cycling of phosphorus in rivers: A review.
885 Sci. Total Environ. **400**: 379–395.
886 Wood, S. A., M. J. Prentice, K. Smith, and D. P. Hamilton. 2010. Low dissolved inorganic
887 nitrogen and increased heterocyte frequency: precursors to Anabaena planktonica blooms in

a temperate, eutrophic reservoir. J. Plankton Res. **32**: 1315–1325.

Zeglin, L. H. 2015. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. Front. Microbiol. **6**: 454.

Zwart, G., B. C. Crump, M. P. Kamst-van Agterveld, F. Hagen, and S. K. Han. 2002. Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat. Microb. Ecol. **28**: 141–155.

927 **Tables**

928

Table 1. 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; corr = 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; corr = 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; corr = 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; corr= 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; corr = 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; corr = 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; corr = 0.89, P = < 2e-16
> 2.7 µm	18S	PO_4^{3-}	Red	r = 0.60, P = 2e-08	39	R^2 = 0.618; corr = 0.799, P = < 2e-16

929

930

931 **Figure Legends**

932

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

939 **Figure 2.** Non-metric multidimensional scaling (NMDS) results for whole community with
940 correlations to environmental parameters (phosphate, nitrate, nitrite, ammonia, distance
941 (km), water temperature, turbidity (cm), river speed (mph)) and the top ten OTUs based on
942 significance (P) and strength of correlation (r). The four plots represent > 2.7 µm (A, C)
943 and 0.2-2.7 µm (B, D) fractions for the 16S (A, B) and 18S (C, D) rRNA gene
944 communities. Vector length is proportional to the strength of the correlation.

945 **Figure 3.** Relative abundance, by phylum, according to transect distance, for phyla accounting
946 for > 0.1% of the total reads for the 16S rRNA gene > 2.7 µm (A) and 0.2-2.7 µm (B)
947 communities. Non-linear regressions with 95% CI (gray shading) are provided for
948 reference.

949 **Figure 4.** Relative abundance, by phylum, according to transect distance, for phyla accounting
950 for > 0.1% of the total reads for the 18S rRNA gene > 2.7 µm (A) and 0.2-2.7 µm (B)
951 communities. Non-linear regressions with 95% CI (gray shading) are provided for
952 reference.

953 **Figure 5.** Core microbiome aggregate abundance for the 16S (A) and 18S (B) rRNA gene. In
954 each, triangles and circles points represent 0.2-2.7 µm and > 2.7 µm fractions,

955 respectively. Non-linear regressions with 95% confidence intervals (CI) (gray shading) are
956 provided for reference.

957 **Figure 6.** PLS results for the 16S rRNA gene community submodules most associated with
958 phosphate (A) and nitrate (B). OTU correlation with a given nutrient is indicated on the y-
959 axis according to the number of co-correlations (node centrality) on the x-axis.
960 Community fractions: 0.2-2.7 µm (A) and > 2.7 µm (B). Circle size is proportional to VIP
961 scores, with top 10 VIP scoring and top node centrality OTUs labeled with their highest-
962 resolution taxonomic classification and OTU number. Colors represent the taxonomic
963 classification the phylum level.

964 **Figure 7.** PLS results for the 18S rRNA gene community submodules most associated with
965 phosphate and nitrate. OTU correlation with a given nutrient is indicated on the y-axis
966 according to the number of co-correlations (node centrality) on the x-axis. Community
967 fractions: 0.2-2.7 µm (A) and > 2.7 µm (B,C). Circle size is proportional to VIP scores,
968 with top 10 VIP scoring and top node centrality OTUs labeled with their highest-
969 resolution taxonomic classification and OTU number. Colors represent the taxonomic
970 classification the phylum level.

971
972
973
974
975
976
977
978
979
980
981
982
983
984

985 **Supplemental Figure Legends**
986
987 **Figure S1.** NMDS results for the 16S (A) and 18S (B) rRNA gene communities. In each,
988 circles and triangles represent the > 2.7 µm and 0.2-2.7 µm fractions, respectively.
989 **Figure S2.** Hclust results for the 16S (A) and 18S (B) rRNA gene communities.
990 **Figure S3.** Richness and Evenness indexes for the 16S > 2.7 µm (A) and 0.2-2.7 µm (B) and
991 18S > 2.7 (C) and 0.2-2.7 µm (D) communities. Non-linear regressions with 95% CI (gray
992 shading) are provided for reference.
993 **Figure S4.** Separated upper and lower MSR core microbiome aggregate abundance for the 16S
994 rRNA gene communities. For the upper and lower river, the core microbiome was defined
995 separately requiring OTUs to have greater than one read in 90% of the samples. > 2.7 µm
996 (A) and 0.2-2.7 µm (B) 16S rRNA gene communities in the upper and lower MSR.
997 **Figure S5.** Relative abundance, by environmental ontology group, according to transect
998 distance for the 16S rRNA gene > 2.7 µm (A) and 0.2-2.7 µm (B) communities. Non-
999 linear regressions with 95% CI (gray shading) are provided for reference.
1000 **Figure S6.** PLS results for the 0.2-2.7 µm and > 2.7 µm 16S (A-D) and 18S (E-H) rRNA gene
1001 community for selected submodules with nitrate and phosphate and a VIP score > 1.
1002 Correlation of submodule OTUs to nitrate and phosphate according to the number of co-
1003 correlations (node centrality) for 0.2-2.7 µm (A,B, E,F) and > 2.7 µm (C,D,G,H) 16S and
1004 18S rRNA gene communities. Circle size is proportional to VIP scores, with top 10 VIP
1005 scoring and top node centrality OTUs labeled with their highest-resolution taxonomic
1006 classification and OTU number. Colors represent the taxonomic classification the phylum
1007 level.
1008 **Figure S7.** WCGNA results for 0.2-2.7 µm 16S rRNA gene community submodules of interest

1009 based on Pearson correlations to nitrate and phosphate (A, B, D) and the resultant PLS
1010 regression of the predicted versus measured nutrient concentrations (C, E). Linear
1011 regressions with 95% CI (gray shading) are provided for reference.

1012 **Figure S8.** WCGNA results for 0.2-2.7 µm 18S rRNA gene community submodules of interest
1013 based on Pearson correlations to nitrate and phosphate (A, B, D) and the resultant PLS
1014 regression of the predicted versus measured nutrient concentrations (C, E). Linear
1015 regressions with 95% CI (gray shading) are provided for reference.

1016 **Figure S9.** WCGNA results for > 2.7 µm 16S rRNA gene community submodules of interest
1017 based on Pearson correlations to nitrate and phosphate (A, B, D) and the resultant PLS
1018 regression of the predicted versus measured nutrient concentrations (C, E). Linear
1019 regressions with 95% CI (gray shading) are provided for reference.

1020 **Figure S10.** WCGNA results for > 2.7 µm 18S rRNA gene community submodules of interest
1021 based on Pearson correlations to nitrate and phosphate (A, B, D) and the resultant PLS
1022 regression of the predicted versus measured nutrient concentrations (C, E). Linear
1023 regressions with 95% CI (gray shading) are provided for reference.

1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034

Supplemental Tables

Supplemental Table S1 is a spreadsheet, TableS1.xlsx. This includes site data, NMDS correlations for environmental data, Relative abundances of the Core taxa, SeqENV data, 16S and 18S rRNA OTU tables, and VIP scores for all samples and datum.

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

Figure 1

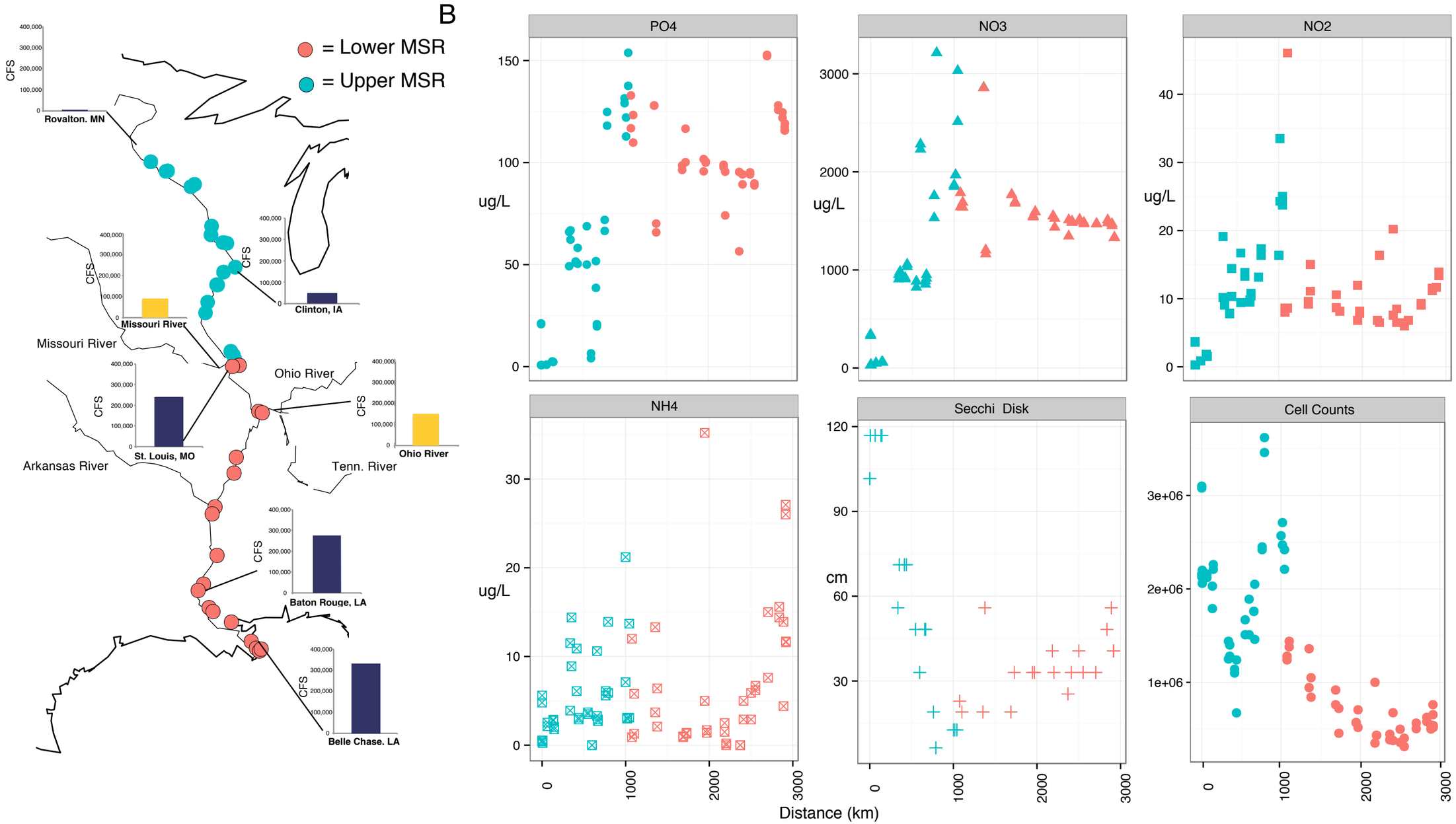
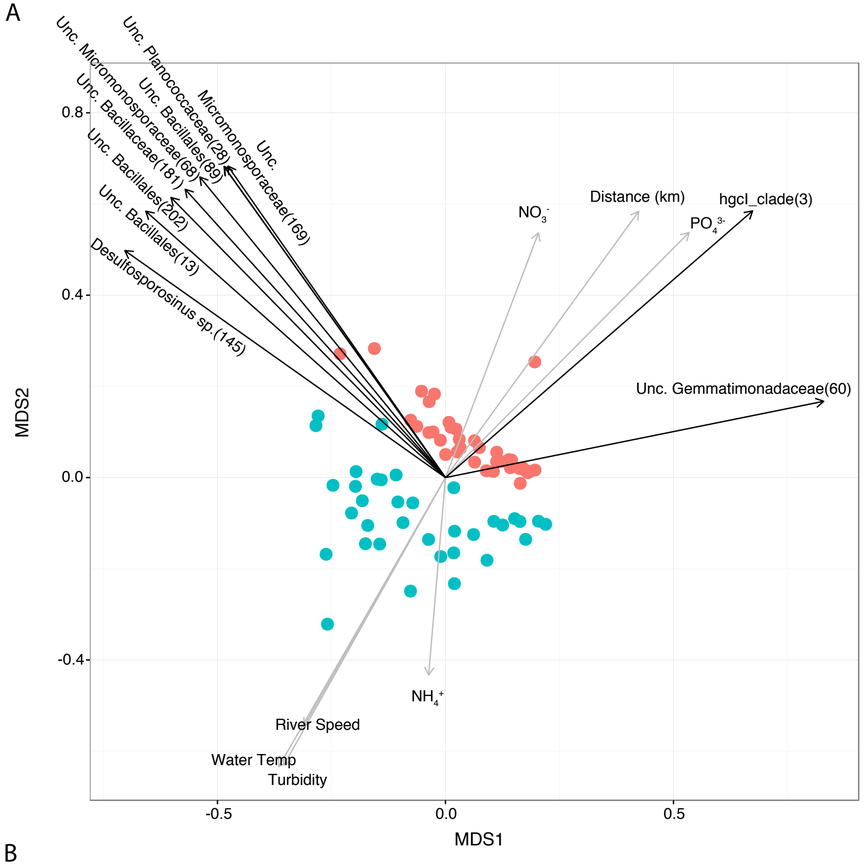
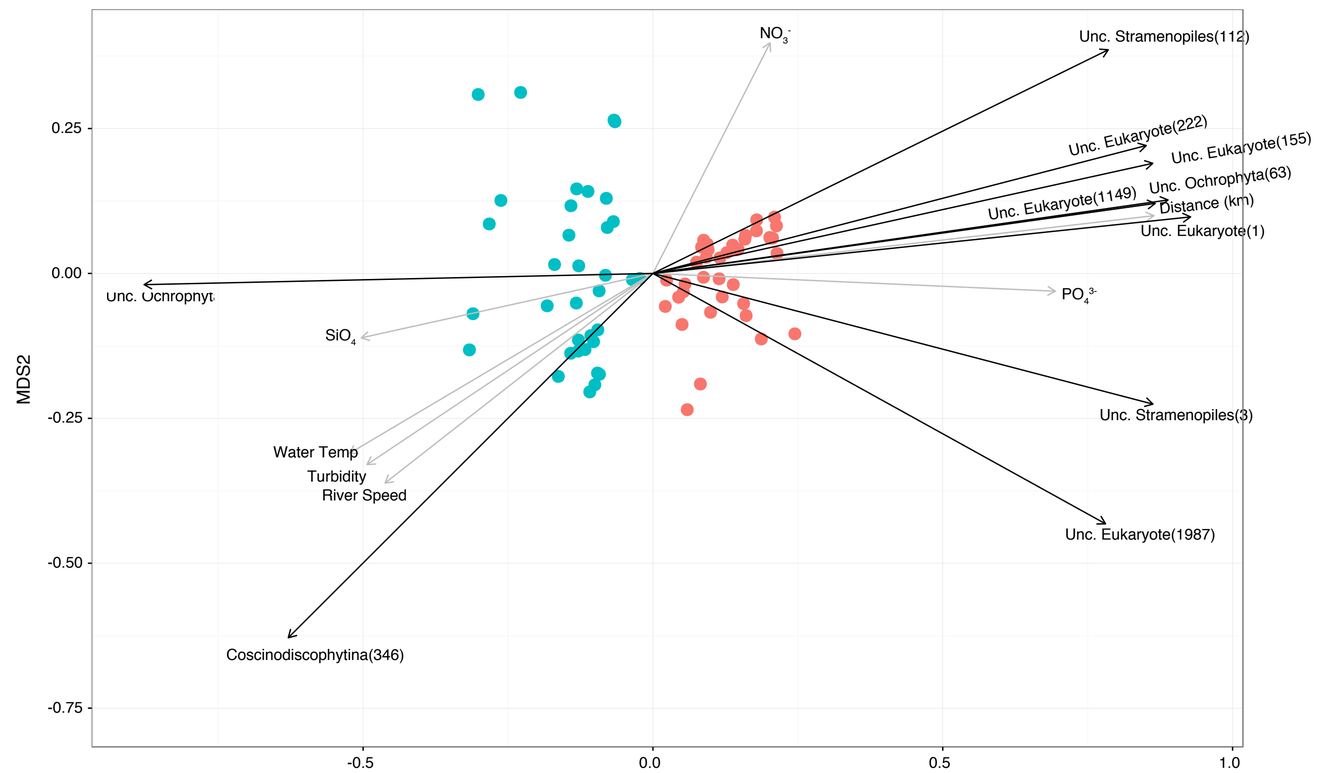

A
B
= Lower MSR
= Upper MSR
CFS
Rovalton. MN
Clinton, IA
Missouri River
Ohio River
St. Louis, MO
Arkansas River
Tenn. River
Baton Rouge, LA
Belle Chase. LA
PO4
NO3
NO2
NH4
Secchi Disk
Cell Counts
ug/L
cm
Distance (km)

Figure 2

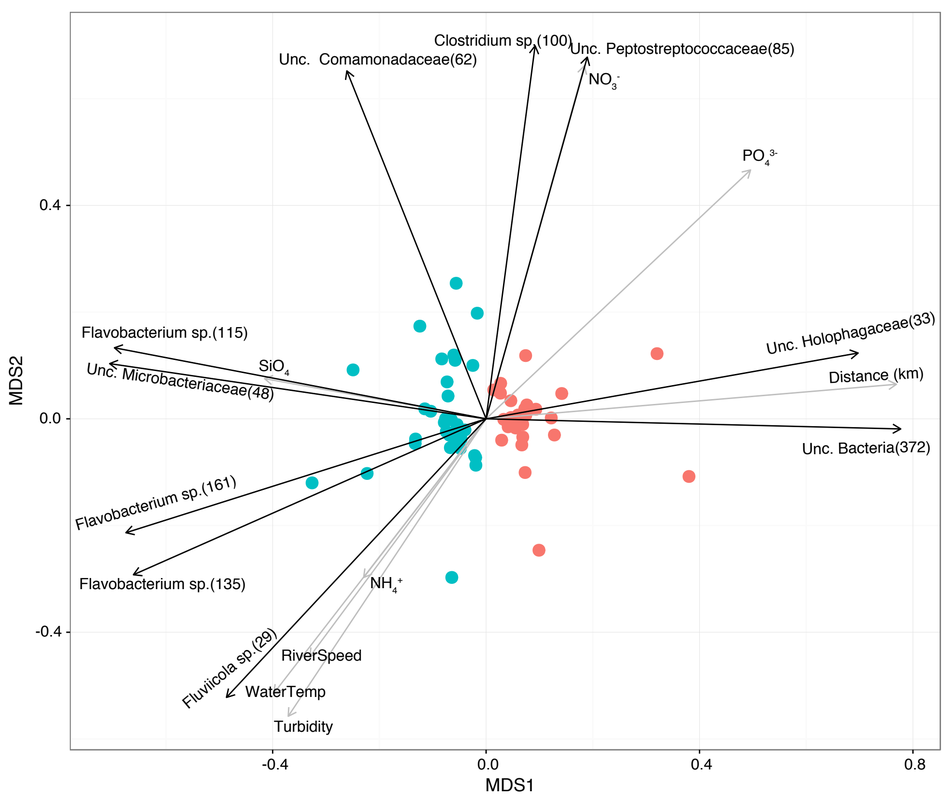
A



C



B



D

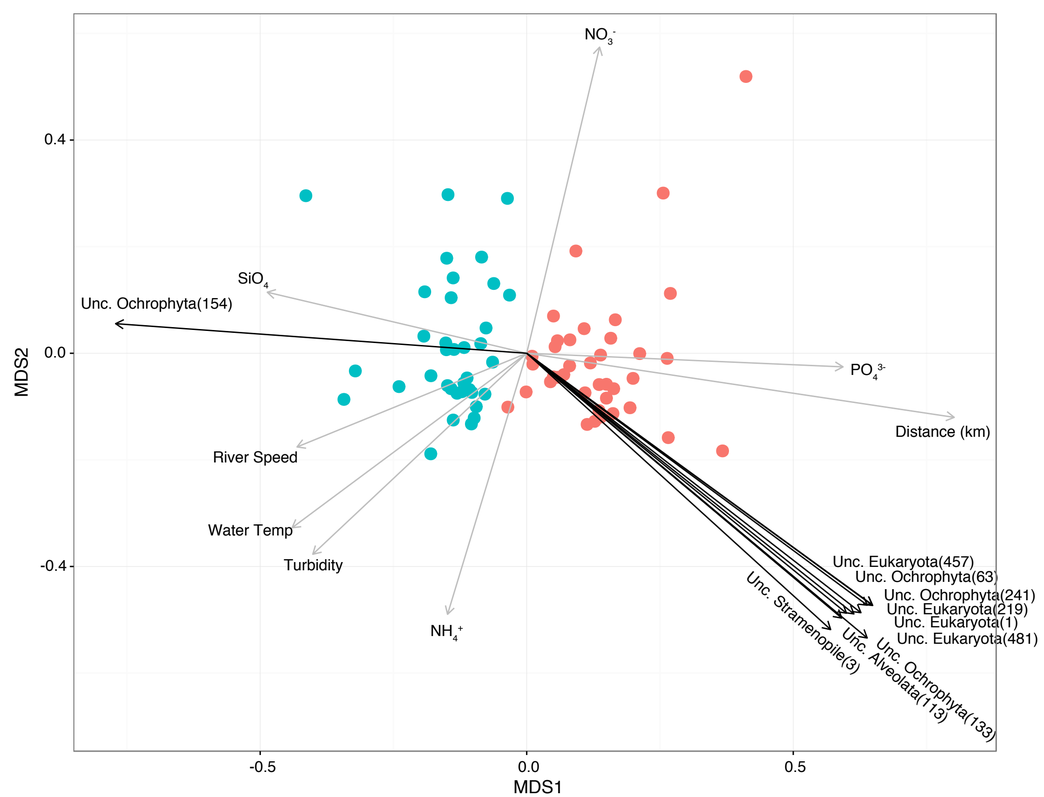
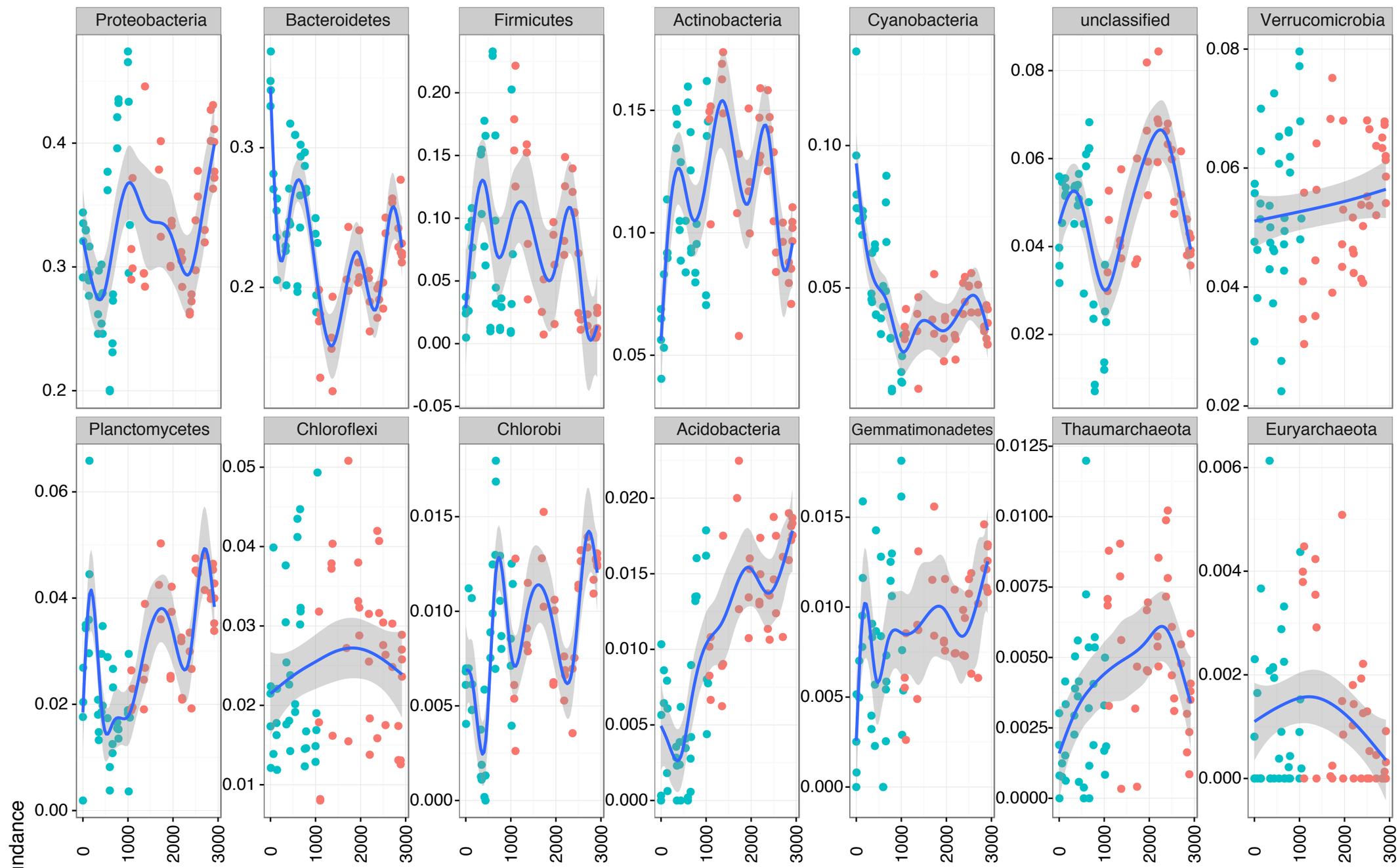


Figure 3 A


Proteobacteria
Bacteroidetes
Firmicutes
Actinobacteria
Cyanobacteria
unclassified
Verrucomicrobia
Planctomycetes
Chloroflexi
Chlorobi
Acidobacteria
Gemmatimonadetes
Thaumarchaeota
Euryarchaeota
0
1000
2000
3000

B

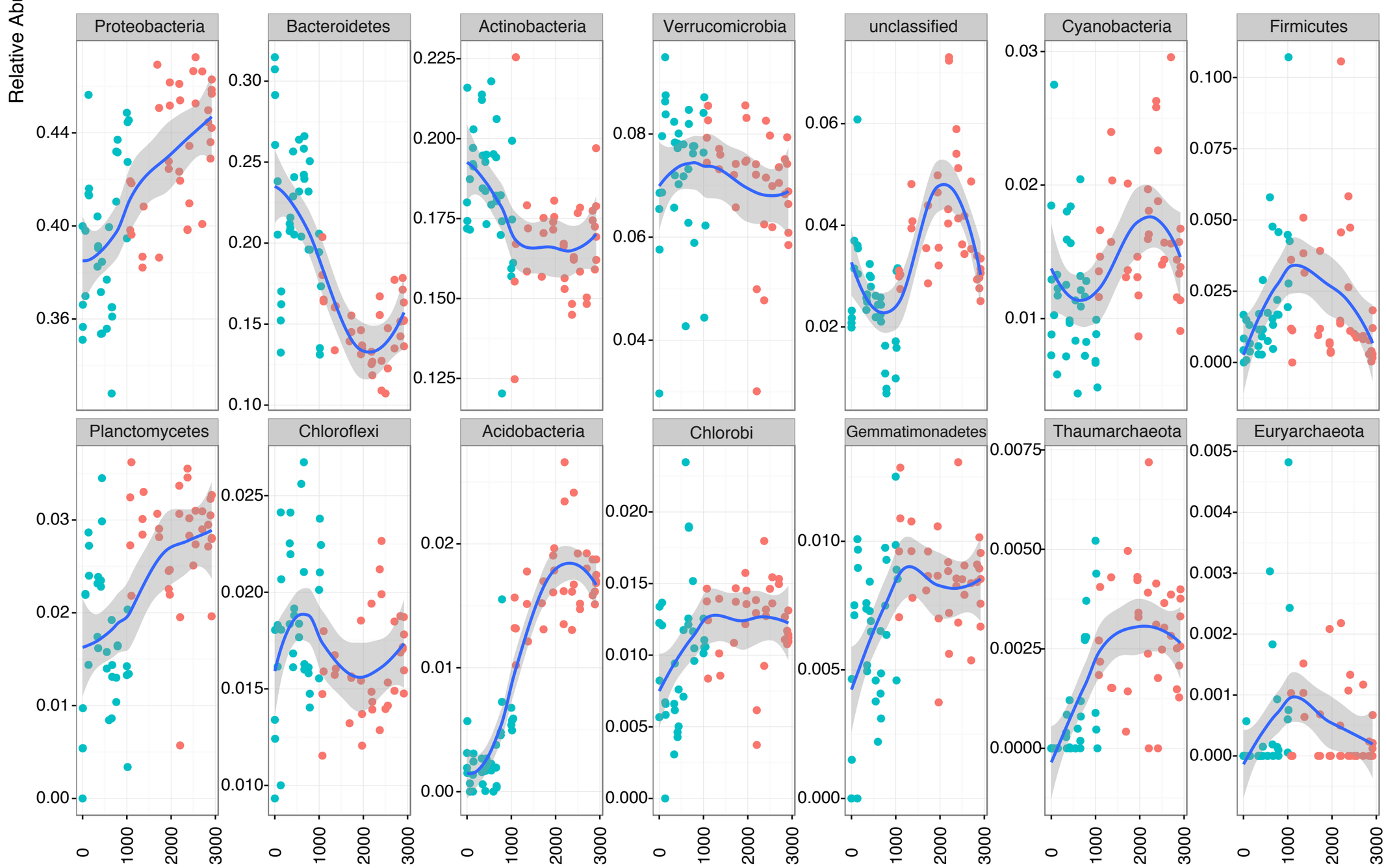

Relative Abundance
Proteobacteria
Bacteroidetes
Actinobacteria
Verrucomicrobia
unclassified
Cyanobacteria
Firmicutes
Planctomycetes
Chloroflexi
Acidobacteria
Chlorobi
Gemmatimonadetes
Thaumarchaeota
Euryarchaeota
0
1000
2000
3000

Figure 4 A

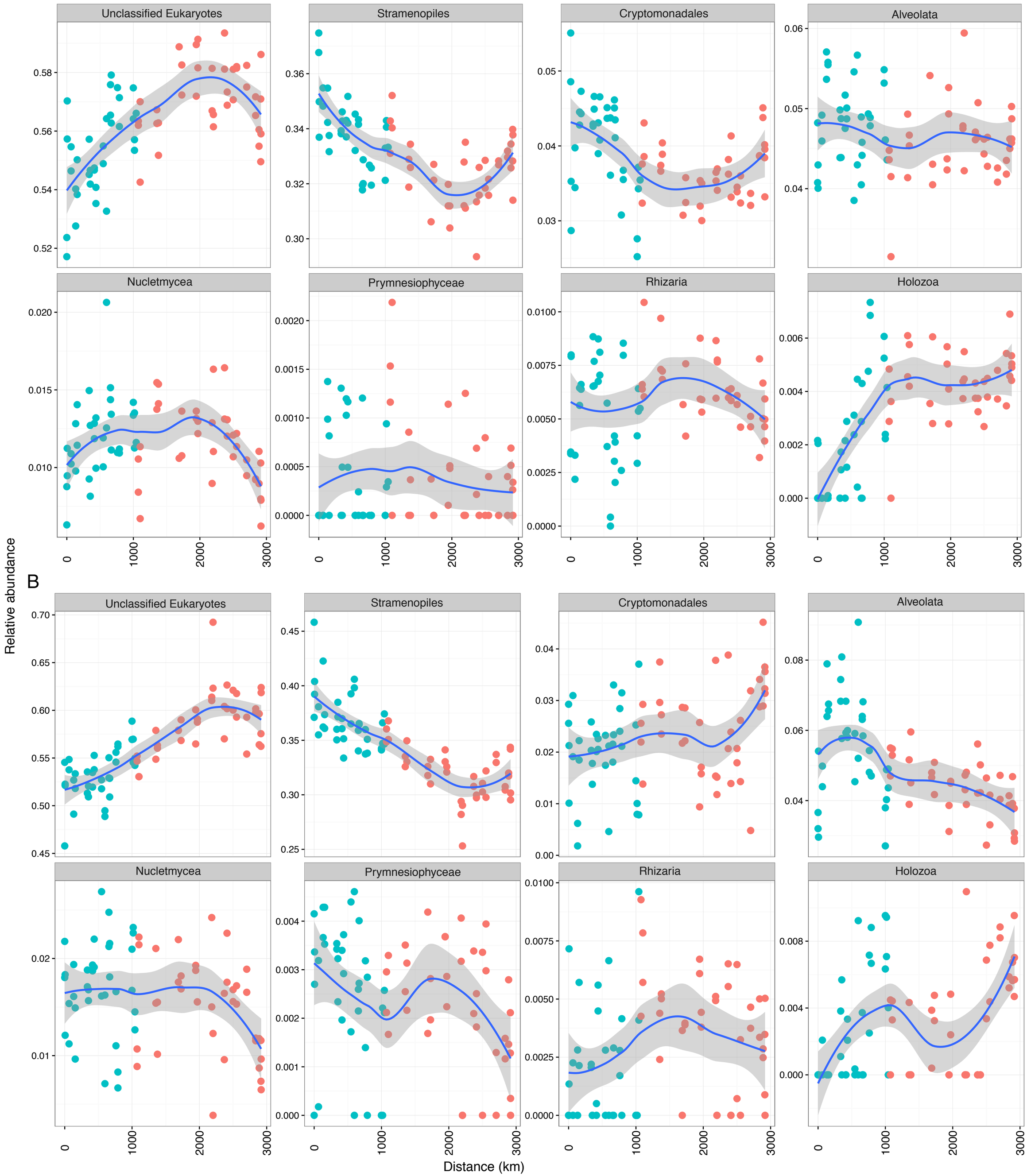
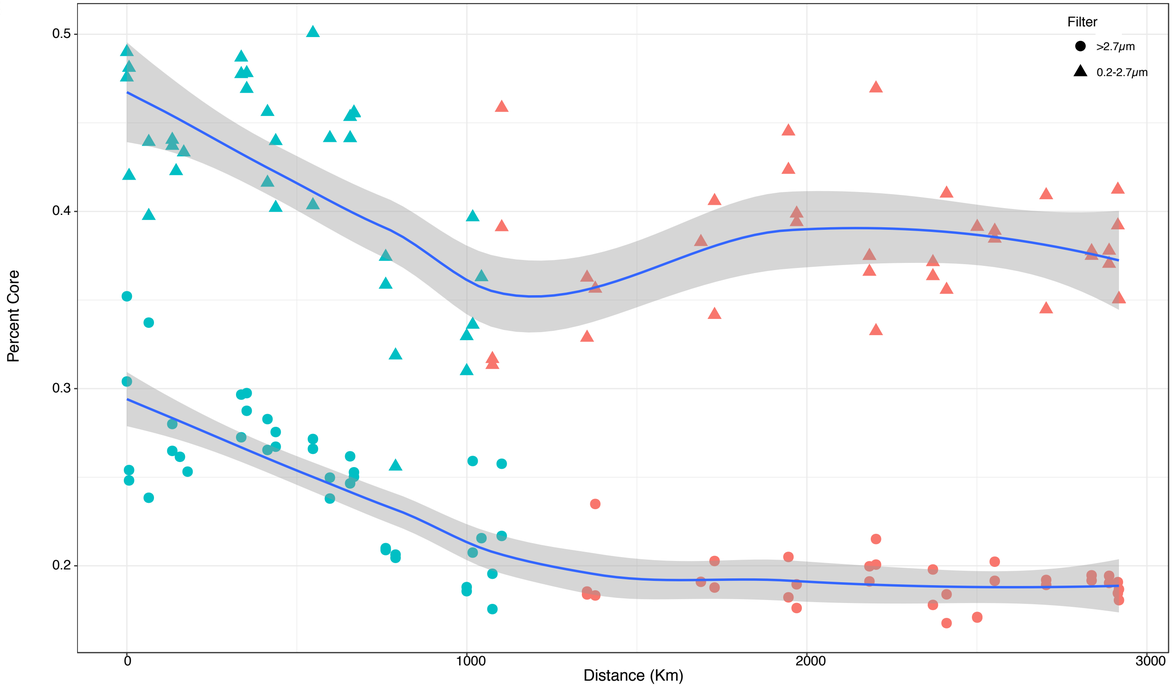
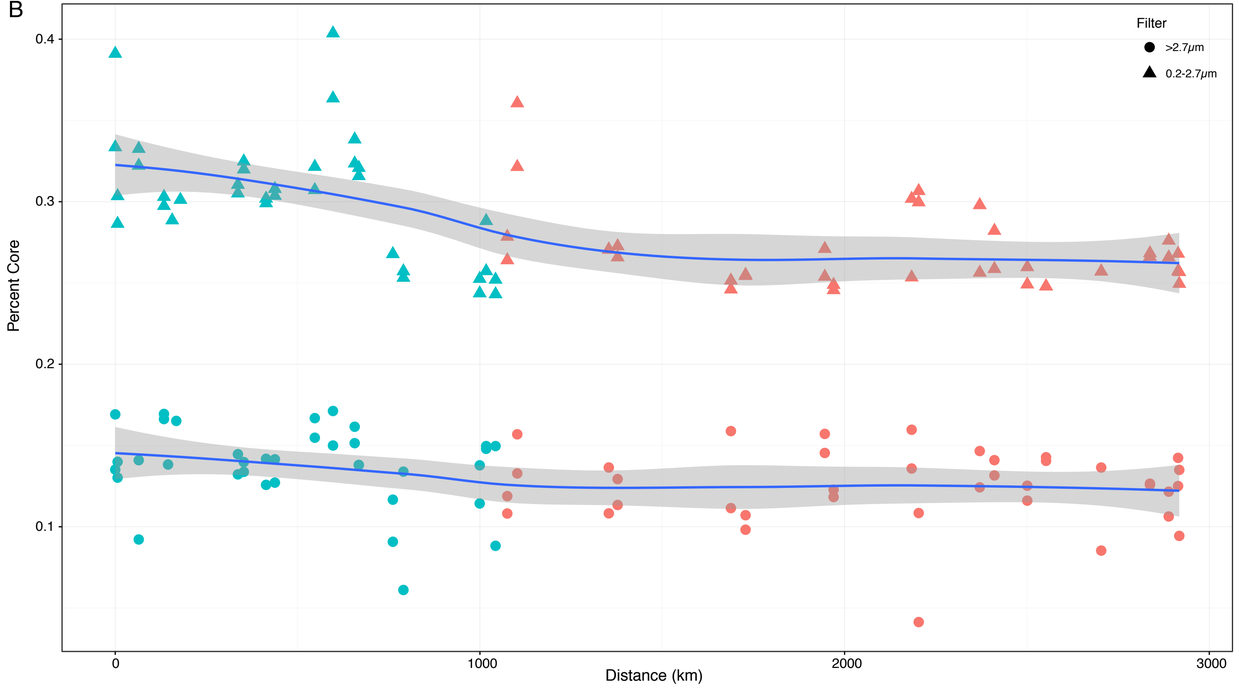
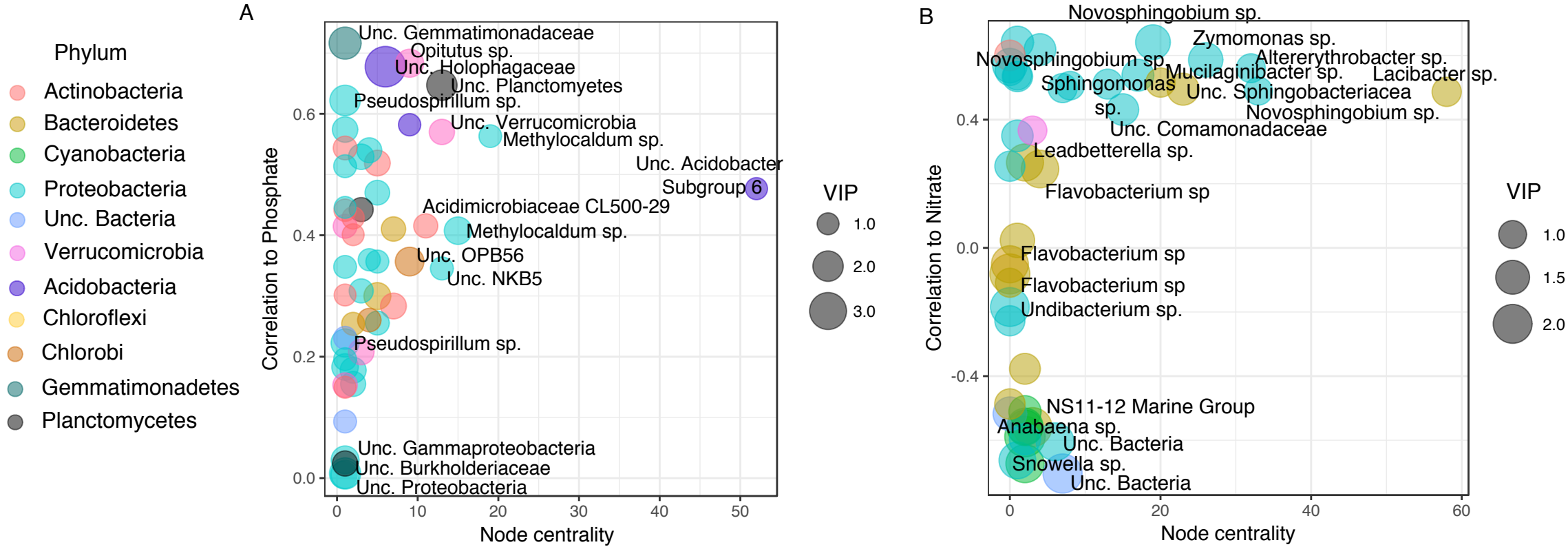

Unclassified Eukaryotes
Stramenopiles
Cryptomonadales
Alveolata
Nucletmycea
Prymnesiophyceae
Rhizaria
Holozoa
B
Relative abundance
Distance (km)

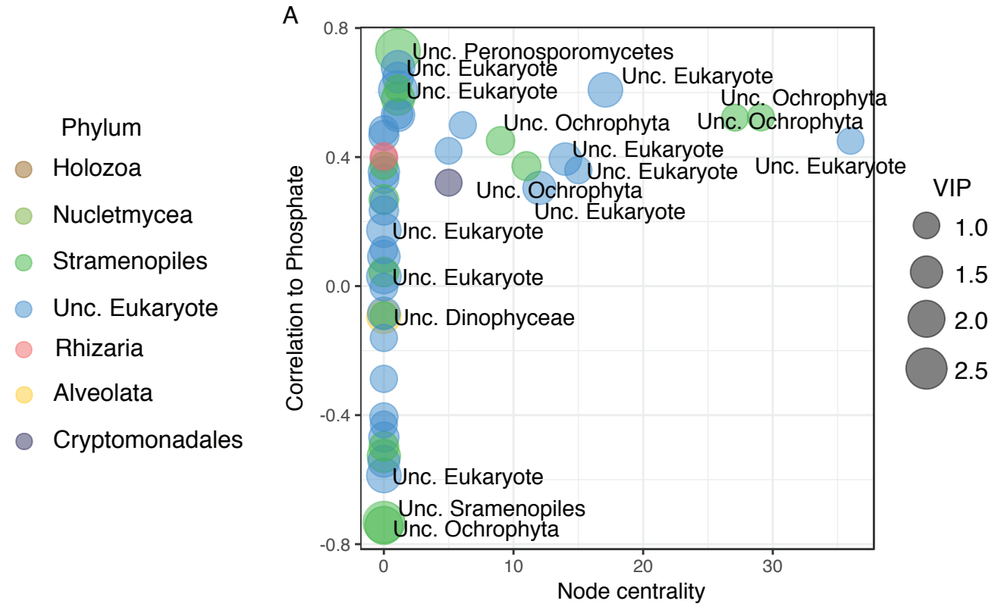
Figure 5

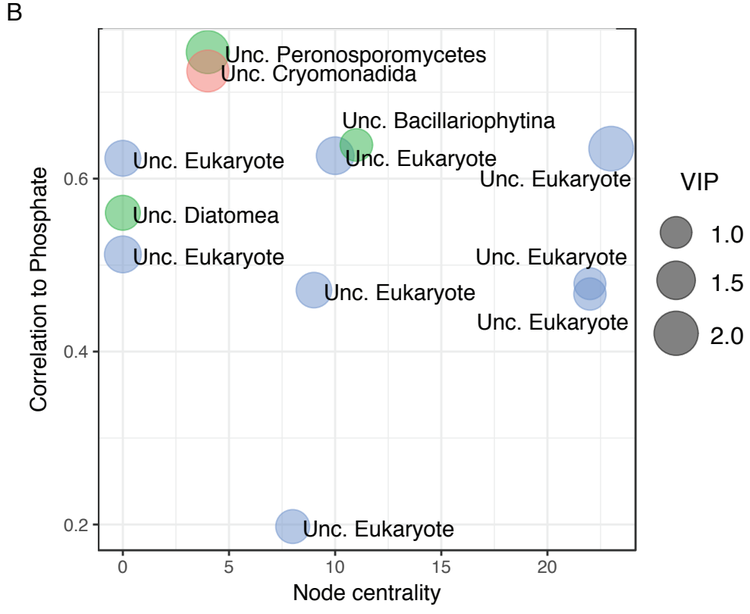

Filter
>2.7μm
0.2-2.7μm
Percent Core
0.5
0.4
0.3
0.2
0
1000
2000
3000
Distance (Km)

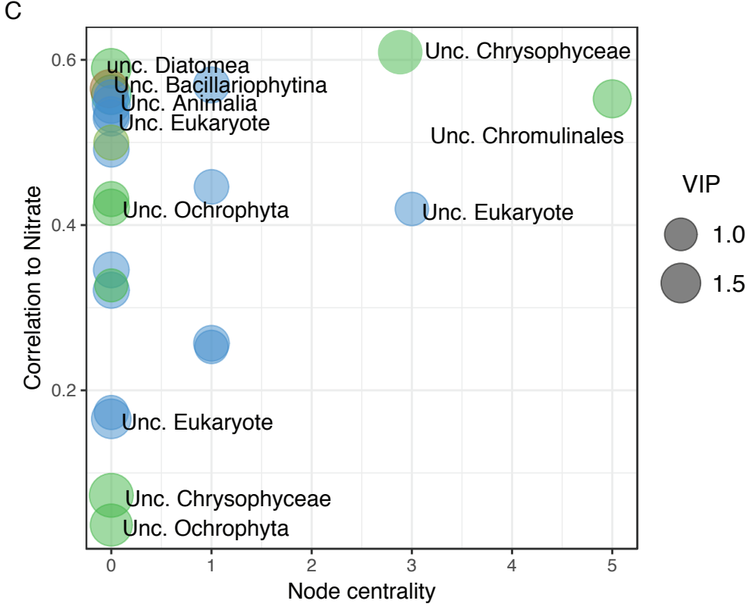
B


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


Phylum
Actinobacteria
Bacteroidetes
Cyanobacteria
Proteobacteria
Unc. Bacteria
Verrucomicrobia
Acidobacteria
Chloroflexi
Chlorobi
Gemmatimonadetes
Planctomycetes
A
Unc. Gemmatimonadaceae
Opitutus sp.
Unc. Holophagaceae
Unc. Planctomyetes
Pseudospirillum sp.
Unc. Verrucomicrobia
Methylocaldum sp.
Unc. Acidobacter
Subgroup 6
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
B
Novosphingobium sp.
Zymomonas sp.
Novosphingobium sp.
Altererythrobacter sp.
Mucilaginibacter sp.
Lacibacter sp.
Sphingomonas
Unc. Sphingobacteriacea
sp.
Novosphingobium sp.
Unc. Comamonadaceae
Leadbetterella sp.
Flavobacterium sp
Flavobacterium sp
Flavobacterium sp
Undibacterium sp.
NS11-12 Marine Group
Anabaena sp.
Unc. Bacteria
Snowella sp.
Unc. Bacteria
Correlation to Nitrate
Node centrality
VIP
1.0
1.5
2.0


Phylum
Holozoa
Nucletmycea
Stramenopiles
Unc. Eukaryote
Rhizaria
Alveolata
Cryptomonadales
A
Correlation to Phosphate
Node centrality
Unc. Peronosporomycetes
Unc. Eukaryote
Unc. Ochrophyta
Unc. Dinophyceae
Unc. Sramenopiles
VIP
1.0
1.5
2.0
2.5


B
Correlation to Phosphate
Node centrality
Unc. Peronosporomycetes
Unc. Cryomonadida
Unc. Bacillariophytina
Unc. Eukaryote
Unc. Diatomea
VIP
1.0
1.5
2.0


C
Correlation to Nitrate
Node centrality
unc. Diatomea
Unc. Bacillariophytina
Unc. Animalia
Unc. Eukaryote
Unc. Chrysophyceae
Unc. Chromulinales
Unc. Ochrophyta
VIP
1.0
1.5