bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # DNA metabarcoding potentially reveals multi-assemblage
2 # eutrophication responses in an eastern North American river

3

4 Joseph M. Craine[1], Matthew V. Cannon[2], Andrew J. Elmore[3], Steven M. Guinn[3], Noah
5 Fierer[1,4,5]

6 [1]Jonah Ventures, Manhattan KS 66502, USA.

7 [2] Institute for Genome Sciences, University of Maryland Baltimore, Baltimore, MD
8 21201, USA

9 [3] University of Maryland Center for Environmental Science, Appalachian Laboratory,
10 Frostburg, MD, 21532, USA.

11 [4]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO
12 80309, USA.

13 [5]Cooperative Institute for Research in Environmental Science, University of
14 Colorado, Boulder, CO 80309.

15

16

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17 # Abstract

18 Freshwater aquatic ecosystems provide a wide range of ecosystem services, yet provision
19 of these services is increasingly threatened by human activities. Directly quantifying
20 freshwater biotic assemblages has long been a proxy for assessing changing
21 environmental conditions, yet traditional aquatic biodiversity assessments are often time
22 consuming, expensive, and limited to only certain habitats and certain taxa. Sequencing
23 aquatic environmental DNA via metabarcoding has the potential to remedy these
24 deficiencies. Such an approach could be used to quantify changes in the relative
25 abundances of a broad suite of taxa along environmental gradients, providing data
26 comparable to that obtained using more traditional bioassessment approaches. To
27 determine the utility of metabarcoding for comprehensive aquatic biodiversity
28 assessments, we sampled aquatic environmental DNA at 25 sites that spanned the full
29 length of the Potomac River from its headwaters to the Potomac estuary. We measured
30 dissolved nutrient concentrations and also sequenced amplified marker genes using
31 primer pairs broadly targeting four taxonomic groups. The relative abundances of
32 bacteria, phytoplankton, invertebrate, and vertebrate taxa were distinctly patterned along
33 the river with significant differences in their abundances across headwaters, the main
34 river, and the estuary. Within the main river, changes in the abundances of these broad
35 taxonomic groups reflected either increasing river size or a higher degree of
36 eutrophication. The larger and more eutrophic regions of the river were defined by high
37 total dissolved phosphorus in the water, a unique suite of bacteria, phytoplankton such as
38 species of the diatom *Nitzschia*, invertebrates like the freshwater snail *Physella acuta*,
39 and high abundance of fish including the common carp (*Cyprinus carpio*). Taxonomic

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40 richness of phytoplankton and vertebrates increased downriver while it consistently
41 decreased for bacteria. Given these results, multi-assemblage aquatic environmental
42 DNA assessment of surface water quality is a viable tool for bioassessment. With
43 minimal sampling effort, we were able to construct the equivalent of a freshwater water
44 quality index, differentiate closely-related taxa, sample places where traditional
45 monitoring would be difficult, quantify species that are difficult to detect with traditional
46 techniques, and census taxa that are generally captured with more traditional
47 bioassessment approaches. To realize the full potential of aquatic environmental DNA for
48 bioassessment, research is still needed on primer development, a geographically broad set
49 of reference sites need to be characterized, and reference libraries need to be further
50 developed to improve taxonomic identification.

51

52 Keywords: metabarcoding; Potomac River; bioassessment; eutrophication; bacteria;
53 phytoplankton, macroinvertebrates; vertebrates; phosphorus; nitrogen

54

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

55

56 # Introduction

57 People rely on freshwater aquatic ecosystems for drinking water, recreation, fisheries,
58 and agriculture. Yet, the quality of the water and the integrity of aquatic communities
59 found in our lakes, rivers, and streams is increasingly threatened by human activities
60 including agriculture, roads, industry, mining, human waste, urbanization, and
61 deforestation [1, 2]. Poor water quality directly reduces quality of life and increases
62 economic costs while reducing economic output [3]. Effective monitoring of water
63 quality and the causes of water quality impairment is a critical step to maintaining our
64 freshwater resources, preventing further degradation, and guiding restoration efforts.

65 Although water quality can be measured directly, water quality can also be quantified
66 through bioassessment—the utilization of species abundances to indicate environmental
67 conditions [4, 5]. As opposed to direct measurements of environmental conditions,
68 bioassessment of water quality provides the benefits of a more robust indicator of water
69 characteristics and integrates over longer temporal and spatial scales than direct point
70 measurements [6]. As different species differ in their responses to physical, biological,
71 and chemical stresses and disturbances, bioassessment can indicate changes in a range of
72 water quality metrics, including nutrients, pollutants, pH, clarity/turbidity, or temperature.
73 Bioassessment uses the composition of biotic assemblages to infer stressors and
74 disturbances with the assumption that individual taxa respond uniquely to these factors
75 and the relative abundances of taxa can be used to infer the relative importance of
76 individual stressors or disturbances [7].

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

77 Bioassessment typically involves the direct collection of organisms with their abundances
78 quantified via visual inspection by trained taxonomists. For example, phytoplankton are
79 typically identified and counted under a microscope to estimate the biovolume of
80 different taxa in a given water sample [8]. Despite the widespread acceptance of these
81 traditional bioassessment approaches [9], traditional visual assessment of the relative
82 abundance of phytoplankton can be expensive, subject to observer bias that restricts
83 comparisons over time and across observers, and constrained by low taxonomic
84 resolution [10]. Macroinvertebrate assessment has similar constraints, but is further
85 constrained by generally being limited to hard-bottom wadable streams [11]. Fish
86 collection tends to be the most intensive sampling, is less effective for larger rivers than
87 smaller rivers, and is less useful for those species that reside at depth or do not float when
88 shocked [12, 13].

89 In contrast to traditional bioassessment, sequencing of aquatic environmental DNA
90 (eDNA) via metabarcoding provides an alternate approach to assess the relative
91 abundances of organisms in a given water body [14-16]. To accomplish this, DNA within
92 a water sample is purified either directly from the water or from filtered particulates.
93 Marker gene regions that are taxonomically informative (i.e. sufficiently variable in
94 nucleotide composition to discriminate between different groups of organisms) are then
95 amplified and sequenced, yielding information on the relative abundance of DNA from
96 different organisms. Compared to traditional bioassessment approaches, sequencing
97 aquatic eDNA is typically less expensive and often yields higher taxonomic resolution
98 than visual assessment. Plus, sampling for eDNA analysis is often logistically easier than
99 traditional assessment, can readily be performed in a wide range of different aquatic

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

100 environments, and can be used to quantify abundances of organisms not traditionally
101 censused. Although the eDNA approach is not bias-free and care must be taken when
102 interpreting the results [17], the limitations of eDNA analysis are potentially offset by its
103 advantages and the fact that the resulting data are not subject to observer bias, yielding
104 datasets that are more reliable and consistent over time and space.

105 The utility of using eDNA-based metabarcoding to quantify aquatic organisms in rivers
106 has already been demonstrated for some taxonomic groups [16, 18-21]. Despite the
107 potential of metabarcoding for bioassessment, the technique has still not been tested
108 extensively and we do not know whether multiple assemblages can simultaneously be
109 assessed with current primer sets to generate biotic indices of environmental conditions.
110 To examine the utility of aquatic eDNA metabarcoding for reconstruction of assemblages
111 and bioassessment of environmental conditions, we sampled water from sites distributed
112 along 475 km of the North Branch of the Potomac River from its headwaters to the
113 Potomac estuary below Washington D.C. over a 3-day period in April 2017. The
114 Potomac was chosen for assessment as it passes through a wide range of land uses from
115 forest to agricultural to urban. Portions of the Potomac River are also considered to have
116 experienced eutrophication due to agricultural and wastewater inputs, which in turn serve
117 as a source of excessive nutrients for the Chesapeake Bay [22]. The Potomac is also the
118 sole source of water for Washington D.C. and is an important recreational river for a
119 large population. Given what is known about the Potomac, we employed metabarcoding
120 to assess the relative abundances of bacteria, phytoplankton, invertebrates, and
121 vertebrates using four primer pairs. At each site, we also analyzed water samples for total
122 dissolved nitrogen and phosphorus as an independent estimate of nutrient availability in

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

123 the water. We then assessed the multivariate correlations among the relative abundances
124 of taxa and nutrient concentrations to test whether broad suites of taxa were responding
125 similarly to changes in environmental conditions.

126 # Methods

127 ## Site selection and sampling

128 Water samples were collected from 25 sites located on the North Branch or main stem of
129 the Potomac River. These sites start at Fairfax Stone and end in the Potomac estuary
130 downstream of Washington D.C., spanning a distance of 475 km. Among the 25 sites, 4
131 sites are considered headwaters (0-40 km), 9 sites were in the Upper Potomac (40-240
132 km), 9 sites were in the Lower Potomac (240-445 km), and 3 sites were in the Potomac
133 estuary (>445 km downstream).

134 Sampling occurred between April 19-21, 2017. At each site, water was drawn into a
135 sterile 60 mL syringe and then pushed through a Whatman Puradisc 25 mm 1µm nylon
136 syringe filter. This process was repeated until no more water could be pushed through the
137 syringe by the user. Across sites, an average of 252 mL was sampled, with a range of 120
138 to 500 mL of water sampled per site. The syringe filter was then placed into a 60 mL
139 specimen cup with silica gel desiccant and stored at either room temperature or at 4°C
140 until DNA was extracted.

141 ## DNA sequencing

142 DNA was extracted from filters with a MoBio PowerSoil DNA kit (MoBio Laboratories,
143 Carlsbad, CA) following the manufacturer's protocol. For the phytoplankton analyses, we

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

144 amplified a region of the 23S rRNA gene using PCR primers designed to amplify this
145 gene region from a broad range of phytoplankton taxa, including Cyanophyta
146 (cyanobacteria), Chlorophyta (green algae), and Bacillariophyta (diatoms) (Sherwood and
147 Presting, 2007). Both primers also contained a 5' adaptor sequence to allow for
148 subsequent indexing and Illumina sequencing. Each DNA sample was amplified in
149 triplicate reactions that were subsequently combined. These PCR reactions included 12.5
150 µL of Promega Mastermix, 0.5 µL of each primer, 1.0 µL of extracted DNA, and 10.5 µL
151 of DNase/RNase-free H_2O. The PCR reaction conditions consisted of an initial
152 denaturation step of 3 min at 94°C, followed by 40 cycles at 94°C (30 seconds), 55°C (45
153 seconds), and 72°C (60 seconds), followed by a final elongation step of 10 minutes at
154 72°C. Similar procedures were used for bacteria, invertebrates, and vertebrates (Table 1).

155 After PCR, the amplicons were visualized on a 2% agarose gel to visually confirm that
156 the PCRs yielded amplicons of the expected size. 20µl of the PCR amplicon was used for
157 PCR clean-up using ExoI/SAP reaction. To index the amplicons with a unique identifier
158 sequence, the first round of PCR was followed by an indexing 8-cycle PCR reaction to
159 attach 10-bp error-correcting barcodes unique to each sample to the pooled amplicons
160 from each site. These products were again visualized on a 2% agarose gel to check for
161 band intensity and amplicon size. PCR products were purified and normalized using the
162 Life Technologies SequalPrep Normalization kit and samples pooled together. Amplicons
163 were sequenced on an Illumina MiSeq at the University of Colorado Next-Generation
164 Sequencing Facility running the paired-end 2x250bp V2 sequencing chemistry.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

165 **Bioinformatic processing**

166 After de-multiplexing the reads, the paired-end reads were merged using fastq_merge
167 pairs [23]. Since merged reads often extended beyond the amplicon region of the
168 sequencing construct, we used fastx_clipper to trim primer and adapter regions from both
169 ends (https://github.com/agordon/fastx_toolkit). Sequences lacking a primer region on
170 both ends of the merged reads were discarded. Sequences were quality trimmed to have a
171 maximum expected number of errors per read of less than 0.1 and only sequences with
172 more than 3 identical replicates were included in downstream analyses. BLASTN 2.2.30+
173 was run locally, with a representative sequence for each operational taxonomic unit
174 (OTU) as the query and the current National Center for Biotechnology Information
175 (NCBI) nt nucleotide and taxonomy database as the reference. The tabular BLAST hit
176 tables for each OTU representative were then parsed so only hits with > 97% query
177 coverage and identity were kept. Similar procedures were used for the other primer pairs
178 (Table 1).

179 The 23S rRNA sequences were clustered into OTUs at the ≥ 97% sequence similarity
180 level and sequence abundance counts for each OTU were determined using the usearch7
181 approach. The NCBI genus names associated with each hit were used to populate the
182 OTU taxonomy assignment lists. Sequences that did not match over 90% of the query
183 length and did not have at least 85% identity were considered unclassified, otherwise the
184 top BLASTn hit was used for taxonomy assignment. Similar procedures were used for
185 the other primer pairs (Table 1).

186

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

187 For the 23S rRNA gene analyses (phytoplankton), we removed all OTUs that were
188 identified as higher plants or uncultured organisms. The average number of remaining
189 reads per sample was 12454. For the 12S rRNA gene analyses (vertebrates), we removed
190 all taxa except those assigned to Chordata. The average number of remaining reads for
191 12S rRNA was 1096 per sample. For the COI gene analyses (macroinvertebrates), we
192 removed all OTUs except those assigned to Annelida, Arthropoda, Cnidaria, and
193 Mollusca. After removing all other taxa, the average number of remaining COI reads per
194 samples was 125. The reason there were so few reads was that >95% of the reads were
195 from Oomycota (3% were from Rotifera). The COI gene from oomycetes, i.e. water
196 molds, is amplified with the primers we used and oomycetes happen to be highly
197 abundant in the Potomac. As our target taxa were macroinvertebrates, we had made the
198 decision before analyzing the data to exclude all taxa that were not macroinvertebrates
199 and did not include oomycetes in our analyses here.

200 Taxonomic identifications were constrained by the availability of sequences in our
201 reference databases. As such, some OTUs were likely assigned to species that were
202 closely related, but not identical, to those present in the Potomac. For example, one OTU
203 was assigned to *Cottus szanaga*, which is only found in Asia. No other *Cottus* species
204 were identified. More than likely, the DNA was derived from a different *Cottus* species
205 present in the Potomac that has not been sequenced yet such as *Cottus caeruleomentum*
206 or *Cottus girardi*. In all cases, we refer to OTUs based on the species to which they were
207 matched despite these limitations.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

208 **Nutrient analyses**

209 At each site, 250 mL of filtered water was retained in a sterile scintillation vial and kept
210 cold and in the dark. In the laboratory, the water was re-filtered with a 0.45 μm filter and
211 frozen at -20°C for analysis within 28 d. Thawed water was analyzed for total dissolved
212 phosphorus and nitrogen at the University of Maryland Center for Environmental Science,
213 Appalachian Laboratory's Water Chemistry Analytical Lab using offline persulfate
214 digestion followed by colorimetric analysis for orthophosphate and nitrate+nitrite on a
215 Lachat QuikChem 8000 Flow Injection Analyzer.

216 **Statistical analyses**

217 With 3-5 lab replicates for each sample, all data for the lab replicates were summed for
218 each sample. With multiple OTUs often assigned to the same taxon, all data on the
219 number of reads was summed for vertebrate, macroinvertebrate, and phytoplankton data
220 based on taxonomic identity. Bacterial 16S rRNA data were retained at the OTU level.

221 To estimate taxonomic richness, the number of reads for each sample was rarefied to a set
222 number of randomly selected reads per sample to control for differences in sequencing
223 depth. 16S rRNA data were rarefied to 22698 reads per sample, 23S rRNA data were
224 rarefied to 6564 reads, 12S rRNA data to 762 reads. Rarefied richness was not calculated
225 for COI data due to there being too few reads per samples with these primers.

226 To assess the general patterns of the abundances of taxa with equal weighting among the
227 major taxonomic groups, we ran principal components analyses (PCA) with the top 30
228 taxa for each primer pair. We also ran a single PCA that included the top 30 taxa for each

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

229 primer pair together as well as rarefied richness for taxa identified with 16S rRNA, 23S
230 rRNA, and 12S rRNA primer pairs.

231 Sørensen's index of similarity was calculated for all pairs of samples for each of four
232 groups of taxa using the *betadiver* command of *vegan* package in R. Similarity indices
233 between sampling points were compared with hydrologic distances along the river with a
234 Mantel test from the *vegan* package [24]. P values for the index of similarity were
235 calculated as 1-p where p is the likelihood of a randomization permutation resulting in a
236 matrix becoming more dissimilar with hydrologic distance. For each taxon, we also
237 calculated the average distance each taxon was found along the river by calculating the
238 average distance of all samples weighted by the relative read abundance of that taxon.

239 All analyses were conducted in R version 3.3.2 except for the PCAs, which were
240 conducted in JMP 13.0.0 (SAS Institute Inc., Cary, NC, USA).

241 **Results**

242 **Taxa-level patterns**

243 For bacteria sampled across the Potomac, 52% of the reads were assigned to
244 Proteobacteria, 22% to Bacteroidetes, 14% to Actinobacteria, and 6% to Verrucomicrobia.
245 Proteobacteria averaged 60% of the reads through 300 km and then declined in the Lower
246 Potomac to as low as 30%. In contrast, Bacteriodetes, Actinobacteria, and
247 Verrucomicrobia all increased in relative abundance with increasing distance downstream
248 ($P < 0.001$ for all), peaking just before the estuary. Individual bacterial OTUs were
249 uniquely distributed in ways that mirrored phylum-level patterns, but not always. For

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

250 example, the most abundant bacterial OTU (OTU2, Proteobacteria, Sphingomonadales)
251 was found on average 275 km downstream, while another Proteobacteria (OTU 25,
252 Hyphomonadaceae) was found on average 114 km downstream. Average rarefied
253 richness declined with distance downstream ($P < 0.001$), declining at a rate of 2.66 ± 0.49
254 OTUs km^{-1}.

255 Across the phytoplankton dataset, 78% of the reads were assigned to Bacillariophyta,
256 3.4% to Eustigmatophyta, 2.0% to Cyanobacteria, and 1.8% to Chlorophyta. Diatom
257 (Bacillariophyta) read abundance was ~20% for the first 10 km and then increased in
258 abundance with distance downriver, dominating the rest of the river and estuary,
259 representing 85% of the reads, on average, after 50 km (Figure 2). Examining the
260 abundance-weighted distances of all diatom reads, diatoms were most abundant 252 km
261 downstream. In contrast, eustigmatophytes dominated reads from the headwaters, with a
262 given eustigmatophyte located on average at just 59 km downstream (Figure 2).

263 Individual phytoplankton taxa revealed distinct patterning along the Potomac. For
264 example, the most abundant phytoplankton taxon on average was the diatom *Navicula*
265 *salinicola,* and was most abundant in the upper estuary (Figure 3). On average, it was
266 located 367 km downriver. In contrast, the diatom Nitzschia sp. [BOLD:AAX5147] was
267 the third most abundant phytoplankton and more abundant in Upper Potomac, found on
268 average 148 km downriver (Figure 3). Observed phytoplankton species of general interest
269 included *Didymosphenia geminate,* a.k.a. “rock snot”, which can form large mats in
270 nutrient-poor waters and can bloom to nuisance levels. It was found on average 103 km
271 downriver where it accounted for >10% of the phytoplankton sequences, with the highest
272 relative abundance in the middle of the sampled reach. Cyanobacteria species of the

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

273 genus *Synechococcus* were more abundant in the lower Potomac, but were relatively rare
274 (< 5% of the reads). Rarefied richness of phytoplankton taxa increased with increasing
275 distance and dropped dramatically in the estuary (Figure 4).

276 For the macroinvertebrates identified via COI sequencing, 87% of the reads were
277 assigned to Arthropoda, 7% were to Mollusca, and 6% to Annelida. Of the reads assigned
278 to Arthropoda, 61% were Diptera, 20% Ephemeroptera, and 5% Plecoptera. Like the
279 phytoplankton, invertebrate taxa were differentially abundant along the Potomac. For
280 example, the most abundant invertebrate taxa, the Diptera *Cricotopus trifascia*, was
281 found towards the middle region of the Potomac, while the second most abundant
282 invertebrate taxa, the mollusk *Physella acuta* was found closest to the mouth of the
283 Potomac (Figure 3).

284 Of the 80 vertebrate taxa identified, 39% of the reads were Mammalia, including
285 sequences identified to humans, mice, pigs, cattle, white-tailed deer, beaver, raccoon,
286 bank voles, and muskrat. 56% of the reads were Actinopterygii including sculpin (*Cottus*),
287 darters (*Etheostoma*), shad (*Alosa*), sunfish (*Lepomis*), suckers (*Catostomus*), trout
288 (*Salmo* and *Oncorhynchus*), and carp (*Cyprinus carpio*). 5% of the reads were assigned to
289 Amphibia and Reptilia, including the American toad (*Anaxyrus americanus*), the
290 American bullfrog (*Rana catesbeiana*), snapping turtle (*Chelydra serpentina*), and the
291 northern dusky salamander (*Desmognathus fuscus*). Human DNA was the most abundant
292 vertebrate DNA recovered from the water samples, averaging 21.7% of the 12S rRNA
293 gene reads with human DNA generally most abundant in the upper Potomac (Figure 3).
294 The second most abundant vertebrate taxon identified was *Alosa aestivalis*, but likely
295 represented multiple *Alosa* species, if not related species. Approximately 10% of the

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

296 reads were assigned to *Alosa aestivalis* and this taxon was dominant in the estuarine
297 samples ($P < 0.001$; Figure 3). When all samples were rarefied to 762 reads, vertebrate
298 diversity averaged 11.5 species and peaked at 25 species ~390 km downriver (Figure 4).

299 Assessing the Sørensen's index of community similarity as a function of distance for
300 each taxonomic group, adjacent samples were projected to have similarity values of
301 55.7% for bacteria, 65.3% for phytoplankton, 21% for invertebrates, 50.1% for
302 vertebrates. Similarity declined with distance for all four groups ($P < 0.001$; Figure 5).
303 The rate of decline in similarity was greatest for vertebrates (7.1% $100km^{-1}$); intermediate
304 for bacteria and phytoplankton (4.7% and 4.6% $100km^{-1}$), and lowest for invertebrates
305 (2.6% $100km^{-1}$), which had the lowest degree of similarity for geographically adjacent
306 assemblages.

307 Nutrients

308 As with the biotic communities, nutrient concentrations also showed distinct patterns
309 with distance downriver (Figure 6). Total dissolved phosphorus concentrations were
310 generally the lowest in the headwaters and gradually increased until peaking in the Lower
311 Potomac. In contrast, total dissolved nitrate concentrations were high in the headwaters,
312 were lowest in the Upper Potomac, and were high in Lower Potomac and estuary samples
313 (Figure 6).

314 Multivariate patterns in overall assemblage composition

315 We examined multivariate patterns of relative abundance with principal component
316 analysis to assess the broad patterns of changes in relative abundance among taxa without
317 *a priori* assumptions of what factors might be structuring the observed patterns.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

318 Examining the results of the principal component analysis for each primer pair, the first
319 multivariate axis in each PCA appeared to capture either increasing phosphorus
320 availability, i.e. eutrophication, or increasing river size except for COI (Figure 7). A PCA
321 analysis with all surveyed taxa combined revealed a stronger pattern of shifts in taxa
322 abundances with either downstream eutrophication or river size. In the multi-assemblage
323 PCA, Axis 1 explained 20.1% of the data, which is 23.4 times more than expected by
324 chance. Axis 1 increased with river distance peaking just before Great Falls where the
325 river becomes tidal (Figure 7). Axis 1 correlated with distance ($r = 0.85$, $P < 0.001$) and
326 total dissolved phosphorus ($r = 0.78$, $P < 0.001$), but not total dissolved nitrogen ($P =$
327 0.09).

328 We next examined the coefficients of the variables in the PCA to assess which taxa were
329 the most important contributors to the general patterns described above. Of the top 10
330 taxa with the highest coefficients for the multi-assemblage Axis 1, 8 were bacteria (Table
331 S1). These included the Actinomycete of ACK-M1 group and the Bacteriodetes
332 *Sedimentibacteria*. The other taxa were the common carp (*Cyprinus carpio*) and a diatom
333 indicative of eutrophic conditions (*Bacillaria paxillifer*). Other select taxa with high Axis
334 1 coefficients include redspotted sunfish (*Lepomis miniatus*), spotfin shiner (*Cyprinella*
335 *spiloptera)*, which is often found in poor quality waters, a mayfly that tends to inhabit
336 large rivers, *Anthopotamus verticis*, the snail *Physella acuta*, which often is found in
337 degraded waters, and a diatom *Nitzschia* indicative of eutrophic waters.

338 Axis 1 also distinguished among taxa that were found in higher abundance in the Upper
339 Potomac, but not exclusively in the headwaters (Table S1). Of the 10 taxa with the lowest
340 coefficients on Axis 1, five were Proteobacteria including a Hyphomonadaceae OTU and

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

341 a Rhodobacter OTU. Other taxa included the diatoms *Didymosphenia geminata* and
342 *Sellaphora pupula* and fish species similar to the emerald shiner (*Notropis atherinoides)*
343 (Table S1). Human DNA was also more abundant in Upper Potomac and had a low
344 coefficient on Axis 1 (Table S1). Although many of the strongest taxa driving the
345 separation of the communities along Axis 1 were associated with bacteria, the scores of
346 sites on a separate PCA with the phytoplankton, macroinvertebrate, and vertebrate taxa,
347 but without bacteria, were strongly correlated with those generated with all four sets of
348 taxa ($r = 0.96$, $P < 0.001$).

349 The second axis of the multi-assemblage PCA primarily separated out taxa that were
350 more abundant in the estuary (Figure 7). Bacteria that were more abundant in the estuary
351 included multiple cyanobacterial OTUs. Phytoplankton that were more abundant in the
352 estuary were diatoms such as *Cyclotella* sp. [WC03_2] and taxa similar to *Thalassiosira*
353 *pseudonana*, a marine diatom (Table S1). Invertebrates that were relatively abundant
354 included a *Cricotopus* species (a dipteran) and the Asian clam, *Corbicula fluminea*,
355 which accounted for >10% of the reads in the most downstream site (Table 1). Fish such
356 as white perch (*Morone americana*), american eel (*Anguilla rostrata*), shad (*Alosa*
357 *aestivalis*), and gizzard shad (*Dorosoma cepedianum*) were dominant in the estuarine
358 samples (Table 1, Supplemental).

359 Axis 3 primarily distinguished headwater sites, which harbored a unique set of taxa
360 (Table 1). Headwater bacteria included the Proteobacteria *Mycoplana* and *Rhodobacter*.
361 Headwater phytoplankton species were generally taxa other than diatoms such as the
362 eustigmatophyte, *Nannochloropsis salina,* the chlorophyte *Choricystis parasitica,* and the
363 green alga, *Actinotaenium cruciferum* (Table S1). Arthropods included the mayfly,

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

364 *Ephemerella dorothea* and the Diptera *Parametriocnemus* sp. BOLD-2016 (Table S1).
365 Fish characteristic of the headwater samples included the white sucker (*Catostomus*
366 *commersonii*), and sculpin (identified as *Cottus szanaga*).

367 ## Discussion

368 By analyzing eDNA collected from the Potomac River, we were able to demonstrate the
369 potential utility of eDNA for assessing aquatic assemblages and constructing multi-taxa
370 indices that reflect environmental conditions. Here, using four primer pairs, we were able
371 to quantify the DNA of resident bacteria, phytoplankton, invertebrates, and vertebrates
372 across the length of a major river. Although better reference databases are needed to
373 improve taxonomic identifications, the taxa observed corresponded well to taxa observed
374 in previous surveys of rivers in the region [25-28]. Abundant bacterial taxa observed here,
375 e.g. *Sphingomonodales*, *Flectobacillus*, and *Flavobacterium succinicans*) are typical of
376 rivers [27, 29, 30]. Previous work has shown that riverine and estuarine phytoplankton
377 communities share many of the same taxa observed here, e.g. *Navicula*, *Thalassiosira*,
378 and *Nitzschia* [31-33]. Likewise, the more abundant invertebrate and vertebrate species
379 we detected (Figure 1) are taxa we would expect in these types of river systems [25, 34-
380 36].

381 In addition to identifying the composition of the biotic assemblages, we were able to
382 assess correlations in taxon abundances across sites. We found that the biotic
383 assemblages were spatially patterned and reflected changing environmental conditions
384 across the length of the Potomac River. We were also able to clearly delineate changes in
385 assemblages across the headwaters, the main river, and the estuary with the strongest

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

386 shift in biotic assemblages observed between the headwaters and the estuary. Most likely,
387 this gradient represents some combination of taxa responding to eutrophication and river
388 size, which could not cleanly be separated here given the general increases in phosphorus
389 concentrations with distance downstream. Along this gradient, the river increased in size
390 from a small spring with a discharge of <1 $m^3 s^{-1}$ at the Fairfax Stone, to a mean
391 discharge of 322 $m^3 s^{-1}$ at 440 km (Little Falls). We found stronger correlations between
392 assemblage composition and distance along the river than with total dissolved
393 phosphorus. Although phosphorus availability was measured with spring baseflows,
394 which are considered good indicators of annual discharge-weighted mean concentrations
395 [37], nutrient concentrations were single point measurements and might not necessarily
396 represent integrated annual availability, to which the organisms are more likely
397 responding. Alternatively, the observed patterns may be driven by differences in
398 eutrophication levels with distance along the river, an interpretation supported by the
399 observation that many of the taxa driving the patterns in assemblage composition evident
400 in Figure 7 are taxa frequently associated with eutrophication status. Many *Nitzschia* and
401 *Bacillaria* taxa are considered indicators of eutrophication [38, 39]. Fish such as
402 *Cyprinus carpio* and *Cyprinella spiloptera,* as well as the snail *Physella acuta* are also
403 associated with waters degraded to different degrees [40-42]. Many of the taxa that
404 scored low on Axis 1 are associated with low nutrient conditions, such as the diatom
405 *Didymosphenia geminate* [43, 44]. Yet, in contrast, one mayfly taxa, *Anthopotamus*
406 *verticis,* was associated with high Axis 1 scores. Most mayfly species are considered
407 indicators of high, not low, water quality, but *Anthopotamus* species often occupy waters
408 with intermediate rather than low P availability [45].

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

409 Despite the relative success at this attempt to generate multi-assemblage indices of
410 environmental conditions, more research and development is needed in a number of areas
411 before the use of environmental DNA for bioassessment can be widely adopted. The
412 uncertainty of the interpretation of assemblage compositions highlights the need for a
413 comprehensive process to calibrate indices generated with eDNA. Multiple reference
414 sites will need to be established that allow for separation of covariates such as river size
415 and eutrophication. For example, multiple large rivers of different nutrient status will
416 have to be sampled to separate out river size and nutrient conditions as drivers of taxon
417 abundances. River systems that span gradients in other environmental variables, such as
418 pH and temperature, will have to be surveyed to further separate out other covariates.
419 More research on individual taxa will also be needed to link taxon abundances to
420 environmental conditions. Although indices can be generated independent of taxonomy
421 [46], building ecological understanding of individual taxa will assist in interpreting multi-
422 taxa indices of environmental conditions.

423 Our results also highlight the need for the development of more effective primer pairs. To
424 assess macroinvertebrates (namely arthropods), we used a primer pair that is considered
425 to be relatively specific for arthropods [47]. However, 95% of the reads with these COI
426 primers were assigned to Oomycota taxa. Oomycetes are commonly known as water
427 molds, include plant and animal pathogens as well as taxa that are important decomposers
428 of organic matter in water [48, 49]. A full assessment of the patterns of Oomycota in the
429 Potomac was outside the scope of this research, but is warranted given their importance
430 in ecosystem function and their potential roles as pathogens. In the future, assessing
431 invertebrates in water using eDNA will either require greater sequencing depth, blocking

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

432 primers, or the development of different primers that are more selective for arthropods.
433 As an additional note, we also observed relatively few vertebrate reads per sample, but
434 this result was likely due to interactions during sequencing that disfavor longer 12S
435 rRNA amplicons when multiplexed with amplicons generated with other primers.
436 Sequencing only 12S rRNA amplicons on a given sequencing run should relieve this
437 limitation.

438 The benefits of using eDNA to reconstruct assemblages appear to extend beyond the
439 ability to reconstruct of environmental conditions. For example, as in a previous study of
440 the Cuyahoga River [21], we were able to detect different amounts of mammalian DNA
441 (including human, pig, and cattle DNA) with these results representing a potential
442 opportunity to identify sources of fecal contamination in water. Current approaches often
443 rely on the bacteria that indicate vertebrate sources of fecal coliform bacteria [50]. Yet, it
444 might be feasible to simply sequence host DNA directly from collected water samples.
445 Sequencing eDNA also has the potential to identify non-native and migratory species
446 [51]. Here, we detected migratory fish such as shad and eel in the estuary, although some
447 species known to be in the Potomac, such as northern snakehead (*Channa argus*), were
448 not detected here. Greater sequencing depth and replication would likely provider greater
449 sensitivity than our preliminary assessment, but probabilities of detection have yet to be
450 resolved for individual taxa of interest.

451 Using eDNA to reconstruct assemblages also opens a new line of research for
452 development of ecological theory. Although the river continuum concept was a crucial
453 development for beginning to understand changes in ecosystem function along the length
454 of a river [52, 53], theoretical development of the regulation of biodiversity within rivers

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

455 has lagged [54, 55]. Data here raise interesting questions. For example, why did bacterial
456 diversity decline downriver while it increased for phytoplankton and vertebrates? Are
457 bacterial assemblages being driven by environmental chemistry as observed in other
458 rivers [56]? Questions about accumulation of species, changes in habitats, nestedness of
459 assemblages, and assemblage turnover are all relevant to understanding the ecologies of
460 river systems that could be addressed by leveraging eDNA approaches to obtain large
461 amounts of relatively low-cost standardized data for rivers.

462 Acknowledgments

463 Amy Beckley carried out the lab work for this research, Mike Robeson was responsible
464 for bioinformatics. Kristy Deiner provided helpful comments on previous drafts of this
465 manuscript.

466 References

467 1. Malmqvist B, Rundle S. Threats to the running water ecosystems of the world.
468 Environ Conserv. 2002;29(2):134-53.
469 2. US Environmental Protection Agency. A compilation of cost data associated with
470 the impacts and control of nutrient pollution. US EPA Office of Water; 2015.
471 3. U.S. Environmental Protection Agency. A Compilation of Cost Data Associated
472 With the Impacts and Control of Nutrient Pollution. Washington, DC.: U.S.
473 Environmental Protection Agency, Office of Water, ; 2015. p. EPA 820-F-15-096.
474 4. Young RA, Loomis JB. Determining the economic value of water: concepts and
475 methods: Routledge; 2014.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

476 5. Palaniappan M, Gleick P, Allen L, Cohen M, Christian-Smith J, Smith C.
477 Clearing the waters: A focus on water quality solutions. Nairobi, Kenya: UNEP/Pacific
478 Institute. 2010.
479 6. Schoolmaster DR, Grace JB, William Schweiger E. A general theory of
480 multimetric indices and their properties. Methods in Ecology and Evolution.
481 2012;3(4):773-81. doi: 10.1111/j.2041-210X.2012.00200.x.
482 7. Karr JR. Defining and measuring river health. Freshw Biol. 1999;41.2:221-34.
483 8. Stevenson RJ, Pan Y. Assessing environmental conditions in rivers and streams
484 with diatoms. The diatoms: applications for the environmental and earth sciences.
485 1999;1(1):4.
486 9. American Public Health Association, American Water Works Association.
487 Standard methods for the examination of water and wastewater. Washington DC:
488 American Public Health Association, American Water Works Association and Water
489 Environment Federation; 2012.
490 10. Zimmermann J, Abarca N, Enke N, Skibbe O, Kusber WH, Jahn R. Taxonomic
491 reference libraries for environmental barcoding: a best practice example from diatom
492 research. PloS one. 2014;9(9):e108793. doi: 10.1371/journal.pone.0108793. PubMed
493 PMID: 25265556; PubMed Central PMCID: PMCPMC4180937.
494 11. Buss DF, Carlisle DM, Chon T-S, Culp J, Harding JS, Keizer-Vlek HE, et al.
495 Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-
496 scale programs. Environ Monit Assess. 2015;187(1):1.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

497 12. Goffaux D, Grenouillet G, Kestemont P. Electrofishing versus gillnet sampling
498 for the assessment of fish assemblages in large rivers. Archiv für Hydrobiologie.
499 2005;162(1):73-90.
500 13. Reyjol Y, Loot G, Lek S. Estimating sampling bias when using electrofishing to
501 catch stone loach. J Fish Biol. 2005;66(2):589-91.
502 14. Barnes MA, Turner CR. The ecology of environmental DNA and implications for
503 conservation genetics. Conservation Genetics. 2015;17(1):1-17. doi: 10.1007/s10592-
504 015-0775-4.
505 15. Mächler E, Deiner K, Steinmann P, Altermatt F. Utility of environmental DNA
506 for monitoring rare and indicator macroinvertebrate species. Freshwater Science.
507 2014;33(4):1174-83. doi: 10.1086/678128.
508 16. Stein ED, White BP, Mazor RD, Jackson JK, Battle JM, Miller PE, et al. Does
509 DNA barcoding improve performance of traditional stream bioassessment metrics?
510 Freshwater Science. 2014;33(1):302-11. doi: 10.1086/674782.
511 17. Elbrecht V, Leese F. Can DNA-Based Ecosystem Assessments Quantify Species
512 Abundance? Testing Primer Bias and Biomass--Sequence Relationships with an
513 Innovative Metabarcoding Protocol. PloS one. 2015;10(7):e0130324. doi:
514 10.1371/journal.pone.0130324. PubMed PMID: 26154168; PubMed Central PMCID:
515 PMCPMC4496048.
516 18. Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, et al. Estimating
517 species richness using environmental DNA. Ecology and evolution. 2016;6(12):4214-26.
518 doi: 10.1002/ece3.2186.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

519 19. Deiner K, Fronhofer EA, Machler E, Walser JC, Altermatt F. Environmental
520 DNA reveals that rivers are conveyer belts of biodiversity information. Nature
521 communications. 2016;7:12544. doi: 10.1038/ncomms12544. PubMed PMID: 27572523;
522 PubMed Central PMCID: PMCPMC5013555.
523 20. Jackson CR, Millar JJ, Payne JT, Ochs CA. Free-Living and Particle-Associated
524 Bacterioplankton in Large Rivers of the Mississippi River Basin Demonstrate
525 Biogeographic Patterns. Appl Environ Microbiol. 2014;80(23):7186-95. doi:
526 10.1128/AEM.01844-14. PubMed PMID: 25217018; PubMed Central PMCID:
527 PMCPMC4249191.
528 21. Cannon MV, Hester J, Shalkhauser A, Chan ER, Logue K, Small ST, et al. In
529 silico assessment of primers for eDNA studies using PrimerTree and application to
530 characterize the biodiversity surrounding the Cuyahoga River. Scientific reports.
531 2016;6:22908. doi: 10.1038/srep22908. PubMed PMID: 26965911; PubMed Central
532 PMCID: PMCPMC4786790.
533 22. Boynton W, Garber J, Summers R, Kemp W. Inputs, transformations, and
534 transport of nitrogen and phosphorus in Chesapeake Bay and selected tributaries.
535 Estuaries and Coasts. 1995;18(1):285-314.
536 23. Edgar RC. Search and clustering orders of magnitude faster than BLAST.
537 Bioinformatics. 2010;26(19):2460-1.
538 24. Dexter KG, Terborgh JW, Cunningham CW. Historical effects on beta diversity
539 and community assembly in Amazonian trees. Proceedings of the National Academy of
540 Sciences. 2012;109(20):7787-92.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

541 25. Cummins J. Fishes of the freshwater Potomac. The Interstate Commission on the
542 Potomac River Basin; 2013.
543 26. Fierer N, Morse JL, Berthrong ST, Bernhardt ES, Jackson RB. Environmental
544 controls on the landscape-scale biogeography of stream bacterial communities. Ecology.
545 2007;88(9):2162-73.
546 27. Payne JT, Millar JJ, Jackson CR, Ochs CA. Patterns of variation in diversity of
547 the Mississippi river microbiome over 1,300 kilometers. PloS one. 2017;12(3):e0174890.
548 doi: 10.1371/journal.pone.0174890. PubMed PMID: 28350888; PubMed Central
549 PMCID: PMCPMC5370145.
550 28. Utz RM, Hilderbrand RH, Boward DM. Identifying regional differences in
551 threshold responses of aquatic invertebrates to land cover gradients. Ecological Indicators.
552 2009;9(3):556-67. doi: 10.1016/j.ecolind.2008.08.008.
553 29. Bernardet J-F, Segers P, Vancanneyt M, Berthe F, Kersters K, Vandamme P.
554 Cutting a Gordian knot: emended classification and description of the genus
555 Flavobacterium, emended description of the family Flavobacteriaceae, and proposal of
556 Flavobacterium hydatis nom. nov.(basonym, Cytophaga aquatilis Strohl and Tait 1978).
557 Int J Syst Evol Microbiol. 1996;46(1):128-48.
558 30. Portillo MC, Anderson SP, Fierer N. Temporal variability in the diversity and
559 composition of stream bacterioplankton communities. Environmental microbiology.
560 2012;14(9):2417-28. doi: 10.1111/j.1462-2920.2012.02785.x. PubMed PMID: 22626459.
561 31. Zimmermann J, Glockner G, Jahn R, Enke N, Gemeinholzer B. Metabarcoding vs.
562 morphological identification to assess diatom diversity in environmental studies. Mol

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

563 Ecol Resour. 2015;15(3):526-42. doi: 10.1111/1755-0998.12336. PubMed PMID:
564 25270047.
565 32. Chariton AA, Stephenson S, Morgan MJ, Steven AD, Colloff MJ, Court LN, et al.
566 Metabarcoding of benthic eukaryote communities predicts the ecological condition of
567 estuaries. Environ Pollut. 2015;203:165-74. doi: 10.1016/j.envpol.2015.03.047. PubMed
568 PMID: 25909325.
569 33. Vasselon V, Rimet F, Tapolczai K, Bouchez A. Assessing ecological status with
570 diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island,
571 France). Ecological Indicators. 2017;82:1-12. doi: 10.1016/j.ecolind.2017.06.024.
572 34. Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA. Environmental
573 monitoring using next generation sequencing: rapid identification of macroinvertebrate
574 bioindicator species. Frontiers in zoology. 2013;10(1):1.
575 35. Phelps HL. The Asiatic clam (Corbicula fluminea) invasion and system-level
576 ecological change in the Potomac River estuary near Washington, DC. Estuaries and
577 Coasts. 1994;17(3):614-21.
578 36. Killgore KJ, Morgan RP, Rybicki NB. Distribution and abundance of fishes
579 associated with submersed aquatic plants in the Potomac River. North American Journal
580 of Fisheries Management. 1989;9(1):101-11.
581 37. Kline KM, Eshleman KN, Garlitz JE, U'Ren SH. Long-term response of surface
582 water acid neutralizing capacity in a central Appalachian (USA) river basin to declining
583 acid deposition. Atmos Environ. 2016;146:195-205. doi:
584 10.1016/j.atmosenv.2016.07.034.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

585 38. Kelly M, Whitton B. The trophic diatom index: a new index for monitoring
586 eutrophication in rivers. J Appl Phycol. 1995;7(4):433-44.
587 39. Potapova M, Charles DF. Diatom metrics for monitoring eutrophication in rivers
588 of the United States. Ecol Indic. 2007;7(1):48-70. doi: 10.1016/j.ecolind.2005.10.001.
589 40. Lougheed VL, Crosbie B, Chow-Fraser P. Predictions on the effect of common
590 carp (Cyprinus carpio) exclusion on water quality, zooplankton, and submergent
591 macrophytes in a Great Lakes wetland. Can J Fish Aquat Sci. 1998;55(5):1189-97.
592 41. Barbour MT, Gerritsen J, Snyder BD, Stribling JB. Rapid bioassessment protocols
593 for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish:
594 US Environmental Protection Agency, Office of Water Washington, DC; 1999.
595 42. Mouthon J, Charvet S, editors. Compared sensitivity of species, genera and
596 families of Molluscs to biodegradable pollution. Annales de Limnologie-International
597 Journal of Limnology; 1999: EDP Sciences.
598 43. Bothwell ML, Taylor BW, Kilroy C. The Didymo story: the role of low dissolved
599 phosphorus in the formation of Didymosphenia geminata blooms. Diatom Research.
600 2014;29(3):229-36.
601 44. Keller SR, Hilderbrand RH, Shank MK, Potapova M, Zhan A. Environmental
602 DNA genetic monitoring of the nuisance freshwater diatom,Didymosphenia geminata, in
603 eastern North American streams. Diversity and Distributions. 2017;23(4):381-93. doi:
604 10.1111/ddi.12536.
605 45. Smith AJ, Bode RW, Kleppel GS. A nutrient biotic index (NBI) for use with
606 benthic macroinvertebrate communities. Ecol Indic. 2007;7(2):371-86. doi:
607 10.1016/j.ecolind.2006.03.001.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

608 46. Apotheloz-Perret-Gentil L, Cordonier A, Straub F, Iseli J, Esling P, Pawlowski J.
609 Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Mol
610 Ecol Resour. 2017. doi: 10.1111/1755-0998.12668. PubMed PMID: 28296259.
611 47. Zeale MR, Butlin RK, Barker GL, Lees DC, Jones G. Taxon-specific PCR for
612 DNA barcoding arthropod prey in bat faeces. Molecular ecology resources.
613 2011;11(2):236-44. Epub 2011/03/25. doi: 10.1111/j.1755-0998.2010.02920.x. PubMed
614 PMID: 21429129.
615 48. Gessner MO, Van Ryckegem G. Water fungi as decomposers in freshwater
616 ecosystems. Encyclopedia of environmental microbiology. 2003.
617 49. van West P. Saprolegnia parasitica, an oomycete pathogen with a fishy appetite:
618 new challenges for an old problem. Mycologist. 2006;20(3):99-104. doi:
619 http://dx.doi.org/10.1016/j.mycol.2006.06.004.
620 50. Tan B, Ng C, Nshimyimana JP, Loh LL, Gin KY, Thompson JR. Next-generation
621 sequencing (NGS) for assessment of microbial water quality: current progress, challenges,
622 and future opportunities. Front Microbiol. 2015;6:1027. doi: 10.3389/fmicb.2015.01027.
623 PubMed PMID: 26441948; PubMed Central PMCID: PMCPMC4585245.
624 51. Stoeckle MY, Soboleva L, Charlop-Powers Z. Aquatic environmental DNA
625 detects seasonal fish abundance and habitat preference in an urban estuary. PloS one.
626 2017;12(4):e0175186. doi: 10.1371/journal.pone.0175186. PubMed PMID: 28403183.
627 52. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The river
628 continuum concept. Can J Fish Aquat Sci. 1980;37(1):130-7.
629 53. Statzner B, Higler B. Questions and comments on the river continuum concept.
630 Can J Fish Aquat Sci. 1985;42(5):1038-44.

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

631 54. Muneepeerakul R, Bertuzzo E, Lynch HJ, Fagan WF, Rinaldo A, Rodriguez-
632 Iturbe I. Neutral metacommunity models predict fish diversity patterns in Mississippi-
633 Missouri basin. Nature. 2008;453(7192):220-2. doi: 10.1038/nature06813. PubMed
634 PMID: 18464742.
635 55. Gutiérrez-Cánovas C, Millán A, Velasco J, Vaughan IP, Ormerod SJ. Contrasting
636 effects of natural and anthropogenic stressors on beta diversity in river organisms. Glob
637 Ecol Biogeogr. 2013;22(7):796-805. doi: 10.1111/geb.12060.
638 56. Nino-Garcia JP, Ruiz-Gonzalez C, Del Giorgio PA. Interactions between
639 hydrology and water chemistry shape bacterioplankton biogeography across boreal
640 freshwater networks. ISME J. 2016;10(7):1755-66. doi: 10.1038/ismej.2015.226.
641 PubMed PMID: 26849312; PubMed Central PMCID: PMCPMC4918434.
642 57. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh
643 PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per
644 sample. Proceedings of the National Academy of Sciences. 2011;108(Supplement
645 1):4516-22. doi: 10.1073/pnas.1000080107.
646 58. Sherwood AR, Presting GG. Universal primers amplify a 23S rDNA plastid
647 marker in eukaryotic algae and cyanobacteria. J Phycol. 2007;43(3):605-8.
648 59. Hamad I, Delaporte E, Raoult D, Bittar F. Detection of termites and other insects
649 consumed by African great apes using molecular fecal analysis. Scientific reports.
650 2014;4:4478. doi: 10.1038/srep04478. PubMed PMID: 24675424; PubMed Central
651 PMCID: PMCPMC3967517.
652 60. Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, et al.
653 Quantification of mesocosm fish and amphibian species diversity via environmental

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

654 DNA metabarcoding. Mol Ecol Resour. 2016;16(1):29-41. doi: 10.1111/1755-

655 0998.12433. PubMed PMID: 26032773; PubMed Central PMCID: PMCPMC4744776.

656

657

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

658 # Tables

659 Table 1. Details on unique laboratory methods and bioinformatic processing of data for
660 different primer pairs.

661

	16S rRNA	23S rRNA	COI	12S rRNA
Target	Bacteria	Phytoplankton	Invertebrates	Vertebrates
Forward (5' – 3')	GTGYCAGCMGCCGCGGTAA	GGACAGAAAGACCCTATGAA	AGATATTGGAACWTTATATTTTATTTTTGG	ACTGGGATTAGATACCCCACTATG
Reverse (5' – 3')	GGACTACNVGGGTWTCTAAT	TGAGTGACGGCCTTTCCACT	WACTAATCAATTWCCAAATCCTCC	GAGAGTGACGGGCGGTGT
Citation	Caporaso et al. 2011	Sherwood and Presting 2007	Hamad et al. 2014	Evans et al. 2016
Denaturation	5 min @ 95°C	3 min @ 94°C	5 min @ 94°C	3 min @ 94°C
# Cycles	35	40	45	40
Cycle conditions	95°C (45s), 50°C (60 s), 72°C (90 s)	94°C (30s), 55°C (45 s), 72°C (60 s)	94°C (30s), 45°C (45 s), 72°C (45 s)	94°C (30s), 55°C (30 s), 72°C (60 s)
	10 min @ 72°C	10 min @ 72°C	10 min @ 72°C	10 min @ 72°C
OTU clustering level	97%	97%	99%	99%
Reference database	Silva v128	GenBank	GenBank	GenBank

662

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

663

664

665 Table S1.

Variable	Axis 1	Axis 2	Axis 3
OTU8.Actinobacteria; Actinobacteria; o__Actinomycetales; f__ACK-M1; g__; s__	0.19	0.02	0.04
OTU17.Bacteroidetes; [Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__Sediminibacterium; s__	0.18	0.01	0.09
OTU9.Verrucomicrobia; Opitutae; o__[Cerasicoccales]; f__[Cerasicoccaceae]; g__; s__	0.18	0.02	0.03
Cyprinus carpio	0.17	0.01	0.05
OTU5.Bacteroidetes; Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__	0.17	-0.06	0.01
OTU85.Actinobacteria; Actinobacteria; o__Actinomycetales; f__ACK-M1; g__; s__	0.17	0.07	-0.04
OTU19.Actinobacteria; Actinobacteria; o__Actinomycetales; f__Microbacteriaceae; g__Candidatus Rhodoluna; s__	0.17	0.05	0.00
OTU3.Bacteroidetes; Cytophagia; o__Cytophagales; f__Cytophagaceae; g__Flectobacillus; s__	0.17	0.03	0.05
OTU16.Bacteroidetes; Flavobacteriia; o__Flavobacteriales; f__Cryomorphaceae; g__Fluviicola; s__	0.16	0.05	0.02
OTU25.Bacillariophyta.Bacillaria paxillifer	0.16	0.12	-0.05
Lepomis gibbosus	0.15	0.01	0.01
Arthropoda Ephemeroptera Anthopotamus verticis	0.15	0.08	-0.02
OTU1.Actinobacteria; Actinobacteria; o__Actinomycetales; f__ACK-M1; g__; s__	0.15	0.04	0.15
Lepomis miniatus	0.14	0.06	0.03
OTU10.Bacteroidetes; Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__	0.14	0.00	0.03
OTU52.Proteobacteria; Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__	0.14	0.02	-0.04
OTU7.Bacillariophyta.Nitzschia sp. CCMP2626	0.14	0.11	-0.03
Cyprinella spiloptera	0.14	0.06	0.03
Mollusca None Physella acuta	0.14	0.01	0.01
OTU35.Bacillariophyta.Fistulifera pelliculosa	0.14	0.14	-0.01
Ictalurus punctatus	0.14	0.06	0.00
Lepomis megalotis	0.14	0.05	0.04
12S rRNA.Richness	0.13	0.05	0.11
Hybognathus regius	0.13	0.03	0.03
Pimephales notatus	0.13	0.04	0.03
Macrhybopsis gelida	0.12	0.00	0.04
OTU398.Bacillariophyta.uncultured Nitzschia	0.11	0.12	-0.10
OTU11.Bacillariophyta.Navicula salinicola	0.11	-0.03	-0.02
23S rRNA.Richness	0.11	0.18	0.02
OTU11.Bacteroidetes; Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__succinicans	0.11	-0.03	-0.04
Fundulus diaphanus	0.10	-0.03	0.01
Carpiodes carpio	0.10	0.00	0.00
OTU33.Bacillariophyta.uncultured Bacillariophyceae	0.10	0.08	-0.11
OTU46.Bacillariophyta.Lithodesmium undulatum	0.10	0.11	-0.05
OTU242.Bacillariophyta.Nitzschia longissima	0.09	0.10	-0.05
Moxostoma carinatum	0.09	0.05	-0.02
OTU310.Dinophyceae.Peridiniopsis niei	0.08	-0.18	0.01

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Arthropoda Diptera Rheotanytarsus sp. BOLD:ACJ8793	0.07	0.04	0.02
Etheostoma blennioides	0.07	0.04	-0.03
Micropterus salmoides	0.06	-0.05	0.13
OTU6.Bacillariophyta.Thalassiosira pseudonana	0.06	-0.22	-0.01
Hypentelium nigricans	0.05	0.04	-0.03
OTU19.Cryptophyta.Falcomonas sp. PR-2015	0.05	-0.12	0.04
Lepomis macrochirus	0.05	-0.13	0.04
OTU2.Proteobacteria; Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__	0.05	0.08	-0.05
Arthropoda Diptera Polypedilum convictum	0.05	-0.13	0.04
Arthropoda Diptera Parakiefferiella sp. BOLD-2016	0.04	0.04	-0.04
Anguilla rostrata	0.03	-0.24	0.00
Arthropoda Ephemeroptera Stenonema sp. AMI 1	0.02	0.03	-0.07
Alosa aestivalis	0.02	-0.27	0.00
OTU3502.Cyanobacteria; Chloroplast; o__Stramenopiles; f__; g__; s__	0.02	-0.24	0.01
Arthropoda Diptera Eukiefferiella sp. BOLD-2016	0.02	0.05	-0.05
OTU7.Cyanobacteria; Chloroplast; o__Stramenopiles; f__; g__; s__	0.02	-0.10	0.00
Arthropoda Diptera Chironominae sp. BOLD-2016	0.02	-0.14	0.01
Arthropoda Diptera Cricotopus sp. BOLD-2016	0.02	-0.21	0.00
Dorosoma cepedianum	0.02	-0.16	0.00
Arthropoda Podocopida Podocopida sp. BOLD:AAH0908	0.02	-0.09	0.00
OTU22.Cyanobacteria; Chloroplast; o__Stramenopiles; f__; g__; s__	0.01	-0.24	-0.02
Etheostoma vitreum	0.01	0.01	0.14
Morone americana	0.01	-0.21	-0.01
OTU365.Bacillariophyta.Bacillariophyceae sp. 3 AS-2014	0.01	0.16	0.13
OTU433.Bacillariophyta.Cyclotella sp. WC03_2	0.00	-0.21	-0.01
Mollusca Veneroida Corbicula fluminea	0.00	-0.20	-0.02
OTU13.Cyanobacteria; ; o__; f__; g__; s__	0.00	-0.20	-0.01
Micropterus dolomieu	0.00	0.04	-0.09
Arthropoda Diptera Cricotopus tremulus	0.00	0.03	-0.03
Arthropoda Diptera Orthocladius sp. BOLD-2016	0.00	0.03	-0.07
OTU83.Bacillariophyta.Asterionella formosa	0.00	0.05	-0.09
OTU30.Proteobacteria; Gammaproteobacteria; o__Alteromonadales; f__OM60; g__; s__	-0.01	0.09	-0.18
Arthropoda Diptera Cricotopus sp. 18ES	-0.01	0.06	0.15
Arthropoda Diptera Rheotanytarsus pellucidus	-0.01	0.03	-0.05
Arthropoda Diptera Cricotopus trifascia	-0.02	0.02	-0.08
OTU214.Bacillariophyta.Sellaphora pupula	-0.02	0.07	0.15
Arthropoda Ephemeroptera Isonychia sp. BOLD:AAA9229	-0.02	0.04	-0.03
Arthropoda Diptera Orthocladiinae sp. BOLD-2016	-0.02	0.01	0.08
Annelida Haplotaxida Martiodrilus sp. 5 DP-2015	-0.03	0.01	-0.09
OTU36.Bacteroidetes; Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__	-0.03	0.05	-0.08
Arthropoda None Eupodidae sp. BOLD:AAF9191	-0.03	0.03	-0.05
Arthropoda Diptera Chironomidae sp. BOLD:ACW5282	-0.03	0.03	0.01
OTU16.Chrysophyceae.Chromulina sp. SAG 17.97	-0.03	-0.02	0.03
Arthropoda Plecoptera Ostrocerca sp. BOLD-2016	-0.04	-0.02	0.06

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Arthropoda Ephemeroptera Baetis sp. BOLD:AAA4715	-0.04	0.00	-0.05
Arthropoda Diptera Tanypodinae sp. BOLD-2016	-0.04	0.05	0.18
OTU4.Eustigmatophyceae.Nannochloropsis salina	-0.04	0.02	0.19
Sus scrofa	-0.04	0.00	0.04
OTU34.Proteobacteria; Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__Rhodobacter; s__	-0.04	-0.03	0.22
Arthropoda Diptera Chironomidae sp. BOLD-2016	-0.05	0.02	-0.03
OTU10.Streptophyta.Closterium lunula	-0.05	0.07	0.03
OTU130.Streptophyta.Cosmarium punctulatum	-0.05	0.00	0.06
Arthropoda Ephemeroptera Isonychia sp. LJ1	-0.05	0.00	0.00
OTU17.Cyanobacteria.Cyanobium sp. PCC 7009	-0.05	0.03	0.07
Arthropoda Ephemeroptera Ephemerella dorothea	-0.05	0.03	0.23
Odocoileus virginianus	-0.05	0.05	0.12
Arthropoda Plecoptera Plecoptera sp. BOLD:AAC1689	-0.06	0.00	0.11
OTU15.Chlorophyta.Choricystis parasitica	-0.06	0.02	0.20
OTU37.Bacillariophyta.Asterionella formosa	-0.06	0.01	0.06
Catostomus commersonii	-0.07	0.04	0.23
OTU5.Bacillariophyta.Melosira tropica	-0.07	0.01	-0.13
Cottus szanaga	-0.07	0.03	0.22
Arthropoda Diptera Parametriocnemus sp. BOLD-2016	-0.08	0.05	0.23
OTU270.Bacillariophyta.Eunotia naegelii	-0.08	0.02	0.17
OTU12.Proteobacteria; Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Mycoplana; s__	-0.08	0.04	0.24
OTU339.Streptophyta.Actinotaenium cruciferum	-0.08	0.06	0.20
OTU24.Proteobacteria; Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Rhodoferax; s__	-0.08	0.03	-0.14
OTU4.Proteobacteria; Deltaproteobacteria; o__Bdellovibrionales; f__Bacteriovoracaceae; g__Peredibacter; s__starrii	-0.08	0.05	0.02
OTU64.Proteobacteria; Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__	-0.09	0.07	-0.16
Mus musculus	-0.09	0.03	-0.15
OTU164.Bacillariophyta.Didymosphenia geminata	-0.09	0.06	0.03
OTU2.Bacillariophyta.Nitzschia sp. BOLD:AAX5147	-0.09	0.05	-0.14
OTU75.Bacillariophyta.Sellaphora pupula	-0.10	0.11	-0.05
OTU3.Bacillariophyta.Bacillariophyceae sp. 1 AS-2014	-0.10	0.02	-0.11
OTU778.Proteobacteria; Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__	-0.10	0.08	-0.12
Richness	-0.11	0.09	0.07
Homo sapiens	-0.11	0.05	-0.09
OTU41.Proteobacteria; Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Aquabacterium; s__	-0.11	0.03	-0.04
Notropis atherinoides	-0.11	0.10	0.01
OTU6.Proteobacteria; Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Caulobacter; s__henricii	-0.11	0.08	-0.04
OTU23.Proteobacteria; Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__Rhodobacter; s__	-0.11	0.04	-0.04
OTU25.Proteobacteria; Alphaproteobacteria; o__Rhodobacterales; f__Hyphomonadaceae; g__; s__	-0.15	0.07	-0.03

666

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

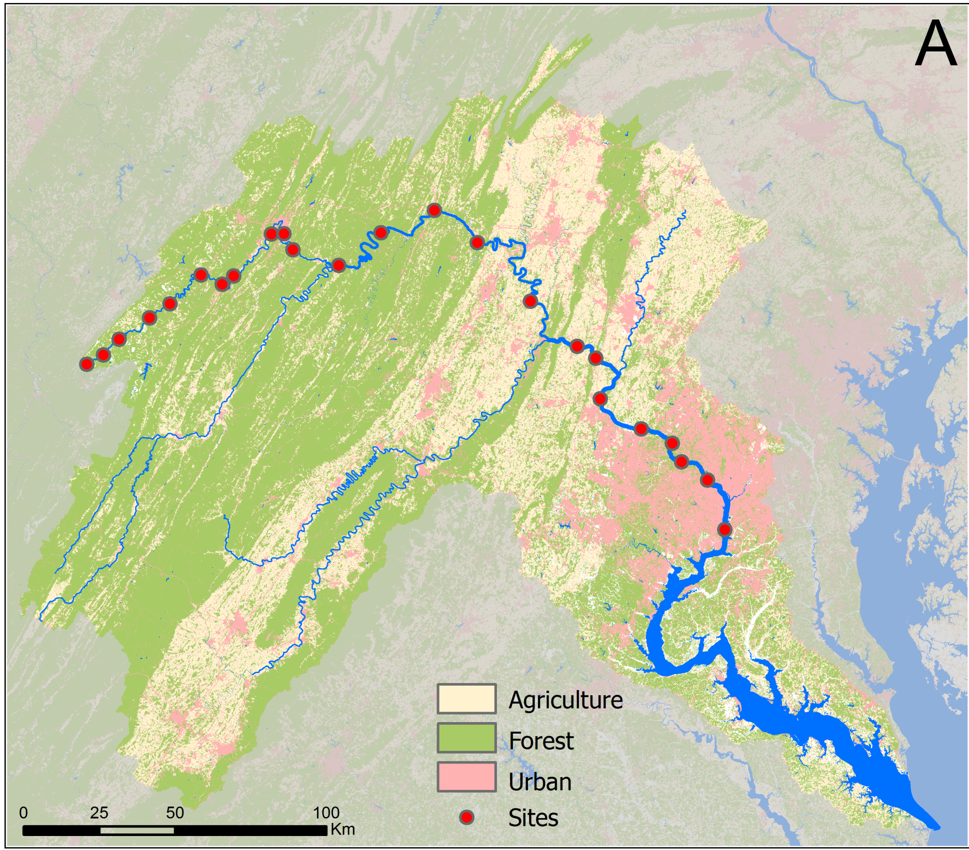
667 # Figure captions

668

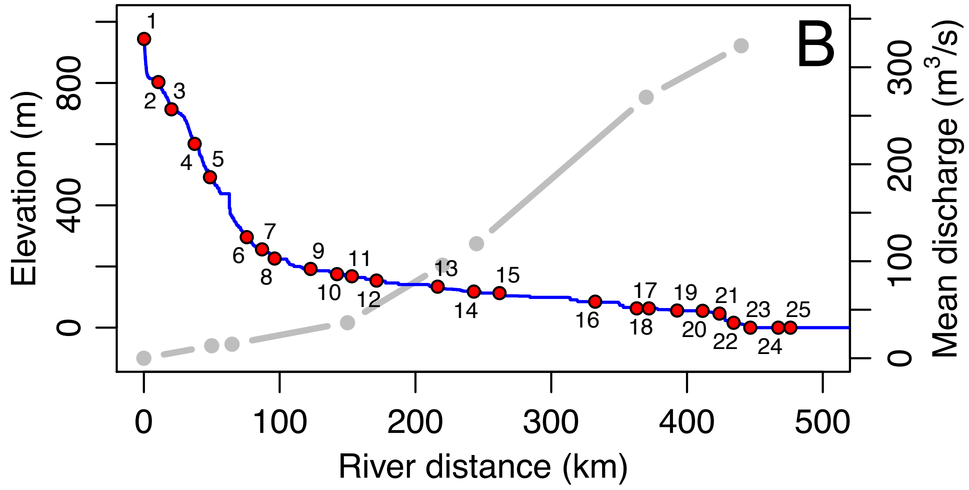
669 Figure 1. Map of the locations of the 25 sites sampled for this study (A), the elevation of
670 each site (B), and mean annual discharge for selected points from USGS streamflow data
671 (B).
672 Figure 2. Percent relative read abundance (%RRA) of sequences assigned to higher order
673 taxonomic groups as a function of distance downriver. Included are abundances for
674 bacteria (A-E), phytoplankton (F-J), macroinvertebrates (K-N), and vertebrates (O-Q)
675 primers.
676 Figure 3. Percent relative read abundance (%RRA) of the ten most abundant taxa for
677 bacteria (first column), phytoplankton (second column), macroinvertebrates (third
678 column), and vertebrates (fourth column) primers plotted versus distance downriver.
679 Figure 4. Rarefied richness of taxa as a function of distance downriver for bacteria (A),
680 phytoplankton (b), and vertebrates (c).
681 Figure 5. Sørensen's index of similarity for bacteria (A), phytoplankton (B), invertebrates
682 (C), and vertebrates (D) as a function of distance between 2 of the 25 water samples
683 collected on the Potomac River. Lower values indicate communities that are more
684 distinct in composition. Best fit linear regressions shown with Mantel test used to assess
685 significance of correlation (ρ = -0.67, -0.64, -0.32, -0.64, respectively; $P < 0.001$ for all).
686 Figure 6. Availability of phosphorus (A) and nitrogen (B) as a function of distance on the
687 Potomac River.
688 Figure 7. Scores of sites on first three axes of principal components analyses (PCA) for
689 bacteria (A-C), phytoplankton (D-F), macroinvertebrates (G-I), vertebrates (J-L), and all

bioRxiv preprint doi: https://doi.org/10.1101/186452; this version posted September 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

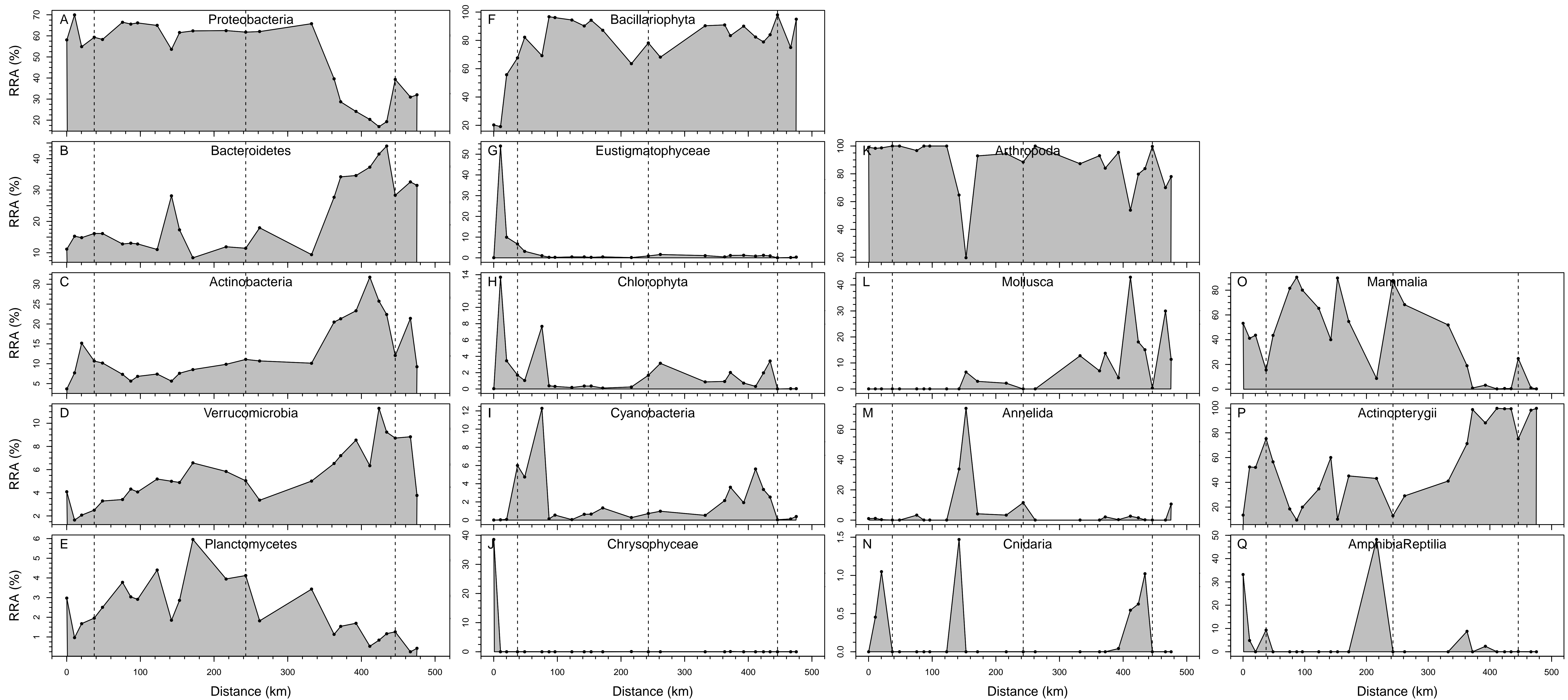
690 4 primer pairs together (M-O). The specific y-axis scores simply represent the position of
691 each assemblage along each individual PCA axis with assemblages that are more similar
692 in composition having more similar axis scores.
693
694

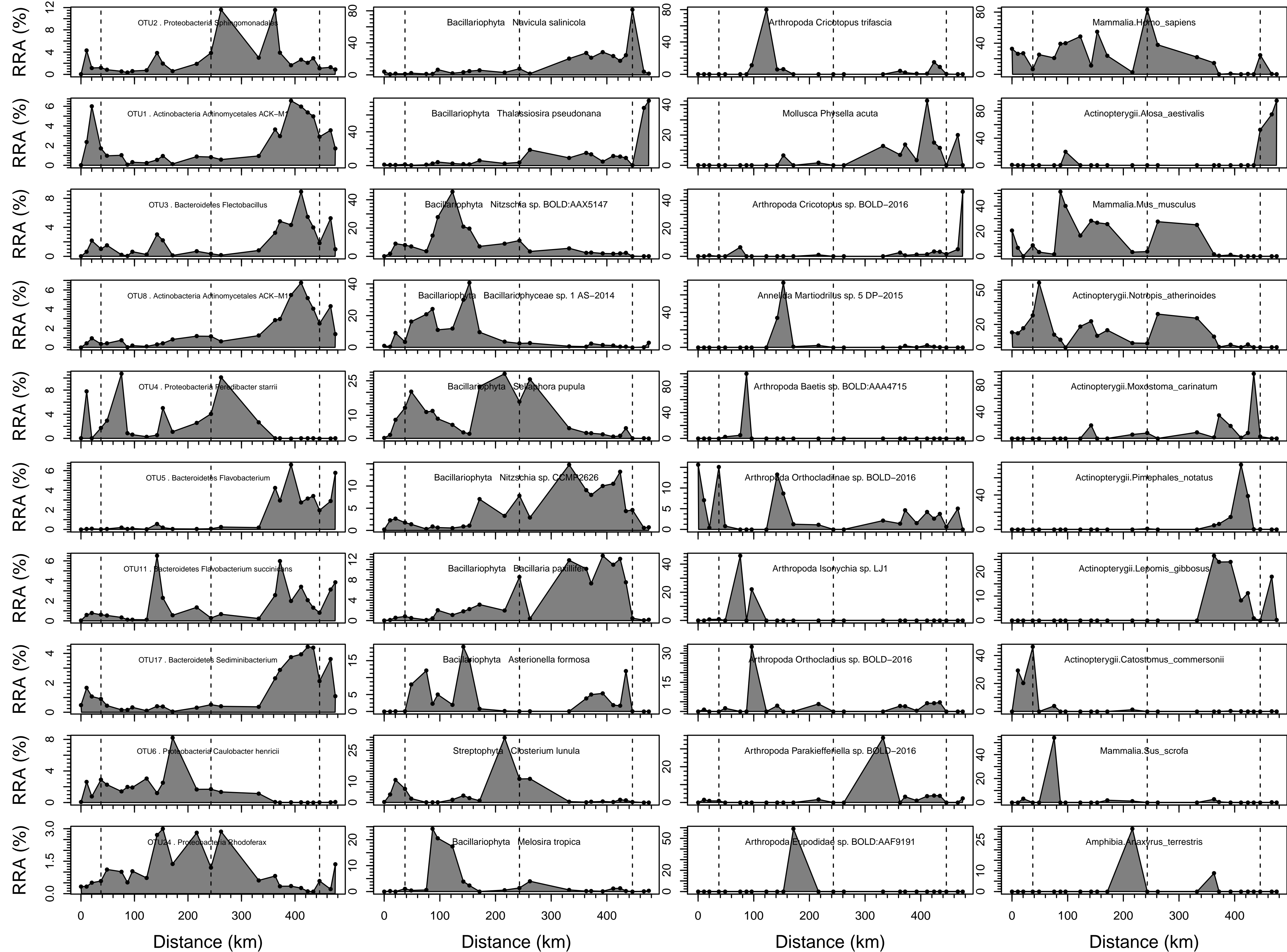


A
Agriculture
Forest
Urban
Sites
0
25
50
100
Km

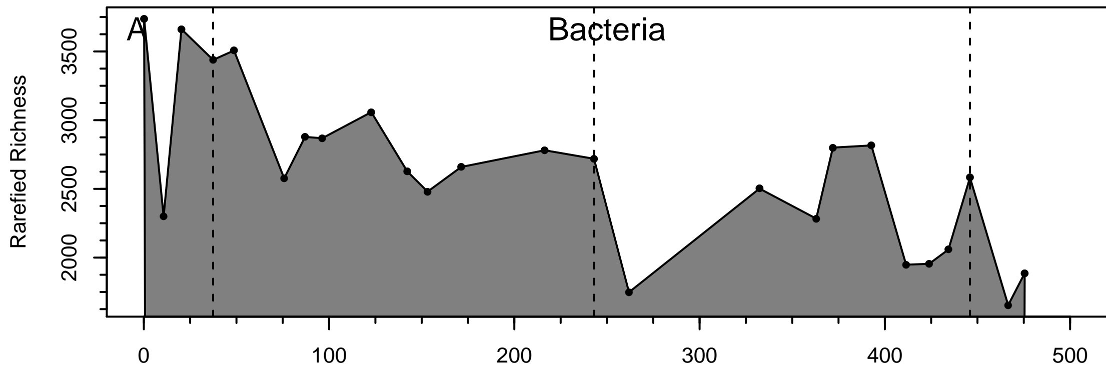


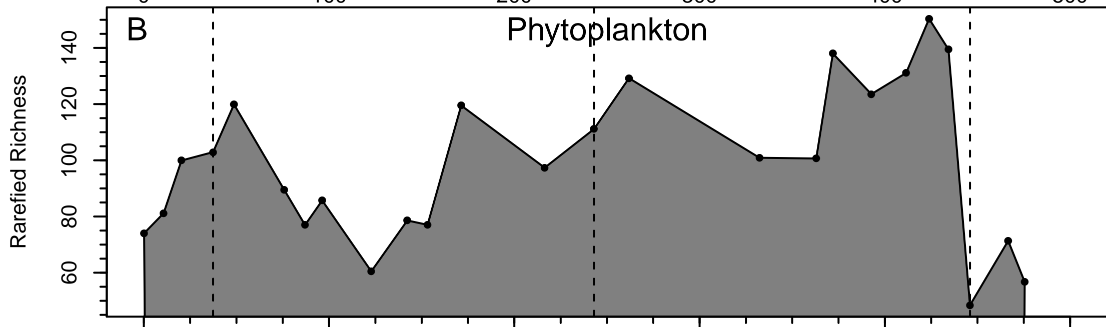
B
Elevation (m)
Mean discharge (m^3/s)
River distance (km)
0
400
800
0
100
200
300
0
100
200
300
400
500
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

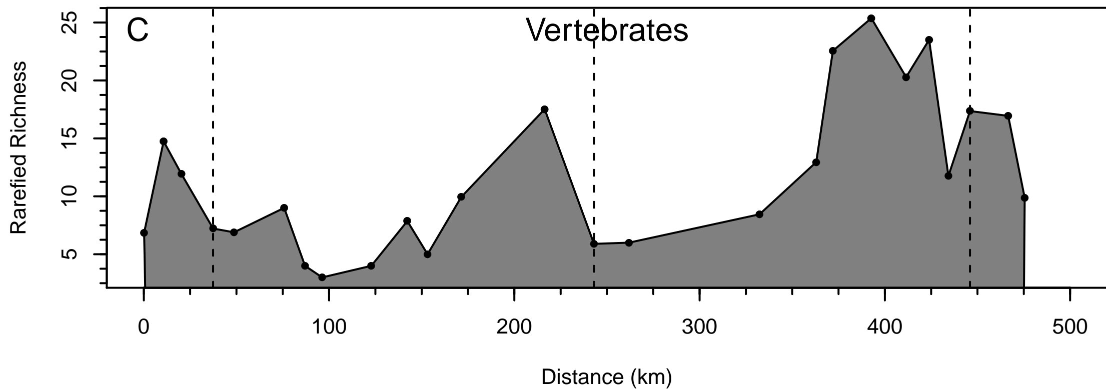

A
Proteobacteria
B
Bacteroidetes
C
Actinobacteria
D
Verrucomicrobia
E
Planctomycetes
F
Bacillariophyta
G
Eustigmatophyceae
H
Chlorophyta
I
Cyanobacteria
J
Chrysophyceae
K
Arthropoda
L
Mollusca
M
Annelida
N
Cnidaria
O
Mammalia
P
Actinopterygii
Q
AmphibiaReptilia
RRA (%)
Distance (km)

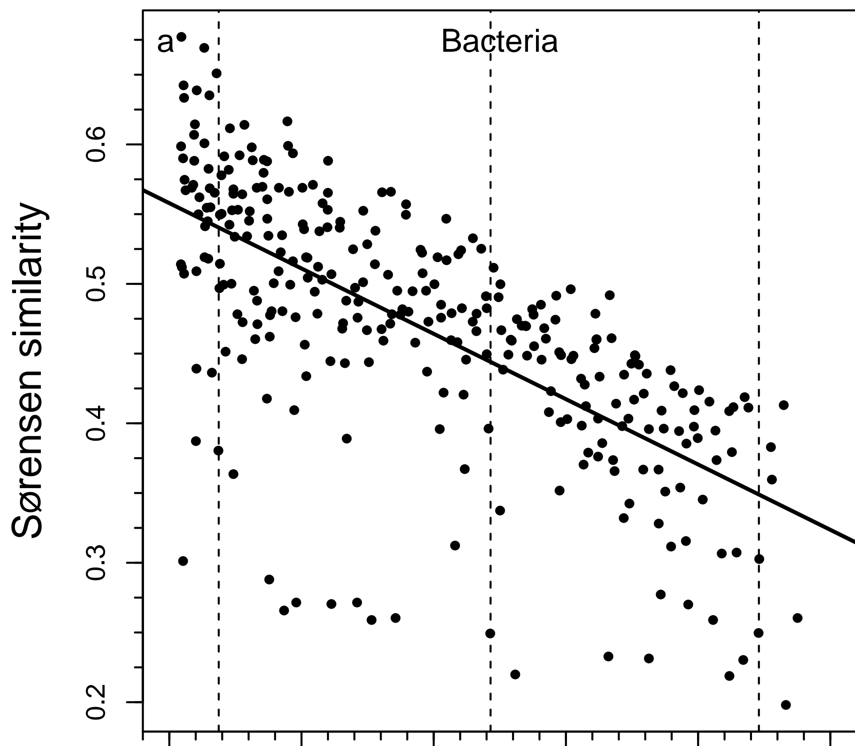


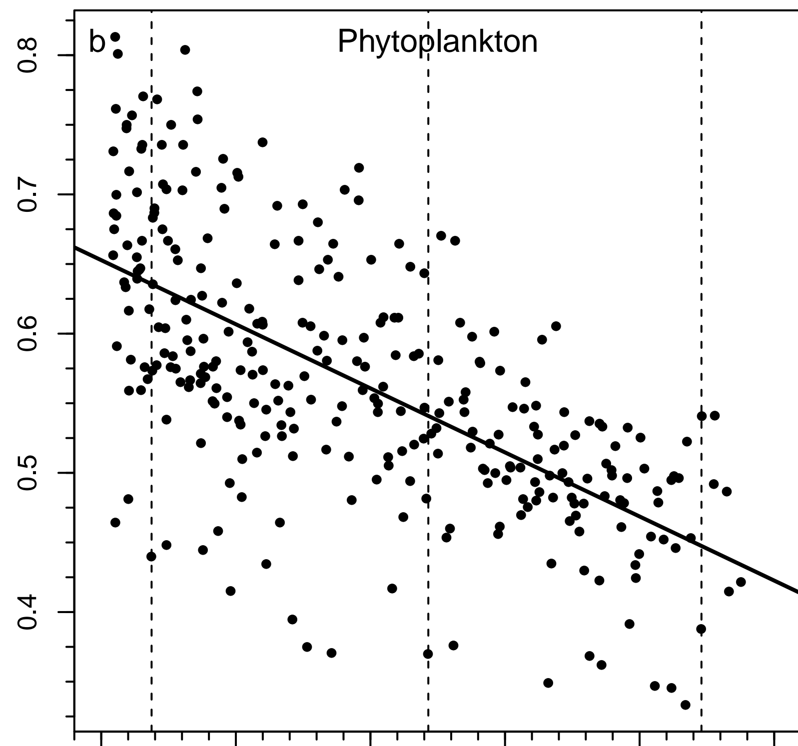
OTU2 . Proteobacteria Sphingomonadales
OTU1 . Actinobacteria Actinomycetales ACK−M1
OTU3 . Bacteroidetes Flectobacillus
OTU8 . Actinobacteria Actinomycetales ACK−M1
OTU4 . Proteobacteria Teredibacter starrii
OTU5 . Bacteroidetes Flavobacterium
OTU11 . Bacteroidetes Flavobacterium succinicans
OTU17 . Bacteroidetes Sediminibacterium
OTU6 . Proteobacteria Caulobacter henricii
OTU24 . Proteobacteria Rhodoferax
Bacillariophyta Navicula salinicola
Bacillariophyta Thalassiosira pseudonana
Bacillariophyta Nitzschia sp. BOLD:AAX5147
Bacillariophyta Bacillariophyceae sp. 1 AS−2014
Bacillariophyta Sellaphora pupula
Bacillariophyta Nitzschia sp. CCMP2626
Bacillariophyta Bacillaria paxillifer
Bacillariophyta Asterionella formosa
Streptophyta Closterium lunula
Bacillariophyta Melosira tropica
Arthropoda Cricotopus trifascia
Mollusca Physella acuta
Arthropoda Cricotopus sp. BOLD−2016
Annelida Martiodrilus sp. 5 DP−2015
Arthropoda Baetis sp. BOLD:AAA4715
Arthropoda Orthocladiinae sp. BOLD−2016
Arthropoda Isonychia sp. LJ1
Arthropoda Orthocladius sp. BOLD−2016
Arthropoda Parakiefferiella sp. BOLD−2016
Arthropoda Eupodidae sp. BOLD:AAF9191
Mammalia.Homo_sapiens
Actinopterygii.Alosa_aestivalis
Mammalia.Mus_musculus
Actinopterygii.Notropis_atherinoides
Actinopterygii.Moxostoma_carinatum
Actinopterygii.Pimephales_notatus
Actinopterygii.Lepomis_gibbosus
Actinopterygii.Catostomus_commersonii
Mammalia.Sus_scrofa
Amphibia.Anaxyrus_terrestris
RRA (%)
Distance (km)

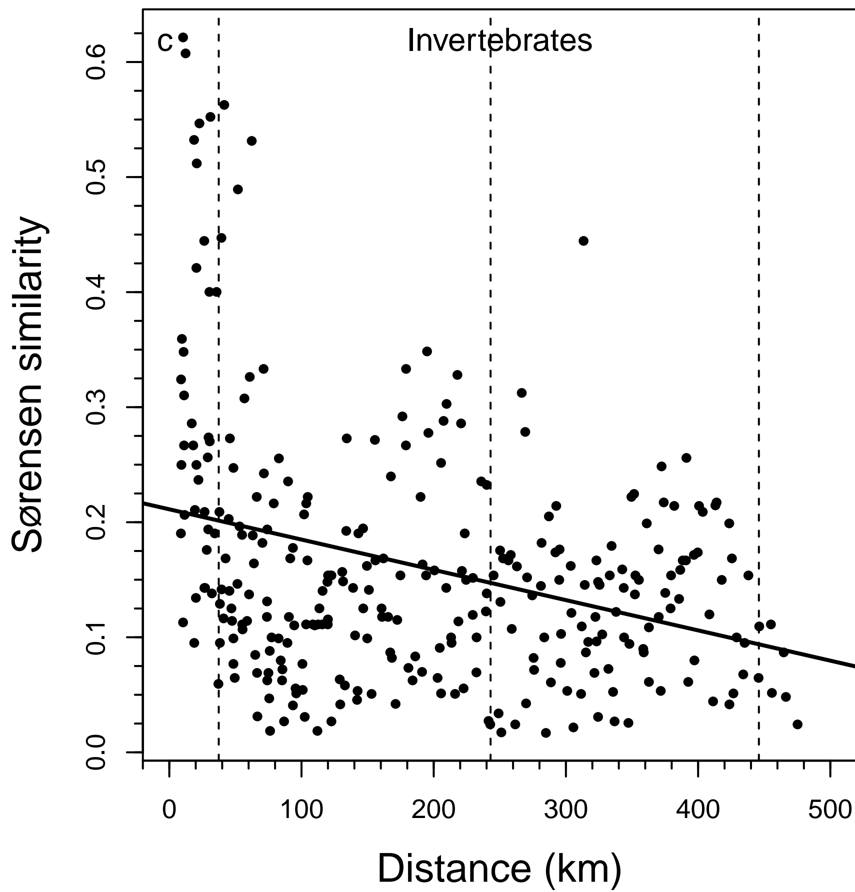

A
Bacteria
Rarefied Richness
2000
2500
3000
3500
0
100
200
300
400
500

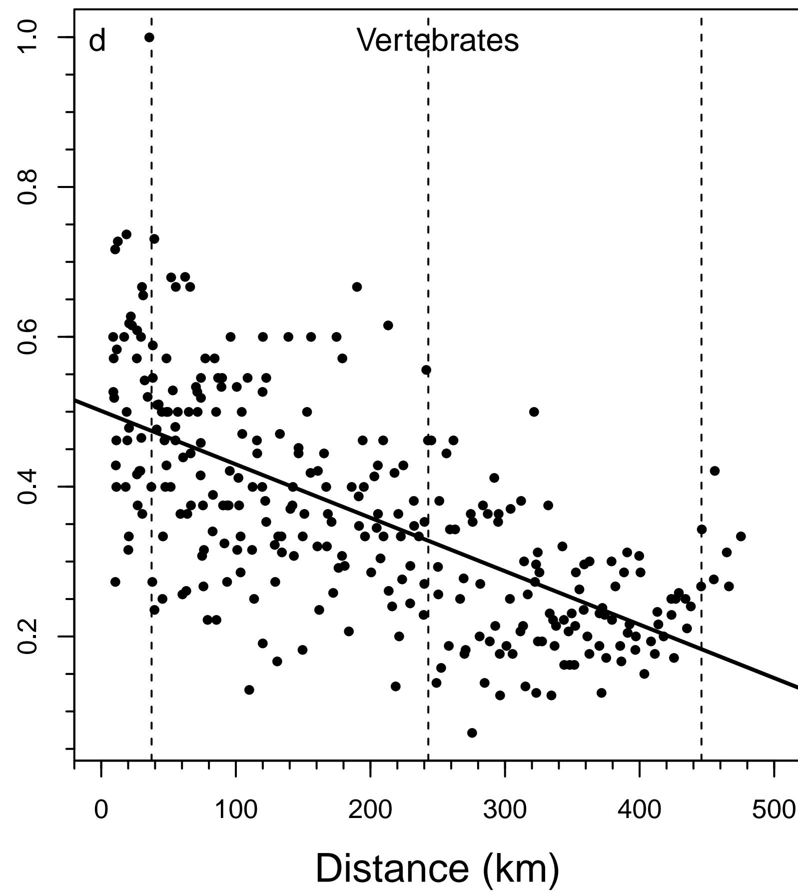

B
Phytoplankton
Rarefied Richness
60
80
100
120
140


C
Vertebrates
Rarefied Richness
5
10
15
20
25
0
100
200
300
400
500
Distance (km)


a
Bacteria
Sørensen similarity
0.2
0.3
0.4
0.5
0.6


b
Phytoplankton
0.4
0.5
0.6
0.7
0.8


c
Invertebrates
Sørensen similarity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
100
200
300
400
500
Distance (km)


d
Vertebrates
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
Distance (km)

Total Dissolved Phosphorus

A

Conc. (mg L^{-1})

0.005

0.010

0.015

0 100 200 300 400 500

Total Dissolved Nitrogen

B

Conc. (mg L^{-1})

0.4

0.6

0.8

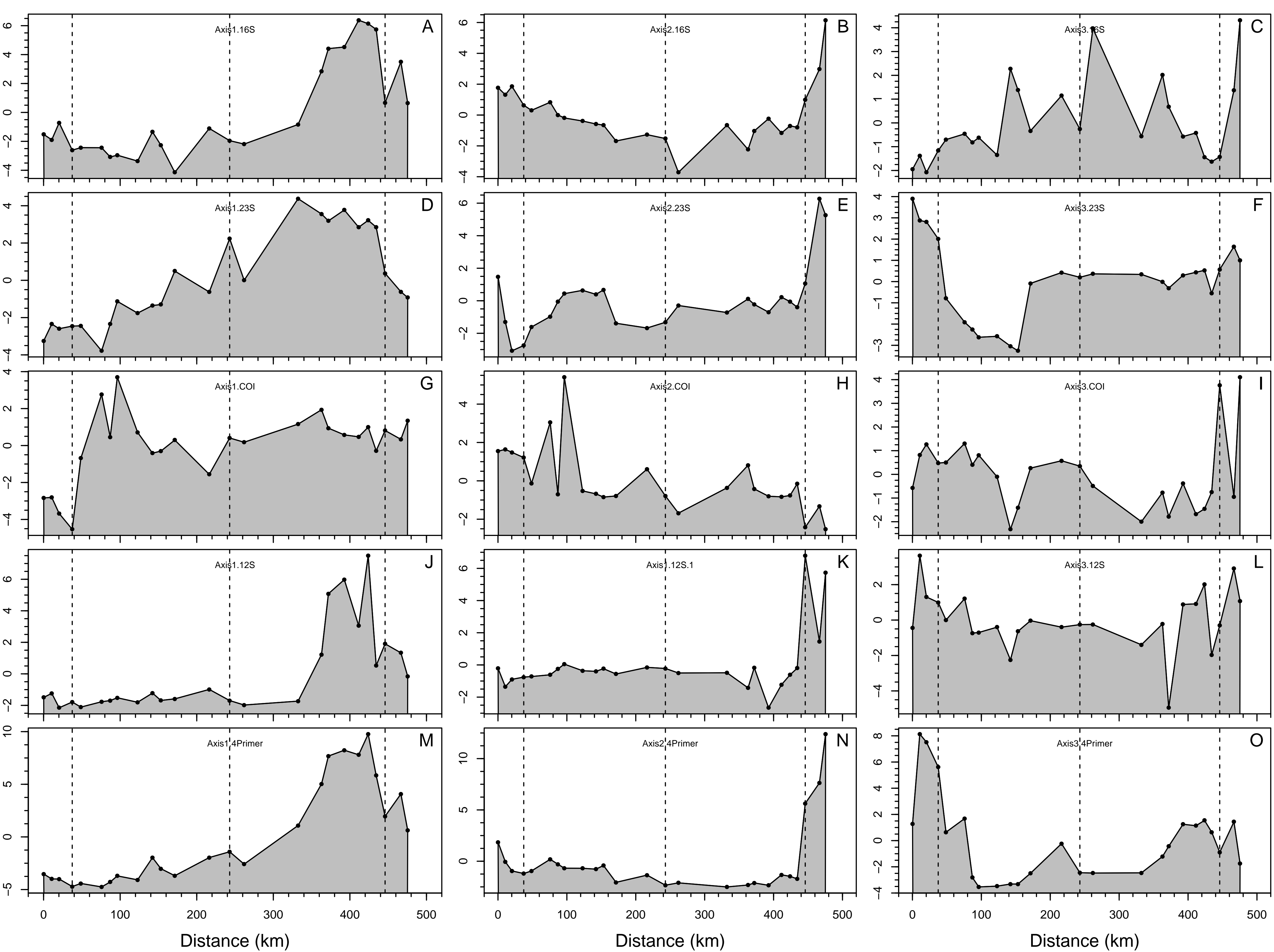
1.0

1.2

1.4

0 100 200 300 400 500

Distance (km)



Axis1.16S
A
Axis2.16S
B
Axis3.16S
C
Axis1.23S
D
Axis2.23S
E
Axis3.23S
F
Axis1.COI
G
Axis2.COI
H
Axis3.COI
I
Axis1.12S
J
Axis1.12S.1
K
Axis3.12S
L
Axis1.4Primer
M
Axis2.4Primer
N
Axis3.4Primer
O
Distance (km)
Distance (km)
Distance (km)