

1 **Title**

2 Wastewater treatment plant effluent introduces recoverable shifts in microbial community
3 composition in receiving streams
4

5 **Author names and affiliations**

6 Jacob R. Price^a, Sarah H. Ledford^b, Michael O. Ryan^a, Laura Toran^b, and Christopher M. Sales^{a*}
7

8 ^a Civil, Architectural, and Environmental Engineering, Drexel University, 3141 Chestnut Street,
9 Philadelphia, Pennsylvania 19104, United States of America

10 ^b Earth and Environmental Science, Temple University, 1901 N. 13th St, Philadelphia,
11 Pennsylvania, 19122, United States of America

12 * Corresponding Author. E-mail: chris.sales@drexel.edu. Phone: (215) 895-2155. Fax: (215)
13 895-1363.
14

15 **Corresponding author**

16 Christopher M. Sales

17 Civil, Architectural, and Environmental Engineering

18 Drexel University

19 chris.sales@drexel.edu

20 Phone: (215) 895-2155

21 Fax: (215) 525-4332
22

23 **Abstract**

24 Through a combined approach using analytical chemistry, real-time quantitative polymerase
25 chain reaction (qPCR), and targeted amplicon sequencing, we studied the impact of wastewater
26 treatment plant effluent sources at six sites on two sampling dates on the chemical and microbial
27 population regimes within the Wissahickon Creek, and its tributary, Sandy Run, in Montgomery
28 County, Pennsylvania, USA. These water bodies contribute flow to the Schuylkill River, one of
29 the major drinking water sources for Philadelphia, Pennsylvania. Effluent was observed to be a
30 significant source of nutrients, human and non-specific fecal associated taxa. There was an
31 observed increase in the alpha diversity at locations immediately below effluent outflows, which
32 contributed many taxa involved in wastewater treatment processes and nutrient cycling to the
33 stream's microbial community. Unexpectedly, modeling of microbial community shifts along the
34 stream was not controlled by concentrations of measured nutrients. Furthermore, partial
35 recovery, in the form of decreasing abundances of bacteria and nutrients associated with
36 wastewater treatment plant processes, nutrient cycling bacteria, and taxa associated with fecal
37 and sewage sources, was observed between effluent sources, which we hypothesize is controlled
38 by distance from effluent source. Antecedent moisture conditions were observed to impact
39 overall microbial community diversity, with higher diversity occurring after rainfall. Finally, the
40 efficacy of using a subset of the microbial community including the orders of Bifidobacteriales,
41 Bacteroidales, and Clostridiales to estimate the degree of influence due to sewage and fecal
42 sources was explored and verified.

43

44 **Keywords**

45 Amplicon Sequencing
46 Microbial Community Analysis
47 Water chemistry
48 Urban stream
49 Nutrients

50

51 **Abbreviations**

52 Antecedent Moisture Conditions (AMC)
53 Million gallons per day (MGD)
54 Operational Taxonomic Unit (OTU)
55 Real-Time Polymerase Chain Reaction (qPCR)
56 Ribosomal Sequence Variant (RSV)
57 Variance Stabilizing Transformation (VST)
58 Wastewater treatment plant (WWTP)

59

60 **1. Introduction**

61 Eutrophication of inland and coastal waters of the United States is well documented, and there
62 has been a major push to minimize nutrient loading from agricultural and urban areas to these
63 water bodies (Carpenter et al., 1998). As the percentage of the world population continues to
64 increase in urban areas, tracking the impacts of urbanization, including nutrient cycling, are more
65 vital than ever. At baseflow, point-sources play a large role in stream processes, especially in
66 headwater streams where wastewater treatment plant (WWTP) effluent can be a majority of
67 stream discharge (Marti et al., 2004). Increased regulation on WWTPs in the United States has
68 decreased the concentration of organic carbon and ammonia in effluent discharge, but limits on
69 nitrate and phosphate are currently less common (Carey and Migliaccio, 2009). Microorganisms
70 could be key in identifying shifts in ecosystem integrity in streams due to environmental
71 perturbations (both positive and negative), due to their sensitivity and short lifespan. While many
72 papers have looked at nitrification below WWTPs (Gücker et al., 2006; Merbt et al., 2015;
73 Merseburger et al., 2005; Sonthiphand et al., 2013), studies are needed to look at the subsequent
74 shift in nutrient processing and microbial communities below point-source pollutants to streams
75 as plants shift away from discharging nitrogen in the form of ammonia.

76
77 Even WWTPs that complete secondary treatment produce effluents with total nitrogen (TN)
78 ranging from 15-25 mg/L and total phosphorous (TP) ranging from 4-10 mg/L, and their
79 discharges often result in significant inputs of nutrients to streams (Carey and Migliaccio, 2009).
80 These large loads can either spur chemolithotrophic and heterotrophic respiration due to an
81 influx of ammonia and organic matter, respectively, or saturate the environment with nutrients,
82 lowering nutrient retention (Aristi et al., 2015; Marti et al., 2004). It is well documented that
83 WWTP effluent overfertilizes streams and subsequently reduces nutrient uptake efficiency
84 (Aristi et al., 2015; Haggard et al., 2005; Haggard et al., 2001; Marti et al., 2004), although some
85 studies have seen differences between phosphate and nitrate uptake impacts below WWTP
86 effluent outfalls (Gücker et al., 2006). Nitrification dominates nitrogen cycling below WWTP
87 outfalls (Merseburger et al., 2005; Ribot et al., 2012; Sonthiphand et al., 2013), but increased
88 regulation of ammonia in WWTP effluents may alter these cycles.

89
90 A variety of tools from micro- and molecular biology have been successfully used to identify and
91 model the influence of environmental factors on the microbial communities in aquatic
92 environments, including culturing and plate counts in selective media (Drury et al., 2013; Harry
93 et al., 2016), flow cytometry (Harry et al., 2016), denaturing gradient gel electrophoresis
94 (DGGE) (Wakelin et al., 2008), and quantifying microbial abundance with real-time polymerase
95 chain reaction (qPCR, also known as quantitative polymerase chain reaction) (Halliday et al.,
96 2014; Savichtcheva and Okabe, 2006). Sequencing clonally amplified genomic DNA (Van der
97 Gucht et al., 2005; Wakelin et al., 2008; Zwart et al., 2002) was an early method of sequencing
98 community DNA, but the advancement of sequencing technologies has eliminated the need for
99 clonal amplification. More recent methods for investigating the microbial structure of aquatic
100 systems are focused on targeted amplicon sequencing (Deiner et al., 2016; Drury et al., 2013;
101 Marti and Balcazar, 2014; Wang et al., 2016), such as those targeting the 16S or 18S rRNA
102 genes for studying taxonomic composition, and metagenomic sequencing (Hladilek et al., 2016)
103 which aims to investigate both the taxonomic composition and functional capabilities of an entire
104 microbial community.

105

106 Most previous studies that use environmental DNA (eDNA) to investigate the impact of WWTP
107 effluent on streams have focused specifically on ammonia-oxidizing assemblages (Merbt et al.,
108 2015; Sonthiphand et al., 2013), denitrifying bacterial communities (Rahm et al., 2016),
109 sediment bacterial communities (Drury et al., 2013), or focus on different pollutant sources
110 (Ibekwe et al., 2016). In this study, we aim to integrate information from multiple sources
111 including traditional stream water chemistry, real-time polymerase chain reaction (qPCR), and
112 targeted amplicon sequencing to get a more complete, though qualitative due to limited sample
113 size, microbial overview of the impact of WWTP discharge on urban streams. To do this we
114 targeted an urban stream, Wissahickon Creek, outside of Philadelphia, PA, in an urban and
115 suburban setting with four WWTPs discharging effluent. Water quality in the Wissahickon is
116 important to the City of Philadelphia as one of the city's drinking water intake pipes is located
117 approximately half of a mile downstream of the confluence of the Schuylkill River and
118 Wissahickon Creek (PWD, 2007). Samples were collected above and below the plants on two
119 dates and analyzed for water chemistry, qPCR, and targeted amplicon sequencing.

120

121 **2. Materials and methods**

122 **2.1. Site description**

123 Wissahickon Creek is a third order, headwater stream that flows through Montgomery and
124 Philadelphia counties, Pennsylvania, USA, before discharging into the Schuylkill River. The
125 main stem is 43.5 km in length, with a total system drainage length of 184.6 km that drains 164.9
126 km² of suburban and urban land use (PWD, 2007). There are four WWTPs that discharge
127 effluent into the Wissahickon: Upper Gwynedd, Ambler, Upper Dublin, and Abington. They
128 vary in volume from an average of 0.76 to 3.64 MGD, although some are permitted to discharge
129 up to 6.5 MGD (Figure 1, Table S1). All four plants complete secondary treatment on their
130 effluent before discharge. Six sampling locations were established above and below the
131 treatment plants for this study: upstream Upper Gwynedd (1.USUG) and downstream Upper
132 Gwynedd (2.DSUG) were each approximately 200 m above and below the Upper Gwynedd
133 WWTP effluent channel; upstream Ambler (3.USAmb) and downstream Ambler (4.DSAmb)
134 were 400 m above and 600 m below the Ambler WWTP effluent channel; and below the
135 confluence of the main stem with the largest tributary, Sandy Run (5.BC) to provide a sample
136 point on the main stem downstream of WWTPs on Sandy Run. Additionally, samples were
137 collected on Sandy Run (6.SR), located 1.2 km below Upper Dublin WWTP and 5.1 km below
138 Abington WWTP. Sampling dates can be differentiated throughout the rest of this manuscript by
139 the number at the end of the site code, with the May 10, 2016 sample indicated by '.1' and the
140 May 17, 2016 sample indicated by '.2'.

141

142 The contributing area to each sampling site has similar land use (Table S2), with approximately
143 65% in various levels of development and an additional 20% from deciduous forests (Homer et
144 al., 2015). There is an established riparian buffer of 50-100 m width along the majority of the
145 main stem of the Wissahickon, limiting the immediate quantity of overland runoff during storms.
146 At baseflow, the only sources of water to the stream are groundwater, wastewater treatment plant
147 effluent, and tributary discharge. For example, effluent is estimated to consist of 30% of total
148 discharge at 2.DSUG. Although five tributaries enter the main stem between 2.DSUG and
149 3.USAmb, sampling indicates that at baseflow they contribute negligible discharge, and thus
150 have minimal impact on the chemistry and microbial community in the Wissahickon.

151

152 **2.2. Field sampling and water quality analysis**

153 Samples were collected from each site on the mornings of May 10 and May 17, 2016. While all
154 samples were collected at baseflow, antecedent moisture conditions (AMC) were wetter on the
155 May 10, 2016 (Figure S1). The antecedent precipitation index (API) can be used to contrast the
156 two dates, where $API = \sum_{t=-1}^{-i} P_t k^{-t}$ (Ali et al., 2010). Taking t as the 10 days prior, P as the
157 amount of precipitation on each day, and using a recession constant, k, of 0.9, May 10 has an
158 API of 110 and May 17 has an API of only 12. Water samples for chemical analysis were
159 collected from the thalweg in two acid-washed HDPE bottles, after being filtered through 0.45
160 μm Millipore cellulose filters. One bottle was acidified to approximately $\text{pH} = 2$ with nitric acid
161 and neither bottle had head space. Samples were stored at 4°C and analyzed within one week.
162 Ions and elements were chosen for analysis due to their common occurrence in natural waters
163 (SO_4^{2-} , Cl^- , NO_3^{2-} , Ca, Mg, Na, and K), potential to be tracers of wastewater (F^- , Sr, Cu, and Mn),
164 or known to be necessary for microbial growth (Fe, NO_2^- , TDP, and Si). Fluoride (F^-), chloride
165 (Cl^-), bromide (Br^-), nitrite (NO_2^- -N), nitrate (NO_3^- -N), and sulfate (SO_4^{3-}) in the non-acidified
166 samples were analyzed on a Dionex ICS-1000 ion chromatograph using three standards, while
167 calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), total iron (Fe), total dissolved
168 phosphorus (TDP), silicon (Si), copper (Cu), manganese (Mn), and strontium (Sr) were measured
169 in elemental quantities in the acidified samples on a Thermo Scientific iCAP 7000 ICP-OES
170 using three standards. Previous work in the field site indicated that at baseflow orthophosphate
171 comprised all of the TDP and there was minimal particulate phosphorus in the stream water (data
172 not shown). National Pollutant Discharge Elimination System (NPDES) permit reporting data
173 was available from Upper Gwynedd WWTP for orthophosphate on both sample days. NPDES
174 nutrient data for the other plants were not available on the sampling days so an average of
175 WWTP effluent collected and analyzed as described above on three different dates (August 28,
176 2016; October 15, 2016; February 25, 2017) was reported. A USGS gage, station number
177 01473900, was established at the 5.BC site and was used for discharge information (Fig S1).

178
179 Water samples were also collected to characterize the microbial community in the streams. Three
180 1L replicate samples were collected at each site in autoclaved polypropylene bottles. Samples
181 were collected from the thalweg, facing upstream, and while wearing gloves. The bottles were
182 rinsed with stream water three times prior to sample collection to remove any carryover that may
183 have occurred during cleaning or in transit. Samples were transported to the laboratory within six
184 hours for DNA extraction.

185 186 **2.3. DNA extraction and quantification**

187 Genomic DNA was extracted from each of the samples using the QIAamp DNA Stool Mini Kit
188 (Qiagen) with modifications to the kit's pathogen detection method and optimized for our
189 protocol (Ryan et al., 2013). 250 to 500 mL of stream sample was filtered through 0.45 μm
190 gamma-sterilized, individually wrapped disposable filter kits containing cellulose nitrate
191 membrane. Membranes were then placed in InhibitEX buffer and incubated for 5 min at 95°C
192 while being homogenized at 900 rpm using a Thermomixer to ensure that cells were fully lysed.
193 Sample tubes were then centrifuged for 1 minute at 14,000 rpm and 200 μl of the supernatant
194 was transferred to the Spin Column. The remaining extraction steps were carried out
195 autonomously via QiaCube according to the manufacturer's instructions. Resulting gDNA
196 concentrations were assessed via a QuBit 2.0 fluorometer. Laboratory blanks were included at
197 each step, and were found to be negative for all steps.

198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243

2.4. Real-time PCR (qPCR)

The qPCR MST protocol outlined by Ryan (2012) was used in this study to evaluate human specific pollution. This protocol used forward HF68, (5'-GGC AGC ATG GTC TTA GCT TG-3') and reverse HF183rc (5' – CGG ACA TGT GAA CTC ATG AT – 3') primers in a SYBR Green qPCR assay to assess concentrations of *Bacteroides dorei*. The genus *Bacteroides* was shown to be selective for and present in very high abundance in humans (Ahmed et al., 2009) and *B. dorei* to be one of the more selective specie (Ryan et al., 2013). A universal *Bacteroides* spp. protocol was used to assess total enteric pollution (Shanks et al., 2010; Siefring et al., 2008). In all cases, the 16s ribosomal RNA gene was targeted due to its conservative nature. qPCR primers and targets are summarized in Table S3.

A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. All assays used the Roche LC480 SYBR Green Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20 μ L. Each reaction mixture contained 5 μ L of template DNA, and 0.5 μ M of the forward and reverse primers. The program employed: pre-incubation for 5 min at 95°C; 40 amplification cycles of 30 s of annealing at 60°C and 10 s of denaturing at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, and each included no template (negative) controls along with positive controls. Qiagen PCR Cloning Plus kits were used to produce plasmids with inserts of amplicons positively identified by DNA sequencing as *B. dori* isolated from assays of WWTP influent samples. These were used as the positive controls.

2.5. Amplicon library preparation and sequencing

Amplicon libraries, targeting 16S rDNA, were created for each of the samples following the Earth Microbiome Project's (Gilbert et al., 2010) 16s rRNA amplification protocol (Caporaso et al., 2012). Primers 515F-Y (Parada et al., 2015; Quince et al., 2011) and 926R (Parada et al., 2015; Quince et al., 2011) were selected to provide coverage of two hypervariable regions (V4 and V5) within the 16s gene. PCR reactions for amplicon library preparations were carried out using a single PCR reaction for each extracted DNA aliquot. Each 25 μ L PCR reaction volume was comprised of 12.5 μ L HotStarTaq Plus Master Mix, 1.5 μ L of primer pre-mix containing both forward and reverse primers at a concentration of 5 μ M, 1 μ L of molecular biology-grade water, and 10 μ L of gDNA. The thermocycler program entailed a 5 minute at 95° C heat-activation step; 25 replication cycles of 94° C for 45 seconds, 50° C for 45 seconds, and 72° C for 90 seconds; a final extension was carried out at 72° C for 5 minutes. PCR triplicates, three aliquots per sample, were pooled after the amplification step, then 4 μ L from each of the 12 pools were then run on a 1% agarose gel to test for amplification and correct amplicon product size. The pools were purified and concentrated with a Qiagen QIAquick Purification kit and the product quantified via a QuBit 2.0 fluorometer. Equal product masses for each sample were combined into a single tube and submitted for paired-end sequencing (2 x 300 bp) on an Illumina MiSeq Sequencer. Raw read files were uploaded to the National Center for Biotechnology Information (NCBI) Sequencing Read Archive (SRA) and are accessible under the SRA Study Accession ID SRP103534 and BioProject Accession ID PRJNA382371; Table S4 contains a list of sample names and their corresponding BioSample Accession IDs.

2.6. Bioinformatics pipeline

244 Data acquired through targeted amplicon sequencing were analyzed using a newly described
245 bioinformatics pathway (Callahan et al., 2016b) exploiting a variety of packages available in the
246 R programming environment (version 3.3.1) (R Development Core Team, 2015). Raw reads
247 were subjected to trimming and filtering, forward and reverse reads were truncated at base pair
248 275 and 175 respectively; the first 10 bp of each read was also removed from both forward and
249 reverse reads. After trimming, the reads were filtered, during which reads containing ambiguous
250 bases (“N”) or an expected error (EE) exceeding 2 were removed. Filtering for PhiX
251 contamination was also carried out during this step.

252
253 The R package dada2 (version 1.3.0) (Callahan et al., 2016a) was used to dereplicate the
254 sequences passing filter, carry out error model parameter learning, and the inference of true
255 ribosomal sequence variants (RSV) for each of the samples. Operational taxonomic units (OTU)
256 are traditionally generated by clustering reads above a certain identity (often 97% for species-
257 level assignment). While RSVs and OTUs represent different approaches to the problem of
258 organism identification for OTU table building, we use the terms interchangeably herein.
259 Forward and reverse sequence variants were merged; any forward/reverse pair that contained a
260 mismatch in their overlapping region was removed. Chimeras were detected and removed from
261 the resulting sequences. Taxonomy was assigned to each of the RSVs using the naïve bayes
262 classifier (Wang et al., 2007) aligning to the SILVA database (release 123) (Pruesse et al., 2007).
263 Taxonomic assignments for the ranks of kingdom through genus required a minimum bootstrap
264 support of 80 (out of 100) to be retained. Taxonomic assignment at the species rank required an
265 exact match to the SILVA database and the (previously) annotated genus was required to match
266 the genus of the exact match. In the event that a sequence was an exact match for multiple taxa
267 within the same genus, the species names of all matching taxa were concatenated together. A
268 phylogenetic tree was constructed using the DECIPHER (version 2.0.2) (Wright, 2015) and
269 phangorn (version 2.0.4) (Schliep, 2011) packages as described by Callahan et al. (2016b).

270
271 To place the samples into ecological context, a tab delimited file was created containing
272 metadata describing the samples, their location, the water chemistry in the stream water, and
273 descriptors of the land use of the contributing area to each sampling location determined from
274 2011 National Land Cover Database (Homer et al., 2015). This metadata file was imported into
275 R, and formatted for use, including setting factor levels and carrying out transformation of
276 metadata values including the log10 transformation of chemical concentrations, following
277 Palmer (1993). To prevent complications arising from taking the log of zero, measurements of
278 nitrite and bromide below the minimum detection limit (MDL) for each method were assigned a
279 value of half the MDL of the individual IC run. The four products from this bioinformatics
280 pipeline, the OTU table (more precisely, RSV table), taxa table, phylogenetic tree, and sample
281 data frame were then merged into a phyloseq object (version 1.16.2) (McMurdie and Holmes,
282 2012).

283
284 All scripts used for bioinformatics and ecological analyses are available at the author’s GitHub
285 repository (located at https://github.com/JacobRPrice/WWTP_Impact_on_Stream).

286 287 **2.7. Statistical analysis**

288 Preprocessing and filtering of the phyloseq object was carried out to reduce error and noise in the
289 dataset by removing taxa that met any of the following three criteria: 1) taxa with zero counts; 2)

290 taxa with ambiguous taxonomic assignment at the Kingdom or Phylum level; 3) taxa that were
291 seen less than 2 times in 2 or more samples.

292
293 Exploratory (unconstrained) ordination with principal component analysis (PCA) (Hotelling,
294 1933; Pearson, 1901) and double principal coordinate analysis (DPCoA) (Pavoine et al., 2004;
295 Purdom, 2011) was carried out to identify potential patterns and relationships within the
296 microbial community and the location at which the sample was collected. DPCoA was selected
297 because it enables the inclusion of phylogenetic distance between taxa. Prior to ordination, a
298 variance stabilizing transformation (VST) (Love et al., 2014; McMurdie and Holmes, 2014) was
299 carried out on OTU counts. The axes of the ordination plots within this manuscript have been
300 scaled to more accurately represent the distances between samples and/or taxa, as described by
301 Callahan et al. (2016b).

302
303 Identification of microbes with differential abundance was carried out using the DESeq2 (version
304 1.12.4) (Anders and Huber, 2010; Love et al., 2014; McMurdie and Holmes, 2014) R package.
305 To account for potential false positives, p values for differentially abundant taxa were adjusted
306 via the Benjamini and Hochberg false discovery rate (FDR) correction for multiple testing
307 (Benjamini and Hochberg, 1995). FDR correction is carried out by default within the DESeq2
308 functionality. Taxa were considered to have significantly different abundance if their adjusted p
309 value was below 0.05 (5%) and their FDR was below 0.10 (10%).

310
311 Microbial community composition was linked to stream water chemistry parameters using the
312 construction of a distance-based redundancy model (db-RDA) (Legendre and Anderson, 1999;
313 McArdle and Anderson, 2001). Variables were selected for the model using automated forward
314 selection (Blanchet et al., 2008) methods available in the vegan R package (version 2.4.1)
315 (Oksanen et al., 2016); a p-value < 0.05 was required for inclusion in the model. The significance
316 of the overall model as well as each individual term was determined through ANOVA-like
317 permutation tests (with 999 permutations) available as the `cca.anova()` function within the vegan
318 R package.

319
320 Finally, the BIOENV procedure (Clarke and Ainsworth, 1993) was used to identify potential (or
321 optimal) subsets of environmental variables which best account for the observed variation in
322 community composition. Carrying out variable selection using two different methods, forward
323 selection db-RDA and the BIOENV procedure, allows for comparison of the resulting models.

324
325 A second round of analysis was carried out to evaluate the contribution of WWTP effluent
326 outflows on community composition. RSV's falling under the orders of Bifidobacteriales,
327 Bacteroidales, and Clostridiales were subjected to the same analysis outlined above. These three
328 orders have been suggested as indicators of human fecal contamination (McLellan et al., 2010)
329 and have been applied to studying natural (environmental) and anthropogenic influences on fecal
330 indicator bacteria (FIB) exceedance events (Halliday et al., 2014; Savichtcheva and Okabe,
331 2006).

332 333 **3. Results**

334 **3.1. Downstream dilution in water chemistry**

335 Nitrate concentrations in the surface water ranged from 0.9 at 1.USUG on May 17, 2016 to 10.2
336 mg N/L at 2.DSUG on May 17, 2016, while average WWTP effluent concentrations were much
337 higher, averaging 17.1 mg N/L at Abington to 25.6 mg N/L at Upper Gwynedd (Figure 2A).
338 WWTP discharge on the main stem resulted in an increase in nitrate in surface waters, but the
339 longer distance between the treatment plants on Sandy Run and the sampling site allowed for
340 retention or removal of nitrate before 6.SR. The decrease in concentration between 2.DSUG and
341 3.USAmb is indicative of a combination of dilution from incoming tributaries and groundwater,
342 assimilatory nitrate uptake, and denitrification along this reach. The decrease with distance
343 continued below the confluence of Sandy Run with the main stem on both dates, but is primarily
344 controlled by mixing between the two sources. May 17, 2016 had higher surface water nitrate
345 concentrations, most likely due to lower discharge and drier AMC (Figure S1). Although there
346 is some variation in nutrient concentrations coming from plant effluent, this system has been in
347 equilibrium for a while (none of the infrastructure is new or has changed in the recent past).
348 Thus, the system is likely at relative temporal equilibrium.

349
350 Total dissolved phosphorus concentrations in grab samples ranged from 0.04 to 0.31 mg/L
351 (Figure 2B). Treatment plant effluent had much higher concentrations of TDP, with larger
352 variation through time than nitrate, however the effluent had lower P concentrations than average
353 secondary treatment plants as reported in Carey and Migliaccio (2009). Upper Gwynedd WWTP
354 had 0.15 mg/L of orthophosphate on May 10, 2016 and 0.19 mg/L on May 17 and thus did not
355 result in a large increase in surface water TDP at 2.DSUG. Ambler WWTP had higher average
356 TDP of 0.81 mg/L, and both sampling dates had a more pronounced increase in TDP at
357 4.DSAmb. Upper Dublin and Abington WWTPs had similar average concentrations, between 0.9
358 and 1.4 mg/L, and while there was some attenuation of P along Sandy Run, 6.SR consistently
359 had the highest surface water P of 0.3 mg/L. During the first sample collection, an unknown
360 source of water, sediment sorption, or biological removal decreased P by 5.BC, but the expected
361 increase in P from mixing at the same station was observed on May 17, 2016. Drier antecedent
362 moisture conditions result in slightly higher TDP concentrations at most sites on May 17, 2016.

363
364 Chloride concentrations in the stream increased below each WWTP outfall (Figure S2). The
365 cities contributing to the treatment plants do not fluoridate their drinking water, explaining the
366 lack of increase in fluoride below effluent outfalls. Silicon increases between 2.DSUG and
367 3.USAmb, where small tributaries flowing into the main stem have been observed to have silicon
368 concentrations as high as 8.7 mg/L. Ambler is the only WWTP that is a source of silicon to the
369 stream. Silicon concentrations decrease with drier antecedent moisture conditions. In the
370 upstream reach, strontium concentrations are controlled by Upper Gwynedd, which has high
371 effluent concentrations. Below this plant, tributaries have varying strontium concentrations, from
372 0.18 to 0.51 mg/L, resulting in a small decrease in concentration with distance. Strontium
373 concentrations increase with dry antecedent moisture conditions due to concentrative effects.
374 Sulfate concentrations are controlled by Upper Gwynedd, which has very high (>400 mg/L)
375 effluent concentrations. This causes a large increase in sulfate at 2.DSUG, and the rest of the
376 distance along the stream is dilution of this signal as discharge increases.

377 378 **3.2. Human indicators in real-time PCR**

379 As anticipated, the real-time PCR (qPCR) results revealed elevated concentrations of *B. dorei* at
380 sites located immediately downstream of wastewater effluent outflows (2.DSUG, 4.DSAmb) in

381 comparison to their upstream counterparts (1.USUG, 3.USAmb) (Figure 3A). *Bacteroides* spp.
382 have been frequently used in culture-independent, 16S rRNA gene based methods for microbial
383 source tracking (MST) (Ahmed et al., 2008; Bae and Wuertz, 2012; Bernhard and Field, 2000;
384 Shanks et al., 2007). Ryan et al. (2013) showed the *B. dorei* specie assay to be highly specific to
385 human host pollution. A universal/non-human-specific *Bacteroides* qPCR protocol was also used
386 to determine total enteric pollution. The non-human-specific assay mirrors the general trend
387 shown by the human specific assay (Figure 3B). Welch two sample T-test's (with unequal
388 variances) confirmed that the abundance of *B. dorei* ($p < 0.001$) and *Bacteroides* spp. ($p < 0.001$)
389 were significantly higher in the downstream sites (2.DSUG, 4.DSAmb in comparison to
390 1.USUG, 3.USAmb). The 5.BC site shows the influence of the 6.SR input in that there are higher
391 concentrations during dry weather. However, during wet weather, there was little observed
392 influence from the 6.SR input to the mainstem. While both the Upper Gwynedd and Ambler
393 outflows impacted *Bacteroides* spp. abundances, Upper Gwynedd had a larger effect. These
394 results are most easily explained by contribution of wastewater effluent to the stream flows due
395 to the differences in volumetric flow rates at each of these sites; both sites have similar average
396 daily discharge (Upper Gwynedd = 2.31 MGD and Ambler = 3.64 MGD, Table S1), but Upper
397 Gwynedd is located in the headwaters of the stream, where effluent makes up a larger percentage
398 of total flow at baseflow. The Wissahickon has somewhat higher discharge at Ambler due to
399 approximately six small tributaries that flow into the main stem during baseflow between Upper
400 Gwynedd and Ambler, plus an unquantified addition of groundwater. Furthermore, the 2.DSUG
401 site was closer to the Upper Gwynedd WWTP than the other downstream sites (200 m
402 downstream in contrast to 600 m and 1 to 5 km downstream for the 4.DSAmb and 5.SR sites).
403 Antecedent moisture conditions seem to have the strongest impact at 2.DSUG, with higher copy
404 numbers during drier conditions, although *Bacteroides* spp. shows the same temporal increase at
405 5.BC and 6.SR, although to a lesser degree.

406

407 **3.3. Influence of effluent on the diversity and phylogenetic abundance of the microbial** 408 **community**

409 Sequencing depth ranged from 74,000-120,000 raw reads per sample, and averaged 100,500
410 reads (Table S5). Rarefaction curve plots indicate that each of the samples has been well sampled
411 (Figure S3). To prevent underestimation, alpha diversity was assessed prior to any filtering/pre-
412 processing. Alpha diversity was quantified using (un-rarified) observed taxa (species richness),
413 Chao1 (Chao, 1984), and Shannon (Shannon, 1948) diversity indices (Figure 4, Table S6).

414

415 Comparing upstream/downstream pairs (1.USUG and 2.DSUG; 3.USAmb and 4.DSAmb),
416 effluent sources appear to increase diversity within the stream (Figure 4, Table S6, Chao1
417 diversity values). Averaging both days, the Upper Gwynedd WWTP appears to have a larger
418 impact on diversity increasing the stream's alpha diversity by 628 taxa, while the Ambler
419 WWTP caused an increase of 113 taxa. A second pattern emerges when comparing the alpha
420 diversity between the two WWTPs (between 2.DSUG 3.USAmb), where diversity decreases by
421 approximately 854 taxa between the two effluent sources. This reduction in diversity, 854 taxa
422 between 2.DSUG and 3.USAmb, exceeds the additional 628 taxa contributed to the stream by
423 the Upper Gwynedd WWTP, indicating that, even with the additional taxa being contributed to
424 the stream by the wastewater effluent, diversity was lost as a result of the effluent source.
425 Samples from Sandy Run (6.SR) had some of the highest diversities for all three measures. This
426 high diversity appears to account for the increase in diversity between 4.Amb and 5.BC.

427
428 With the exception of the Shannon diversity measure for 1.USUG, the samples from the second
429 sampling date were observed to have lower diversity than the first sampling date. This may be
430 explained by the wetter antecedent moisture condition for the first date (Figure S1), as runoff
431 may carry additional microbial sources into the stream and contribute to the disturbance of
432 sediments within the streambed.

433
434 A total of 5,539 RSVs were present in the 12 samples prior to any preprocessing or filtering
435 steps, including 44 singletons. Un-annotated taxa, to some degree, are not unexpected in
436 environmental samples, but the absence of taxonomic assignment makes it difficult to interpret
437 results. To avoid this, RSVs that were not annotated at the kingdom or phylum level (481 RSVs)
438 were removed from the dataset. In an effort to reduce noise and error, only taxa seen more than
439 twice in two independent samples were kept in the final dataset; 3,925 RSV's remained in the
440 final set.

441
442 At the phylum level, Bacteroidetes and Proteobacteria comprise the vast majority of the
443 microbial community, representing roughly 75 to 85% of the reads (Figure 5). Proteobacteria
444 appear to dominate in sites with lower discharge (1.USUG, 2.DSUG, and 6.SR) and
445 Bacteroidetes is more abundant in sections with higher discharge (3.USAmb, 4.DSAmb, and
446 5.BC). Cyanobacteria and Verrucomicrobia were observed to generally be the 3rd and 4th most
447 abundant phyla, respectively. Cyanobacteria were more abundant at Upper Gwynedd, both up
448 and downstream of the WWTP outflow, and greater than in any other location.

449
450 Although they do not track perfectly, the relative abundance of reads attributed to *B. dorei*
451 (Figure 3C) and *Bacteroides* spp. (Figure 3D) generally followed the abundances found via
452 qPCR (Figure 3A and 3B). The differences observed between the qPCR and amplicon results
453 may be due to primer bias either during qPCR or amplicon library preparation or limitations of
454 the taxonomic database used for assignment. It should be noted that absolute quantification of
455 target loci using qPCR is a more accurate measure of the abundance of specific taxa in
456 comparison to relative abundances obtained from sequencing efforts. Effluent outflows (between
457 1.USUG and 2.DSUG and between 3.USAmb and 4.DSAmb) appear to be the source of both
458 human associated and general fecal-associated bacteria. Between 2.DSUG and 3.USAmb, the
459 abundance of these taxa is reduced a significant extent. 2.DSUG shows the strongest increase
460 with drier antecedent moisture conditions, but unlike the qPCR results, most other sites have
461 higher relative abundances on the wetter sampling date. Both of these observations complement
462 conclusions drawn from the qPCR results. Samples from Sandy Run (6.SR) have a much higher
463 relative abundances of *B. dorei* and *Bacteroides* spp. than those found below the confluence of
464 the Wissahickon Creek and Sandy Run (5.BC), suggesting that the larger flow of the
465 Wissahickon is diluting their abundance.

466
467 **3.4. Contrasts between wet and dry weather samples using exploratory ordination**
468 Shifts in sample placement are observed in the PCA plot, indicating relative influence of WWTP
469 inputs (Figure 6). Within the PCA plot, the influence of WWTP effluent outflows appeared to
470 shift the sample placement in the same direction (to the left and upwards). The wet and dry
471 conditions shifted about the same at Upper Gwynedd but the shift was lower under dry
472 conditions for Ambler. The influence of Sandy Run also changes with antecedent moisture

473 conditions. The 5.BC.2 taxa fell between 6.SR.2 and 4.DSAmb.2 in dry conditions, suggesting a
474 mixing relationship. In wet conditions, there was no such mixing relationship and the 5.BC.1
475 sample falls to the right of 4.DSAmb.1. The 4.DSAmb sample drives this shift and it differs
476 between the two sampling dates although they are only 7 days apart. Microbial cohorts within the
477 three most abundant phyla drove the distinctions between samples, as shown in double principal
478 coordinate analysis (DPCoA) (Figure S4), and, as in Figure 6, sample placement appears to be at
479 least partly determined by the influence of WWTP effluent outflows.

480

481 **3.5. Differential abundance testing**

482 **3.5.1. Differential abundance within the microbial community up and downstream of** 483 **effluent sources**

484 Differential abundance confirmed the influence of the WWTPs on taxa observed. The R package
485 DESeq2 (Anders and Huber, 2010; Love et al., 2014) was used to test for differential taxa
486 abundance between upstream and downstream positions, while controlling for location. For this
487 analysis, only samples located at Upper Gwynedd and Ambler were used, as upstream samples
488 were not collected for the Sandy Run site (SR.6) and possible noise induced from Sandy Run's
489 contribution below the confluence (BC.5) could confound the results. 137 RSVs were found to
490 be differentially abundant (Figure S5); Bacteroidetes (36), Cyanobacteria (8), and Proteobacteria
491 (76) account for the vast majority of these taxa. The increase in Bacteroidetes and Proteobacteria
492 are consistent with the trends identified in the relative abundance (Figure 5) and DPCoA analysis
493 (Figure S4).

494

495 Conceptualizing log₂-fold changes can be difficult, especially for multiple taxa. To assist
496 interpretation, the relative abundance of the differentially abundant taxa is presented in Figure 7
497 and the decimal and percentage values are presented in Table S7. The differences in abundances
498 between upstream (1.USUG and 3.USAmb) and downstream (2.DSUG and 4.DSAmb) sites are
499 striking. At the Upper Gwynedd WWTP the differentially abundant cohort was approximately 7
500 times more abundant at the downstream site (Figure 7, Table S7). At Ambler WWTP the
501 response was more variable, but the abundance increased by an average of 11.6 times. It was not
502 clear how much the population at 5.BC was influenced by 4.DSAmb versus 6.SR.

503

504 Some of the taxa with the largest increases in abundance are related to wastewater treatment. For
505 example *Zoogloea ramigera* has been observed to comprise up to 10% of the bacterial
506 community within activated sludge systems (Rossello-Mora et al., 1995), and *Zoogloea caeni*
507 was first isolated from activated sludge systems (Shao et al., 2009). *Malikia granosa* was also
508 first isolated from activated sludge, and has been shown to be a phosphate-accumulating
509 bacterium (Spring et al., 2005).

510

511 Wastewater effluent is known to contain excess nutrients, which can influence the health and
512 community structure of receiving streams (Carey and Migliaccio, 2009; Ibekwe et al., 2016;
513 Merseburger et al., 2005). In this study, this manifested itself in the form of increased abundance
514 of nutrient cycling organisms within the stream. Perhaps most notably, abundances of
515 *Nitrosomonas* spp., an ammonia oxidizing bacteria (AOB), had an observed log₂-fold change of
516 4.92 in downstream of effluent sources, while nitrite oxidizing bacteria (NOB), including
517 *Nitrospira defluvii* and another *Nitrospira* spp., were observed to have log₂-fold changes of
518 approximately 3-4. Both *Nitrosomonas* and *Nitrospira* can be found in WWTPs carrying out

519 nitrification/denitrification steps to remove excess nitrogen (Saunders et al., 2016). Several of the
520 taxa have been confirmed to be active denitrifiers within WWTPs, including *Dechloromonas* sp.
521 (1 taxa), *Haliangium* sp. (5 taxa), *Thermomonas brevis* (1 taxa) (McIlroy et al., 2016). For the
522 current study, it is unclear whether the increase in nitrifiers is due to their addition by WWTP
523 effluent (as active or dead biomass). Lastly, 8 taxa falling within Cyanobacteria were found to be
524 differentially abundant; 7 of the 8 were found to have an average log₂-fold change of 3.4, while
525 one was found to decrease in abundance. An increase in Cyanobacteria could be attributed to the
526 influx of nutrients from the wastewater effluent (Vaquer-Sunyer et al., 2016).

527
528 In addition to WWTP and nutrient cycling bacteria, 11 of the 130 differentially abundant RSVs
529 fall into two of the three Orders designated as indicators: 5 from Bacteroidales and 6 from
530 Clostridiales. Two of the differentially abundant taxa fall within the *Bacteroides* genus.
531 Differentially abundant Bacteroidales taxa had an average log₂-fold change of 4.32, while
532 Clostridiales taxa had an average log₂-fold change of 4.67.

533 534 **3.5.2. Differential abundance within the microbial community with distance below Upper** 535 **Gwynedd WWTP**

536 The analytical techniques used thus far have indicated that the microbial community composition
537 shifted back toward pre-effluent compositions with distance downstream of WWTP, i.e. with
538 distance below Upper Gwynedd WWTP (2.DSUG to 3. USAmb), including the large changes in
539 fecal indicator copy number (Figure 3), alpha diversity measures (Figure 4, Table S6),
540 abundance of dominant Phyla (Figure 5), and the DPCoA plot (Figure S4). These observations
541 prompted an additional round of testing to identify taxa that were differentially abundant
542 between 2.DSUG and 3.USAmb. 205 RSVs were found to be differentially abundant (Figure
543 S6).

544
545 The phyla of Actinobacteria (21), Bacteroidetes (55), Cyanobacteria (34), and Proteobacteria
546 (77) account for the majority of these differentially abundant taxa. Roughly half of the
547 differentially abundant taxa had higher abundance at 3.USAmb (102, positive log₂-fold change
548 in Figure S6), and the remaining taxa were lower in abundance (103, negative log₂-fold change
549 in Figure S6). The majority of Actinobacteria (19 of 21) increased in abundance, while taxa
550 falling within Bacteroidetes had mixed responses with 36 (of 55 differentially abundant) taxa
551 having higher abundance and the remaining 19 with lower abundance. The Actinobacteria and
552 Bacteroidetes clusters occur on opposing ends of the first DPCoA axis (Figure S4). The mixed
553 signals in Bacteroidetes response may explain why the transitions along the first DPCoA axis
554 between the two sampling days (2.DSUG.1 to 3.USAmb.1, and 2.DSUG.2 to 3.USAmb.2) are
555 inconsistent. Taxa falling within Cyanobacteria also showed mixed signals with 20 taxa (of 34)
556 having increased abundance and the remaining 14 differentially abundant Cyanobacteria taxa
557 having decreased abundance. Proteobacteria make up the majority of taxa (56 of the 103) with
558 decreased abundance at 3.USAmb (although an additional 21 Proteobacteria were observed to
559 increase in abundance), and may explain the shifts along the second DPCoA axis in Figure S4.
560 The abundance of the WWTP-linked taxa *M. granosa*, and three *Zoogloea* species were found to
561 be significantly lower at 3.USAmb than at 2.DSUG, as were fecal indicators *B. dorei* and *B.*
562 *vulgatus*, two nitrite oxidizing bacteria belonging to the *Nitrospira* genus, as well as 5
563 *Haliangium* species.

564

565 A total of 53 taxa are found to be in common when comparing the taxa identified as
566 differentially abundant between samples collected immediately up- and downstream of effluent
567 sources (1.USUG and 3.USAmb; 2.DSUG and 4.DSAmb; Figure S5) to those identified as
568 differentially abundant downstream of the Upper Gwynedd WWTP (between 2.DSUG and
569 3.USAmb, Figure S6). All 53 taxa fit the pattern that (i) they have increased abundance at sites
570 immediately downstream of effluent sources (2.DSUG and 4.DSAmb) and (ii) they were
571 observed to have lower abundance at 3.USAmb in comparison to 2.DSUG.
572

573 **3.6. Environmental influences on the composition of the microbial community**

574 **3.6.1. Distance-based redundancy analysis (db-RDA)**

575 Stream water chemistry composition was related to community structure using distance-based
576 redundancy analysis (db-RDA) (Legendre and Anderson, 1999; McArdle and Anderson, 2001)
577 with Bray-Curtis distances (Bray and Curtis, 1957). The final db-RDA model included the
578 effects of Cl^- , Si, Sr, and F^- (all log-scaled), had an adjusted r^2 value of 0.3351 (unadjusted r^2
579 value of 0.5769), and was found to be statistically significant at the 5% significance level ($p =$
580 0.001); each of the individual terms were also found to be significant when analyzed separately
581 (Cl^- $p = 0.001$, Si $p = 0.001$, Sr $p = 0.002$, F^- $p = 0.026$). As in previous ordinations, displacement
582 between upstream/downstream sample pairs occurred over both axes (Figure S7), and the db-
583 RDA model indicates that this displacement is correlated with the concentration of Cl^- . The
584 relationship between chemical dilution and change in microbial community indicate what portion
585 of the variation can be related to dilution.
586

587 **3.6.2. BIOENV procedure**

588 The BIOENV procedure (Clarke and Ainsworth, 1993) was applied to identify environmental
589 variables or combination of variables as an alternative method to forward selection db-RDA. The
590 optimal correlational model identified for the full microbial community included (log-
591 transformed) Cl^- and SO_4^{2-} and had a BIOENV correlation of 0.6003 (Table S8). Cl^- was
592 determined to be the best single parameter for explaining variation in OTU abundance, and
593 appeared in all of the investigated models (Table S8).
594

595 **3.7. Indicator community results**

596 In addition to investigating the overall microbial community, RSV's within the orders of
597 Bacteroidales, Bifidobacteriales, and Clostridiales were selected for additional study due to their
598 association with human fecal contamination (Halliday et al., 2014; McLellan et al., 2010). The
599 results of this round of analysis mirror those found for the full microbial community. In brief,
600 alpha diversity within these three orders were observed to be higher in samples collected
601 immediately downstream of effluent sources than those collected upstream (Figure S8) and that
602 the relative abundance of these taxa also increased up to downstream (Figure S9), indicating that
603 effluent outflows are sources of new species found within these three orders. Roughly 15% of
604 these taxa were found to be significantly higher in abundance at downstream sites (Figure S12).
605 These differentially abundant taxa showed similar, but more variable, patterns in changes in
606 relative abundance (Figure S13, Table S7) to those observed for the whole microbial community
607 (Figure 7, Table S7). In this study, we used forward variable selection to build db-RDA models
608 and applied the BIOENV procedure to identify the chemical parameters that have the most
609 significant correlation with the microbial community structure. The db-RDA models had
610 adjusted r^2 values of 33-35% (unadjusted r^2 of 57.69% for the full microbial community and

611 52.52% for the indicator community). BIOENV produced models that resulted in unadjusted r^2
612 values of 60.03% and 68.04% for the full and indicator communities respectively (Table S8 and
613 Table S9). Full details of the results from this round of analysis are presented in Section 1 of the
614 Supplemental Material

615

616 **4. Discussion**

617 Surface waters such as rivers and streams have been described as conveyers of biodiversity
618 information in the form of environmental DNA (eDNA), both for the species inhabiting those
619 water bodies but also terrestrial plants and animals (Deiner et al., 2016). In cases where WWTP
620 effluent outflows exist, their contribution to the composition of receiving streams must be taken
621 into account. In this study, we used methods in analytical chemistry and molecular ecology to
622 observe how the impact of wastewater effluent on the microbial consortia varies with distance
623 from and composition of sources. Using a variety of statistical and analytical tools, we were able
624 to reveal a variety of both direct and indirect effects of WWTP effluent sources and how the
625 microbial community responds. These observations include: (i) changes in species richness and
626 microbial community composition below effluent discharges, (ii) partial attenuation of microbial
627 community downstream of effluent contribution, and (iii) correlation between overall microbial
628 diversity and the three Orders proposed to track WWTP effluent by McLellan et al. (2010).
629 While this investigation has resulted in a number of interesting observations, the relatively small
630 sample size of this study limits the generalizability of some of its findings, and consequently, the
631 analysis is intended to only describe the sites and dates within the current dataset. Expansion of
632 the dataset to include samples collected during all seasons would provide insight about the
633 seasonal and temporal stability of the trends reported here and also enables the application of
634 new analytical approaches.

635

636 **4.1. Direct impacts of effluent sources on surface water microbial communities**

637 In comparing samples immediately upstream (1.USUG, 3.USAmb) to those immediately
638 downstream (2.DSUG, 4.DSAmb), our results indicate that effluent outflows are significant
639 sources of the fecal indicators *B. dorei* and *Bacteroides* spp. from both absolute quantitative and
640 relative abundance perspectives (Figure 3). The sequencing results show that effluent sources
641 contribute additional diversity to the microbial community (Figure 4, Table S6) in addition to
642 shifting the overall composition of the community (Figure 6, Figure S7). Differential abundance
643 testing was used to identify 137 taxa that were more abundant in downstream samples than in
644 those collected upstream of an effluent source (Figure S5). The proportion of reads attributed to
645 these 137 taxa increased an extraordinary 7 to 11 times (Figure 7, Table S7). Some of the largest
646 increases belonged to those taxa found within Bacteroidetes and Proteobacteria, and the vast
647 majority of the differentially abundant taxa are related to WWTP processes, nutrient cycling, or
648 are fecal-associated in nature. Analysis of the effluent, which was not possible for this study,
649 would be beneficial from a microbial source tracking perspective, and would also enable the use
650 of tools such as SourceTracker (Knights et al., 2011), which enables investigators to determine
651 the relative contributions of multiple microbial sources to a sample's community; such
652 applications have been used to describe the effect of urban watersheds on aquatic community
653 composition (Wang et al., 2016).

654

655 These microbial community alterations may be affected by antecedent moisture conditions
656 (AMC). Analysis of storm events in this watershed show that nutrient concentrations recover

657 within one to two days of precipitation events, with overland flow contributing nutrients to
658 tributaries with less riparian buffering than the main stem. In contrast, water level remains high
659 for three to four days after rain, potentially from storage in saturated and unsaturated storage
660 zones slowly draining to the stream, resulting in higher water levels without high nutrient
661 concentrations. We expected a similar response in microbial communities, with lower diversities
662 a few days after precipitation; however, samples collected after wet AMC showed an increase in
663 microbial community diversity, while those collected after dry conditions showed a decrease in
664 the relative abundance of microbes. The effect of AMC on the microbial communities, however,
665 does seem to be impacted by the discharge at the site. Both 1.USUG and 2.DSUG, where
666 discharge is low, are less impacted by the shift in AMC, while the other sites, which have higher
667 discharge, have larger shifts between sampling dates and AMC. Although a contrast in AMC was
668 observed on the two sample dates only one week apart and collected at the same time of day, the
669 future collection and analysis of samples representing a range of AMC and stream discharge
670 values may enable these trends to be better explored.

671
672 The availability of nutrients, such as N and P species, often shape the microbial community (Lee
673 et al., 2017; Mello et al., 2016), particularly in systems with a long term source such as the
674 WWTPs in our study. Comparing the models obtained from both forward selection and BIOENV
675 emphasizes that chloride and silicon had the most significant correlation with the variation in the
676 communities, while strontium, magnesium, sulfate, and nitrite had smaller correlation. With the
677 caveat that in some cases, Si can be limiting for some photosynthetic organisms, major nutrients
678 such as ammonium, nitrate, or phosphate do not appear to be major drivers of community
679 structure. Instead, the models suggest that the governing forces behind community structure are
680 correlated with stream water constituents that are highly influenced by WWTP outflows or
681 ground and surface water composition. This unexpected result and the small sample size of the
682 current study indicate that additional investigation is needed to fully characterize the extent that
683 nutrient availability has on community composition at these sites. Such work may also enable the
684 inspection of relationships between specific nutrients and microbial groups, for example, the
685 interactions between Si and photosynthetic organisms.

686 687 **4.2. Attenuation with distance below WWTPs**

688 Two distinct patterns emerged when investigating microbial community composition with
689 distance downstream of the Upper Gwynedd WWTP. The first was that, despite the
690 instantaneous increase in diversity below effluent outflows, most likely the result of new taxa
691 being introduced into the stream by the effluent, dramatic reductions in diversity were observed
692 between 2.DSUG and 3.USAmb on both sampling dates. The second pattern that emerged was
693 that roughly 40% (53 of 137) of the taxa found at higher abundances downstream of effluent
694 sites decreased in abundance over the course of the 10 km that separated 2.DSUG and 3.USAmb
695 (Figure 1). Therefore, while these effluent outflows have been demonstrated to be sources of
696 diversity, these effluent-sourced taxa do not appear to proliferate downstream of the effluent
697 source. Moreover, the loss of additional diversity between 2.DSUG and 3.USAmb (on the order
698 of ~ 850 taxa) is much larger in magnitude than what was determined through differential
699 abundance testing; this disparity is most likely the result of a small sample size and the
700 conservative values selected for differential abundance detection. Loss of abundance and
701 diversity in effluent impacted streams and their beds have been attributed to excess nutrients,
702 sediment organic matter within the water column, or the presence of biologically active

703 compounds (Drury et al., 2013). A variety of these bioactive compounds, including
704 pharmaceuticals, hormones and organic wastewater contaminants, have been found in WWTP
705 effluent (Kolpin et al., 2002). Wastewater treatment methods such as the activated sludge process
706 were not initially designed for the removal of such compounds and, as a result, WWTPs can
707 serve as sources for antibiotics (Akiyama and Savin, 2010) and other pharmaceutical compounds
708 (Bartelt-Hunt et al., 2009).

709
710 Another observation that may partly explain these signals comes from the nature of the source
711 material and analysis methods. Culture independent methods such as qPCR and amplicon
712 sequencing do not distinguish between viable live cells, and those that are dead. Rapid
713 degradation of eDNA has been observed in aquatic systems (Barnes et al., 2014; Strickler et al.,
714 2015; Thomsen et al., 2012a; Thomsen et al., 2012b). Sampling immediately downstream of an
715 effluent source may inadvertently cause the detection of DNA from dead bacteria in the effluent.
716 After traveling approximately 10 km further downstream (3.USAmb), these cells will have had
717 the opportunity to lyse and their components, including DNA, to degrade, and thus would not be
718 detected or detected in lower quantities. If the effluent contains live bacteria, the disappearance
719 of those taxa could be attributed to those cells settling, sorbing to particles in the water column
720 (aiding settling), sorbing to the streambed, or the environment in the stream is not suitable for
721 their survival (such as those taxa requiring hosts or specific nutrients). Other potential controls
722 on the decrease in diversity include shifts in hydrologic characteristics, including small tributary
723 contributions and changes in discharge, velocity, channel morphology (i.e. ripple vs. run vs.
724 pool), and hyporheic flow. Regardless of cause, the removal of these taxa represents a partial
725 attenuation of the microbial community within the stream. Including additional sample sites
726 between 2.DSUG and 3.USAmb in future studies may provide an opportunity to explain, or at
727 least better describe, the interesting changes in community structure including the non-
728 proliferation of effluent-sourced bacteria and the loss of diversity between the two WWTPs.

729 730 **4.3. Sequencing as an augmentative tool for microbial source tracking and identification**

731 Two additional components of this study were (i) to explore the agreement between results
732 obtained through targeted amplicon sequencing and qPCR and (ii) to investigate the efficacy of
733 using a subset of the microbial community composed of three Orders proposed by McLellan et
734 al. (2010) to predict the degree of influence from WWTP effluent and sewage. In this study, we
735 found the relative abundance of 16s sequences for *B. dorei* and *Bacteroides* spp., commonly used
736 to quantify fecal contamination for human and general sources respectively, track reasonably
737 well with the copy number obtained through qPCR (Figure 3), in particular for 1.USUG,
738 2.DSUG, 3.USAmb, and 4.DSAmb. The indicator subset, which covered three Orders, not just a
739 single genus or species, also followed the general trends obtained through qPCR (Figure S9).
740 Comparing the relative abundance of *B. dorei* and *Bacteroides* spp. (Figure 3), with that of the
741 indicator community (Figure S9), there appears to be a large degree of correspondence,
742 supporting the conclusions put forth by McLellan et al. (2010) about the suitability of these
743 orders for detecting fecal or sewage contamination.

744
745 Beyond subjective impressions, utilizing only the indicator subset allowed the identification of
746 26 differentially abundant taxa, in comparison to the 11 taxa falling within these Orders found
747 when using the entire microbial community dataset. The detection of 15 additional taxa may
748 have been facilitated by the removal of potential noise induced by the non-indicator taxa. In

749 terms of relative abundance, the magnitude of the differentially abundant taxa within the
750 indicator subset displayed similar, albeit more variable, increases in magnitude (Figure S13) in
751 comparison to the taxa identified as differentially abundant in the full community (Figure 7).

752
753 Finally, it may be useful to augment the orders or taxa selected as indicators, as some taxa outside
754 of these orders are highly associated with fecal sources. For example Wu et al. (2010) observed
755 that the ratio of Bacilli, Bacteroidetes, and Actinobacteria to α -Proteobacteria was useful for
756 identifying sources that were impacted by gut-flora and sewage sources. Another potential
757 addition could be to include, as indicators, taxa known to be highly abundant in WWTP
758 processes, and relatively rare in the environment. The identification or selection of such species
759 may require considerable effort to prevent masking or occluding natural signal in the results.

760

761 **5. Conclusions**

762 In this study, we combined information from stream water chemistry, qPCR, and amplicon
763 sequencing to identify the response of the microbial population structure to WWTP effluent
764 sources. Our data demonstrate the extent that wastewater effluents shape both the chemical and
765 biological composition of their receiving streams. We observed that organisms linked to WWTP
766 processes, nutrient cycling, and fecal indicators were significantly more abundant in sites
767 immediately downstream of effluent sources. These factors vary with antecedent moisture
768 conditions. Many of these organisms, in particular the WWTP linked and fecal indicator taxa,
769 were found to decrease with distance downstream of those effluent sources. Species richness was
770 also observed to decrease with distance downstream of WWTP effluent sources, perhaps due to
771 chemical stressors present in the effluent sources (Drury et al., 2013; Kolpin et al., 2002).
772 Surprisingly, nutrient availability was not observed to have a significant effect on the microbial
773 community composition.

774

775 This study confirms and contributes to the work investigating anthropogenic impacts on the
776 natural environment by confirming observations on changes in absolute and relative abundance
777 of fecal indicators such as *B. dorei* and *Bacteroides* spp., changes in alpha diversity and species
778 richness, and the usefulness of using subsets of microbial communities as indicators for
779 determining the impact of fecal and effluent sources. We also demonstrated the application of
780 recently developed methods such as those for the error correction of amplicon sequencing results
781 (the dada2 R package) and the determination of differentially abundant taxa (the DESeq2 R
782 package). Additionally, we have made public all data and scripts needed to replicate the results
783 of this investigation to encourage other investigators to explore and expand on these techniques
784 with their own datasets.

785

786

787 **Acknowledgements**

788 The authors would like to acknowledge Shana McDevitt and Dylan Smith for their technical
789 assistance during MiSeq Sequencing. This work was supported under NSF grant number
790 1245632 to Drexel University and by the William Penn Foundation as part of the Delaware River
791 Watershed Initiative grant to Temple University. This work used the Vincent J. Coates Genomics
792 Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants
793 S10RR029668 and S10RR027303. The authors would like to acknowledge their use of the
794 Drexel University Research Computing Facility. The authors thank the editor and anonymous
795 reviewers for their comments and feedback.

796
797 **References**

- 798
799 Ahmed W, Goonetilleke A, Powell D, Chauhan K, Gardner T. Comparison of molecular markers
800 to detect fresh sewage in environmental waters. *Water Res* 2009; 43: 4908-17.
- 801 Ahmed W, Stewart J, Gardner T, Powell D. A real-time polymerase chain reaction assay for
802 quantitative detection of the human-specific enterococci surface protein marker in sewage
803 and environmental waters. *Environ Microbiol* 2008; 10: 3255-64.
- 804 Akiyama T, Savin MC. Populations of antibiotic-resistant coliform bacteria change rapidly in a
805 wastewater effluent dominated stream. *Sci Total Environ* 2010; 408: 6192-201.
- 806 Ali S, Ghosh NC, Singh R. Rainfall-runoff simulation using a normalized antecedent
807 precipitation index. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*
808 2010; 55: 266-274.
- 809 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*
810 2010; 11: R106.
- 811 Aristi I, von Schiller D, Arroita M, Barcelo D, Ponsati L, Garcia-Galan MJ, et al. Mixed effects
812 of effluents from a wastewater treatment plant on river ecosystem metabolism: subsidy or
813 stress? *Freshwater Biology* 2015; 60: 1398-1410.
- 814 Bae S, Wuertz S. Survival of host-associated bacteroidales cells and their relationship with
815 *Enterococcus* spp., *Campylobacter jejuni*, *Salmonella enterica* serovar Typhimurium, and
816 adenovirus in freshwater microcosms as measured by propidium monoazide-quantitative
817 PCR. *Appl Environ Microbiol* 2012; 78: 922-32.
- 818 Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. Environmental
819 conditions influence eDNA persistence in aquatic systems. *Environ Sci Technol* 2014;
820 48: 1819-27.
- 821 Bartelt-Hunt SL, Snow DD, Damon T, Shockley J, Hoagland K. The occurrence of illicit and
822 therapeutic pharmaceuticals in wastewater effluent and surface waters in Nebraska.
823 *Environ Pollut* 2009; 157: 786-91.
- 824 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful
825 Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-*
826 *Methodological* 1995; 57: 289-300.
- 827 Bernhard AE, Field KG. A PCR assay To discriminate human and ruminant feces on the basis of
828 host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl Environ*
829 *Microbiol* 2000; 66: 4571-4.
- 830 Blanchet FG, Legendre P, Borcard D. Forward selection of explanatory variables. *Ecology* 2008;
831 89: 2623-32.

832 Bray RJ, Curtis JT. An ordination of the upland forest communities of southern Wisconsin.
833 Ecological monographs 1957; 27: 325-349.

834 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-
835 resolution sample inference from Illumina amplicon data. Nat Methods 2016a; 13: 581-3.

836 Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor Workflow for
837 Microbiome Data Analysis: from raw reads to community analyses. F1000Res 2016b; 5:
838 1492.

839 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-
840 throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.
841 ISME J 2012; 6: 1621-4.

842 Carey RO, Migliaccio KW. Contribution of wastewater treatment plant effluents to nutrient
843 dynamics in aquatic systems: a review. Environ Manage 2009; 44: 205-17.

844 Carpenter SR, Caraco NF, Correll DL. Nonpoint pollution of surface waters with phosphorus and
845 nitrogen. Nonpoint pollution of surface waters with phosphorus and nitrogen 1998.

846 Chao A. Nonparametric-Estimation of the Number of Classes in a Population. Scandinavian
847 Journal of Statistics 1984; 11: 265-270.

848 Clarke KR, Ainsworth M. A Method of Linking Multivariate Community Structure to
849 Environmental Variables. Marine Ecology Progress Series 1993; 92: 205-219.

850 Deiner K, Fronhofer EA, Meachler E, Walser J-C, Altermatt F. Environmental DNA reveals that
851 rivers are conveyor belts of biodiversity information. bioRxiv 2016: 20800.

852 Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and
853 diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ
854 Microbiol 2013; 79: 1897-905.

855 Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, et al. The Earth Microbiome Project:
856 Meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne
857 National Laboratory October 6 2010. Stand Genomic Sci 2010; 3: 249-53.

858 Gücker B, Brauns M, Pusch MT. Effects of wastewater treatment plant discharge on ecosystem
859 structure and function of lowland streams. Journal of the North American ... 2006.

860 Haggard BE, Stanley EH, Storm DE. Nutrient retention in a point-source-enriched stream.
861 Journal of the North American Benthological Society 2005; 24: 29-47.

862 Haggard BE, Storm DE, Stanley EH. Effect of a point source input on stream nutrient retention.
863 Journal of the American Water Resources Association 2001; 37: 1291-1299.

864 Halliday E, McLellan SL, Amaral-Zettler LA, Sogin ML, Gast RJ. Comparison of bacterial
865 communities in sands and water at beaches with bacterial water quality violations. PLoS
866 One 2014; 9: e90815.

867 Harry IS, Ameh E, Coulon F, Nocker A. Impact of Treated Sewage Effluent on the Microbiology
868 of a Small Brook Using Flow Cytometry as a Diagnostic Tool. Water Air Soil Pollut
869 2016; 227: 57.

870 Hladilek MD, Gaines KF, Novak JM, Collard DA, Johnson DB, Canam T. Microbial community
871 structure of a freshwater system receiving wastewater effluent. Environ Monit Assess
872 2016; 188: 626.

873 Homer C, Dewitz J, Yang LM, Jin S, Danielson P, Xian G, et al. Completion of the 2011
874 National Land Cover Database for the Conterminous United States - Representing a
875 Decade of Land Cover Change Information. Photogrammetric Engineering and Remote
876 Sensing 2015; 81: 345-354.

877 Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of*
878 *Educational Psychology* 1933; 24: 417-441.

879 Ibekwe AM, Ma J, Murinda SE. Bacterial community composition and structure in an Urban
880 River impacted by different pollutant sources. *Sci Total Environ* 2016; 566-567: 1176-85.

881 Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian
882 community-wide culture-independent microbial source tracking. *Nat Methods* 2011; 8:
883 761-3.

884 Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, et al.
885 Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams,
886 1999-2000: a national reconnaissance. *Environ Sci Technol* 2002; 36: 1202-11.

887 Lee ZM, Poret-Peterson AT, Siefert JL, Kaul D, Moustafa A, Allen AE, et al. Nutrient
888 Stoichiometry Shapes Microbial Community Structure in an Evaporitic Shallow Pond.
889 *Front Microbiol* 2017; 8: 949.

890 Legendre P, Anderson MJ. Distance - based redundancy analysis: testing multispecies responses
891 in multifactorial ecological experiments. *Ecological monographs* 1999.

892 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
893 data with DESeq2. *Genome Biol* 2014; 15: 550.

894 Marti E, Aumatell J, Gode L, Poch M, Sabater F. Nutrient retention efficiency in streams
895 receiving inputs from wastewater treatment plants. *J Environ Qual* 2004; 33: 285-93.

896 Marti E, Balcazar JL. Use of pyrosequencing to explore the benthic bacterial community
897 structure in a river impacted by wastewater treatment plant discharges. *Res Microbiol*
898 2014; 165: 468-71.

899 McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on
900 distance-based redundancy analysis. *Ecology* 2001; 82: 290-297.

901 McIlroy SJ, Starnawska A, Starnawski P, Saunders AM, Nierychlo M, Nielsen PH, et al.
902 Identification of active denitrifiers in full-scale nutrient removal wastewater treatment
903 systems. *Environ Microbiol* 2016; 18: 50-64.

904 McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. Diversity and
905 population structure of sewage-derived microorganisms in wastewater treatment plant
906 influent. *Environ Microbiol* 2010; 12: 378-92.

907 McMurdie PJ, Holmes S. Phyloseq: a bioconductor package for handling and analysis of high-
908 throughput phylogenetic sequence data. *Pac Symp Biocomput* 2012: 235-46.

909 McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible.
910 *PLoS Comput Biol* 2014; 10: e1003531.

911 Mello BL, Alessi AM, McQueen-Mason S, Bruce NC, Polikarpov I. Nutrient availability shapes
912 the microbial community structure in sugarcane bagasse compost-derived consortia. *Sci*
913 *Rep* 2016; 6: 38781.

914 Merbt SN, Auguet JC, Blesa A, Marti E, Casamayor EO. Wastewater treatment plant effluents
915 change abundance and composition of ammonia-oxidizing microorganisms in
916 mediterranean urban stream biofilms. *Microb Ecol* 2015; 69: 66-74.

917 Merseburger GC, Marti E, Sabater F. Net changes in nutrient concentrations below a point
918 source input in two streams draining catchments with contrasting land uses. *Sci Total*
919 *Environ* 2005; 347: 217-29.

920 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. *vegan: Community*
921 *Ecology Package*, 2016.

922 Palmer MW. Putting Things in Even Better Order - the Advantages of Canonical
923 Correspondence-Analysis. *Ecology* 1993; 74: 2215-2230.

924 Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA
925 primers for marine microbiomes with mock communities, time series and global field
926 samples. *Environ Microbiol* 2015.

927 Pavoine S, Dufour AB, Chessel D. From dissimilarities among species to dissimilarities among
928 communities: a double principal coordinate analysis. *Journal of theoretical biology* 2004.

929 Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical*
930 *Magazine Series* 6 1901; 2: 559-572.

931 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive
932 online resource for quality checked and aligned ribosomal RNA sequence data
933 compatible with ARB. *Nucleic Acids Res* 2007; 35: 7188-96.

934 Purdom E. Analysis of a Data Matrix and a Graph: Metagenomic Data and the Phylogenetic
935 Tree. *Annals of Applied Statistics* 2011; 5: 2326-2358.

936 PWD. Wissahickon Creek Watershed Comprehensive Characterization Report. Philadelphia
937 Water Department, 2007.

938 Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced
939 amplicons. *BMC Bioinformatics* 2011; 12: 38.

940 R Development Core Team. R: A Language and Environment for Statistical Computing. R
941 Foundation for Statistical Computing, Vienna, Austria, 2015.

942 Rahm BG, Hill NB, Shaw SB, Riha SJ. Nitrate Dynamics in Two Streams Impacted by
943 Wastewater Treatment Plant Discharge: Point Sources or Sinks? *Journal of the American*
944 *Water Resources Association* 2016; 52: 592-604.

945 Ribot M, Marti E, von Schiller D, Sabater F, Daims H, Battin TJ. Nitrogen processing and the
946 role of epilithic biofilms downstream of a wastewater treatment plant. *Freshwater Science*
947 2012; 31: 1057-1069.

948 Rossello-Mora RA, Wagner M, Amann R, Schleifer KH. The abundance of *Zoogloea ramigera*
949 in sewage treatment plants. *Appl Environ Microbiol* 1995; 61: 702-7.

950 Ryan MO. Microbial source tracking of human and animal waste pollution of diverse watersheds
951 and of urban drainage systems using molecular methods. *Civil, Architectural, and*
952 *Environmental Engineering*. Drexel University, Philadelphia, PA, 2012.

953 Ryan MO, Gurian PL, Haas CN, Rose JB, Duzinski PJ. Acceptable microbial risk: Cost-benefit
954 analysis of a boil water order for *Cryptosporidium*. *Journal - American Water Works*
955 *Association* 2013; 105: E189-E194.

956 Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a
957 core community of abundant organisms. *ISME J* 2016; 10: 11-20.

958 Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: relations with pathogens and
959 conventional indicators, current methodologies for direct pathogen monitoring and future
960 application perspectives. *Water Res* 2006; 40: 2463-76.

961 Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011; 27: 592-3.

962 Shanks OC, Domingo JW, Lu J, Kelty CA, Graham JE. Identification of bacterial DNA markers
963 for the detection of human fecal pollution in water. *Appl Environ Microbiol* 2007; 73:
964 2416-22.

965 Shanks OC, White K, Kelty CA, Hayes S, Sivaganesan M, Jenkins M, et al. Performance
966 assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal
967 pollution. *Appl Environ Microbiol* 2010; 76: 1359-66.

968 Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal 1948;
 969 27: 379-423.
 970 Shao Y, Chung BS, Lee SS, Park W, Lee SS, Jeon CO. Zoogloea caeni sp. nov., a floc-forming
 971 bacterium isolated from activated sludge. Int J Syst Evol Microbiol 2009; 59: 526-30.
 972 Siefring S, Varma M, Atikovic E, Wymer L, Haugland RA. Improved real-time PCR assays for
 973 the detection of fecal indicator bacteria in surface waters with different instrument and
 974 reagent systems. J Water Health 2008; 6: 225-37.
 975 Sonthiphand P, Cejudo E, Schiff SL, Neufeld JD. Wastewater effluent impacts ammonia-
 976 oxidizing prokaryotes of the Grand River, Canada. Appl Environ Microbiol 2013; 79:
 977 7454-65.
 978 Spring S, Wagner M, Schumann P, Kampfer P. Malikia granosa gen. nov., sp. nov., a novel
 979 polyhydroxyalkanoate- and polyphosphate-accumulating bacterium isolated from
 980 activated sludge, and reclassification of Pseudomonas spinosa as Malikia spinosa comb.
 981 nov. Int J Syst Evol Microbiol 2005; 55: 621-9.
 982 Strickler KM, Fremier AK, Goldberg CS. Quantifying effects of UV-B, temperature, and pH on
 983 eDNA degradation in aquatic microcosms. Biological Conservation 2015; 183: 85-92.
 984 Thomsen PF, Kielgast J, Iversen LL, Moller PR, Rasmussen M, Willerslev E. Detection of a
 985 diverse marine fish fauna using environmental DNA from seawater samples. PLoS One
 986 2012a; 7: e41732.
 987 Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MT, et al. Monitoring
 988 endangered freshwater biodiversity using environmental DNA. Mol Ecol 2012b; 21:
 989 2565-73.
 990 Van der Gucht K, Vandekerckhove T, Vloemans N, Cousin S, Muylaert K, Sabbe K, et al.
 991 Characterization of bacterial communities in four freshwater lakes differing in nutrient
 992 load and food web structure. FEMS Microbiol Ecol 2005; 53: 205-20.
 993 Vaquer-Sunyer R, Reader HE, Muthusamy S, Lindh MV, Pinhassi J, Conley DJ, et al. Effects of
 994 wastewater treatment plant effluent inputs on planktonic metabolic rates and microbial
 995 community composition in the Baltic Sea. Biogeosciences 2016; 13: 4751-4765.
 996 Wakelin SA, Colloff MJ, Kookana RS. Effect of wastewater treatment plant effluent on
 997 microbial function and community structure in the sediment of a freshwater stream with
 998 variable seasonal flow. Appl Environ Microbiol 2008; 74: 2659-68.
 999 Wang P, Chen B, Yuan R, Li C, Li Y. Characteristics of aquatic bacterial community and the
 1000 influencing factors in an urban river. Science of The Total Environment 2016; 569-570:
 1001 382-389.
 1002 Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of
 1003 rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007; 73:
 1004 5261-7.
 1005 Wright ES. DECIPHER: harnessing local sequence context to improve protein multiple sequence
 1006 alignment. BMC Bioinformatics 2015; 16: 322.
 1007 Wu CH, Sercu B, Van de Werfhorst LC, Wong J, DeSantis TZ, Brodie EL, et al.
 1008 Characterization of coastal urban watershed bacterial communities leads to alternative
 1009 community-based indicators. PLoS One 2010; 5: e11285.
 1010 Zwart G, Crump BC, Agterveld MPKV, Hagen F, Han SK. Typical freshwater bacteria: an
 1011 analysis of available 16S rRNA gene sequences from plankton of lakes and rivers.
 1012 Aquatic Microbial Ecology 2002; 28: 141-155.
 1013

1015 **List of figure captions**

1016 Figure 1: Wissahickon Creek is a third-order stream that discharges into the Schuylkill River in
1017 Montgomery and Philadelphia counties, Pennsylvania. There are four wastewater treatment
1018 plants (WWTPs) on the stream, and samples were taken above and below the main stem plants
1019 (1.USUG, 2.DSUG, 3.USAmb, and 4.DSAmb), along with below the confluence of the main
1020 stem with the largest tributary, Sandy Run (5.BC). Additionally, samples were collected on
1021 Sandy Run below two additional WWTPs (6.SR). WWTPs are indicated with circles and sample
1022 locations are indicated with triangles.

1023
1024 Figure 2: (A) Nitrate-N concentrations and (B) total dissolved phosphorus and orthophosphate
1025 concentrations with distance along the stream relative to the stream's headwaters. Upper
1026 Gwynedd WWTP samples from NPDES report on the two sampling dates are reported as
1027 orthophosphate and all other samples are reported as TDP. Samples collected on May 10, 2016
1028 are shown with circles while samples collected on May 17, 2016 are shown in triangles; for
1029 clarity, lines connect the samples collected on the same day and do not imply any trend. Error
1030 bars indicate analytical error. Average WWTP concentrations for the other three treatment plants
1031 from three different samples are shown in squares. Error bars for these points indicate the
1032 standard deviation among the three samples.

1033
1034 Figure 3: Observed abundances of human-source fecal indicators (*B. dorei*) and non-specific
1035 fecal indicators (*Bacteroides spp.*): The top panels present the Copy Number (CN) concentration
1036 of *B. dorei* (A) and *Bacteroides spp.* (B), as determined by qPCR. The bottom panels present the
1037 relative abundances of *B. dorei* (C) and *Bacteroides spp.* (B), determined through sequencing.
1038 Grey bars indicate taxa that were assigned to the genus *Bacteroides* but were not assigned at the
1039 species level. The reader should note that the ordinal axes in panels A and B are log-scaled while
1040 the axes in panels C and D are linear in nature.

1041
1042 Figure 4: Alpha Diversity Measures for the full microbial community. Samples are grouped by
1043 their site number, found in the first character in the sample name in Figure 1. Error bars for
1044 Chao1 represent standard error (SE).

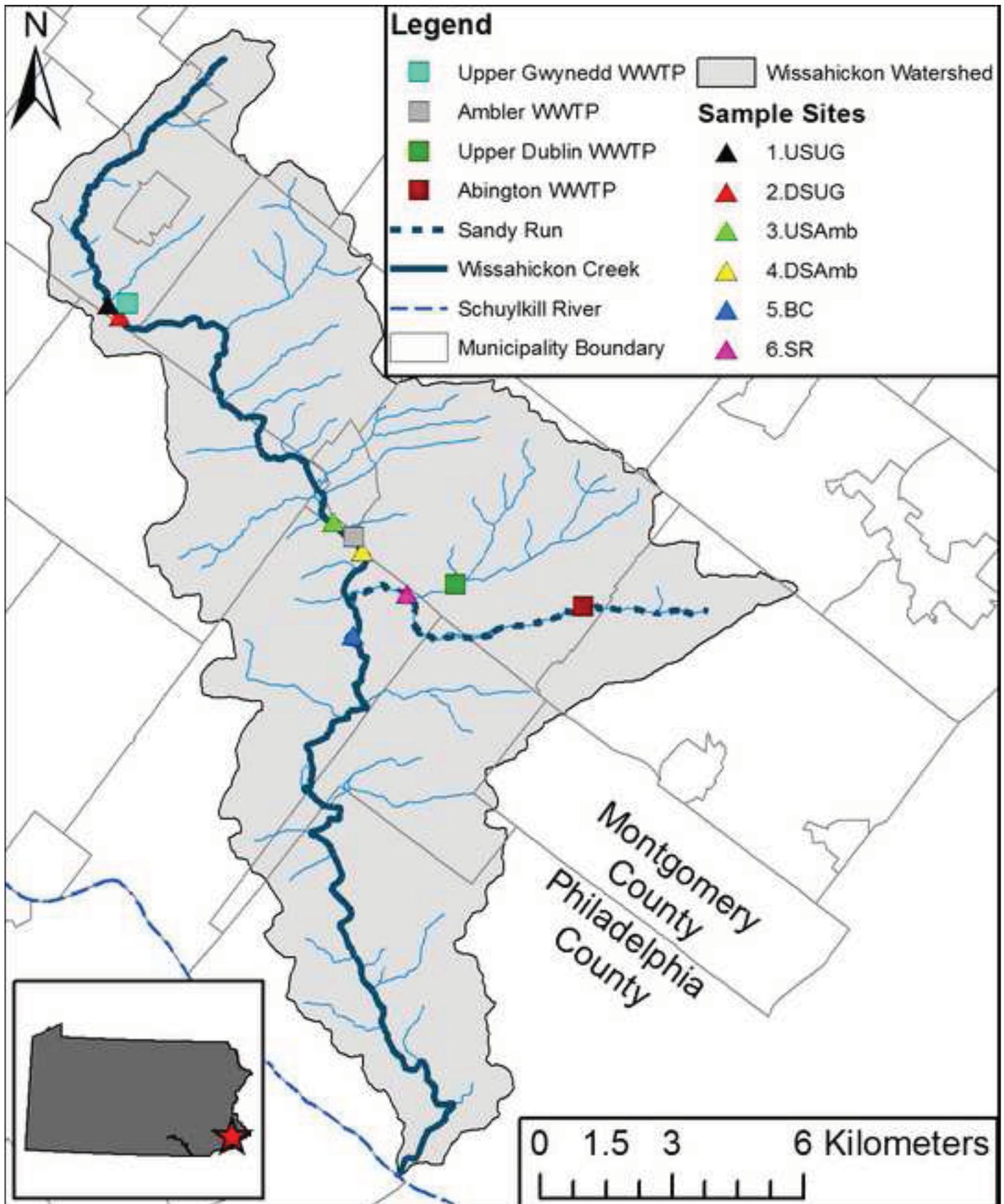
1045
1046 Figure 5: Barcharts representing the relative abundance of the 7 most abundant Phyla in the
1047 microbial community.

1048
1049 Figure 6: A plot of the PCA results for the complete microbial community.

1050
1051 Figure 7: Relative abundance of differentially abundant taxa, observed during the analysis of the
1052 whole microbial community. Samples from sampling sites 5.BC and 6.SR are included only for
1053 comparison and were not included in the differential abundance testing.

1054
1055

Figure1
[Click here to download high resolution image](#)



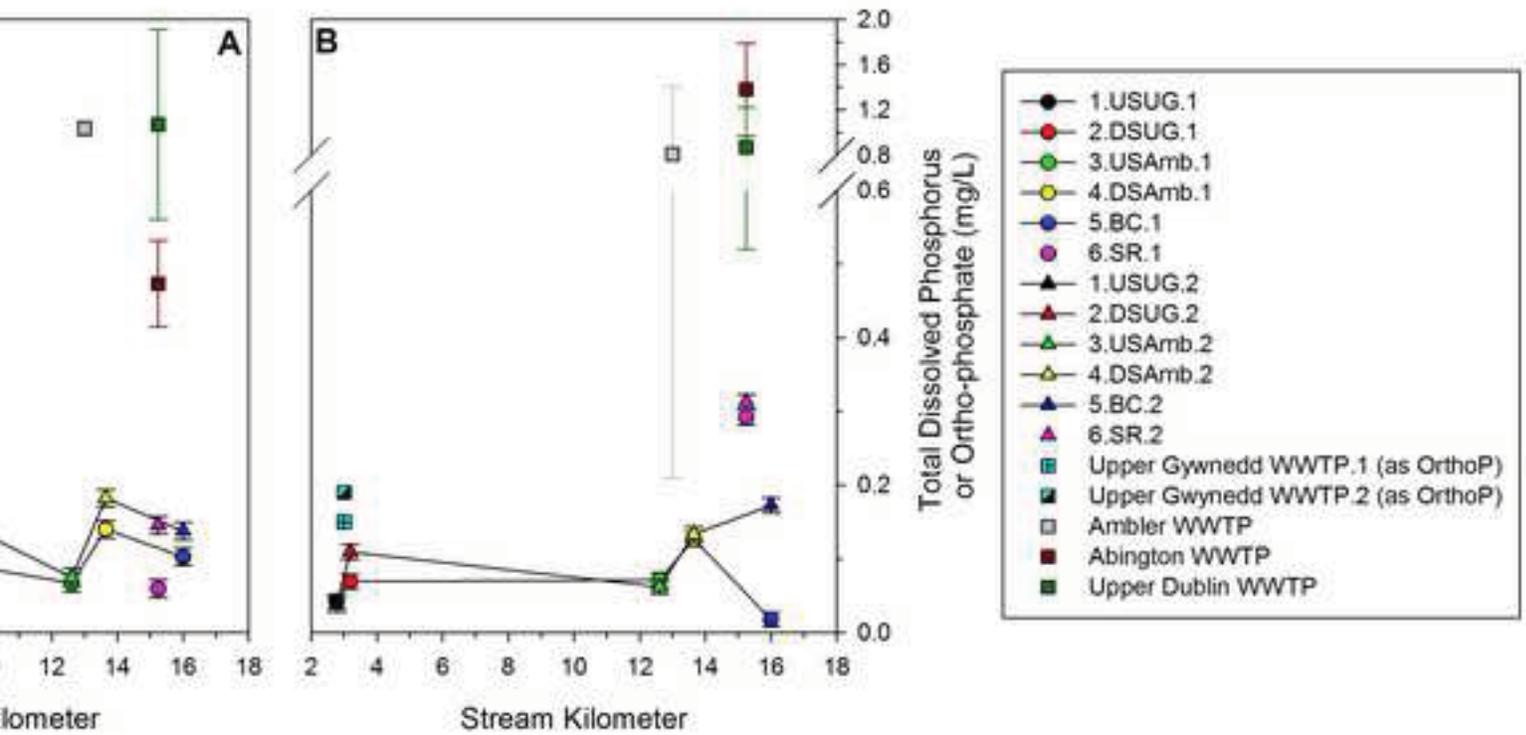
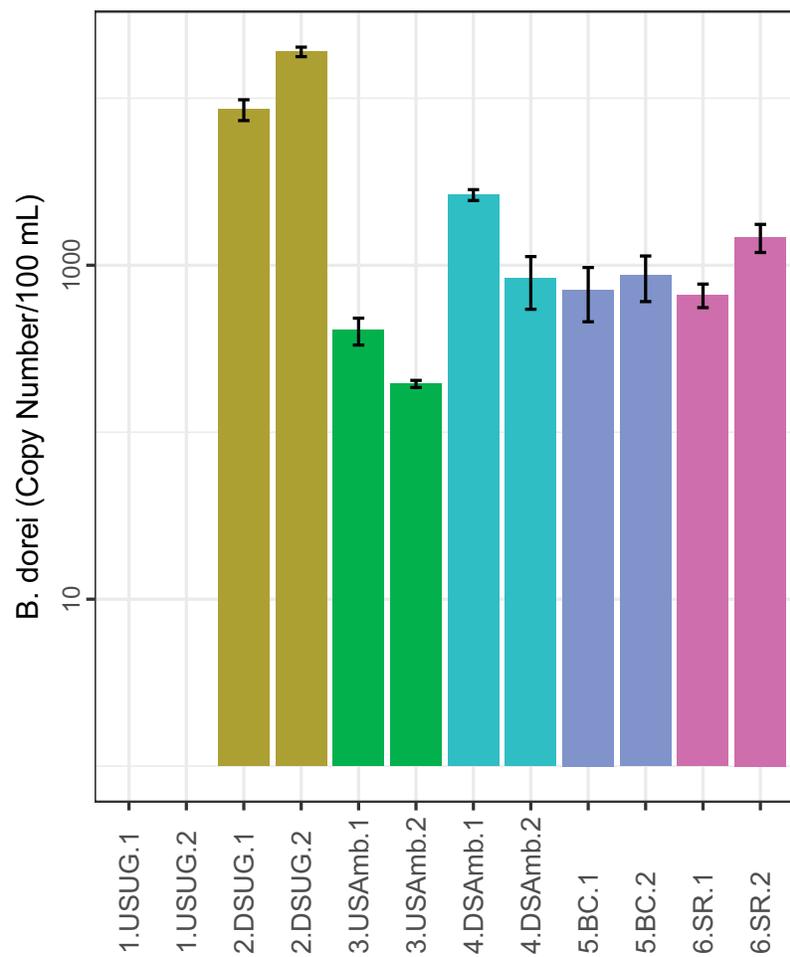


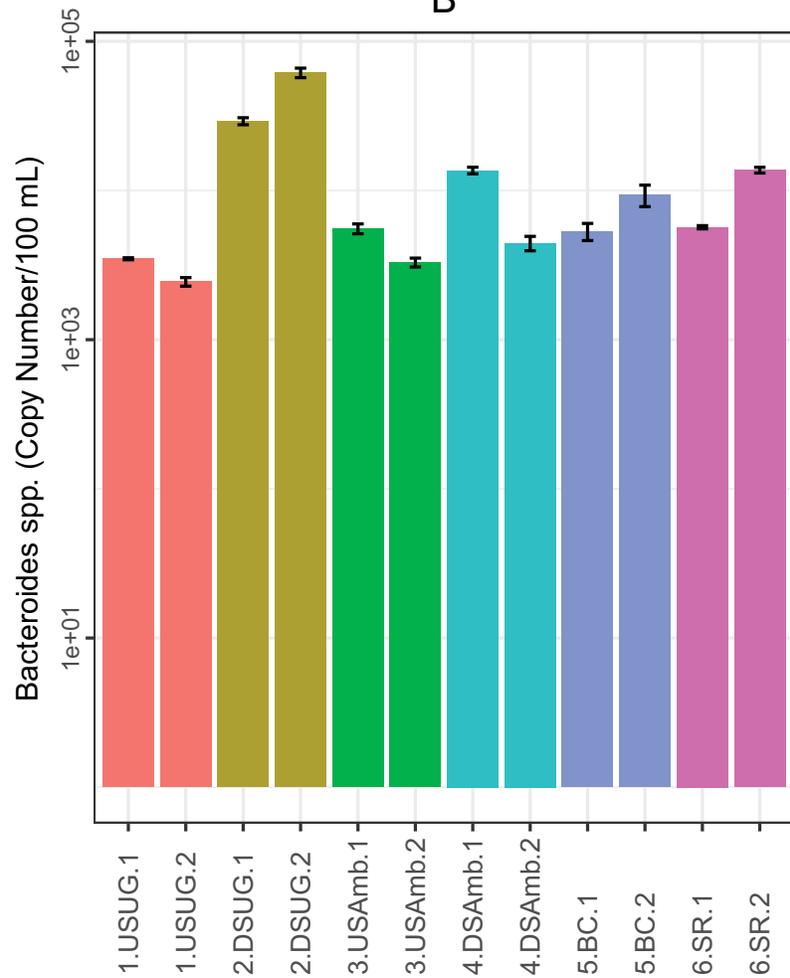
Figure3

[Click here to download Figure: qPCR-&-Bact-&-Bdorei_Figure3.eps](#)

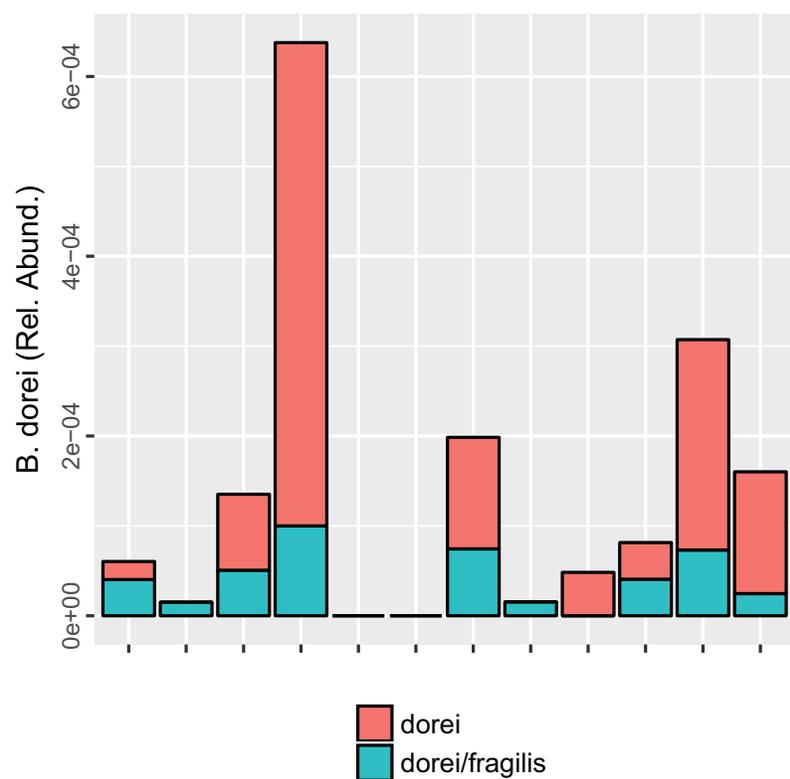
A



B



C



D

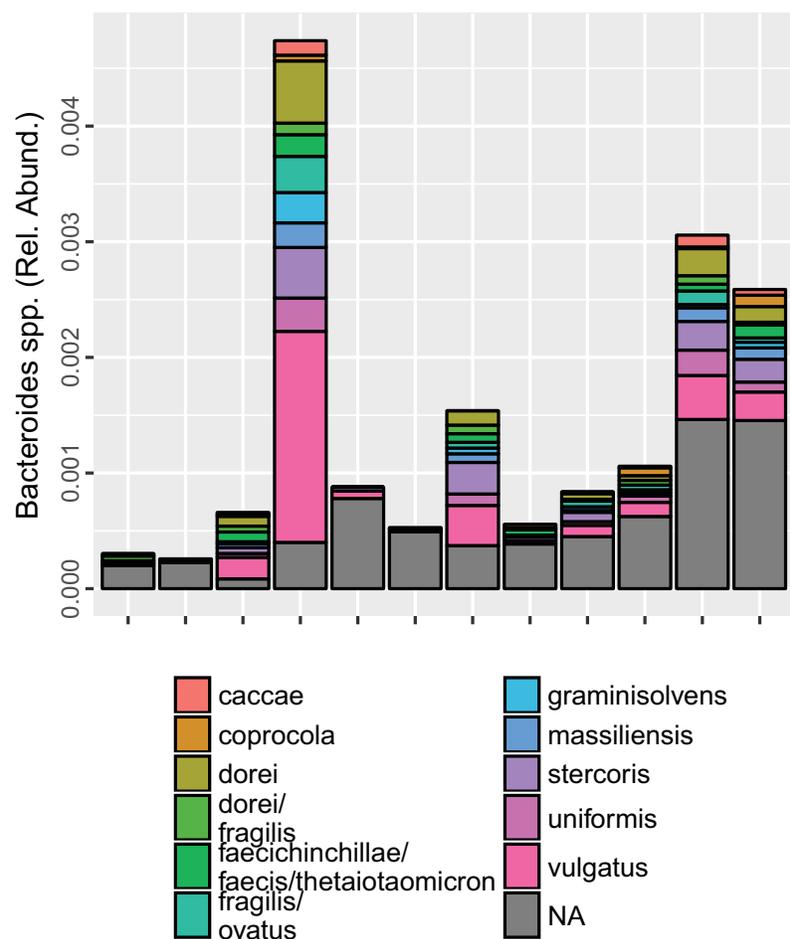
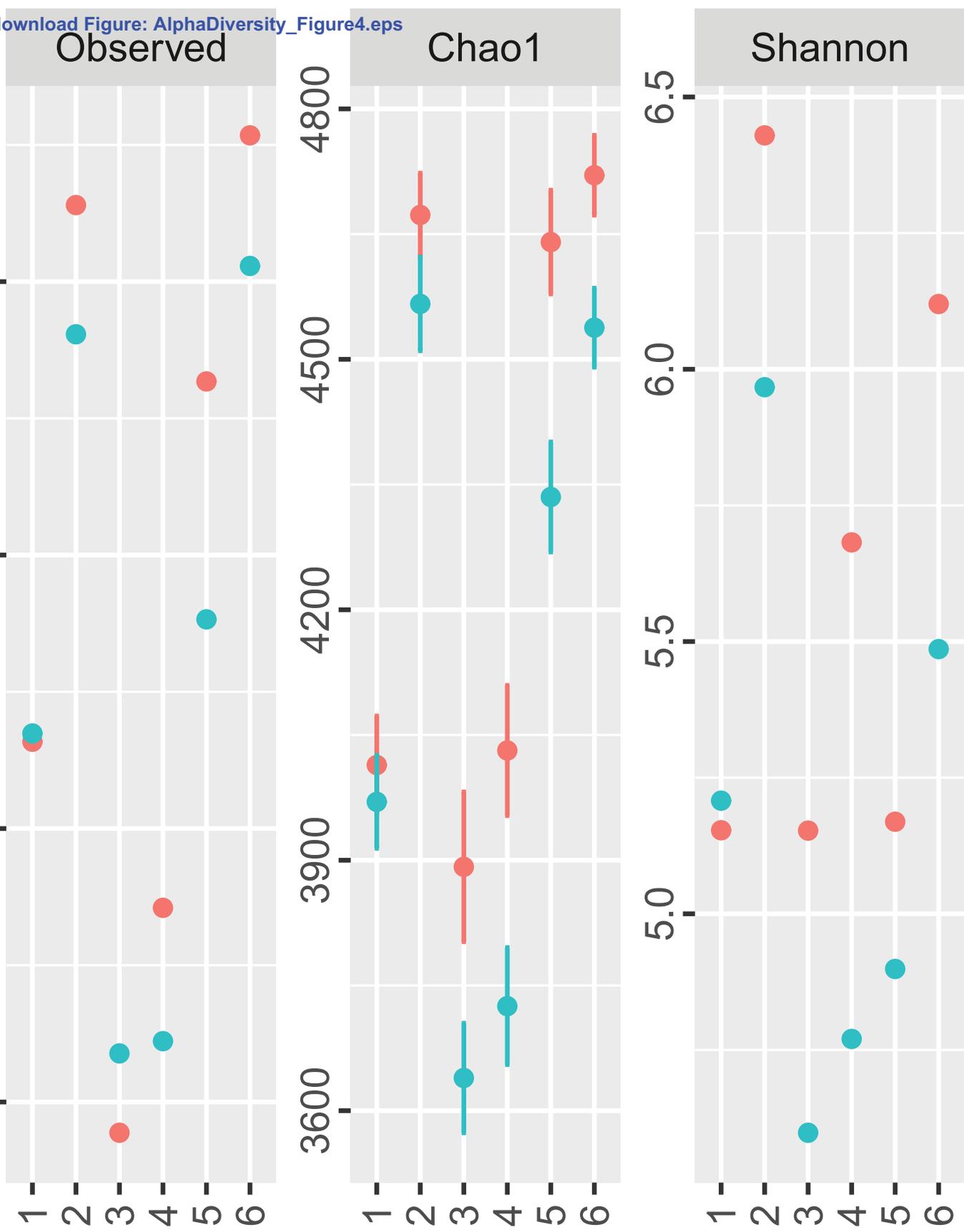


Figure4
[Click here to download Figure: AlphaDiversity_Figure4.eps](#)

Alpha Diversity Measure



Day 1 2

Figure5

[Click here to download Figure: TopLevelRelAbund.f_Figure5.eps](#)

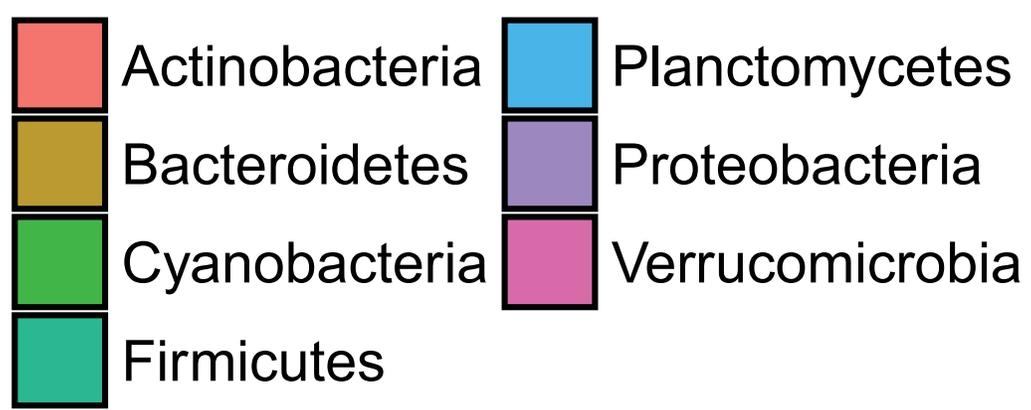
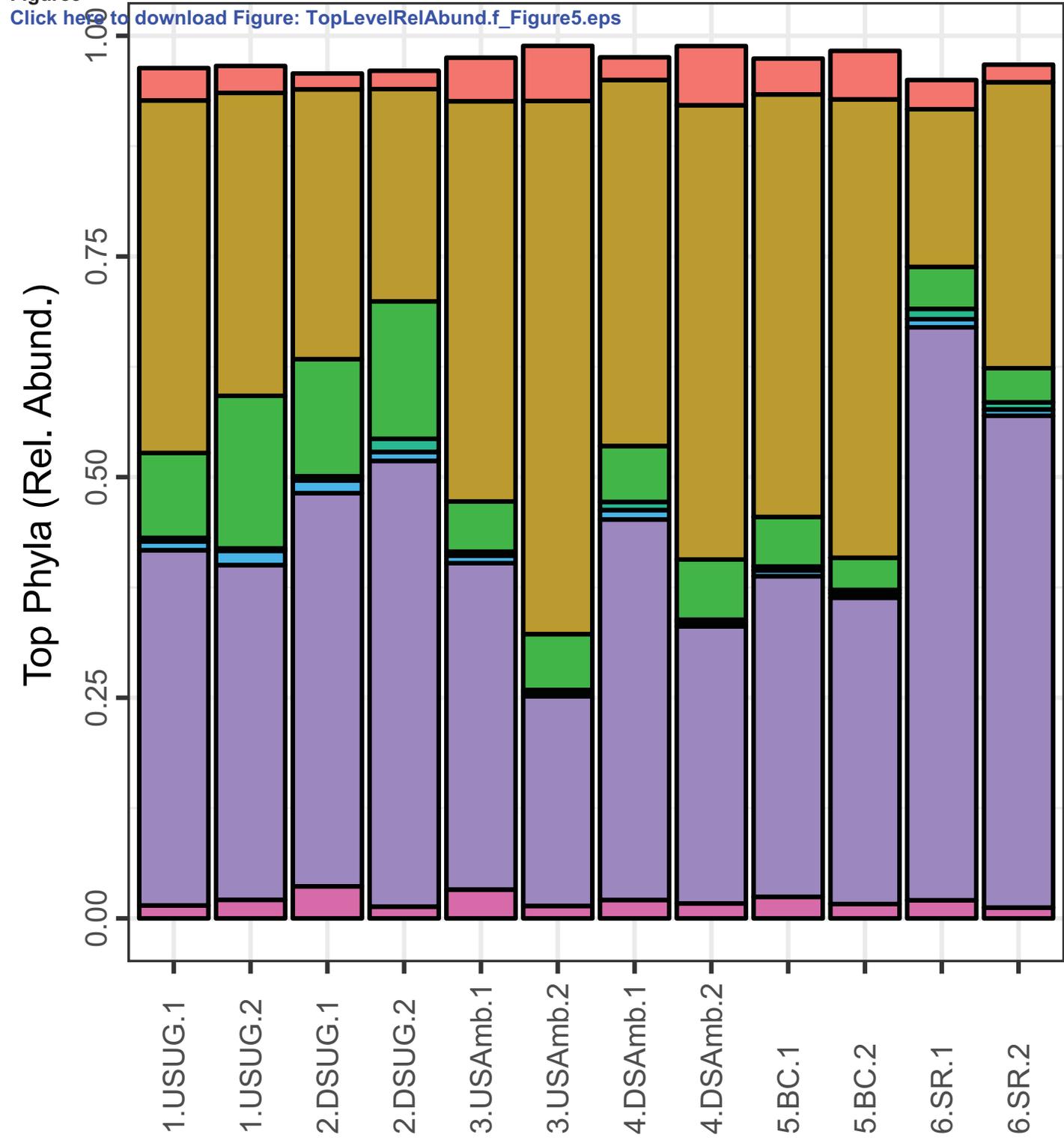


Figure6
Click here to download Figure: PCA_Figure6.eps

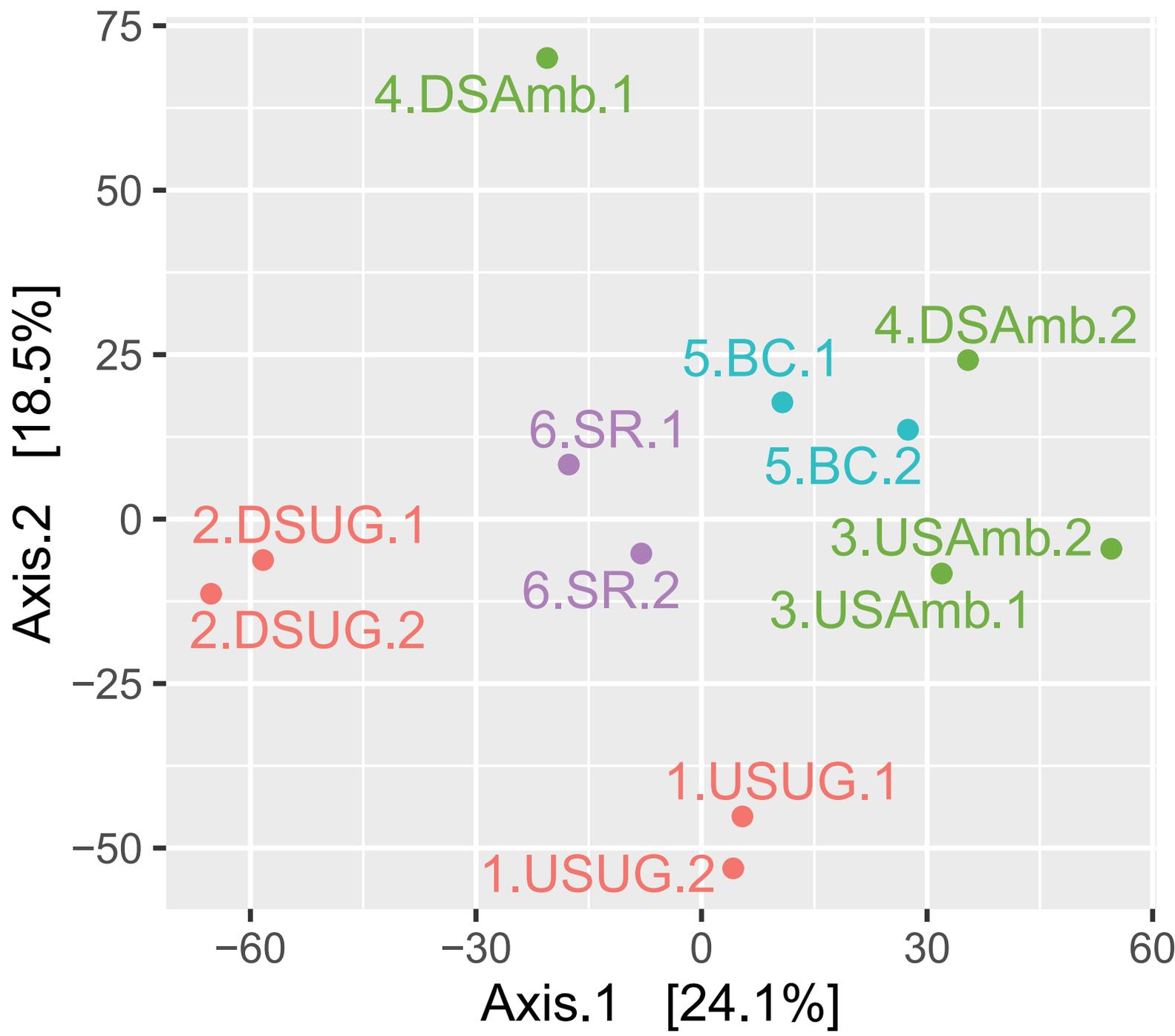
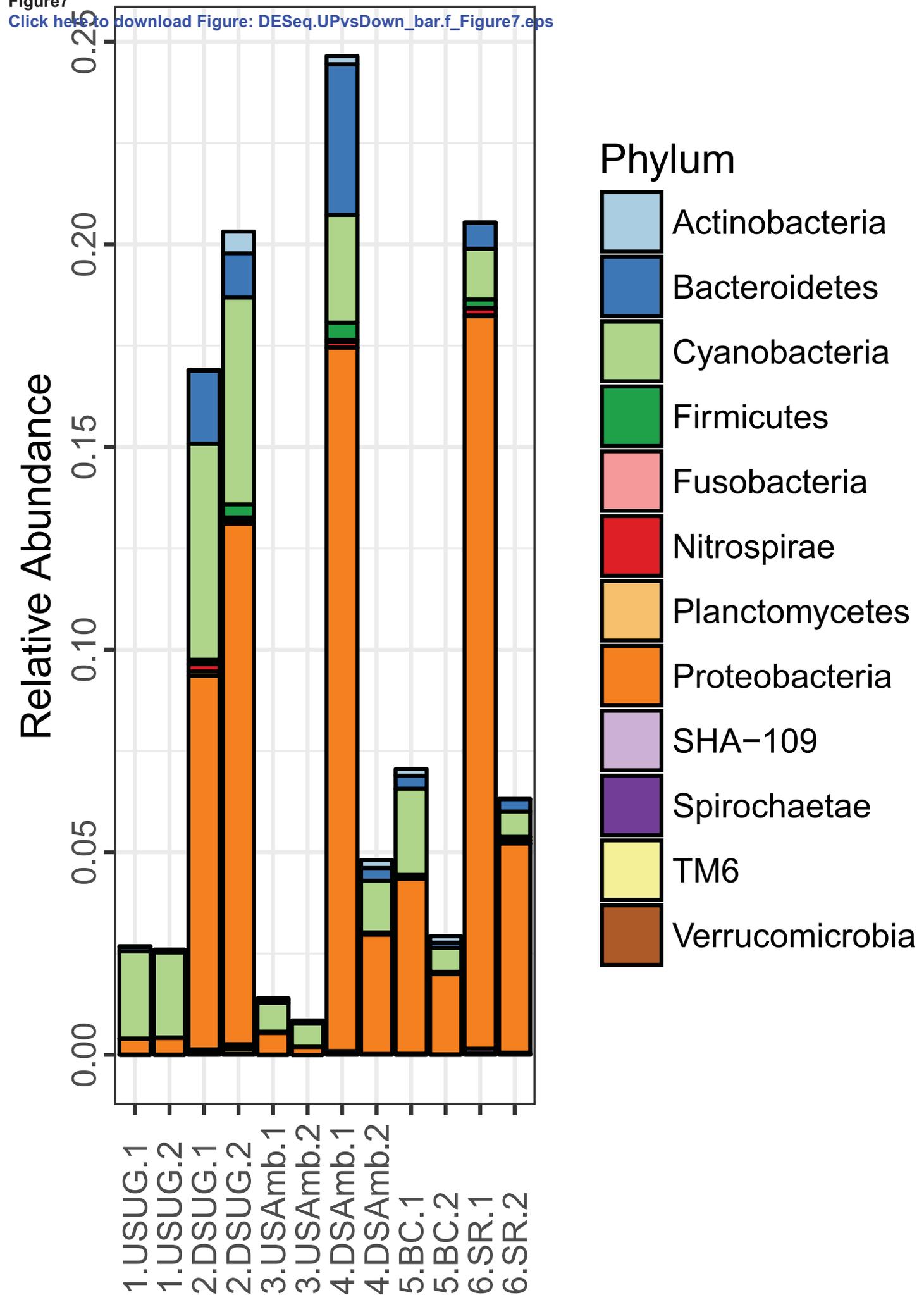


Figure7

[Click here to download Figure: DESeq.UPvsDown_bar.f_Figure7.eps](#)



Supplementary material for on-line publication only

[Click here to download Supplementary material for on-line publication only: SuppMat_Revision_Final_Clean_Merged.pdf](#)

Conflict of Interest

The authors declare no conflict of interest.