

1 Microbial community dynamics of surface water in British 2 Columbia, Canada

3
4
5
6 Miguel I. Uyaguari-Diaz^{1,2*,#a}, Matthew A. Croxen^{1,2*,#b,#c}, Kirby Cronin^{1,#d,#e}, Zhiyao Luo^{2,#f},
7 Judith Isaac-Renton^{1,2}, Natalie A. Prystajecky^{1,2^}, Patrick Tang^{1,2,#g,#h^}

8
9
10 ¹ Department of Pathology and Laboratory Medicine, Faculty of Medicine, The University of
11 British Columbia

12 ² British Columbia Centre for Disease Control Public Health Laboratory

13 ^{#a}Current address: Department of Microbiology, University of Manitoba, Winnipeg, Manitoba,
14 Canada R3T 2N2

16 ^{#b}Current address: Provincial Laboratory for Public Health, Edmonton, Alberta, Canada T6G 2J2

17 ^{#c}Current address: Department of Laboratory Medicine and Pathology, University of Alberta,
18 Edmonton, Alberta, Canada, T6G 2R7

20 ^{#d}Current address: Public Health Ontario Laboratories, Toronto, Ontario, Canada M5G 1M1

21 ^{#e}Current address: National Microbiology Laboratory, Public Health Agency of Canada,
22 Winnipeg, Manitoba, Canada R3E 3R2

25 ^{#f}Current address: British Columbia Cancer Agency, [Provincial Health Services Authority](#),
26 Vancouver, Canada V5Z 4E6

27

28 ^{#g}Current address: Department of Pathology, Sidra Medicine, Doha, Qatar

29 ^{#h}Current address: Department of Pathology and Laboratory Medicine, Weill Cornell Medicine,
30 Doha, Qatar

31 ^ Corresponding authors

32 Email: ptang@sidra.org

33 Email: natalie.prstajecky@bccdc.ca

35 * These authors contributed equally to this work.

36

37

38 **Abstract**

39 Traditional methods for monitoring the microbiological quality of water focus on the detection of
40 fecal indicator bacteria such as *Escherichia coli*, often tested as a weekly grab sample. To
41 understand the stability of *E.coli* concentrations over time, we evaluated three approaches to
42 measuring *E. coli* levels in water: microbial culture using Colilert, quantitative PCR for *uidA* and
43 next-generation sequencing of the 16S rRNA gene. Two watersheds, one impacted by
44 agricultural and the other by urban activities, were repeatedly sampled over a simultaneous ten-
45 hour period during each of the four seasons. Based on 16S rRNA gene deep sequencing, each
46 watershed showed different microbial community profiles. The bacterial microbiomes varied
47 with season, but less so within each 10-hour sampling period. *Enterobacteriaceae* comprised
48 only a small fraction (<1%) of the total community. The qPCR assay detected significantly
49 higher quantities of *E. coli* compared to the Colilert assay and there was also variability in the
50 Colilert measurements compared to Health Canada's recommendations for recreational water
51 quality. From the 16S data, other bacteria such as *Prevotella* and *Bacteroides* showed promise as
52 alternative indicators of fecal contamination. A better understanding of temporal changes in
53 watershed microbiomes will be important in assessing the utility of current biomarkers of fecal
54 contamination, determining the best timing for sample collection, as well as searching for
55 additional microbial indicators of the health of a watershed.

56

57 **Introduction**

58 The use of *Escherichia coli* as a fecal bacterial indicator in water is widely-adopted but also
59 contentious (Ferguson & Signoretto 2011; Isobe et al. 2004; Rochelle-Newall et al. 2015). As a
60 species, *E. coli* is extremely diverse in its ability to colonize a wide range of hosts and in the
61 range of diseases it can cause (Croxen & Finlay 2010; Kaper et al. 2004). Since *E. coli* colonizes
62 the gastrointestinal tract of most birds and animals, and is readily culturable from feces, it seems
63 to be well suited as an indicator of fecal contamination. Many regulatory bodies around the
64 world use *E. coli* as an indicator of water quality and the amount of *E. coli* cultured from water is
65 used as an actionable metric to enact decisions on boil water advisories or beach closures. A
66 major problem with *E. coli* as an indicator organism, however, is the limited amount of
67 information that can be derived from a positive- or negative-culture result. While a positive
68 result may be indicative of recent fecal pollution or also represent environmental re-growth from
69 non-recent fecal deposition, it does not provide information on the source of contamination or on
70 the risk to health. A negative result does not rule out other pathogens such as viruses or protozoa,
71 exemplified by an outbreak in Wisconsin, where over an estimated 400,000 people fell ill due to
72 *Cryptosporidium parvum* oocysts present in municipal waters; this despite treated water samples
73 testing negative for coliforms (Mac Kenzie et al. 1994). Furthermore, while the ability of *E. coli*
74 to survive in the environment appears poor under simulated conditions (van Elsas et al. 2011),
75 there have reports of *Escherichia* spp. naturally present in the environment (phenotypically
76 identical to *E. coli*) (Byappanahalli & Fujioka 1998; Clermont et al. 2011; Walk et al. 2009).
77 These issues raise questions about the usefulness of *E. coli* for routine water quality testing
78 purposes.

79 Routine water quality testing involves the collection of a water sample (“grab sample”) and
80 culturing it for a fecal bacterial indicator such as *E. coli*. Culture-positive results are then
81 reported either as a single maximum concentration or as a geometric mean over a defined
82 sampling period (BC 2017). How representative one result, from one sampling event or averaged
83 from previous week’s results, is uncertain. Culture of two separate 100 ml water samples
84 collected at the same time may yield different *E. coli* results (Wohlsen et al. 2006). Whether this
85 is merely representative of a minor, culturable community member, such as *E. coli*, or true for
86 the entire community, is not yet known.

87 It is important to protect watersheds as deterioration of their health impacts the environment and
88 inter-connected ecosystems. A large metagenomics project was carried out to develop
89 biomarkers to better measure water health compared to current methods
90 (www.watersheddiscovery.ca). The work presented here provides a foundation to better
91 understand the dynamics of source water microbial communities over the course of a day, and
92 during four quarters throughout the year.

93 Materials and methods

94 Water sample collection

95 Two source waters were chosen from the lower mainland area of British Columbia. The first site
96 is a slough that runs through dairy, poultry and produce farms and is impacted by agricultural
97 land use. The second site is impacted by urban land use; it runs through an urban development
98 and includes parkland that is used for recreational purposes. Sampling days were on January 29th,
99 2013 (Quarter 1; Q1), April 29th, 2013 (Q2), July 31st, 2013 (Q3), and October 29th, 2013 (Q4)

100 with typical weather patterns for the area. The first set of samples was collected at 08:00 Pacific
101 Standard time (PST), followed by a second set collected 10 minutes later (08:10 PST). Collection
102 of water in this manner occurred every two hours until 18:10 PST. At each time point, four
103 samples were collected in succession; two 250 ml samples were collected for filtration in 250 ml
104 wide mouth translucent HDPE bottles. Two additional samples were collected for Colilert
105 (IDEXX, Westbrook, ME) in narrow mouth 250 ml high-density polyethylene bottles containing
106 sodium thiosulfate (Systems Plus, Baden, ON). Samples were kept on ice throughout the day,
107 stored at 4 °C overnight, and processed within 24 hours. In total, 48 samples (24 for filtration
108 and 24 for Colilert) were collected at each site, per day for an overall total of 196 filtration
109 samples (24 samples x 2 sites x 4 sampling days) and 196 Colilert samples.

110 **Water filtration, coliform and *E. coli* counts**

111 One hundred milliliters (ml) of water were filtered using 0.2-µm 47-mm Supor-200
112 polyethersulfone membrane disc filters (Pall Corporation, Ann Harbor, MI) inserted into 300 ml
113 Pall 47 mm magnetic polyphenylsulfone filter funnels (Pall Corporation, Ann Harbor, MI). The
114 filters were rolled up and stored at -20°C until needed for nucleic acid extraction. A negative
115 filtration control was also processed using MilliQ water (Millipore Corporation, Billerica, MA).
116 Dilutions were conducted from the bottles containing sodium thiosulfate (1:10 for the urban site,
117 1:100 for the agricultural site) and 100 ml of diluted sample was poured into a Colilert Quanti-
118 Tray/2000 (IDEXX). Trays were sealed according to manufacturer's instructions and incubated
119 at 37°C. Yellow (coliforms) and blue fluorescent wells (*E. coli*) were enumerated after 24 hours
120 of incubation, and used to calculate the most probable number (MPN) according to the table
121 provided with the Colilert kit. MilliQ water was used as a negative control.

122 Nucleic acid extraction, 16S rRNA gene amplification and sequencing

123 Following filtration polyethersulfone membrane disc filters were cut into four pieces. Nucleic
124 acids were extracted from each quarter filter using the MO BIO PowerLyzer PowerSoil DNA
125 Extraction Kit (MoBio, Carlsbad, CA) according the manufacturer's instructions. Extracted
126 DNA samples from each quarter filter were pooled and further concentrated using 0.1 volumes
127 of 3 M sodium acetate, two volumes of 100 % ethanol, and 5 µl of 5 µg/µl linear acrylamide.
128 Eluents were stored at –80 °C overnight, and then centrifuged at 17,000×g for 30 min at 4 °C.
129 Supernatants were discarded, and pellets were washed with 70 % ice-cold ethanol, air dried, and
130 resuspended in 34 µl of 10 mM Tris, pH 8.5. One µl of DNA from each extract was used to
131 amplify the V4 region (515F: 5'- GTGCCAGCMGCCGCGTAA-3' and 806R: 5'-
132 GGAATCTAATGGGTWTCTAAT-3') of the 16S ribosomal RNA (ca. 253 bp) using GoLay
133 primers as previously described (Caporaso et al. 2012). These GoLay primers add compatible
134 adapters and barcodes for the Illumina MiSeq and HiSeq sequencing platforms. The negative
135 processing control was also included in the PCR amplification, as well as a mock community
136 that consisted of *E. coli*, *Pseudomonas putida*, *P. aeruginosa*, *P. fluorescens*, *Burkholderia*
137 *cenocepacia*, *Bacillus amyloliquefaciens*, *B. cereus*, *Rhodobacter capsulatus*, *Streptomyces*
138 *coelicolor*, *Micrococcus luteus*, and *Frankia* sp. CcI3 as described earlier (Peabody et al. 2015).

139 Amplicons were purified using the QIAgen QIAquick PCR purification kit according to
140 manufacturer's instructions. Purified PCR products were quantified using Qubit dsDNA HS
141 Assay Kit (Invitrogen). Each amplicon from a single sampling day (24 agricultural, 24 urban, 1
142 negative control and, 1 mock community) was pooled together for a final equimolar
143 concentration of 4 nM. Each pool (4 total) was diluted to a final loading concentration of 11.5
144 pM, and PhiX at 8 pM was spiked in at 30%. Each pool was run individually on an Illumina

145 MiSeq (Illumina, Inc., San Diego, CA) using the 300-cycle (2 x 150 bp) MiSeq Reagent Kit v2
146 (Illumina). Immediately prior to sequencing, additional primers (100 μ M) were added to the
147 MiSeq reagent cartridge as follows: 3.4 μ l of Read 1 sequencing primer to reservoir 12; 3.4 μ l of
148 Index sequencing primer to reservoir 13; 3.4 μ l of Read 2 sequencing primer to reservoir 14
149 (Caporaso et al. 2012).

150 **Quantitative PCR for *E. coli***

151 Quantitative PCR (qPCR) was used to target the β -glucuronidase gene (*uidA*) of *E. coli* and
152 *Shigella* spp., and amplified an 84 bp product. The oligonucleotides 784F (5'-
153 GTGTGATATCTACCCGCTTCGC-3'), 866R (5'-GAGAACGGTTGTGGTTAACAGGA-
154 3') and TaqMan probe EC807 (5'-FAM-TCGGCATCCGGTCAGTGGCAGT-BHQ1-3') have
155 been previously described (Frahm & Obst 2003), except TAMRA was replaced with the BHQ1
156 quencher. All oligonucleotides were purchased from Integrated DNA Technologies (IDT, Inc.,
157 Coralville, IA).

158 Each qPCR reaction was done in a 20 μ l volume which consisted of 2 μ l of template, 1X
159 PerfeCTa qPCR ToughMix, UNG, Low ROX (Quanta), 0.4 μ M 784F, and 0.4 μ M 866R.
160 Template DNA was diluted 10-fold for agricultural samples, while template DNA from urban
161 samples were not diluted. Amplification was performed on MicroAmp Fast Optical 96-Well
162 Reaction Plates (Life Technologies, Carlsbad, CA) on an Applied Biosystems 7500 Real-Time
163 PCR System (Applied Biosystems). The following cycling conditions were used: 45 °C for 5
164 minutes, 95 °C for 3 minutes, then 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. A
165 standard curve was included in each qPCR run on 10-fold serial dilutions (1,740,000 to 174
166 copies) of genomic DNA from *E. coli* ATCC 25922 that was extracted using the QIAgen

167 QIAamp DNA Mini Kit (Qiagen Sciences, Inc., Germantown, MD) according to the
168 manufacturer's protocol. Each sample was run in duplicate, while standards were run in
169 triplicate. Gene copy numbers were normalized per 100 ml of sample.

170 **16S rRNA gene qPCR assay**

171 To estimate bacterial copy number in water samples, copy numbers of a 16S rRNA gene
172 fragment (~352 bp) of were calculated as described by Ritalahti et al. (2006). Primers
173 Bac1055YF (5'- ATGGYTGTCGTCAAGCT-3') (Ferris et al. 1996; Ritalahti et al. 2006) and
174 Bac1392R (5' - ACGGGCGGTGTGTAC-3') (Lane 1991) were used in combination with Probe
175 Bac1115 containing a 5' 6-FAM dye (CAACGAGCGCAACCC) (Harms et al. 2003; Lane 1991)
176 with an internal ZEN quencher and a 3' Iowa Black fluorescent quencher (Life Technologies,
177 Carlsbad, CA). Due to multiple copy nature of the 16S rRNA gene in a bacterium, copy numbers
178 were normalized by a factor of 4.3 (Lee et al. 2009) per 100 ml of sample. *E. coli* genomic DNA
179 was used for standard curves for *16S rRNA* gene. Each 20 µl real-time PCR reaction consisted of
180 10 µl of ABI TaqMan universal master mix (Life Technologies, Carlsbad, CA), 0.4 µM of each
181 primer, 0.1 µM of probe and template DNA. DNA samples from agricultural site were diluted
182 1:10 using nuclease free-water (Promega Corporation, Fitchburg, WI), while that 1 to 2 µl of
183 undiluted template DNA were used for urban water samples, and 1 µl. Quantitative PCR
184 reactions were conducted on an Applied Biosystems 7500 Fast real-Time PCR system (Life
185 Technologies, Carlsbad, CA). The thermal cycling conditions consisted of incubation for 2 min
186 at 50 °C, initial denaturation for 10 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 60 s
187 at 60 °C. Standards were run in triplicate, while that environmental samples were run in
188 duplicate.

189 **Sequence analysis**

190 Demultiplexed forward and reverse reads were error corrected using BayesHammer (Nikolenko
191 et al. 2013), followed by adapter and primer trimming, and quality filtering using Trimmomatic
192 v0.32 (Bolger et al. 2014). Trimmed and filtered reads were assembled using PANDAseq v2.2
193 (Masella et al. 2012), and any assembled sequences under 220 nucleotides were discarded. The
194 QIIME v1.9.1 package (Caporaso et al. 2010b) was used to compare the microbial communities
195 based on 16S sequences. We used UCLUST (Edgar 2010) to pick Operational Taxonomic Units
196 (OTU) by following a workflow for open-reference clustering (Rideout et al. 2014) at 97% using
197 the GreenGenes v13_8 database (DeSantis et al. 2006). After OTU picking, OTUs with
198 abundances less than 0.005% were removed as recommended by Bokulich et al. (Bokulich et al.
199 2013). ChimeraSlayer (Haas et al. 2011) was used to identify chimeras and the remaining OTUs
200 were aligned using PyNast (Caporaso et al. 2010a) before a phylogenetic tree was constructed
201 using FastTree2 (Price et al. 2010). Bray-Curtis dissimilarity was used to determine beta
202 diversity for each sample. Figures were generated with QIIME, R (ggplots2) (Wickham 2009),
203 and hclust2 (<https://bitbucket.org/nsegata/hclust2>).

204 **Data analysis**

205 All data were \log_{10} transformed for analysis. Spearman's rank correlation analysis was
206 conducted among variables. Paired t-tests were used to assess differences between qPCR and
207 Colilert assays. Data was analyzed using Statistical Analysis System (SAS, version 9.1 for
208 Windows). A p-value of 0.05 was assumed for all tests as a minimum level of significance.

209 **Data availability**

210 All sequences generated for this study are available under NCBI BioProject PRJNA287840,
211 SRA samples SRR2083863 through SRR2084062.

212

213 **Results and discussion**

214 **Sampling sites**

215 The lower mainland of British Columbia (BC) has temperate weather. We chose two study
216 watersheds that are impacted differently; one is impacted by farming land use (Agricultural;
217 AGR), and one has more anthropogenic impact through proximity to residential and recreational
218 land use (Urban; URB). Fig 1 represents AGR and URB sites with land cover. Distance between
219 these non-connected watersheds is ~65 Km.

220 ***E. coli* counts in watershed samples can range depending on sampling event 221 and time**

222 We collected surface water samples during the winter (Q1), spring (Q2), summer (Q3) and fall
223 (Q4). We recognized that due to the single sampling event, these results are still a snapshot in
224 that particular quarter of the year and should not be interpreted as representative of an entire
225 season, or even reflective of a typical season in the study area. As well, we collected samples
226 during a defined daily period. Water samples for assessing water quality in British Columbia are
227 typically collected once a week for *E. coli* culture. To also assess the variability of *E. coli* counts
228 in the agricultural- and urban-impacted watershed study sites, we collected samples in duplicate,
229 followed in duplicate again 10 minutes later every two hours, from 08:00 hrs until 18:00 hrs.

230 Agreement between duplicate samples varied; sometimes samples did not differ, when other
231 times they did. The most extreme example of this were the results of Agricultural Q1 at the 16:10
232 hrs time point. Sample A showed 100 *E. coli* MPN/100 ml, whereas the duplicate sample (B) had
233 12X more culturable *E. coli* cells (Fig 2). Similarly, *E. coli* counts fluctuated during the different
234 time points during the day in both the agricultural and urban watersheds. Health Canada's
235 recommendations for recreational water quality is 400 *E. coli* MPN in 100 ml from a single
236 sampling event (Canada 2012; Levesque & Gauvin 2007). Some *E. coli* counts were variable
237 during the day and there are many instances where the sampling time would have made a
238 difference to the interpretation and concomitant action of public health officials based on Health
239 Canada's guidelines (red dotted line; Fig 2). Because of sampling location and transportation to
240 the laboratory, we acknowledge sample holding times exceeded the preferred time interval of 8
241 hours from collection to examination (APHA 2005; Bartram & Rees 2000). In the present study,
242 samples were processed within 24 hours. No significant differences in counts of *E. coli* and total
243 coliforms have been reported in water samples up to 27 hours since collection (Aulenbach 2010;
244 Maier et al. 2015; Pope et al. 2003).

245

246 Comparison of *E.coli* counts and *uidA* gene quantitative PCR assay

247 We then compared the number of viable *E. coli* cells to the single copy *uidA* gene, detected by
248 qPCR. There was a moderate positive correlation between the culturable *E. coli* and the qPCR
249 results ($r_s = 0.5374$, p-value<0.0001), suggesting that both methods shared an agreement of
250 28.9% of common values. This comparison also indicated a difference of approximately 2 orders
251 of magnitude in test results, with higher numbers detected by the *uidA* gene qPCR assay (p-
252 value<0.0001). While the *uidA* gene is specific for *E. coli* and some *Shigella* spp. (Maheux et al.

253 2011), this qPCR assay detects DNA from viable, viable but non-culturable, as well as dead
254 cells. When compared to other studies (Noble et al. 2010; Oliver et al. 2016), this association
255 seems to be low, however a drop in correlation coefficients have been reported in environmental
256 water samples (Walker et al. 2017).

257

258 **Microbial community diversity differ in the long-term, but not short-term**

259 Since *E. coli* counts varied throughout the day, we wanted to examine the dynamics of bacterial
260 communities. We used the same water samples to conduct deep amplicon sequencing. In
261 agricultural sites and across seasons, clusters clearly stood out from each other suggesting a more
262 or less steady microbial community all day long (Fig 3A). While clusters were also found in
263 urban sites, these data points were more spread out within the groups with some overlap in the
264 community composition (Fig 3B). Factors such as water use, runoff, possible leaking of
265 wastewater or seepage (Vermonden 2010; Vermonden et al. 2009) may explain such
266 perturbations (Fig 3B). Results from the 16S rRNA gene sequencing and top 50 OTUs at the
267 family level are depicted in Fig 4. *Enterobacteriaceae* constituted a relatively small fraction
268 (<1%) of the total microbial community. Relative abundance values of 0.06% and 0.76% were
269 observed for this family in AGR (Fig 4A) and URB (Fig 4B), respectively. Although *E. coli* was
270 detected using both the *uidA* qPCR and testing methods, it could not be detected using deep
271 amplicon sequencing. This observation, not only for *E.coli* but also for other sequences, may be
272 related to amplicon fragment size (254 bp). This may be a limitation to obtaining sufficient
273 genetic information (Calus et al. 2018) to identify bacteria at the genus or species level.

274

275 In the agricultural study site, two clusters were observed for bacterial communities between Q4
276 (fall) and Q3 (summer), Q2 (spring) and Q1 (winter). Within this latter group, 3 sub-groups were
277 identified. In both sampling sites, the most abundant OTUs were in the family *Comamonadaceae*
278 with average values across seasons of 28.1% and 18.9% for AGR and URB, respectively. Similar
279 estimates have been reported for this family in aquatic systems impacted by urbanization or
280 agricultural activities (Griffin et al. 2017; Lopes et al. 2016; Newton & McLellan 2015; Willems
281 2014). Although *Comamonadaceae* was predominant across seasons, different patterns were
282 observed for this family. For example, in AGR, a higher abundance of *Comamonadaceae* was
283 observed during Q3, while members of this family were more abundant during Q4 in URB. A
284 possible explanation for these patterns is the manure application occurring on study site farms
285 during this part of the year (summer) and weather related runoff into the sampling site. In URB,
286 a higher abundance of *Comamonadaceae* was observed in Q4, corresponding to the onset of the
287 rainy season with storm water runoff in the region. While not considered human pathogens, two
288 genera (*Xylophilus* and some species of *Acidovorax*) within this family affect plants (Willems
289 2014). Other relevant major families ($\geq 4\%$) observed in AGR were *Flavobacteriaceae* (6.7%),
290 *Crenotrichaceae* (5.4%) and *Campylobacteraceae* (4.3%). Members of the first two families are
291 widely distributed in the environment (McBride 2014; Siljanen et al. 2011; Urios et al. 2006),
292 with similar estimates being reported in agricultural settings (Pandey et al. 2018; Shawkey et al.
293 2009) and water-sediment interfaces (Chidamba & Korsten 2015; Frindte et al. 2016). On the
294 other hand, some members of *Campylobacteraceae* such as *Campylobacter* and *Arcobacter*
295 contain species that are well-known human pathogens (Lastovica et al. 2014; Lehner et al. 2005).
296 Although *Campylobacteraceae* was detected in both sites, we observed an average of ten-fold
297 difference between AGR (4.1%)(Fig 4A) and URB (0.4%) (data not shown). Interestingly, the

298 relative abundance of *Campylobacteraceae* in AGR was at least twice as high during Q1 (winter)
299 compared to the other seasons (Fig 4A), with no major changes observed in URB. Note that
300 members of *Campylobacteraceae* are livestock and poultry associated (Mughini Gras et al.
301 2012), both of which are farmed in the study area. This *Campylobacteraceae* prevalence,
302 concentration and survival during winter compared to other seasons in AGR, may be associated
303 to with their sensitivity to desiccation during drier months (Smith et al. 2016), occurrence of
304 heavy rainfall (Ahmed et al. 2013; Moriarty et al. 2011), their presence in waterfowl feces
305 (Moriarty et al. 2012), or various farming practices (Rapp et al. 2014). Alternatively this could
306 also represent detection of material from bacterial lysis (Feng et al. 2017). Another family
307 observed with high relative abundance (10-30%) in AGR was *Phormidiaceae* but this family was
308 only abundant during fall (Q4), and completely absent or in very low abundance during other
309 seasons (Fig 4A). Members of this family are common in poor quality streams with precipitation
310 triggering its bloom during fall (Hossain et al. 2012; Kaestli et al. 2017).
311 It is important to mention that in URB fecally-associated bacteria such as *Prevotella* (2.20%),
312 *Bacteroides* (1.29%), *Blautia* (0.76%), *Faecalibacterium* (0.57%) and *Coprococcus* (0.24%)
313 were observed (Figs 4B and 5). These genera have been proposed as potential indicators of
314 sewage and human fecal contamination (Eren et al. 2015; Fisher et al. 2015; Fogarty & Voytek
315 2005; McLellan et al. 2013). Significant positive correlations were detected among these genera
316 ($r \geq 0.5137$, $p < 0.001$). Major peaks on these genera of bacteria in URB occurred in Q1 during 8
317 AM and 2 PM (PST) (Fig 5). Moreover, a relatively high abundance of these genera was
318 observed during Q2, Q3 and Q4 between 2 PM and 4PM. Besides precipitation (17.8 mm) in
319 URB during Q1, these observations may represent water use (i.e. flushing events) occurring
320 during the day; with the latter time reflecting changes associated to daylight saving time. When

321 compared to the agricultural watershed, *Prevotella*, *Bacteroides*, and *Coprococcus* were present
322 in AGR, but in combination these genera made up less than 0.23% of the total microbial
323 community. We observed a positive correlation between *Prevotella* and *Bacteroides* ($r < 0.7479$,
324 $p < 0.001$). Genera such as *Blautia* and *Faecalibacterium* were not detected in AGR using with
325 deep amplicon sequencing.

326 The prevalence of *Prevotella* and *Bacteroides* in URB may also reflect biomarkers of the
327 sampled urban setting (Gorvitovskaia et al. 2016). Another relevant taxa reported to be part of
328 the human gut microbiome (Arumugam et al. 2011) and observed in our study was
329 *Ruminococcus*. However, low average values of 0.07% and 0.10% were found for AGR and
330 URB, respectively. The fact that no major differences between sites were observed across
331 seasons for *Ruminococcus* may indicate either an anthropogenic influence occurring in AGR. We
332 also cannot rule out wildlife as contributor of this genera (Arumugam et al. 2011). Overall,
333 genera such *Prevotella* and *Bacteroides* were present in higher relative abundance in URB
334 compared to the other fecally associated bacteria. These two genera have been proposed as
335 alternative indicator to assess water quality impacted by anthropogenic activities (Fogarty &
336 Voytek 2005).

337 To potentially identify novel biomarkers of water quality more detectable than *E.coli*, additional
338 correlation analyses were conducted with the 10 more abundant bacterial families in each
339 watershed and physico-chemical water quality parameters (Fig 6). In AGR and URB sites there
340 was a moderate to strong positive correlation among variables such as 16S rRNA, *uidA* gene, *E.*
341 *coli* colonies and total coliforms (as measured by Colilert) (Fig 6). Among these variables, 16S
342 rRNA gene copy numbers were positively correlated with *Comamonadaceae*, the most abundant
343 bacteria detected in AGR and URB by deep amplicon sequencing. Moreover, water quality

344 parameters related with conductivity (specific conductivity, salinity, total dissolved solids, and
345 oxidation-reduction potential) were positively ($p<0.0001$) correlated with *Comamonadaceae*. In
346 contrast, this family was negatively correlated with dissolved oxygen (DO). This distinguishable
347 pattern has also been reported for *Comamonadaceae* in anthropogenic impacted aquatic systems
348 (Aguirre et al. 2017).

349 Temperature had a positive correlation with *Comamonadaceae* and a negative correlation with
350 *Crenotrichaceae* in both watersheds. In AGR, pH was significantly correlated with seven of the
351 most abundant families including *Crenotrichaceae*, *Phormidiaceae*, *Actinomycetales*,
352 *Oxalobacteraceae*, *Helicobacteraceae*, *Gallionellaceae*, *Methylophilaceae*. Moderate changes in
353 pH may cause community shifts especially in less abundant groups of bacteria or pH-susceptible
354 groups (Krause et al. 2012) as the ones described in our study. Similar to the observations
355 reported by Krause et al. 2012, pH did not have a significant effect on the overall bacterial
356 abundance in both watersheds as estimated by the 16S rRNA gene (Fig 6).

357 Dissolved oxygen and precipitation were parameters highly correlated between them in both
358 watersheds. This association was reflected by members of the *Crenotrichaceae* family in AGR
359 and URB. Other families associated to these 2 variables and mostly positively associated
360 between them included *Flavobacteriaceae*, *Prevotellaceae*, *Lachnospiraceae*, a proposed novel
361 bacterial division BD4-9 (possibly from the phylum Verrucomicrobia) (Briee et al. 2007;
362 Derakshani et al. 2001), *Bacteroidaceae* and *Moraxellaceae* in URB. These latter two families
363 were positively and strongly correlated between each other in URB ($r_s=0.8103$, $p\text{-value}<0.0001$)
364 and AGR ($r_s=0.8319$, $p\text{-value}<0.0001$). Additional positive associations between bacteria from
365 fecal origin in URB such as *Lachnospiraceae*, *Bacteroidaceae* and *Prevotellaceae* were also
366 determined by correlation analysis (Fig 6B), and may reflect the microbiome of human

367 populations and in particular, sewage (McLellan et al. 2013; Newton & McLellan 2015).

368 Although multiple correlations were observed in the present study, we only described relevant
369 associations that governed the most abundant microbes in AGR and URB (Fig 6).

370

371 Conclusions

372 Metagenomic analysis of bacterial communities indicated fairly stable patterns, while *E. coli*
373 counts varied widely over time. A moderate correlation was detected between culturable *E. coli*
374 (as measured by Colilert) and *uidA* gene. Nevertheless, a difference of 2 orders of magnitude was
375 observed between these approaches. The structure of microbial communities did not change
376 significantly on an hourly basis, however dramatic changes were observed across seasons in the
377 two study locations. Although each watershed exhibited different microbial profile patterns,
378 *Comamonadaceae* was the most abundant bacteria detected in both watersheds by 16S rRNA
379 gene deep sequencing. Relative abundance of *Campylobacteraceae* indicated a ten-fold
380 difference between AGR (4.1%) and URB (0.4%). Moreover, members of the family
381 *Enterobacteriaceae* made up less than 1% of the total microbial community in AGR and URB.
382 In this context, *E. coli* could not even be detected at the genus level. Genera such *Prevotella* and
383 *Bacteroides* were present in higher relative abundance in URB compared to other fecally
384 associated bacteria, which may suggest alternatives in the search for better biomarkers of fecal
385 pollution. Precipitation and DO were positively correlated with seven of the most abundant
386 bacterial families in URB, while that in AGR, pH was positively correlated with the same
387 number of families. Besides a taxa-based approach such as 16S rRNA gene sequencing, the use
388 of metagenomics may be an opportunity to discover better fecal pollution biomarkers. Discovery
389 approaches could include use of taxonomic information but also information at the metabolic and

390 functional level (Simon & Daniel 2011). Although it has been demonstrated that microbial
391 communities may change over short distances (10 km) (Hewson et al. 2006), we found that the
392 study watersheds showed generally stable community structure over a 10-hour period, but we did
393 see notable differences comparing across seasons. As the accepted standard for measuring
394 watershed health, it would be ideal to have a more somewhat less variable (counts differed
395 greatly between samples collected at 10-minute intervals) signal. This supports our belief that
396 biomarkers other than *E. coli* should be sought as better indicators of fecal contamination. Future
397 studies could examine more intermediate time intervals such as sampling per week or per month,
398 to assess possible significant changes in microbial communities at these intervals. All this
399 information will assist in developing better methods of monitoring of water quality.

400 **Author Contributions**

401 **Conceptualization:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen, Natalie A. Prystajecky,
402 Patrick Tang.

403 **Data curation:** Matthew A. Croxen, Miguel I. Uyaguari-Diaz, Kirby Cronin, Zhiyao Luo.

404 **Formal analysis:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

405 **Funding acquisition:** Judith Isaac-Renton, Patrick Tang.

406 **Investigation:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

407 **Methodology:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen, Kirby Cronin, Zhiyao Luo.

408 **Project administration:** Judith Isaac-Renton, Natalie A. Prystajecky, Patrick Tang.

409 **Resources:** Natalie A. Prystajecky, Patrick Tang.

410 **Software:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

411 **Supervision:** Judith Isaac-Renton, Natalie A. Prystajecky, Patrick Tang.

412 **Validation:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

413 **Visualization:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

414 **Writing – original draft:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen.

415 **Writing – review & editing:** Miguel I. Uyaguari-Diaz, Matthew A. Croxen, Kirby Cronin,

416 Zhiyao Luo, Judith Isaac-Renton, Natalie A. Prystajecky, Patrick Tang.

417 **Acknowledgements**

418 We would like to thank Bridget Lee, Michael Peabody, Thea van Rossum, Alvin Xian, Mitchell
419 Burton, Sara Tan and Tyler Nelson for help with the water sampling. We would also like to
420 thank Dr. Fiona Brinkman (Simon Fraser University) for providing the mock community.

421 **References**

422 Aguirre M, Abad D, Albaina A, Cralle L, Goni-Urriza MS, Estonba A, and Zarraonaindia I.
423 2017. Unraveling the environmental and anthropogenic drivers of bacterial community
424 changes in the Estuary of Bilbao and its tributaries. *PLoS One* 12:e0178755.
425 10.1371/journal.pone.0178755

426 Ahmed MF, Schulz J, and Hartung J. 2013. Survival of *Campylobacter jejuni* in naturally and
427 artificially contaminated laying hen feces. *Poult Sci* 92:364-369. 10.3382/ps.2012-02496
428 APHA. 2005. *Standard methods for the examination of water and wastewater*. Washington, DC.:
429 American Public Health Association, American Water Works Association and Water
430 Environment Federation.

431 Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J,
432 Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T,
433 Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh
434 C, Nielsen HB, Nielsen T, Pons N, Poulaing J, Qin J, Sicheritz-Ponten T, Tims S, Torrents
435 D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S,
436 Dore J, Meta HITC, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C,
437 Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss

- 438 C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A,
439 Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo
440 Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N,
441 Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y,
442 Zeller G, Weissenbach J, Ehrlich SD, and Bork P. 2011. Enterotypes of the human gut
443 microbiome. *Nature* 473:174-180. 10.1038/nature09944
- 444 Aulenbach BT. 2010. Bacteria holding times for fecal coliform by mFC agar method and total
445 coliform and Escherichia coli by Colilert-18 Quanti-Tray method. *Environ Monit Assess*
446 161:147-159. 10.1007/s10661-008-0734-3
- 447 Bartram J, and Rees G. 2000. *Monitoring bathing waters: a practical guide to the design and*
448 *implementation of assessments and monitoring programmes*. London, UK: E & FN Spon.
- 449 BC. 2017. BC recreational water quality guidelines: guidelines summary. In: branch
450 WqaasWpas, editor. Victoria, BC: British Columbia Ministry of Environment and
451 Climate Change Strategy. p 13.
- 452 Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, and Caporaso
453 JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon
454 sequencing. *Nat Methods* 10:57-59. 10.1038/nmeth.2276
- 455 Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
456 sequence data. *Bioinformatics* 30:2114-2120. 10.1093/bioinformatics/btu170
- 457 Brieu C, Moreira D, and Lopez-Garcia P. 2007. Archaeal and bacterial community composition
458 of sediment and plankton from a suboxic freshwater pond. *Res Microbiol* 158:213-227.
459 10.1016/j.resmic.2006.12.012
- 460 Byappanahalli MN, and Fujioka RS. 1998. Evidence that tropical soil environment can support
461 the growth of Escherichia coli. *Water Science and Technology* 38:171-174.
[https://doi.org/10.1016/S0273-1223\(98\)00820-8](https://doi.org/10.1016/S0273-1223(98)00820-8)
- 463 Calus ST, Ijaz UZ, and Pinto AJ. 2018. NanoAmpli-Seq: A workflow for amplicon sequencing
464 from mixed microbial communities on the nanopore sequencing platform. *bioRxiv*.
465 10.1101/244517
- 466 Canada H. 2012. Guidelines for Canadian recreational water quality. . Third ed. Ottawa, Ontario:
467 Minister of Health. p 4.
- 468 Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, and Knight R. 2010a.
469 PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*
470 26:266-267. 10.1093/bioinformatics/btp636
- 471 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena
472 AG, Goodrich JK, Gordon JI, Hutley GA, Kelley ST, Knights D, Koenig JE, Ley RE,
473 Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh
474 PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, and Knight R. 2010b. QIIME
475 allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335-336.
476 10.1038/nmeth.f.303
- 477 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J,
478 Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, and Knight R. 2012. Ultra-high-
479 throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.
480 *ISME J* 6:1621-1624. 10.1038/ismej.2012.8
- 481 Chidamba L, and Korsten L. 2015. Pyrosequencing analysis of roof-harvested rainwater and
482 river water used for domestic purposes in Luthengele village in the Eastern Cape
483 Province of South Africa. *Environ Monit Assess* 187:41. 10.1007/s10661-014-4237-0

- 484 Clermont O, Gordon DM, Brisse S, Walk ST, and Denamur E. 2011. Characterization of the
485 cryptic *Escherichia* lineages: rapid identification and prevalence. *Environ Microbiol*
486 13:2468-2477. 10.1111/j.1462-2920.2011.02519.x
- 487 Croxen MA, and Finlay BB. 2010. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 8:26-38. 10.1038/nrmicro2265
- 488 Derakshani M, Lukow T, and Liesack W. 2001. Novel bacterial lineages at the (sub)division
489 level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk
490 soil and rice roots of flooded rice microcosms. *Appl Environ Microbiol* 67:623-631.
491 10.1128/AEM.67.2.623-631.2001
- 492 DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P,
493 and Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and
494 workbench compatible with ARB. *Appl Environ Microbiol* 72:5069-5072.
495 10.1128/AEM.03006-05
- 496 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
497 26:2460-2461. 10.1093/bioinformatics/btq461
- 498 Eren AM, Sogin ML, Morrison HG, Vineis JH, Fisher JC, Newton RJ, and McLellan SL. 2015.
499 A single genus in the gut microbiome reflects host preference and specificity. *ISME J*
500 9:90-100. 10.1038/ismej.2014.97
- 501 Feng J, Ma L, Nie J, Konkel ME, and Lu X. 2017. Environmental stress-induced bacterial lysis
502 and extracellular DNA release contribute to *Campylobacter jejuni* biofilm formation.
503 *Appl Environ Microbiol*. 10.1128/AEM.02068-17
- 504 Ferguson D, and Signoretto C. 2011. Environmental Persistence and Naturalization of Fecal
505 Indicator Organisms. In: Hagedorn C, Blanch AR, and Harwood VJ, eds. *Microbial
506 Source Tracking: Methods, Applications, and Case Studies*. New York, NY: Springer
507 New York, 379-397.
- 508 Ferris MJ, Muyzer G, and Ward DM. 1996. Denaturing gradient gel electrophoresis profiles of
509 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl
510 Environ Microbiol* 62:340-346.
- 511 Fisher JC, Eren AM, Green HC, Shanks OC, Morrison HG, Vineis JH, Sogin ML, and McLellan
512 SL. 2015. Comparison of Sewage and Animal Fecal Microbiomes by Using Oligotyping
513 Reveals Potential Human Fecal Indicators in Multiple Taxonomic Groups. *Appl Environ
514 Microbiol* 81:7023-7033. 10.1128/AEM.01524-15
- 515 Fogarty LR, and Voytek MA. 2005. Comparison of bacteroides-prevotella 16S rRNA genetic
516 markers for fecal samples from different animal species. *Appl Environ Microbiol*
517 71:5999-6007. 10.1128/AEM.71.10.5999-6007.2005
- 518 Frahm E, and Obst U. 2003. Application of the fluorogenic probe technique (TaqMan PCR) to
519 the detection of *Enterococcus* spp. and *Escherichia coli* in water samples. *J Microbiol
520 Methods* 52:123-131.
- 521 Frindte K, Allgaier M, Grossart H-P, and Eckert W. 2016. Redox stability regulates community
522 structure of active microbes at the sediment–water interface. *Environmental
523 Microbiology Reports* 8:798-804. 10.1111/1758-2229.12441
- 524 Gorvitovskaia A, Holmes SP, and Huse SM. 2016. Interpreting Prevotella and Bacteroides as
525 biomarkers of diet and lifestyle. *Microbiome* 4:15. 10.1186/s40168-016-0160-7
- 526 Griffin JS, Lu N, Sangwan N, Li A, Dsouza M, Stumpf AJ, Sevilla T, Culotti A, Keefer LL,
527 Kelly JJ, Gilbert JA, Wells GF, and Packman AI. 2017. Microbial diversity in an

- 529 intensively managed landscape is structured by landscape connectivity. *FEMS Microbiol*
530 *Ecol* 93. 10.1093/femsec/fix120
- 531 Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D,
532 Highlander SK, Sodergren E, Methe B, DeSantis TZ, Human Microbiome C, Petrosino
533 JF, Knight R, and Birren BW. 2011. Chimeric 16S rRNA sequence formation and
534 detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21:494-504.
535 10.1101/gr.112730.110
- 536 Harms G, Layton AC, Dionisi HM, Gregory IR, Garrett VM, Hawkins SA, Robinson KG, and
537 Sayler GS. 2003. Real-time PCR quantification of nitrifying bacteria in a municipal
538 wastewater treatment plant. *Environ Sci Technol* 37:343-351.
- 539 Hewson I, Steele JA, Capone DG, and Fuhrman JA. 2006. Temporal and spatial scales of
540 variation in bacterioplankton assemblages of oligotrophic surface waters. *Marine Ecology
541 Progress Series* 311:67-77.
- 542 Hossain MY, Jewel MAS, Fulanda B, Ahamed F, Rahman S, Jasmine S, and Ohtomi J. 2012.
543 Dynamics of cyanobacteria Planktothrix species (Oscillatoriales: Phormidiaceae) in
544 earthen fish ponds, northwestern Bangladesh.
- 545 Isobe KO, Tarao M, Chiem NH, Minh le Y, and Takada H. 2004. Effect of environmental factors
546 on the relationship between concentrations of coprostanol and fecal indicator bacteria in
547 tropical (Mekong Delta) and temperate (Tokyo) freshwaters. *Appl Environ Microbiol*
548 70:814-821.
- 549 Kaestli M, Skillington A, Kennedy K, Majid M, Williams D, McGuinness K, Munksgaard N,
550 and Gibb K. 2017. Spatial and Temporal Microbial Patterns in a Tropical Macrotidal
551 Estuary Subject to Urbanization. *Front Microbiol* 8:1313. 10.3389/fmicb.2017.01313
- 552 Kaper JB, Nataro JP, and Mobley HL. 2004. Pathogenic Escherichia coli. *Nat Rev Microbiol*
553 2:123-140. 10.1038/nrmicro818
- 554 Krause E, Wichels A, Giménez L, Lunau M, Schilhabel MB, and Gerdts G. 2012. Small Changes
555 in pH Have Direct Effects on Marine Bacterial Community Composition: A Microcosm
556 Approach. *PLoS One* 7:e47035. 10.1371/journal.pone.0047035
- 557 Lane D. 1991. 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics*:125-
558 175.
- 559 Lastovica AJ, On SLW, and Zhang L. 2014. The Family Campylobacteraceae. In: Rosenberg E,
560 DeLong EF, Lory S, Stackebrandt E, and Thompson F, eds. *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*. Berlin, Heidelberg: Springer Berlin
561 Heidelberg, 307-335.
- 562 Lee ZM, Bussema C, 3rd, and Schmidt TM. 2009. rrnDB: documenting the number of rRNA and
563 tRNA genes in bacteria and archaea. *Nucleic Acids Res* 37:D489-493.
564 10.1093/nar/gkn689
- 565 Lehner A, Tasara T, and Stephan R. 2005. Relevant aspects of Arcobacter spp. as potential
566 foodborne pathogen. *Int J Food Microbiol* 102:127-135.
567 10.1016/j.ijfoodmicro.2005.03.003
- 568 Levesque B, and Gauvin D. 2007. Microbiological guideline values for recreational bathing in
569 Canada: Time for change? *Can J Infect Dis Med Microbiol* 18:153-157.
- 570 Lopes FA, Catao EC, Santana RH, Cabral Ade S, Paranhos R, Rangel TP, de Rezende CE,
571 Edwards RA, Thompson CC, Thompson FL, and Kruger RH. 2016. Microbial
572 Community Profile and Water Quality in a Protected Area of the Caatinga Biome. *PLoS
573 One* 11:e0148296. 10.1371/journal.pone.0148296
- 574

- 575 Mac Kenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, Kazmierczak JJ,
576 Addiss DG, Fox KR, Rose JB, and et al. 1994. A massive outbreak in Milwaukee of
577 cryptosporidium infection transmitted through the public water supply. *N Engl J Med*
578 331:161-167. 10.1056/NEJM199407213310304
- 579 Maheux AF, Bissonnette L, Boissinot M, Bernier JL, Huppe V, Picard FJ, Berube E, and
580 Bergeron MG. 2011. Rapid concentration and molecular enrichment approach for
581 sensitive detection of Escherichia coli and Shigella species in potable water samples.
582 *Appl Environ Microbiol* 77:6199-6207. 10.1128/AEM.02337-10
- 583 Maier A, Krolik J, Fan S, Quintin P, McGolrick D, Joyce A, and Majury A. 2015. Evaluating
584 appropriate maximum holding times for private well water samples. *Environmental*
585 *Health Review* 58:35-40. 10.5864/d2015-011
- 586 Masella AP, Bartram AK, Truszkowski JM, Brown DG, and Neufeld JD. 2012. PANDAseq:
587 paired-end assembler for illumina sequences. *BMC Bioinformatics* 13:31. 10.1186/1471-
588 2105-13-31
- 589 McBride MJ. 2014. The Family Flavobacteriaceae. In: Rosenberg E, DeLong EF, Lory S,
590 Stackebrandt E, and Thompson F, eds. *The Prokaryotes: Other Major Lineages of*
591 *Bacteria and The Archaea*. Berlin, Heidelberg: Springer Berlin Heidelberg, 643-676.
- 592 McLellan SL, Newton RJ, Vandewalle JL, Shanks OC, Huse SM, Eren AM, and Sogin ML.
593 2013. Sewage reflects the distribution of human faecal Lachnospiraceae. *Environ*
594 *Microbiol* 15:2213-2227. 10.1111/1462-2920.12092
- 595 Moriarty EM, Mackenzie ML, Karki N, and Sinton LW. 2011. Survival of Escherichia coli,
596 Enterococci, and Campylobacter spp. in sheep feces on pastures. *Appl Environ Microbiol*
597 77:1797-1803. 10.1128/AEM.01329-10
- 598 Moriarty EM, Weaver L, Sinton LW, and Gilpin B. 2012. Survival of Escherichia coli,
599 enterococci and Campylobacter jejuni in Canada goose faeces on pasture. *Zoonoses*
600 *Public Health* 59:490-497. 10.1111/zph.12014
- 601 Mughini Gras L, Smid JH, Wagenaar JA, de Boer AG, Havelaar AH, Friesema IH, French NP,
602 Busani L, and van Pelt W. 2012. Risk factors for campylobacteriosis of chicken,
603 ruminant, and environmental origin: a combined case-control and source attribution
604 analysis. *PLoS One* 7:e42599. 10.1371/journal.pone.0042599
- 605 Newton RJ, and McLellan SL. 2015. A unique assemblage of cosmopolitan freshwater bacteria
606 and higher community diversity differentiate an urbanized estuary from oligotrophic
607 Lake Michigan. *Front Microbiol* 6:1028. 10.3389/fmicb.2015.01028
- 608 Nikolenko SI, Korobeynikov AI, and Alekseyev MA. 2013. BayesHammer: Bayesian clustering
609 for error correction in single-cell sequencing. *BMC Genomics* 14 Suppl 1:S7.
610 10.1186/1471-2164-14-S1-S7
- 611 Noble RT, Blackwood AD, Griffith JF, McGee CD, and Weisberg SB. 2010. Comparison of
612 rapid quantitative PCR-based and conventional culture-based methods for enumeration of
613 Enterococcus spp. and Escherichia coli in recreational waters. *Appl Environ Microbiol*
614 76:7437-7443. 10.1128/AEM.00651-10
- 615 Oliver DM, Bird C, Burd E, and Wyman M. 2016. Quantitative PCR Profiling of Escherichia
616 coli in Livestock Feces Reveals Increased Population Resilience Relative to Culturable
617 Counts under Temperature Extremes. *Environ Sci Technol* 50:9497-9505.
618 10.1021/acs.est.6b02657

- 619 Pandey P, Chiu C, Miao M, Wang Y, Settles M, Del Rio NS, Castillo A, Souza A, Pereira R, and
620 Jeannotte R. 2018. 16S rRNA analysis of diversity of manure microbial community in
621 dairy farm environment. *PLoS One* 13:e0190126. 10.1371/journal.pone.0190126
- 622 Peabody MA, Van Rossum T, Lo R, and Brinkman FS. 2015. Evaluation of shotgun
623 metagenomics sequence classification methods using in silico and in vitro simulated
624 communities. *BMC Bioinformatics* 16:363. 10.1186/s12859-015-0788-5
- 625 Pope ML, Bussen M, Feige MA, Shadix L, Gonder S, Rodgers C, Chambers Y, Pulz J, Miller K,
626 Connell K, and Standridge J. 2003. Assessment of the effects of holding time and
627 temperature on Escherichia coli densities in surface water samples. *Appl Environ
628 Microbiol* 69:6201-6207.
- 629 Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees
630 for large alignments. *PLoS One* 5:e9490. 10.1371/journal.pone.0009490
- 631 Rapp D, Ross CM, Cave V, and Muirhead RW. 2014. Prevalence, concentration and genotypes
632 of *Campylobacter jejuni* in faeces from dairy herds managed in farm systems with or
633 without housing. *J Appl Microbiol* 116:1035-1043. 10.1111/jam.12425
- 634 Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, Chase J, McDonald
635 D, Gonzalez A, Robbins-Pianka A, Clemente JC, Gilbert JA, Huse SM, Zhou HW,
636 Knight R, and Caporaso JG. 2014. Subsampled open-reference clustering creates
637 consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ*
638 2:e545. 10.7717/peerj.545
- 639 Ritalahti KM, Amos BK, Sung Y, Wu Q, Koenigsberg SS, and Loffler FE. 2006. Quantitative
640 PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors
641 multiple *Dehalococcoides* strains. *Appl Environ Microbiol* 72:2765-2774.
642 10.1128/AEM.72.4.2765-2774.2006
- 643 Rochelle-Newall E, Nguyen TM, Le TP, Sengtaeuanghoun O, and Ribolzi O. 2015. A short
644 review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and
645 future directions. *Front Microbiol* 6:308. 10.3389/fmicb.2015.00308
- 646 Shawkey MD, Firestone MK, Brodie EL, and Beissinger SR. 2009. Avian incubation inhibits
647 growth and diversification of bacterial assemblages on eggs. *PLoS One* 4:e4522.
648 10.1371/journal.pone.0004522
- 649 Siljanen HMP, Saari A, Krause S, Lensu A, Abell GCJ, Bodrossy L, Bodelier PLE, and
650 Martikainen PJ. 2011. Hydrology is reflected in the functioning and community
651 composition of methanotrophs in the littoral wetland of a boreal lake. *FEMS
652 Microbiology Ecology* 75:430-445. 10.1111/j.1574-6941.2010.01015.x
- 653 Simon C, and Daniel R. 2011. Metagenomic analyses: past and future trends. *Appl Environ
654 Microbiol* 77:1153-1161. 10.1128/AEM.02345-10
- 655 Smith S, Meade J, Gibbons J, McGill K, Bolton D, and Whyte P. 2016. The impact of
656 environmental conditions on *Campylobacter jejuni* survival in broiler faeces and litter.
657 *Infect Ecol Epidemiol* 6:31685. 10.3402/iee.v6.31685
- 658 Urios L, Agogue H, Lesongeur F, Stackebrandt E, and Lebaron P. 2006. *Balneola vulgaris* gen.
659 nov., sp. nov., a member of the phylum Bacteroidetes from the north-western
660 Mediterranean Sea. *Int J Syst Evol Microbiol* 56:1883-1887. 10.1099/ijsm.0.64285-0
- 661 van Elsas JD, Semenov AV, Costa R, and Trevors JT. 2011. Survival of *Escherichia coli* in the
662 environment: fundamental and public health aspects. *ISME J* 5:173-183.
663 10.1038/ismej.2010.80

- 664 Vermonden K. 2010. Key factors for biodiversity of urban water systems PhD. Radboud
665 University Nijmegen.
666 Vermonden K, Hermus MAA, van Weperen M, Leuven RSEW, van der Velde G, Smolders AJP,
667 Roelofs JGM, and Hendriks AJ. 2009. Does upward seepage of river water and storm
668 water runoff determine water quality of urban drainage systems in lowland areas? A case
669 study for the Rhine–Meuse delta. *Hydrological Processes* 23:3110-3120.
670 10.1002/hyp.7422
671 Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, and Whittam TS. 2009.
672 Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* 75:6534-6544.
673 10.1128/AEM.01262-09
674 Walker DI, McQuillan J, Taiwo M, Parks R, Stenton CA, Morgan H, Mowlem MC, and Lees
675 DN. 2017. A highly specific *Escherichia coli* qPCR and its comparison with existing
676 methods for environmental waters. *Water Res* 126:101-110.
677 10.1016/j.watres.2017.08.032
678 Wickham H. 2009. *ggplot2 : elegant graphics for data analysis*. New York ; London: Springer.
679 Willems A. 2014. The Family Comamonadaceae. In: Rosenberg E, DeLong EF, Lory S,
680 Stackebrandt E, and Thompson F, eds. *The Prokaryotes: Alphaproteobacteria and*
681 *Betaproteobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg, 777-851.
682 Wohlsen T, Bates J, Vesey G, Robinson WA, and Katouli M. 2006. Evaluation of the methods
683 for enumerating coliform bacteria from water samples using precise reference standards.
684 *Lett Appl Microbiol* 42:350-356. 10.1111/j.1472-765X.2006.01854.x
685
686

687 **Figure Legends**

688 **Fig 1. Sampling locations with land cover.** A) Agricultural watershed. B) Urban watershed

689 **Fig 2. Scatterplot of *E. coli* colony counts using Colilert.** The x-axis represents sample
690 collection time in each quarter (Q1-Q4) and the y-axis depicts the *E. coli* most probable number
691 (MPN) per 100 ml of water sample. Full red and empty blue circles depict true sample replicates.
692 Red dotted line set at 400 MPN/100 ml represents Health Canada's limit for recreational water
693 quality.

694 **Fig 3. Principal coordinate analysis based on Bray-Curtis dissimilarity (β -diversity) for the
695 16Sr RNA gene amplicons (n=24 time points) across seasons (Q1-Q4).** A) Agricultural
696 watershed. B) Urban watershed.

697 **Fig 4. Heat map depicting the relative abundance of the top 50 most abundant OTUs in
698 agricultural (A) and urban (B) watersheds.** Hourly data sets were merged using QIIME, based
699 on categorical data for visualization purposes. The color key at the top represents percentages of
700 relative abundance, while that branches depict hierarchical clustering based on similarity.

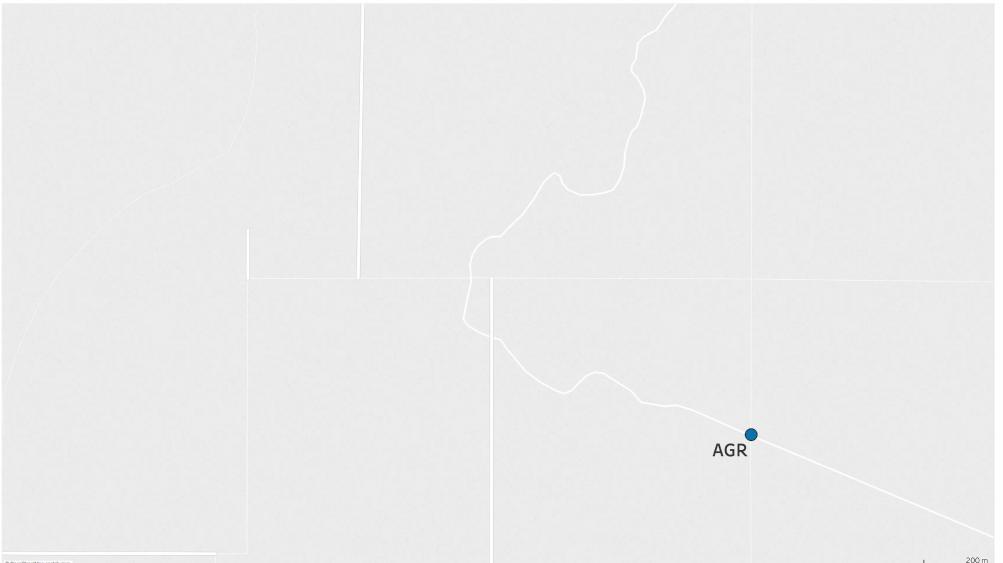
701 **Fig 5. Line plots of fecally-associated bacteria observed in the urban watershed.** The x-axis
702 represents sample collection time in each quarter (Q1-Q4) and the y-axis shows percentages of
703 relative abundance.

704 **Figure 6. Heat map showing the Spearman's rank correlation analysis between the top 10
705 most abundant families and water quality parameters.** A) Agricultural watershed. B) Urban
706 watershed.

707

708

A) Agricultural watershed



B) Urban watershed

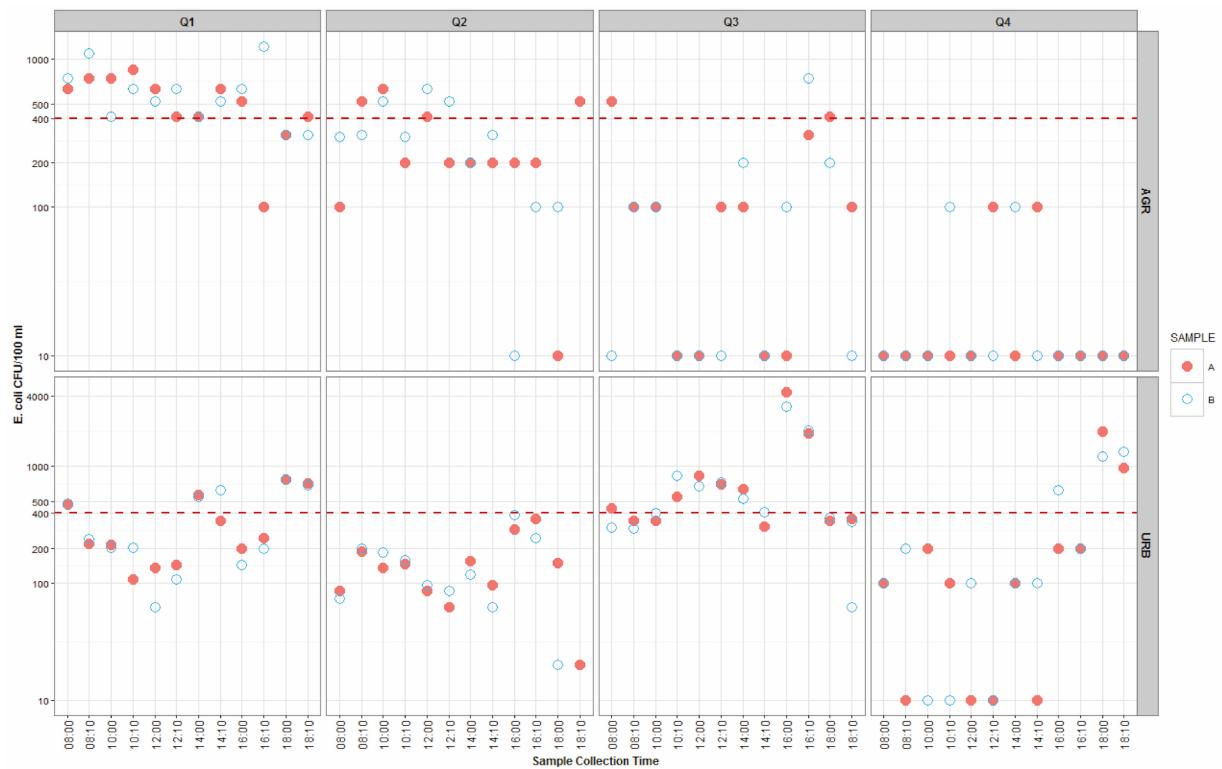


Locations

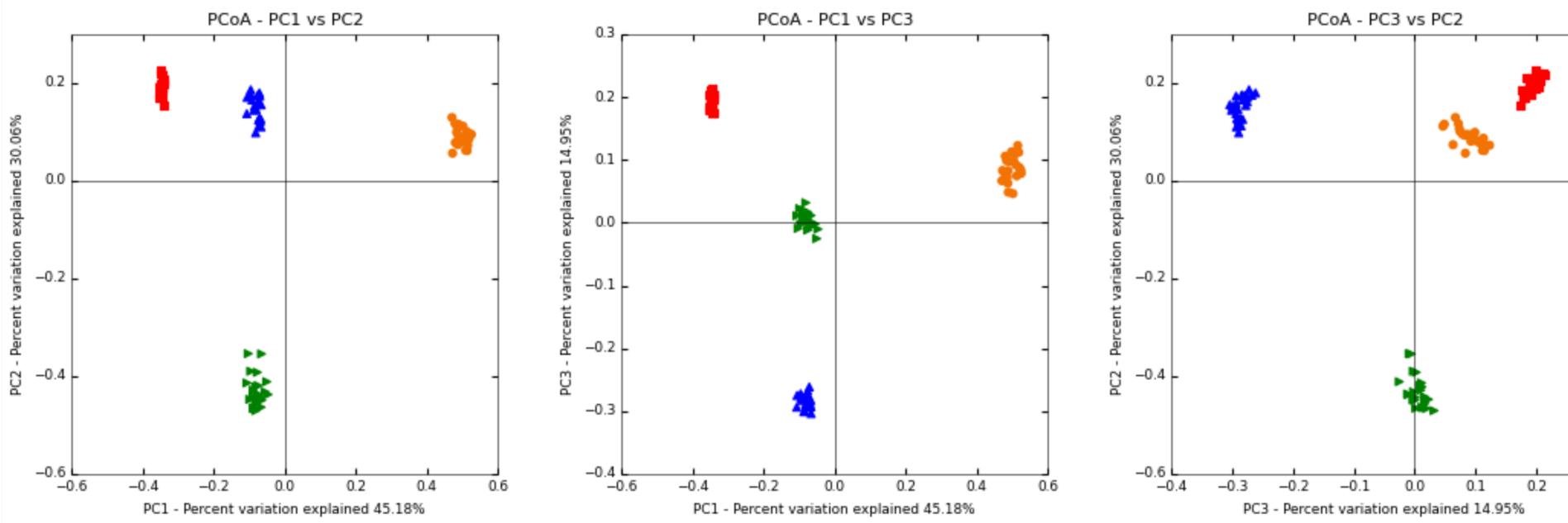
AGR

URB

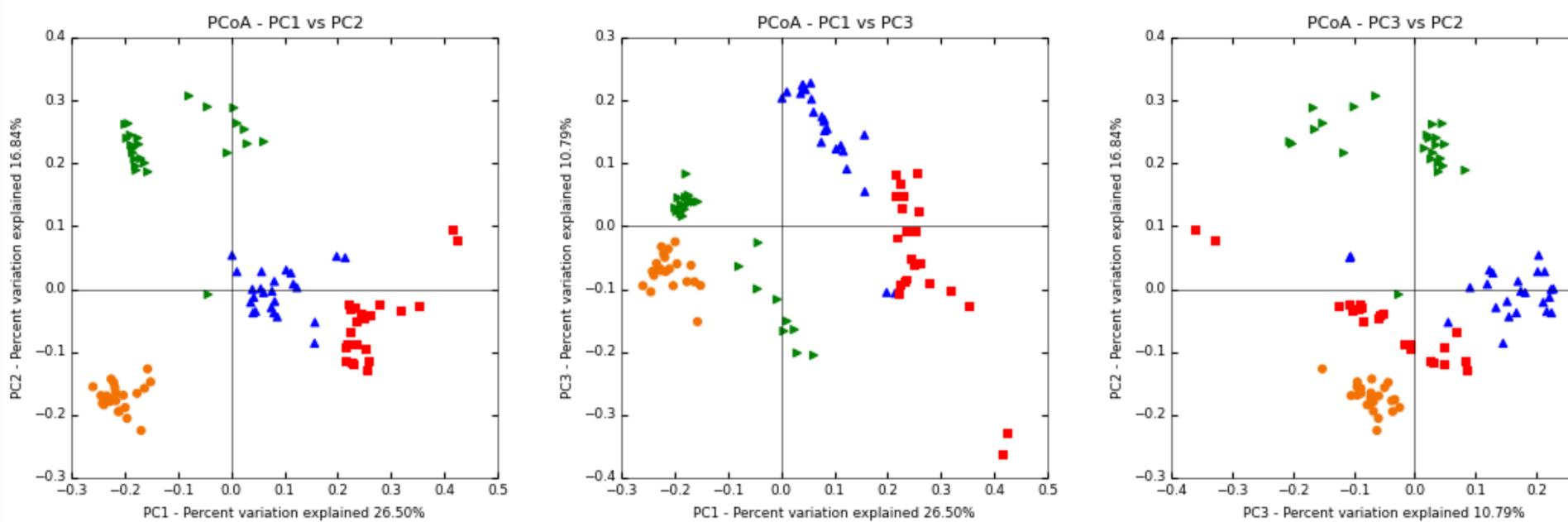
aCC-BY-NC-ND 4.0 International license.



A.



B.



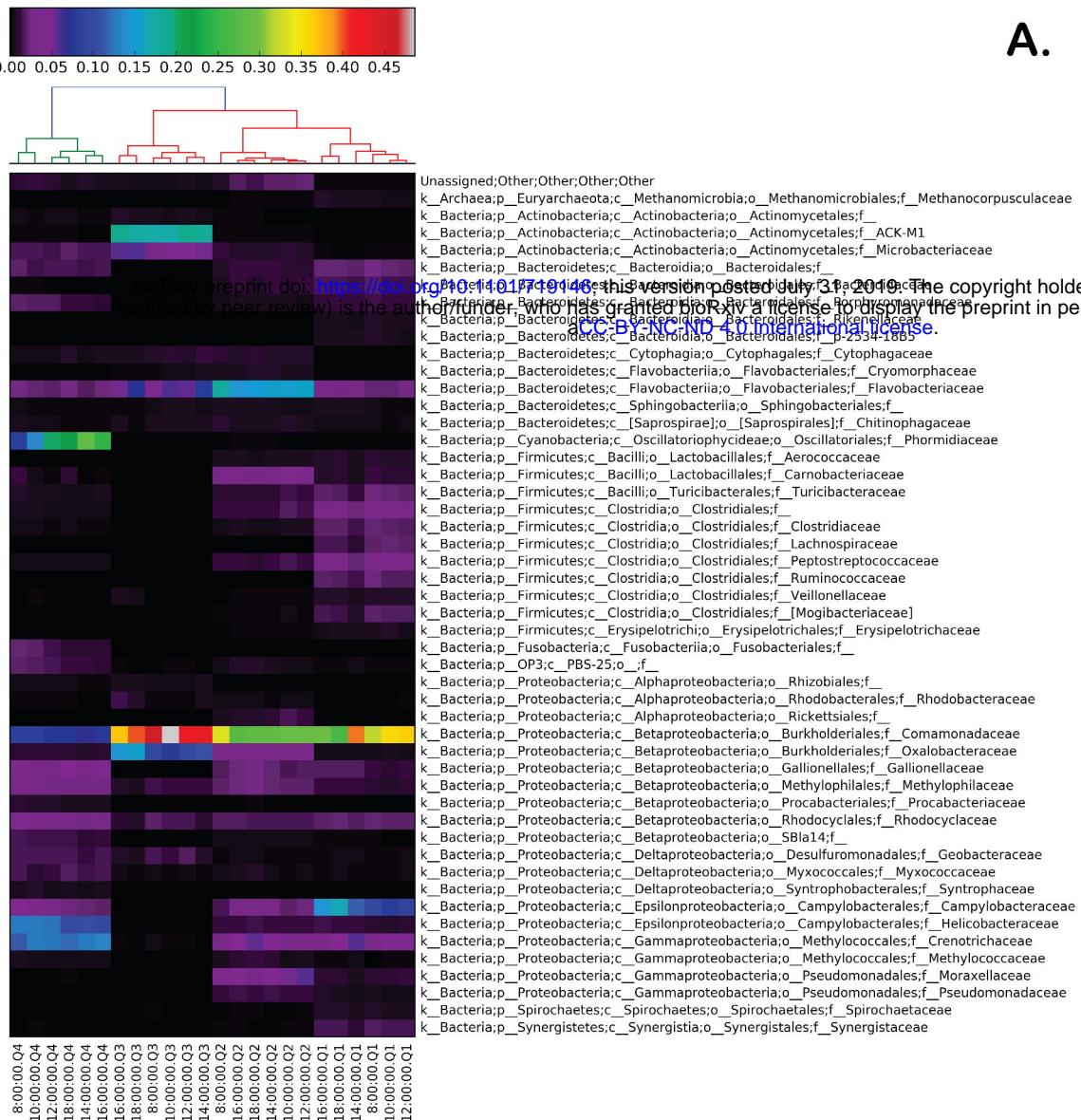
■ Quarter 1

▲ Quarter 2

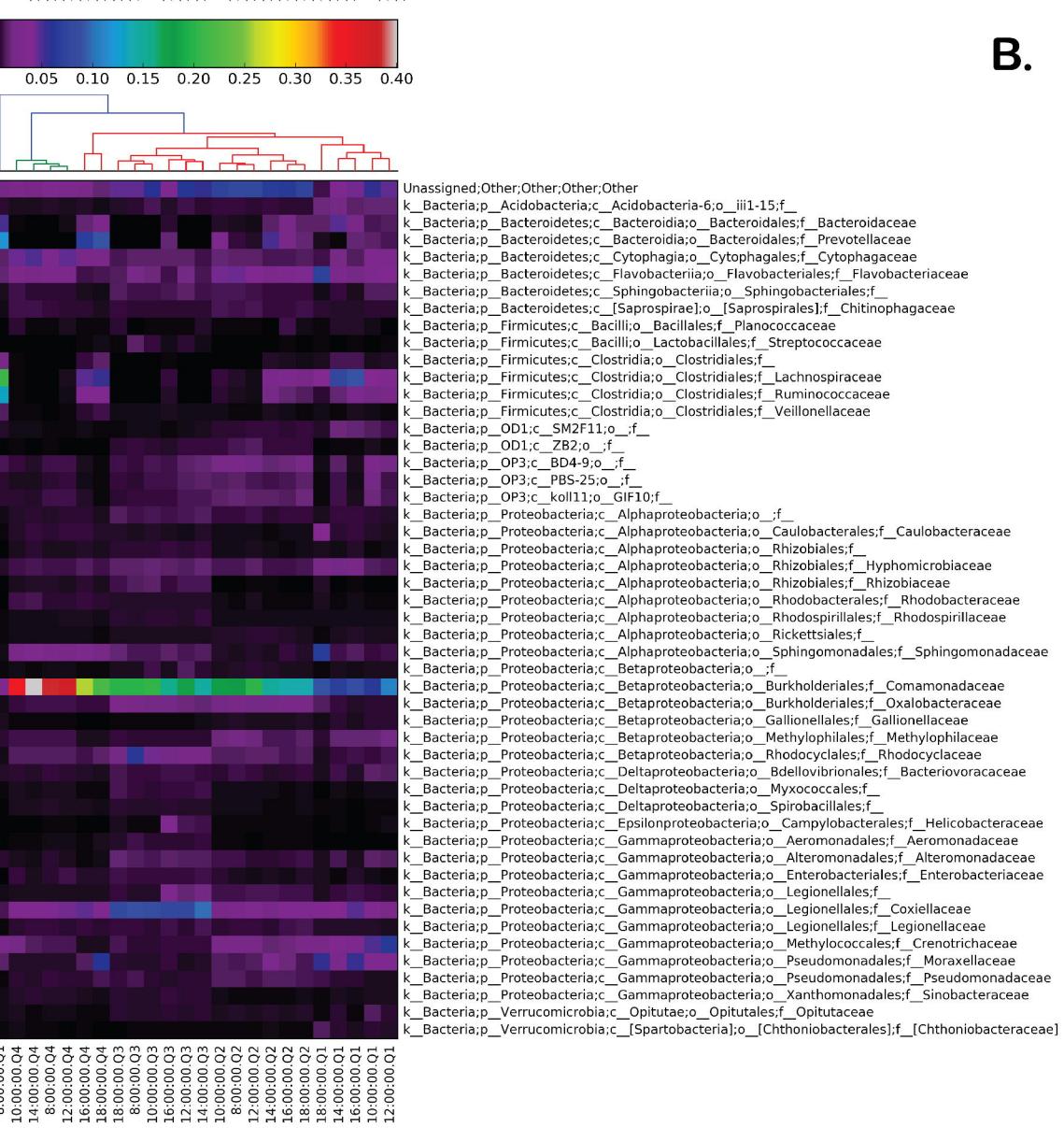
● Quarter 3

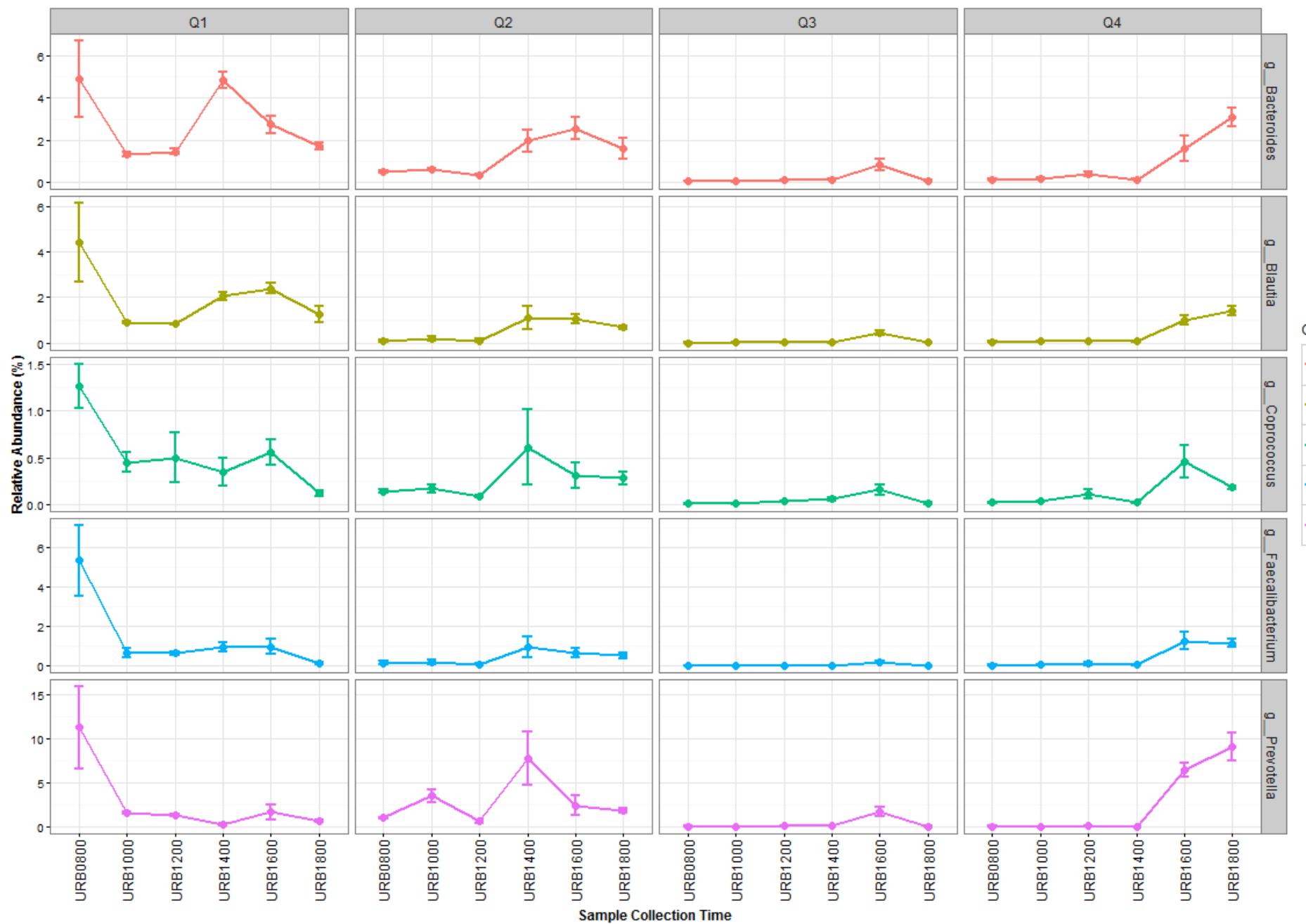
▶ Quarter 4

A.



B.





A) Agricultural watershed



B) Urban watershed

