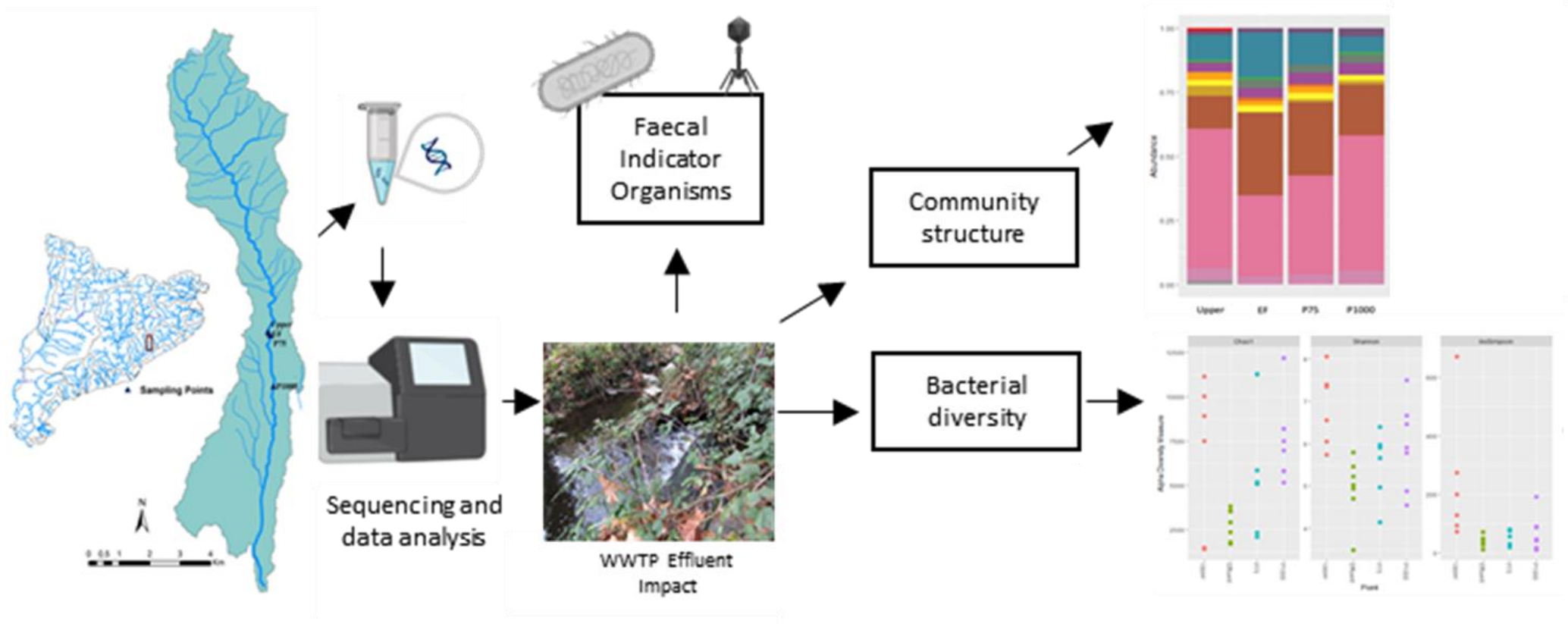

Sampling Points
Faecal Indicator Organisms
Community structure
Bacterial diversity
Sequencing and data analysis
WWTP Effluent Impact
Upper
EF
P75
P1000

Impact of treated sewage effluent on the bacterial community composition in an intermittent Mediterranean stream

Miriam Pascual-Benito[1,2], Elisenda Ballesté[1,2], Toni Monleón-Getino[1,3], Jordi Urmeneta[1,4], Anicet R. Blanch[1,2], Cristina García-Aljaro[1,2*], Francisco Lucena[1,2]

[1]Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Diagonal 643, 08028, Barcelona, Spain.

[2]The Water Research Institute, University of Barcelona, Montalegre 6, 08001 Barcelona, Spain.

[3]BIOST (Research group in biostatistics, bioinformatics and Data Science), GRBIO (Research group in biostatistics and bioinformatics).

[4]Biodiversity Research Institute, University of Barcelona, Barcelona, Spain.

[*]To whom correspondence should be addressed:

Email: crgarcia@ub.edu

Department of Genetics, Microbiology and Statistics

Faculty of Biology. University of Barcelona

Diagonal 643. Prevosti building. Floor 0

08028 Barcelona (Spain)

Tel: +34934021487

ABSTRACT

Water quality monitoring is essential to safeguard human and environmental health. The advent of next-generation sequencing techniques in recent years, which allow a more in-depth study of environmental microbial communities in the environment, could broaden the perspective of water quality monitoring to include impact of faecal pollution bacteria on ecosystem. In this study, 16S rRNA amplicon sequencing was used to evaluate the impact of wastewater treatment plant (WWTP) effluent on autochthonous microbial communities of a temporary Mediterranean stream characterized by high flow seasonality (from 0.02 m^3/s in winter to 0.006 m^3/s in summer).

Seven sampling campaigns were performed under different temperatures and streamflow conditions (winter and summer). Water samples were collected upstream (Upper) of the WWTP, the secondary effluent (EF) discharge and 75 m (P75) and 1000 m (P1000) downstream of the WWTP.

A total of 5,593,724 sequences were obtained, giving rise to 20,650 amplicon sequence variants (ASV), which were further analysed and classified into phylum, class, family and genus. Each sample presented different distribution and abundance of taxa. Although taxon distribution and abundance differed in each sample, the microbial community structure of P75 resembled that of EF samples, and Upper and P1000 samples mostly clustered together. Alpha diversity showed the highest values for Upper and P1000 samples and presented seasonal differences, being higher in winter conditions of high streamflow and low temperature.

Our results suggest the microbial ecology re-establishment, since autochthonous bacterial communities were able to recover from the impact of the WWTP effluent in 1 km. Alpha diversity results indicates a possible influence of environmental factors on the bacterial community structure. This study shows the potential of next-generation sequencing techniques as useful tools in water quality monitoring and management within the climate change scenario.

Keywords: 16S rRNA sequencing, river, faecal pollution, biodiversity, Illumina

INTRODUCTION

Water quality monitoring is essential to ensure public health and protect the environment. Increasing anthropogenic pollution, which is related to the population growth or urban concentration in certain areas, implies a high pressure on water bodies. The development and construction of wastewater treatment plants (WWTPs) has greatly contributed to the improvement of water ecological status in Europe by reducing the concentration of contaminants reaching water bodies (Brion et al., 2015). However, WWTPs effluents still discharge organic matter, nutrients, pollutants and pathogens, which leads to oxygen deficiencies, eutrophication and disruption of biotic communities. This is of particular concern in areas with high population density and water stress, such as the Mediterranean region, which frequently suffers from water shortages. The Mediterranean climate is marked by irregular precipitation, concentrated in spring and autumn, and recurrent episodes of drought and extreme rainfall events (Bonada and Resh, 2013). As a result, strategies are being developed to improve water management, which will be crucial in the coming years, considering that the Mediterranean is one of the areas most vulnerable to climate change (IPCC, 2013).

From a public health point of view, water management has mainly focused on monitoring water quality through the analysis of faecal indicator organisms (FIO) and reference pathogens (WHO, 2001). However, as this approach can only provide a snapshot of water quality at a given moment, modelling faecal pollution dynamics has attracted increasing interest, as it integrates information about the parameters and processes affecting faecal microorganism persistence in the environment.

Water management has also gained importance from an ecological perspective, with the implementation of the Water Framework Directive (EC, 2000) and the adoption of the One Health strategy, which advocates a holistic approach to tackling global health challenges. In this context, a "healthy" river ecosystem is able to restore water status after different impacts through its riparian zone, fauna, and microbiota (Grizzetti et al., 2017; Merlo et al., 2014). Prokaryotes in particular are key players in biogeochemical cycles and ecosystem processes crucial in river self-depuration. However, the resilience of the ecosystem depends on the type and pressure of a given

impact, which may bring about significant changes in biodiversity and weaken the self-depuration capacity. Consequently, studying changes in biodiversity can provide valuable insights into water quality deterioration and help develop suitable water management policies.

Next-generation sequencing (NGS) techniques are being increasingly used in water environmental microbiology and have been proposed as water quality monitoring tools (Chen et al., 2018; Savio et al., 2015; Staley et al., 2013). However, to date their application has been mainly to describe the biodiversity of water ecosystems, including seas, lakes (Llirós et al., 2014) and, less frequently, rivers (Read et al., 2015). The presence of microorganisms able to persist in a dormant state until conditions turn more favourable have been detected in these environments (Caporaso et al., 2011; Gibbons et al., 2013). Accordingly, predictable seasonal variation in bacterial biodiversity has been reported in rivers in polar regions (Crump et al., 2009). Additionally, biotic and abiotic factors such as temperature, radiation, stream flow, sedimentation and predation have been observed to alter the bacterial community structure (Zeglin, 2015). However, to the best of our knowledge, no studies have applied NGS techniques in Mediterranean rivers, which are characterised by continually changing factors.

In a previous work, the presence, dynamics and inactivation of different FIOs along a temporary Mediterranean stream (*Riera de Cànoves*) were assessed, giving rise to the development of a model to assess the microbial water quality along the stream (Ballesté et al., 2019; Pascual-Benito et al., 2020). The stream was affected by the secondary effluent of a WWTP that could constitute up to 100% of the streamflow during the summer period. Such streamflow variation is typical in the Mediterranean regions and, therefore, this catchment is ideal to study the effect of these disturbances on the microbial populations in relation not just from a health related point of view but also from an ecological perspective.

The aims of this research were to: i) assess the seasonal differences in the bacterial diversity and community structure along the stream; ii) study the spatial impact of the effluent on the bacterial diversity and community structure downstream of the WWTP, and iii) evaluate the resilience of the autochthonous populations to overcome the impact of the WWTP effluent.

MATERIAL AND METHODS

Sampling site and sample collection

The study site was located in a 1 km transect along the *Riera de Cànoves*, a temporary Mediterranean water course in Catalonia (Figure 1). The 16.4 km^2 catchment area is mainly forest (77%) and agricultural land (15%).

The stream is characterized by extreme changes in flow between the seasons, ranging from 0.02 m^3/s in winter to 0.006 m^3/s in summer. The studied transect is affected by the effluent of a WWTP serving 9,200 population equivalents. The WWTP treats urban wastewater from small towns, and consists of a pre-treatment and biological treatment system using activated sludge with nitrogen and phosphorous removal as was described in previous studies (Ballesté et al., 2019; Pascual-Benito et al., 2020). The WWTP reduces the biochemical oxygen demand (BOD) from 300 mgO_2/l to <25 mgO_2/l and the suspended solids (SS) from 450 mg/l to <35 mg/l. The contribution of the effluent to the total discharge ranged from 32% in winter to up to 100% in summer, when the stream was completely dry upstream of the WWTP. The streamflow and temperature at the different sampling campaigns are found in (Supplementary Table 1).

Samples were collected in 2016-2018. Seven sampling campaigns were performed in two different periods of the year: 3 in summer and 4 in winter (mean water temperatures of 19.6 °C and 10.1 °C, respectively) to account for differences in temperature and streamflow, which in a previous work were identified as the main environmental drivers for microbial faecal indicators dynamics in this site (Pascual-Benito et al., 2020).

Samples were collected from four different sites: i) 150 m upstream of the WWTP (Upper) (6 samples); ii) directly from the WWTP effluent (EF) (7 samples); iii) 75 m downstream of the WWTP (P75) (6 samples) and iv) 1 km downstream of the WWTP (P1000) (7 samples). These sampling sites were chosen according to our previous work to detect the highest variations in the microbial communities (Pascual-Benito et al., 2020).

Samples were collected from the surface of flowing water in 2 L sterile flasks and transported refrigerated to the laboratory where they were processed in the following 4 hours.

135 **Enumeration of microbial indicators and *Salmonella* spp.**

136 The enumeration of culturable *E. coli* was performed by the pour plate method in
137 Chromocult® agar, as previously described (Astals et al., 2012). Plates were incubated overnight
138 at 44°C and dark blue/purple colonies were counted.

139 Spores of sulphite reducing clostridia (SSRC) were analysed after submitting the samples to a
140 thermal shock at 80°C for 10 minutes. The samples were then cultured by mass inoculation in
141 *Clostridium perfringens* selective agar and incubated overnight at 44°C (Ruiz-Hernando et al.,
142 2014). Black spheres colonies were counted.

143 Somatic coliphages (SOMCPH) and bacteriophages of *Bacteroides thetaiotaomicron* GA17
144 (GA17PH) were enumerated by the double agar layer method according to ISO 10705-2 and ISO
145 10705-4, respectively (ISO, 2001, 2000).

146 In order to enumerate *Salmonella* spp., we adapted the ISO protocol (ISO, 2017) to a most
147 probable number method. Briefly, 500, 50 and 5 ml of each water sample were filtered through a
148 0.45 µm pore-size nitrocellulose membrane. Filters were placed in 10 ml of Buffered Peptone
149 Water (BPW) pH 7 and incubated at 37°C for 24 h. A 0.1 ml volume of the pre-enriched BPW
150 was inoculated in 10 ml of Rappaport-Vassiliadis *Salmonella* enrichment broth and incubated at
151 42°C for 24 h. Then, 0.01 ml was inoculated in SMS® agar in triplicate and incubated at 42°C
152 for 24 h. The presence of *Salmonella* spp. was confirmed by seeding on Hektoen agar, incubating
153 at 37°C for 24 h, followed by incubation in TSA at 37°C for 24 h. Finally, the presence of
154 *Salmonella* spp. was confirmed using the oxidase and API-20E test kits.

155 In the case of bacterial indicators (*E. coli*, SSRC) and *Salmonella* spp., sterile water was
156 used as negative control. Bacterial and viral indicators were enumerated in duplicate.

157

158 **DNA extraction**

159 One litre of each sample was filtered by vacuum filtration through a 3 µm pore- size mixed
160 ester cellulose membrane. The filtrate, which corresponded to microorganisms with a size less
161 than 3 µm, was collected in a sterile glass bottle and subsequently filtered through a 0.22 µm

pore-size polycarbonate membrane. Filters were then placed in a 2 ml screw vial containing glass beads and stored at -80 °C until DNA extraction, which was performed according to previously described methods (Sala-Comorera et al., 2019; Walters et al., 2014) with some modifications. Briefly, 400 µl of phenol, 400 µl of CTAB buffer and 400 µl of chloroform/isoamyl alcohol (24:1 v/v) were added to the screw vial containing the filter and glass beads, mixed by vortex for 15 min and chilled on ice for 1 min. Samples were centrifuged at 13000 x g for 5 min at room temperature (RT) and the supernatant was transferred to a vial containing 500 µl of chloroform/isoamyl alcohol (24:1 v/v), mixed by vortex followed by a centrifugation (13000 x g, 5 min, RT). The supernatant was transferred to a vial containing 270 µl of isopropanol and stored overnight at RT. Tubes were mixed by inversion and centrifuged (13000 x g, 15 min, RT). The supernatant was discarded by decanting and 1 ml of ethanol (70%, ice-cold) was added, followed by centrifugation (13800 x g, 5 min, 4°C). The supernatant was removed and the pellet was dried at 60°C for 30 min. Finally, the pellet was recovered in 100 µl of Elution buffer (Invitrogen, USA). A negative control, which included the reagents used in the DNA extraction, was performed in parallel. Samples were quantified by Qubit (Invitrogen) in order to determine the concentration of DNA obtained from the extraction and stored at -80°C until analysis.

Illumina 16S rRNA amplicon sequencing

Before sequencing, the presence of bacterial DNA in the samples was confirmed by conventional PCR. The V4 hypervariable region from the 16S rRNA gene of the samples was amplified using the degenerated primers 515f (5'- GTGCCAGCMGCCGCGGTAA-3') and 806r (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al., 2011).

A negative control (blank reagents) as well as a positive control (commercial mock microbial community Zymmobiomics, Zymmo Research) was included for the 16S rRNA amplicon sequencing. Illumina sequencing of samples was performed in a single run using the Illumina MiSeq platform at the Research Technology Support Facility of Michigan State University (Michigan, USA). Amplicon libraries of the V4 hypervariable region of the 16S rRNA gene were prepared using the previously described primers 515f and 806r with corresponding adaptors,

following a reported protocol (Kozich et al., 2013). Sequencing was performed in 2 x 250 bp paired end format using a MiSeq v2 reagent cartridge following the manufacturer's instructions (Illumina MiSeq, USA). Base calling was done by Illumina Real Time Analysis (RTA) v1.18.54 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

16S rRNA gene amplicon data analysis

Sequences were processed to amplicon sequence variants (ASV) using the default parameters of the Dada2 workflow (Callahan et al., 2016). Reverse reads were trimmed to 160 bp as recommended to improve downstream processing of the reads (Callahan et al., 2016) and a maximum of 2 errors per read were allowed (maxEE=2). This parameter has been shown to be a better filter than simply averaging quality scores (Edgar and Flyvbjerg, 2015). Taxonomic classifications were assigned to the ASV using the reference SILVA database v132.

Biodiversity analysis

Alpha and beta diversity were analysed using the Phyloseq R package (McMurdie and Holmes, 2013). For alpha diversity analysis, the Chao, Shannon and Inverse Simpson indices were calculated after rarefying the ASV table. For beta diversity analysis, samples were previously transformed to relative proportions, the Bray Curtis index was calculated, and samples were clustered accordingly.

ASV proportions and the origin of the bacterial communities

A binomial test of proportions with adjustment of the p-value (*P*) ("FDR" method) for multiple hypothesis testing to avoid Type I error problems was used to assess differential ASV proportions among the different groups. This test was performed using the function dif.propOTU.between.groups() of the library BDSbiost3 which is based on the Fisher's exact statistical test (Monleón-Getino, 2020).

SourceTracker2 pipeline was used to determine the contribution of the secondary effluent and the Upper to the communities downstream of the WWTP impact, as previously described (Knights et al., 2011).

FISH analysis

FISH analysis was performed to validate the Illumina results and to estimate the number of total bacteria in the stream samples. For this, formaldehyde was added to 100 ml of sample to a final concentration of 2-4% and it was fixed for 1 h at RT. Samples were then filtered through 0.22 µm pore-size polycarbonate membrane filters, which were washed with 20 ml of sterile water. Filters were stored at -20°C in a Petri dish until their analysis according to a described protocol (Pernthaler et al., 2001). Briefly, the filter was cut in sections, each section was placed on a 10 µl drop of probe solution on a microscope slide and covered with another 10 µl of probe solution. The slide was stored in a dark humid chamber, coated with the hybridization buffer and incubated at 46°C for 3 hours. The filter section was then washed with buffer for 10 min at 48°C and with distilled water for 2 min in the dark. The filter section was dried and mounted in Vectashield (Vector Laboratories). Cells were stained with 1.5 $\mu g \cdot ml^{-1}$ DAPI (Sigma, USA) for counting with a Leica TCS SP2 confocal microscope.

Different probes were used to quantify the different phyla and classes (Alm et al., 1996): i) Alphaproteobacteria (5'-GGTAAGGTTCTGCGCGTT-3'), ii) Betaproteobacteria (5'-GCCTCCCCACTTCGTTT-3'), iii) Actinobacteria (5'-TATAGTTACCACCGCCGT-3'), iv) Firmicutes (5'-CCGAAGATTCCCTACTGC-3'), v) Bacteroidetes (5'-GGACCCTTTAAACCCAAT-3'), and vi) Gammaproteobacteria (5'-GCCTCCCCACATCGTTT-3'). All probes were labelled with Cy5 fluorochrome.

Statistical analyses

Statistical analyses were performed using different R functions and libraries (R Core Team, 2016). The BDbiost3 library for R (Monleón-Getino et al., 2017), was used to assess the coverage of the sequenced reads, and for discriminant and exploratory data analysis.

The coverage of the sequenced reads was analysed to assess the representativeness of the obtained ASV, as previously described (Monleón-Getino et al., 2017). For this purpose, the PILI3() function of the BDbiost3 library function was used. PILI3() allowed the computation of the rarefaction curve between the number of reads and the amount of ASV obtained. This function was projected to an infinite rarefaction curve in order to verify its saturation or if it still had a margin to saturate. For exploratory analysis, contingency tables (ASV abundance tables) were obtained separately for the different studied sample groups. These data followed a multinomial distribution (Monleón-Getino and Frías-López, 2020) and allowed us to apply an exploratory dimension reduction technique using the non-metric multidimensional scaling (nMDS). Discriminant analysis was computed using the 20 most abundant ASV, and made possible by the function MDSdbhatta.PAM.Metagen1() of the BDbiost3 library, which allowed the evaluation of 5 different discriminant methods: linear discriminant analysis (LDA), support vector machine (SVM), xboosting (Xboost), kernel discrimination (kernel) and artificial neural nets (ANN). The results obtained also offered final classification accuracy and a confusion matrix as a result of the different discrimination methods performed.

Spearman's correlation coefficients were calculated to assess the significance of the biodiversity changes in relation to the different environmental variables.

RESULTS

Faecal indicator organisms and *Salmonella* spp.

The faecal pollution in the *Riera de Cànoves* was characterized through the analysis of faecal indicator organisms (FIO) and *Salmonella* spp. (Table 1). Upper samples revealed low concentrations of faecal indicators associated with diffuse human faecal pollution and no *Salmonella* spp. were detected. EF samples contained high FIO concentrations, constituting a source of human pollution downstream of the WTTP, and *Salmonella* spp. were detected in some samples but at low concentrations (0.3 $\log_{10}$ (MPN/100ml). The WWTP discharge significantly increased FIO and *Salmonella* spp. concentrations in P75 (P<0.05), where FIO concentrations were about 1-2 $\log_{10}$ higher than in Upper samples and not significantly different compared to the

EF (P>0.05); *Salmonella* spp. were detected and quantified in low concentrations, as in EF samples. Compared to Upper samples, the concentrations of all faecal indicator bacteria in P1000 were not significantly different (P>0.05), whereas those of SOMCPH and GA17PH were significantly higher (P<0.05). *Salmonella* spp. were only detected in one P1000 sample.

General description of the sequencing results

A total of 5,593,724 reads were generated for a total of 26 samples, ranging between 140,897 and 311,146 reads per sample. A total of 4,639,854 reads were selected after quality processing and chimera removal. These reads yielded 23,372 ASV after DADA2 algorithm processing, 20,650 of which were affiliated to Bacteria and kept for further analysis. It should be noted the 2,460 ASV affiliated to Archaea were discarded, because the used primers were targeting the bacterial 16S rRNA gene, in accordance with the study aim. A total of 16,854 ASV were represented by more than 10 reads (>80% of the ASV). The representativeness of the reads with respect to ASV biodiversity was very high, ranging from 92% to 97.5% (Supplementary Fig. 1). The ASV coverage obtained for Upper and P1000 samples was close to 97.5%, indicating that the obtained reads reflected the expected diversity in these samples. In the case of EF and P75, a slightly lower value was obtained (around 92%), showing a lower ASV coverage in these samples compared to others, although the majority of ASV were still represented in the study.

Bacterial communities in the sampling sites

Bacterial communities in Upper

The bacterial communities of the Upper site were represented by 15,791 ASV. A minor number of ASV (144 ASV) was shared by all the samples, whereas 760 ASV were detected in 5 out of 6 samples. At phylum level (Figure 2), 67% of the reads affiliated to Proteobacteria, 11% Bacteroidetes, 8% Epsilonbacteraeota, 7 % Patescibacteria, 1% Actinobacteria, 1% Firmicutes followed by phyla each one with abundance lower than 1% (Dependentiae, Chlamydiae, Fibrobacteres, Cyanobacteria, Verrucomicrobia, Fusobacteria, Synergistetes, Omnitrophicaeota, Elusimicrobia, Nitrospirae, Planctomycetes, Acidobacteria, Spirochaetes, Chloroflexi,

300 Rokubacteria). One percent of the reads were unclassified phyla. Class, order and family
301 affiliation is shown in Supplementary material file (Supplementary Fig. 2 to Fig. 4)
302 *Bacteroides* and *Bifidobacterium*, two of the most abundant genera in human faeces,
303 constituted less than 1% of the Upper genera, whereas pathogenic bacteria with faecal-oral
304 transmission, such as *Campylobacter* or *Salmonella*, were not detected. *Rhodoferax* (16%) was
305 the most abundant genus in both seasons (Figure 3). These results should be interpreted with
306 caution, as each genome is known to carry different copies of the 16S rRNA gene, even in species
307 of the same genus (Větrovský and Baldrian, 2013). The quantification of the Upper samples by
308 FISH analysis resulted in 7.8 $\log_{10}$ (cells/100 ml) and indicated that 89.3% of the bacteria
309 belonged to 4 phyla: Proteobacteria (50.5%), Firmicutes (27.6%), Actinobacteria (8.3%) and
310 Bacteroidetes (2.9%).
311
312 Bacterial communities in EF
313 The bacterial communities of the 7 EF samples were represented by 8,062 ASV, only 227 of
314 which were present in all the samples, whereas 500 ASV were present in 6 out of 7 samples. At
315 the phylum level (Figure 2), reads corresponded to Proteobacteria (35%), Patescibacteria (31%),
316 Bacteroidetes (17%), Epsilonbacteraeota (6%), Actinobacteria (2%), Firmicutes and
317 Dependentiae (1%), and other phyla with an abundance lower than 1% (Chlamydiae,
318 Fibrobacteres, Cyanobacteria, Verrucomicrobia, Fusobacteria, Synergistetes, Tenericutes,
319 Omnitrophicaeota, Elusimicrobia, Nitrospirae, Lentisphaerae , Planctomycetes). Two percent of
320 the reads were unclassified.
321 *Flavobacterium* (11%) was the most abundant genus in EF samples in both seasons (Figure
322 3). Despite being the most abundant genera in faeces, *Bacteroides* and *Bifidobacterium*
323 constituted only 2% and 0.01%, respectively, of the total ASV.
324
325 Bacterial communities in P75
326 The bacterial communities of P75 were represented by 13,252 ASV, only 351 of which were
327 present in all the samples, whereas 918 ASV were present in 5 out of 6 samples. In P75 (Figure

328 2), the reads indicated a predominant affiliation to the phylum Proteobacteria (42%), followed by
329 Patescibacteria (26%), Bacteroidetes (16%), Epsilobacteraeota (6%), Firmicutes (2%),
330 Actinobacteria (1%) and Fibrobacteres (1%), and other phyla with abundances lower than 1%
331 (Chlamydiales, Cyanobacteria, Verrucomicrobiae, Fusobacteriales, Synergistaceae,
332 Omnitrophicaeota, Elusimicrobia, Nitrospiraceae, Lentisphaerae, Planctomycetes, Acidobacteria,
333 Spirochaetes and Chloroflexi). Two percent of the reads were unclassified.
334 *Polynucleobacter* (6%) and *Arcobacter* (6%) were the most abundant genera in both seasons
335 (Figure 3), while *Bacteroides* and *Bifidobacterium* represented 2% and 0.03% of the total ASV,
336 respectively. The quantification of the total bacteria by FISH showed a total of 8.4 $\log_{10}$ (cells/100
337 ml). The analysis performed in p75 samples showed that 4 phyla accounted for 86.7% of the
338 bacteria: Proteobacteria (71.2%), Bacteroidetes (10.7%), Firmicutes (2.7%) and Actinobacteria
339 (2.1%).
340
341 Bacterial communities in P1000
342 The bacterial communities of P1000 were represented by 16,645 ASVs, only 567 of which
343 were present in all the samples, whereas 1624 ASV were present in 6 out of 7 samples. In P1000,
344 the distribution of the reads at phylum level (Figure 2) was mainly in Proteobacteria (62%),
345 followed by Patescibacteria (13%), Bacteroidetes (6%), Epsilobacteraeota (7%), Actinobacteria
346 (3%) and other phyla with abundances lower than 1% (Firmicutes, Chlamydiales, Fibrobacteres,
347 Cyanobacteria, Verrucomicrobiae, Fusobacteriales, Synergistaceae, Omnitrophicaeota,
348 Elusimicrobia, Nitrospiraceae, Lentisphaerae, Planctomycetes, Acidobacteria, Spirochaetes,
349 Gemmatimonadetes, Chloroflexi). Two percent of the reads were unclassified.
350 *Rhodoferax* (9%) was the most abundant genus followed by *Arcobacter* (7%) and
351 *Polynucleobacter* (7%) (Figure 3), while *Bacteroides* and *Bifidobacterium* represented 0.4% and
352 0.0%, respectively. The FISH analyses of P1000 samples showed that 95.7% of the bacteria
353 belonged to the 4 analysed phyla: Proteobacteria (42.7%), Firmicutes (20.4%), Actinobacteria
354 (18.3%) and Bacteroidetes (14.3%). The total amount of bacteria quantified in P1000 was 7.9
355 $\log_{10}$ (cells/100 ml).

356

357 **Tracking the origin of the ASV after the anthropogenic impact**

358 Using Kernel discriminant analysis (Supplementary Table 2 and Supplementary Fig. 5), it
359 was possible to separate samples belonging to the different sampling points with an accuracy of
360 0.692 (CI95%: 0.4821, 0.8567) taking into consideration the 20 most abundant ASV. This result
361 reflected that the discriminant analysis had a limited ability to separate P1000 from Upper samples
362 and P75 from EF samples. However, the accuracy improved if the samples were separated by
363 season (accuracy 0.818 [CI95%:0.4822, 0.9772] and 0.8667 [CI95%:0.5954, 0.9834] in summer
364 and winter, respectively), supporting a seasonal difference of the community structure.

365 Further analysis of the effect of the sewage effluent on the communities downstream of the
366 WWTP (Supplementary Table 3) showed that the EF contribution to the communities in P75 was
367 higher in samples taken in summer (mean contribution of 74.7%) than in winter (mean
368 contribution of 56.8%). The EF had a lower contribution to the P1000 than the P75 communities,
369 but it was also higher in summer (47.6%) than in winter (38.1%). In contrast, the Upper
370 contribution to P75 communities was higher in winter (8.1%) than in summer (3.2%). Moreover,
371 Upper had a greater impact on P1000 than P75, being higher in summer (20.5%) than in winter
372 (16.5%). Certain percentages, which could not be assigned to Upper or EF, were considered as
373 unknown origin (22.1-45.4%).

374

375 **Impact of the sewage effluent on community structure**

376 We analysed the differences in ASV abundances between the samples upstream and
377 downstream of the WWTP according to the season (Supplementary Table 4). Comparing
378 abundance in Upper with downstream (P75 and P1000) sampling points, significant differences
379 were found in 25 ASV in winter and 32 in summer. Among these, 15 ASV (60%) of the winter
380 samples and 11 ASV (34%) of the summer samples showed significant differences between
381 Upper and P75, but not between Upper and P1000, suggesting a recovery of bacterioplankton
382 from the WWTP impact and the microbial ecology re-establishment.

383 The ASV with significant differences between Upper and P75 showed two types of behaviour,
384 increasing or decreasing their proportion. Compared to the Upper site, the proportion of some
385 ASV was significantly higher in P75: 22 of the 25 ASV (88%) in winter and 20 of the 32 (62.5%)
386 ASV in summer. Significant differences in proportion were also observed between Upper and
387 P1000 sites for 5 ASV in winter and 4 ASV in summer. Only 3 ASV (ASV1 ASV6 and ASV22)
388 were significantly lower in P75 compared to Upper samples and showed a recovery in P1000 in
389 both seasons; they corresponded to *Flavobacterium*, an unclassified genus of the
390 order_Absconditabacteriales, and *Rhodoferax*. In contrast, the differences between Upper and
391 P75 samples for ASV10 and ASV14 were maintained in P1000 in both seasons; they belonged to
392 C39 (*Rhodocyclaceae*) and *Flavobacterium*, respectively.
393
394 **Impact of the sewage effluent on the bacterioplankton diversity**
395 In order to assess the impact of the sewage effluent on the bacterioplankton diversity, alpha
396 and beta diversity indices, which are indicative of species richness and overall bacterial
397 community structure, were calculated. Three different indices of alpha diversity were analysed:
398 Chao 1, Shannon and Inverse Simpson (Figure 4, Supplementary Table 5). Bacterioplankton
399 alpha diversity was highest in Upper and P1000 samples and lowest in the sewage effluent,
400 indicating that the effluent discharge reduced the alpha diversity in P75. Similar trends were
401 observed for the three analysed indices and the alpha diversity reduction from Upper to P75
402 samples was statistically significant in values of Shannon and Inverse Simpson indices.
403 Additionally, the alpha diversity values were higher in winter than in summer, suggesting an
404 association between diversity and abiotic factors. To shed further light on this relationship, the
405 possible correlation between the alpha diversity measured by the Shannon index and water
406 temperature, one of the main environmental factors (Supplementary Fig. 6), was studied. Results
407 showed a negative statistically significant correlation ($r=-0.68$, $P<0.01$) between water
408 temperature and alpha diversity, i.e., the lowest alpha diversity values corresponded to the highest
409 temperatures.

Beta diversity, which quantifies the bacterial communities considering the river space, was also analysed through Bray-Curtis dissimilarity (Figure 5). This measure divided the samples in two main clusters: one was constituted mainly by Upper and P1000 samples and the other mainly by EF and P75 samples, which were subclustered according to the season. Clustering of the samples was also observed in the multidimensional scaling of the dissimilarity (Supplementary Fig. 7).

DISCUSSION

Access to high quality water, already a major global problem, is expected to worsen due to urban demographic growth and a concomitant increase in water demand. These pressures can also lead to the functional deterioration of water ecosystems, especially in areas highly vulnerable to the impact of climate change, such as the Mediterranean (IPCC, 2013). Focusing on pollution with a point source rather than diffuse origin, the aim of this study was to analyse the impact of the secondary effluent discharged from a WWTP in a Mediterranean stream with a low and intermittent flow regime.

Although faecal pollution was observed in the *Riera de Cànoves* upstream of the WWTP, the sewage effluent significantly increased the downstream concentration of faecal indicators and pathogens such as *Salmonella* spp. Faecal pollution levels subsequently returned to those of Upper samples (after a distance of 3-15 km) due to the *Riera de Cànoves* self-depuration capacity, previously described and modelled (Ballesté et al., 2019; Pascual-Benito et al., 2020). Similar behaviour has been reported in other streams (Price et al., 2018).

In the *Riera de Cànoves*, 84% of the bacterioplankton community structure consisted of four phyla (Proteobacteria, Bacteroidetes, Patescibacteria and Actinobacteria), with a variable distribution according to the sampling point. Proteobacteria was predominant throughout the studied river transect, whereas Bacteroidetes and Patescibacteria increased after the WWTP effluent discharge, subsequently decreasing in the sampling points downstream. Although with different percentages, the trends for Proteobacteria and Bacteroidetes were confirmed by FISH analysis. NGS techniques have been used to describe bacterioplankton community structure in rivers of different characteristics and geographical areas. Studies on the Danube (Europe),

438 Mississippi (USA), Tama (Japan) and Apies (South Africa) rivers also report Proteobacteria,
439 Bacteroidetes and Actinobacteria as the most abundant phyla, although with varying proportions
440 (Abia et al., 2018; Reza et al., 2018; Savio et al., 2015; Staley et al., 2013).

441 Proteobacteria, one of the most abundant phyla in water ecosystems worldwide (Newton et
442 al., 2011), is divided into different classes such as Alphaproteobacteria and
443 Gammaproteobacteria, whose diverse characteristics drive its ubiquity. Bacteroidetes, a phylum
444 widely distributed in marine and freshwater ecosystems, is highly specialized in organic matter
445 degradation (Traving et al., 2017), which explains its high percentages in the sewage effluent and
446 samples downstream of the WWTP. The phylum Actinobacteria has been proposed as a water
447 quality indicator due to its sensitivity to the conditions that cause cyanobacterial blooms (Ghai et
448 al., 2014). Patescibacteria, which was observed in high percentages in the *Riera de Cànoves*,
449 especially after the sewage effluent discharge, has been reported in other rivers but in lower
450 abundances (Zemskaya et al., 2019). Given the anaerobic nature of Patescibacteria, a likely source
451 is sewage water (Castelle et al., 2017), although another source could be the streambed, as the
452 phylum is associated with mobilizable sediments (Herrmann et al., 2019).

453 Species sorting, a concept that refers to community selection by environmental factors (Logue
454 and Lindström, 2010), could explain the decrease of phyla such as Bacteroidetes and
455 Patescibacteria from P75 to P1000 samples. Most of these bacteria are commensal or related with
456 faecal pollution, and in a freshwater ecosystem selective pressure favours the recovery of
457 autochthonous river communities as they are transported downstream. Although *Bacteroides* and
458 *Bifidobacterium* are described as the most abundant genera in human faeces, they were found in
459 low frequencies in EF samples and decreased from P75 to P1000 samples. Another factor in the
460 reduction of these anaerobic genera is the aerobic conditions found in both the WWTP and the
461 stream.

462 The freshwater *Rhodoferax*, reported in rivers worldwide (Cottrell et al., 2005; Galand et al.,
463 2008), was the most abundant genus found in Upper and P1000 samples. In P75 the most abundant
464 genera were *Polynucleobacter* and *Arcobacter*, also freshwater bacteria. However, *Arcobacter*
465 has been associated with human faeces and sewage water (Lerner et al., 1994), with high

466 abundances reported in WWTP influents, due to a capacity to proliferate in infrastructures such
467 as sewage pipes (Assanta et al., 2002). This behaviour suggests the WWTP effluent is the most
468 plausible source of this genus. *Arcobacter* is important in water quality monitoring, as some
469 species are human pathogens and have been reported to cause waterborne outbreaks (Collado et
470 al., 2010; Prouzet-Mauléon et al., 2006). The most predominant genus in EF samples was
471 *Flavobacterium*, in agreement with previous studies that detected high abundances in eutrophic
472 water and faecally polluted urban streams (Eiler and Bertilsson, 2007). Overall, the abundance of
473 genera with pathogenic species was higher in EF, P75 and P1000 than in Upper samples. For
474 instance, higher abundances of *Escherichia*, *Shigella*, *Klebsiella* and *Enterobacter* were found in
475 P75 and P1000 compared to Upper samples. This result confirms the relevance of secondary
476 effluents in the spread not only of faecal pollutants but also of human pathogens.

477 The impact of the WWTP effluent on the community structure along the stream transect was
478 also observed in proportional changes of ASV in the different testing sites. In P75, most of the
479 ASV whose proportion had increased compared to Upper samples belonged to sewage water-
480 related genera, such as *Flavobacterium, Arcobacter* or *Polynucleobacter*, whereas some of the
481 ASV whose proportions had declined belonged to freshwater genera, such as *Rhodoferax*,
482 *Rhizobacter*, *Limnohabitans*, *Novosphingobium* and *Pseudarcicella*. Changes in ASV behaviour
483 and relative abundance along the transect could help to determine the impact of the WWTP
484 effluent and to identify potential water quality indicator genera.

485 ASV from the Upper reaches and sewage effluent were the source of 71.5% of P75 ASV and
486 60.4% of P1000 ASV, indicating that the bacterial community structure and distribution
487 downstream of the WWTP were mainly determined by the bacterial composition of both Upper
488 and EF. This prediction is lower than the 95% of Mansfeldt and co-workers (Mansfeldt et al.,
489 2019), who also described that the contributions of the WWTP effluents were higher than 50%
490 downstream of the discharge point, similar to the EF contribution obtained here. The EF
491 contributed more to P75 than to P1000, suggesting the sewage effluent lost influence on the
492 community composition with distance. A percentage of sequences of unknown origin (ranging
493 from 16% to 39% in P75 and 19% to 62% in P1000) was determined, possibly related to

494 communities in the streambed sediments, which could constitute a reservoir of mobilizable
495 bacteria. Although the study of the sediments was not addressed in this research, it could provide
496 us more information to understand the bacterial dynamics in the stream. This phenomenon has
497 been widely explored in studies of faecal indicators and pathogens (García-Aljaro et al., 2017;
498 Jamieson et al., 2005) and could play an important role in the diversity of the water column.
499 The high contribution of ASV from EF to P75 reflects the degree of impact of the sewage
500 effluent on P75. The impact was also reflected by the decrease in alpha diversity. The sewage
501 effluent had the lowest alpha diversity values of the study and decreased the biodiversity of the
502 stream immediately after the effluent impact. These results are in agreement with previous studies
503 (Drury et al., 2013; Mansfeldt et al., 2019), where wastewater reduced the alpha diversity of the
504 receiving water body, although an increase in diversity downstream of the WWTP has also been
505 reported by other authors (Marti and Balcázar, 2014; Wakelin et al., 2008). Such contradictory
506 results could be due to differences in the environmental context and the techniques used, including
507 improved next-generation sequencing. The increase in alpha diversity in P1000 suggests a partial
508 recovery from the effluent impact.
509 In addition, a significant negative correlation was found between temperature and alpha
510 diversity, the lowest temperature being related with the highest alpha diversity and vice versa.
511 This trend has been reported previously (Kent et al., 2004; Rubin and Leff, 2007), indicating that
512 environmental factors play an important role in the bacterial community structure and diversity
513 in water ecosystems. In a previous study on faecal pollution in the *Riera de Cànoves*, temperature
514 and streamflow were found to be crucial in the self-depuration capacity downstream of the
515 WWTP, which increased at high temperatures and low streamflow (Pascual-Benito et al., 2020).
516 The results obtained here support the important role of the streamflow in the community structure
517 downstream of the WWTP. During summer, when the streamflow is low and the sewage effluent
518 barely diluted, the contribution of EF bacterial sequences to P75 and P1000 was highest. The high
519 contribution of Upper during the winter caused a highest dilution of the effluent, giving rise to
520 highest alpha diversity downstream of the WWTP. Although the temperature reduced the self-
521 depuration distances, it also reduced the alpha diversity. Biotic factors such as the natural

522 inactivation of microorganisms and predation could be emphasised with the temperature (Ballesté
523 and Blanch, 2010). Therefore, from an ecological point of view, the increase in temperature in a
524 climate change scenario would be detrimental for the *Riera de Cànoves* biodiversity.

525 García-Armisen and collaborators have described the seasonal resilience of bacterial
526 communities, reporting their recovery from the WWTP impact in the driest season (García-
527 Armisen et al., 2014), although the river they studied is not comparable with the *Riera de*
528 *Cànoves*, in terms of dimensions, WWTPs and climate. Overall, the bacterial communities from
529 the *Riera de Cànoves* showed a high resilience to the impact of the sewage effluent, which
530 consisted mainly of organic matter. This was demonstrated by analysing the FIO and pathogen
531 concentrations and the bacterial communities and diversity, all of which showed recovery in only
532 1 km.

533

534 **CONCLUSIONS**

535 The anthropogenic impact of the WWTP secondary effluent in the *Riera de Cànoves* caused
536 an alteration of the community structure and reduced the alpha diversity at the P75 sampling
537 point. However, the results suggest that the autochthonous communities of the *Riera de Cànoves*
538 had partially recovered 1 km downstream of the WWTP effluent, showing a high resilience that
539 was dependant on environmental factors such as temperature. Within the climate change scenario,
540 which predicts an increase in temperature and decrease in streamflow, the resilience of the
541 bacterial communities and diversity may be negatively affected. Modelling the stream resilience
542 under different environmental conditions may therefore be crucial for water management and
543 quality monitoring in the future. This study shows that next-generation sequencing techniques
544 can be useful to monitor bacterial community dynamics and identify water samples with different
545 levels of anthropogenic impact. They therefore have potential application as practical tools in
546 water management to assess the ecological status of rivers.

547

ACKNOWLEDGEMENTS

This work was supported by Spanish Ministerio de Economía y Competitividad MEDSOUL project (CGL2014-59977-C3-1-R), the Catalan government (2017 SGR 170) and the Water Research Institute. M Pascual-Benito was supported by a FPI grant of the Spanish Ministerio de Economía y Competitividad (BES-2015-072112). We thank Manel Bosch (Advanced Microscopy Unit of Scientific and Technological Center of the University of Barcelona) and Robert Benaiges-Fernández for their help in FISH analyses.

DATA AVAILABILITY

The research data are available at: https://data.mendeley.com/datasets/gvs398nfzb/draft?a=29e1bcab-8ae2-4824-8f88-5560e2a6b19a

REFERENCES

Abia, A.L.K., Alisoltani, A., Keshri, J., Ubomba-Jaswa, E., 2018. Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. Sci. Total Environ. 616–617, 326–334. https://doi.org/10.1016/j.scitotenv.2017.10.322

Alm, E.W., Oerther, D.B., Larsen, N., Stahl, D.A., Raskin, L., 1996. The oligonucleotide probe database. Appl. Environ. Microbiol. 62, 3557–3559.

Assanta, M.A., Roy, D., Lemay, M.J., Montpetit, D., 2002. Attachment of Arcobacter butzleri, a new waterborne pathogen, to water distribution pipe surfaces. J. Food Prot. 65, 1240–1247. https://doi.org/10.4315/0362-028X-65.8.1240

Astals, S., Venegas, C., Peces, M., Jofre, J., Lucena, F., Mata-Alvarez, J., 2012. Balancing hygienization and anaerobic digestion of raw sewage sludge. Water Res. 46, 6218–6227. https://doi.org/10.1016/j.watres.2012.07.035

Ballesté, E., Blanch, A.R., 2010. Persistence of Bacteroides species populations in a river as measured by molecular and culture techniques. Appl. Environ. Microbiol. 76, 7608–7616. https://doi.org/10.1128/AEM.00883-10

576 Ballesté, E., Pascual-Benito, M., Martín-Díaz, J., Blanch, A.R., Lucena, F., Muniesa, M., Jofre,
577 J., García-Aljaro, C., 2019. Dynamics of crAssphage as a human source tracking marker in
578 potentially faecally polluted environments. Water Res. 155, 233–244.
579 https://doi.org/10.1016/j.watres.2019.02.042
580 Bonada, N., Resh, V.H., 2013. Mediterranean-climate streams and rivers: Geographically
581 separated but ecologically comparable freshwater systems. Hydrobiologia 719, 1–29.
582 https://doi.org/10.1007/s10750-013-1634-2
583 Brion, N., Verbanck, M.A., Bauwens, W., Elskens, M., Chen, M., Servais, P., 2015. Assessing
584 the impacts of wastewater treatment implementation on the water quality of a small urban
585 river over the past 40 years. Environ. Sci. Pollut. Res. 22, 12720–12736.
586 https://doi.org/10.1007/s11356-015-4493-8
587 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016.
588 DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13,
589 581–583. https://doi.org/10.1038/nmeth.3869
590 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J.,
591 Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions
592 of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108, 4516–4522.
593 https://doi.org/10.1073/pnas.1000080107
594 Castelle, C.J., Brown, C.T., Thomas, B.C., Williams, K.H., Banfield, J.F., 2017. Unusual
595 respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate
596 Phyla Radiation. Sci. Rep. 7, 1–12. https://doi.org/10.1038/srep40101
597 Chen, W., Wilkes, G., Khan, I.U.H., Pintar, K.D.M., Thomas, J.L., Lévesque, C.A., Chapados,
598 J.T., Topp, E., Lapen, D.R., 2018. Aquatic Bacterial Communities Associated With Land
599 Use and Environmental Factors in Agricultural Landscapes Using a Metabarcoding
600 Approach. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02301
601 Collado, L., Kasimir, G., Perez, U., Bosch, A., Pinto, R., Saucedo, G., Huguet, J.M., Figueras,
602 M.J., 2010. Occurrence and diversity of Arcobacter spp. along the Llobregat River
603 catchment, at sewage effluents and in a drinking water treatment plant. Water Res. 44, 3696–

604 3702. https://doi.org/10.1016/j.watres.2010.04.002
605 Cottrell, M.T., Waidner, L.A., Yu, L., Kirchman, D.L., 2005. Bacterial diversity of metagenomic
606 and PCR libraries from the Delaware River. Environ. Microbiol. 7, 1883–1895.
607 https://doi.org/10.1111/j.1462-2920.2005.00762.x
608 Crump, B.C., Peterson, B.J., Raymond, P.A., Amon, R.M.W., Rinehart, A., McClelland, J.W.,
609 Holmes, R.M., 2009. Circumpolar synchrony in big river bacterioplankton. Proc. Natl.
610 Acad. Sci. U. S. A. 106, 21208–21212. https://doi.org/10.1073/pnas.0906149106
611 Drury, B., Rosi-Marshall, E., Kelly, J.J., 2013. Wastewater Treatment Effluent Reduces the
612 Abundance and Diversity of Benthic Bacterial Communities in Urban and Suburban Rivers.
613 Appl. Environ. Microbiol. 79, 1897–1905. https://doi.org/10.1128/aem.03527-12
614 EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council eatblishing a
615 framework for Community action in the field of water policy.
616 https://doi.org/10.1039/AP9842100196
617 Edgar, R.C., Flyvbjerg, H., 2015. Error filtering, pair assembly and error correction for next-
618 generation sequencing reads. Bioinformatics 31, 3476–3482.
619 https://doi.org/10.1093/bioinformatics/btv401
620 Eiler, A., Bertilsson, S., 2007. Flavobacteria blooms in four eutrophic lakes: Linking population
621 dynamics of freshwater bacterioplankton to resource availability. Appl. Environ. Microbiol.
622 73, 3511–3518. https://doi.org/10.1128/AEM.02534-06
623 Galand, P.E., Lovejoy, C., Pouliot, J., Garneau, M.È., Vincent, W.F., 2008. Microbial community
624 diversity and heterotrophic production in a coastal Arctic ecosystem: A stamukhi lake and
625 its source waters. Limnol. Oceanogr. 53, 813–823.
626 https://doi.org/10.4319/lo.2008.53.2.0813
627 García-Aljaro, C., Martín-Díaz, J., Viñas-Balada, E., Calero-Cáeres, W., Lucena, F., Blanch,
628 A.R., 2017. Mobilisation of microbial indicators, microbial source tracking markers and
629 pathogens after rainfall events. Water Res. 112, 248–253.
630 https://doi.org/10.1016/j.watres.2017.02.003
631 García-Armisen, T., Inceo□lu, Ö., Ouattara, N.K., Anzil, A., Verbanck, M.A., Brion, N., Servais,

632 P., 2014. Seasonal variations and resilience of bacterial communities in a sewage polluted
633 urban river. PLoS One 9. https://doi.org/10.1371/journal.pone.0092579
634 Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., Rodriguez-Valera, F., 2014. Key roles for
635 freshwater Actinobacteria revealed by deep metagenomic sequencing. Mol. Ecol. 23, 6073–
636 6090. https://doi.org/10.1111/mec.12985
637 Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., Gilbert, J.A., 2013. Evidence
638 for a persistent microbial seed bank throughout the global ocean. Proc. Natl. Acad. Sci. U.
639 S. A. 110, 4651–4655. https://doi.org/10.1073/pnas.1217767110
640 Grizzetti, B., Pistocchi, A., Liquete, C., Udias, A., Bouraoui, F., van de Bund, W., 2017. Erratum:
641 Human pressures and ecological status of European rivers. Sci. Rep. 7, 6941.
642 https://doi.org/10.1038/s41598-017-04857-5
643 Herrmann, M., Wegner, C.E., Taubert, M., Geesink, P., Lehmann, K., Yan, L., Lehmann, R.,
644 Totsche, K.U., Küsel, K., 2019. Predominance of Cand. Patescibacteria in groundwater is
645 caused by their preferential mobilization from soils and flourishing under oligotrophic
646 conditions. Front. Microbiol. 10, 1–15. https://doi.org/10.3389/fmicb.2019.01407
647 IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group
648 I to the Fifth Assessment Report of the Intergovern- mental Panel on Climate Change
649 [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y.
650 Xi, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to
651 the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
652 ISO, 2017. International Standard ISO 6579: Microbiology of the food chain -Horizontal method
653 for the detection, enumeration and serotyping of Salmonella -Part 1: Detection of
654 Salmonella spp.
655 ISO, 2001. International Standard ISO 10705-4: Water Quality - Detection and Enumeration of
656 Bacteriophages. Part 4: Enumeration of bacteriophages infecting Bacteroides fragilis.
657 ISO, 2000. International Standard ISO 10705-2: Water Quality - Detection and Enumeration of
658 Bacteriophages. Part 2: Enumeration of somatic coliphages.
659 Jamieson, R.C., Joy, D.M., Lee, H., Kostaschuk, R., Gordon, R.J., 2005. Resuspension of

660 sediment-associated Escherichia coli in a natural stream. J. Environ. Qual. 34, 581–589.
661 https://doi.org/10.2134/jeq2005.0581
662 Kent, A.D., Jones, S.E., Yannarell, A.C., Graham, J.M., Lauster, G.H., Kratz, T.K., Triplett, E.W.,
663 2004. Annual patterns in bacterioplankton community variability in a Humic Lake. Microb.
664 Ecol. 48, 550–560. https://doi.org/10.1007/s00248-004-0244-y
665 Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman,
666 F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent
667 microbial source tracking. Nat. Methods 8, 761–765. https://doi.org/10.1038/nmeth.1650
668 Lerner, J., Brumberger, V., Preac-Mursic, V., 1994. Severe diarrhea associated with Arcobacter
669 butzleri. Eur. J. Clin. Microbiol. Infect. Dis. 13, 660–662.
670 https://doi.org/10.1007/BF01973994
671 Llirós, M., Inceoglu, Ö., García-Armisen, T., Anzil, A., Leporcq, B., Pigneur, L.M., Viroux, L.,
672 Darchambeau, F., Descy, J.P., Servais, P., 2014. Bacterial community composition in three
673 freshwater reservoirs of different alkalinity and trophic status. PLoS One 9, 1–27.
674 https://doi.org/10.1371/journal.pone.0116145
675 Logue, J.B., Lindström, E.S., 2010. Species sorting affects bacterioplankton community
676 composition as determined by 16S rDNA and 16S rRNA fingerprints. ISME J. 4, 729–738.
677 https://doi.org/10.1038/ismej.2009.156
678 Mansfeldt, C., Deiner, K., Mächler, E., Fenner, K., Eggen, R.I.L., Stamm, C., Schönenberger, U.,
679 Walser, J.-C., Altermatt, F., 2019. Microbial community shifts in streams receiving treated
680 wastewater effluent. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.135727
681 Marti, E., Balcázar, J.L., 2014. Use of pyrosequencing to explore the benthic bacterial community
682 structure in a river impacted by wastewater treatment plant discharges. Res. Microbiol. 165,
683 468–471. https://doi.org/10.1016/j.resmic.2014.04.002
684 Merlo, C., Reyna, L., Abril, A., Amé, M.V., Genti-Raimondi, S., 2014. Environmental factors
685 associated with heterotrophic nitrogen-fixing bacteria in water, sediment, and riparian soil
686 of Suquía River. Limnologica 48, 71–79. https://doi.org/10.1016/j.limno.2014.06.004
687 Monleón-Getino, A., Rodríguez-Casado, C., Méndez-Viera, J., 2017. How to calculate number

688 of samples in the design of pre / pro-biotics studies (metagenomic studies), in: How to
689 Calculate Number of Samples in the Design of Pre/pro-Biotics Studies. Seville.
690 https://doi.org/10.13140/RG.2.2.27611.67367
691 Monleón-Getino, T., 2020. Library for R BDSBIOST3: Machine learning and advanced statistical
692 methods for omic, categorical analysis and others [WWW Document]. URL
693 https://github.com/amonleong/BDSbiost3
694 Monleón-Getino, T., Frías-López, J., 2020. A priori estimation of sequencing effort in complex
695 microbial metatranscriptomes. Pending Publ.
696 Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A Guide to the Natural
697 History of Freshwater Lake Bacteria, Microbiology and Molecular Biology Reviews.
698 https://doi.org/10.1128/mmbr.00028-10
699 Pascual-Benito, M., Nadal-Sala, D., Tobella, M., Ballesté, E., García-Aljaro, C., Sabaté, S.,
700 Sabater, F., Martí, E., Gracia, C.A., Blanch, A.R., Lucena, F., 2020. Modelling the seasonal
701 impacts of a wastewater treatment plant on water quality in a Mediterranean stream using
702 microbial indicators. J. Environ. Manage. 261.
703 https://doi.org/10.1016/j.jenvman.2020.110220
704 Pernthaler, J., Glöckner, F.-O., Schönhuber, W., Amann, R., 2001. Fluorescence in situ
705 hybridization (FISH) with rRNA-targeted oligonucleotide probes. Methods Microbiol. 30,
706 207–226. https://doi.org/10.1016/s0580-9517(01)30046-6
707 Price, J.R., Ledford, S.H., Ryan, M.O., Toran, L., Sales, C.M., 2018. Wastewater treatment plant
708 effluent introduces recoverable shifts in microbial community composition in receiving
709 streams. Sci. Total Environ. 613–614, 1104–1116.
710 https://doi.org/10.1016/j.scitotenv.2017.09.162
711 Prouzet-Mauléon, V., Labadi, L., Bouges, N., Ménard, A., Mégraud, F., 2006. Arcobacter
712 butzleri: Underestimated Enteropathogen. Emerg. Infect. Dis. 12, 307–309.
713 R Core Team, 2016. A language and environment for statistical computing. R Found. Stat.
714 Comput.
715 Read, D.S., Gweon, H.S., Bowes, M.J., Newbold, L.K., Field, D., Bailey, M.J., Griffiths, R.I.,

716 2015. Catchment-scale biogeography of riverine bacterioplankton. ISME J. 9, 516–526.
717 https://doi.org/10.1038/ismej.2014.166
718 Reza, M.S., Mizusawa, N., Kumano, A., Oikawa, C., Ouchi, D., Kobiyama, A., Yamada, Y.,
719 Ikeda, Y., Ikeda, D., Ikeo, K., Sato, S., Ogata, T., Kudo, T., Jimbo, M., Yasumoto, K.,
720 Yoshitake, K., Watabe, S., 2018. Metagenomic analysis using 16S ribosomal RNA genes of
721 a bacterial community in an urban stream, the Tama River, Tokyo. Fish. Sci. 84, 563–577.
722 https://doi.org/10.1007/s12562-018-1193-6
723 Rubin, M.A., Leff, L.G., 2007. Nutrients and other abiotic factors affecting bacterial communities
724 in an Ohio River (USA). Microb. Ecol. 54, 374–383. https://doi.org/10.1007/s00248-007-
725 9209-2
726 Ruiz-Hernando, M., Martín-Díaz, J., Labanda, J., Mata-Alvarez, J., Llorens, J., Lucena, F., Astals,
727 S., 2014. Effect of ultrasound, low-temperature thermal and alkali pre-treatments on waste
728 activated sludge rheology, hygienization and methane potential. Water Res. 61, 119–129.
729 https://doi.org/10.1016/j.watres.2014.05.012
730 Sala-Comorera, L., Blanch, A.R., Casanovas-Massana, A., Monleón-Getino, A., García-Aljaro,
731 C., 2019. Traceability of different brands of bottled mineral water during shelf life, using
732 PCR-DGGE and next generation sequencing techniques. Food Microbiol. 82, 1–10.
733 https://doi.org/10.1016/j.fm.2019.01.006
734 Savio, D., Sinclair, L., Ijaz, U.Z., Parajka, J., Reischer, G.H., Stadler, P., Blaschke, A.P., Blöschl,
735 G., Mach, R.L., Kirschner, A.K.T., Farnleitner, A.H., Eiler, A., 2015. Bacterial diversity
736 along a 2600km river continuum. Environ. Microbiol. 17, 4994–5007.
737 https://doi.org/10.1111/1462-2920.12886
738 Staley, C., Unno, T., Gould, T.J., Jarvis, B., Phillips, J., Cotner, J.B., Sadowsky, M.J., 2013.
739 Application of Illumina next-generation sequencing to characterize the bacterial community
740 of the Upper Mississippi River. J. Appl. Microbiol. 115, 1147–1158.
741 https://doi.org/10.1111/jam.12323
742 Traving, S.J., Rowe, O., Jakobsen, N.M., Sørensen, H., Dinasquet, J., Stedmon, C.A., Andersson,
743 A., Riemann, L., 2017. The effect of increased loads of dissolved organic matter on estuarine

744 microbial community composition and function. Front. Microbiol. 8, 1–15.
745 https://doi.org/10.3389/fmicb.2017.00351
746 Větrovský, T., Baldrian, P., 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes
747 and Its Consequences for Bacterial Community Analyses. PLoS One 8, 1–10.
748 https://doi.org/10.1371/journal.pone.0057923
749 Wakelin, S.A., Colloff, M.J., Kookana, R.S., 2008. Effect of wastewater treatment plant effluent
750 on microbial function and community structure in the sediment of a freshwater stream with
751 variable seasonal flow. Appl. Environ. Microbiol. 74, 2659–2668.
752 https://doi.org/10.1128/AEM.02348-07
753 Walters, E., Kätzl, K., Schwarzwälder, K., Rutschmann, P., Müller, E., Horn, H., 2014.
754 Persistence of fecal indicator bacteria in sediment of an oligotrophic river: Comparing large
755 and lab-scale flume systems. Water Res. 61, 276–287.
756 https://doi.org/10.1016/j.watres.2014.05.007
757 WHO, 2001. Guidelines , Standards and Health : Assessment of risk and risk management for
758 water-related infectious disease 1–431.
759 Zeglin, L.H., 2015. Stream microbial diversity in response to environmental changes: Review and
760 synthesis of existing research. Front. Microbiol. 6, 1–15.
761 https://doi.org/10.3389/fmicb.2015.00454
762 Zemskaya, T.I., Bukin, S.V., Zakharenko, A.S., Chernitsyna, S.M., Shubenkova, O.V., 2019.
763 Microbial communities in the estuarine water areas of the rivers in the southeastern part of
764 Lake Baikal. Limnol. Freshw. Biol. 2019, 259–265. https://doi.org/10.31951/2658-3518-
765 2019-a-4-259
766

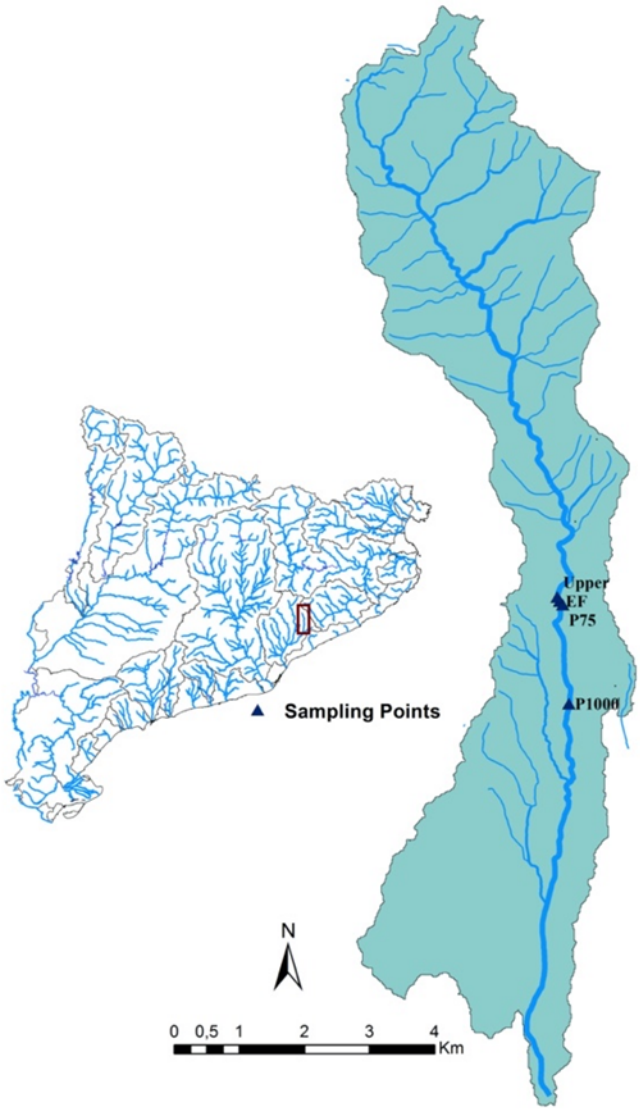
TABLES

Table 1: Mean concentration and standard deviation of *E. coli*, spores of sulphite reducing clostridia (SSRC) (in $\log_{10}$ (CFU/100ml), somatic coliphages (SOMCPH), bacteriophages of *Bacteroides thetaiotaomicron* GA17 (GA17PH) (in $\log_{10}$ (PFU/100ml) and *Salmonella* spp. ($\log_{10}$ (MPN/100ml)).

	E. coli	SSRC	SOMCPH	GA17PH	*Salmonella* spp.
Upper	3.4 ± 0.8	2.3 ± 0.4	2.2 ± 0.4	0.3 ± 0.6	<-0.7
EF	4.2 ± 0.4	3.7 ± 0.2	4.3 ± 0.2	1.7 ± 0.6	0.3 ± 0.6
P75	4.1 ± 0.3	3.4 ± 0.2	4.1 ± 0.2	1.6 ± 0.7	0.2 ± 0.6
P1000	3.1 ± 0.4	2.6 ± 0.6	3.2 ± 0.7	0.9 ± 0.4	<-0.7

FIGURES

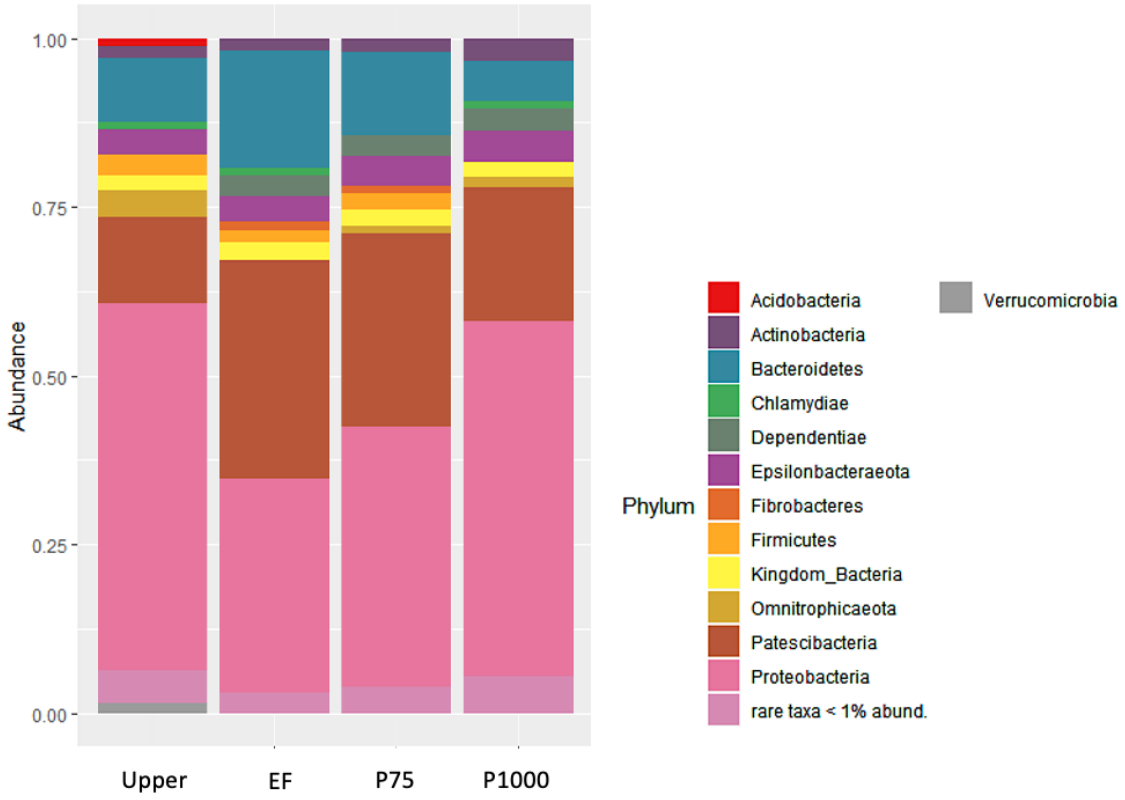
Figure 1: Study site. Upper is the sampling site located upstream of the WWTP, EF is the secondary effluent of the WWTP, P75 is located 75 m downstream of the WWTP and P1000 is 1 km downstream of the WWTP.



Site	Point_X	Point_Y	Latitude	Longitude
UPPER	2.355749	41.682615	N41° 40' 57,414"	E2° 21' 20,696"
EF	2.35532	41.682778	N41° 40' 58,001"	E2° 21' 19,152"
P75	2.355918	41.682294	N41° 40' 56,435"	E2° 21' 21,316"
P1000	2.357199	41.66862	N41° 40' 7,032"	E2° 21' 25,916"

778 **Figure 2:** Average distribution of phyla in each sampling point. Rare taxa grouped the phyla

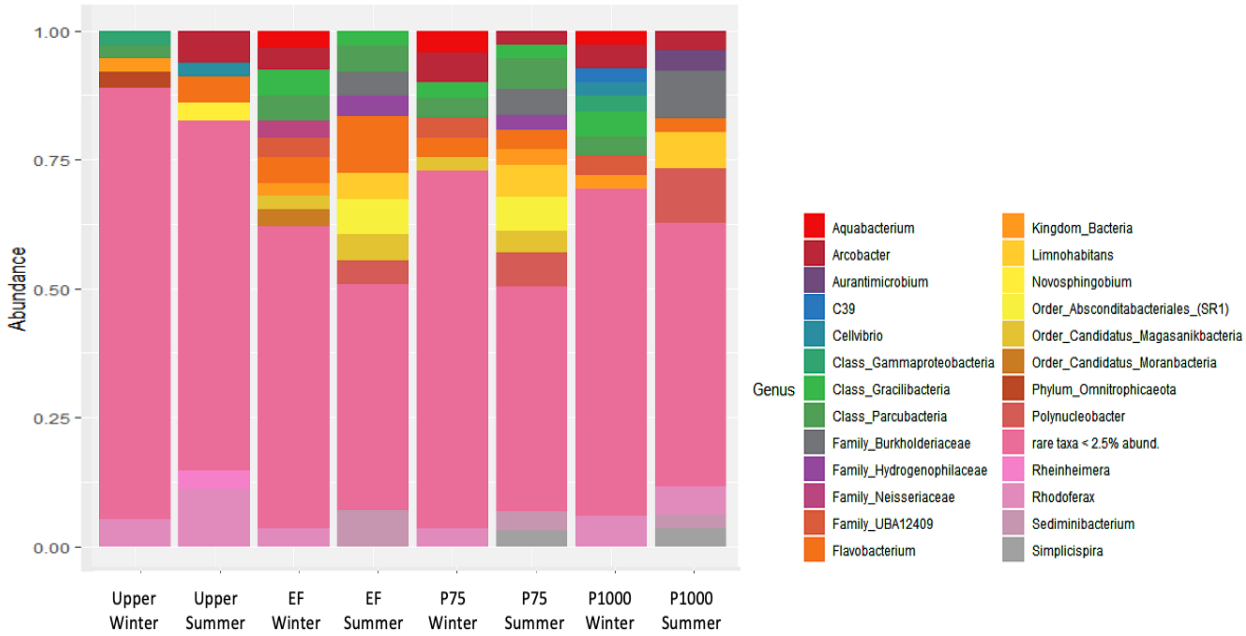
779 with proportions <1%.



780

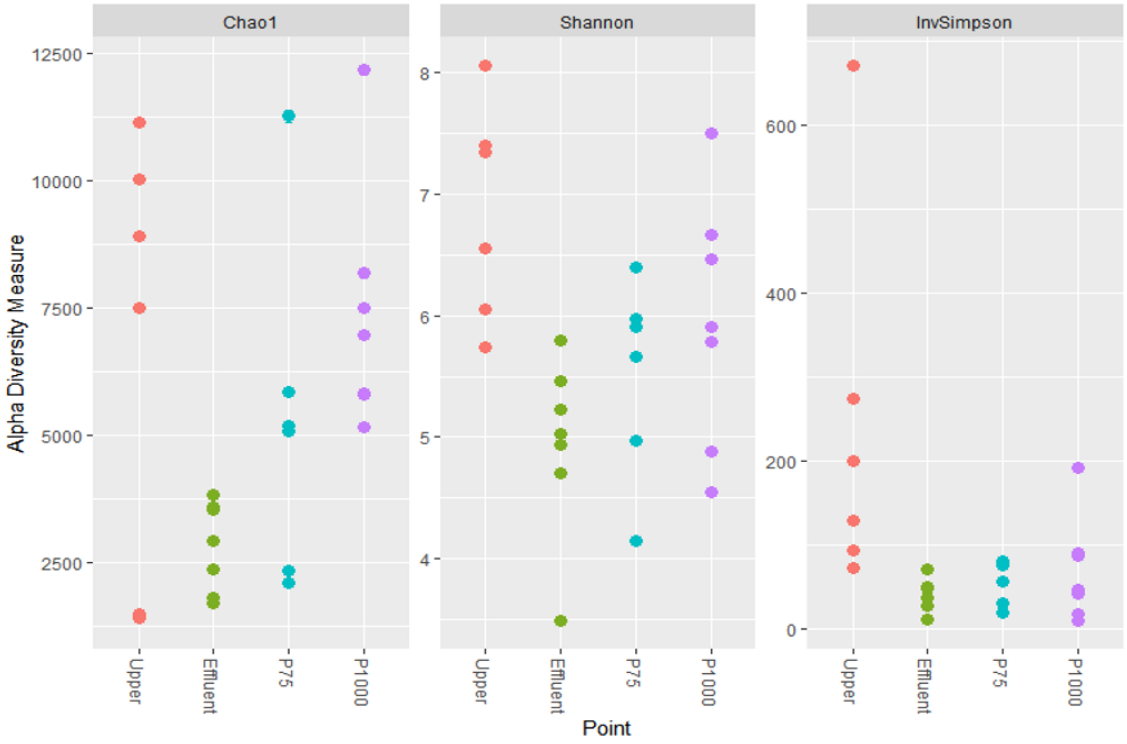
781 **Figure 3:** Distribution of genera in each sampling point and separated by season (summer and

782 winter). Rare taxa grouped the genera with proportions <2.5%.



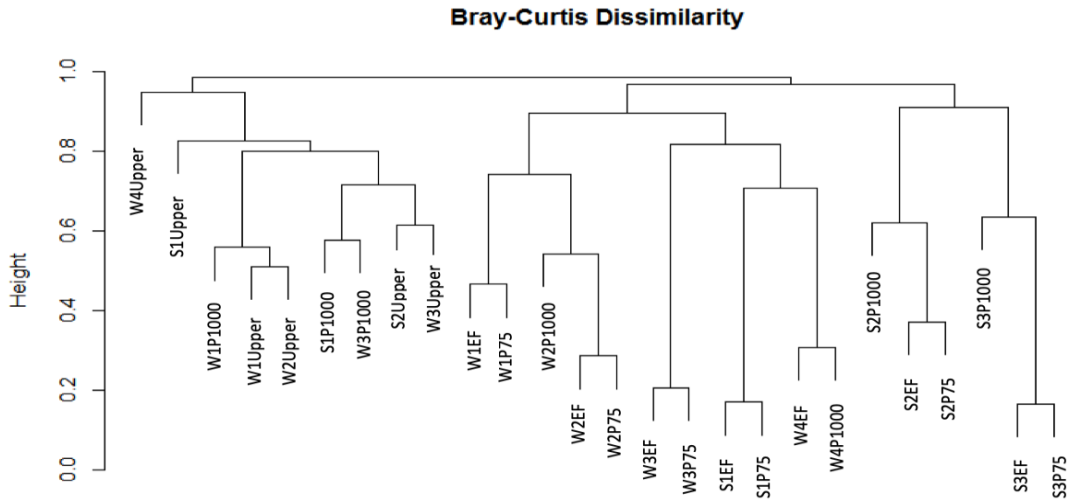
783

Figure 4: Alpha diversity in each sampling point expressed by Chao 1, Shannon and inverse Simpson indices, including the mean and standard deviation.



Upper	6838.5 ± 4319.2	Upper	6.9 ± 0.9	Upper	165.8 ± 124.4
Effluent	2997.0 ± 919.1	Effluent	5.0 ± 0.7	Effluent	40.5 ± 18.9
P75	5413.3 ± 3320.0	P75	5.5 ± 0.8	P75	56.9 ± 26.5
P1000	7547.4 ± 2434.1	P1000	6.0 ± 1.0	P1000	70.3 ± 62.8

Figure 5: Clustering of samples according to Bray-Curtis Dissimilarity. Seasonality is indicated in each sample with W or S, corresponding to winter or summer, respectively. Sampling campaign is indicated with a number from 1 to 4 depending on the season, before the sampling site name.



792 **SUPPLEMENTARY MATERIAL**

793 **Supplementary Table 1:** Streamflow (Q, in m^3/s) in Upper, EF and downstream of the WWTP
794 (P75) and main water temperature (T, in ºC) in each sampling campaign. Seasonality is indicated
795 in each sample with W or S, corresponding to winter or summer, respectively and sampling
796 campaign is indicated with a number from 1 to 4 depending on the season.

Sampling campaign	Q_{Upper}	Q_{EF}	Q_{P75}	T
S1	0.003	0.006	0.009	17.1
S2	0.001	0.007	0.008	20.7
S3	0.000	0.006	0.006	21.1
W1	0.001	0.006	0.007	8.6
W2	0.003	0.006	0.009	9.7
W3	0.004	0.010	0.014	12.5
W4	0.008	0.001	0.018	9.5

797

798

Supplementary Table 2: Statistical descriptors (sensitivity analysis and confusion matrix) of Kernel discriminant analysis used to separate samples according to the distribution of the most abundant 20 ASVs and confusion matrix.

a) All samples together

	Upper	Effluent	P75	P1000
Sensitivity	1.0000	0.8571	0.5000	0.4286
Specificity	0.9000	0.7895	0.9000	1.0000
Positive Prediction Value	0.7500	0.6000	0.6000	1.0000
Negative Prediction Value	1.0000	0.9375	0.8571	0.8261

Prediction/Reference	Upper	Effluent	P75	P1000
Upper	6	0	0	2
Effluent		6	3	1
P75		1	3	1
P1000				3

b) Summer samples

	Upper	Effluent	P75	P1000
Sensitivity	1.0000	0.6667	0.6667	1.0000
Specificity	1.0000	0.8750	0.8750	1.0000
Positive Prediction Value	1.0000	0.6667	0.6667	1.0000
Negative Prediction Value	1.0000	0.8750	0.8750	1.0000

Prediction/Reference	Upper	Effluent	P75	P1000
Upper	2	0	0	0
Effluent		2	1	0
P75		1	2	0
P1000				3

808 c) Winter samples

	Upper	Effluent	P75	P1000
Sensitivity	1.0000	0.9091	0.6667	0.7500
Specificity	1.0000	0.8750	0.9167	1.0000
Positive Prediction Value	1.0000	0.8000	0.6667	1.0000
Negative Prediction Value	1.0000	1.0000	0.9167	0.9167

809

Prediction/Reference	Upper	Effluent	P75	P1000
Upper	4	0	0	0
Effluent		4	1	0
P75			2	1
P1000			0	3

810

Supplementary Table 3: Contribution (%) of the river (Upper), sewage effluent (EF) and unknown origin to the reads in P75 and P1000 in winter (W) and summer (S).

	Upper	EF	Unknown
WP75	8.2	56.8	35.0
SP75	3.2	74.7	22.1
WP1000	16.5	38.1	45.4
SP1000	20.5	47.6	31.9

Supplementary Table 4: Statistically significant differences in ASV relative abundance in the different sampling sites. In red, ASV showing an increase with respect to Upper, and in white there is no difference.

	Winter		Summer		
	P75	P1000	P75	P1000	Genus
ASV1					*Flavobacterium*
ASV2					*Rhodoferax*
ASV3/ASV27					*Limnohabitans*
ASV4/ASV34					*Polynucleobacter*
ASV5					Family_Burkholderiaceae
ASV6					Order_Absconditabacteriales_(SR1)
ASV7/ASV18					*Rhodoferax*
ASV8					*Sediminibacterium*
ASV9					Family_Neisseriaceae
ASV10					C39
ASV11					*Arcobacter*
ASV12					*Simplicispira*
ASV13					*Arcobacter*
ASV14					*Flavobacterium*
ASV15					Class_Gracilibacteria
ASV16					Family_UBA12409
ASV17					*Aquabacterium*
ASV19					Family_Hydrogenophilaceae
ASV20					Order_Candidatus_Magasanikbacteria
ASV21					*Cellvibrio*
ASV22					*Rhodoferax*
ASV24					Order_Candidatus_Moranbacteria
ASV26					*Aurantimicrobium*
ASV28/ASV53					Order_Absconditabacteriales_(SR1)
ASV29					*Rhodoferax*
ASV30					Class_Gracilibacteria
ASV32					Class_ABY1
ASV35					*Flavobacterium*
ASV36					Class_Gammaproteobacteria
ASV37					*Cavicella*
ASV39					Family_UBA12409
ASV42					*Polynucleobacter*
ASV43					*hgcI_clade*
ASV44					*Acidovorax*
ASV45					*Arcobacter*
ASV46/ASV84					*Novosphingobium*
ASV47					Class_Parcubacteria
ASV50					*Sediminibacterium*
ASV54					*Prevotella_9*
ASV55					*Rheinheimera*
ASV57					Class_Gracilibacteria
ASV58					Phylum_Proteobacteria
ASV60					Family_Fibrobacteraceae
ASV68					*Rhizobacter*
ASV93					*Pseudarcicella*

Supplementary Table 5: *P* values obtained from the comparison of alpha diversity in each sampling point for a) Chao1 index, b) Shannon index and c) Inverse Simpson index. Statistically significant differences are showed in bold font.

a) Chao1 index

		EF	P75	P1000
Upper		0.08	0.54	0.72
	EF		0.13	**0.001**
			P75	0.21

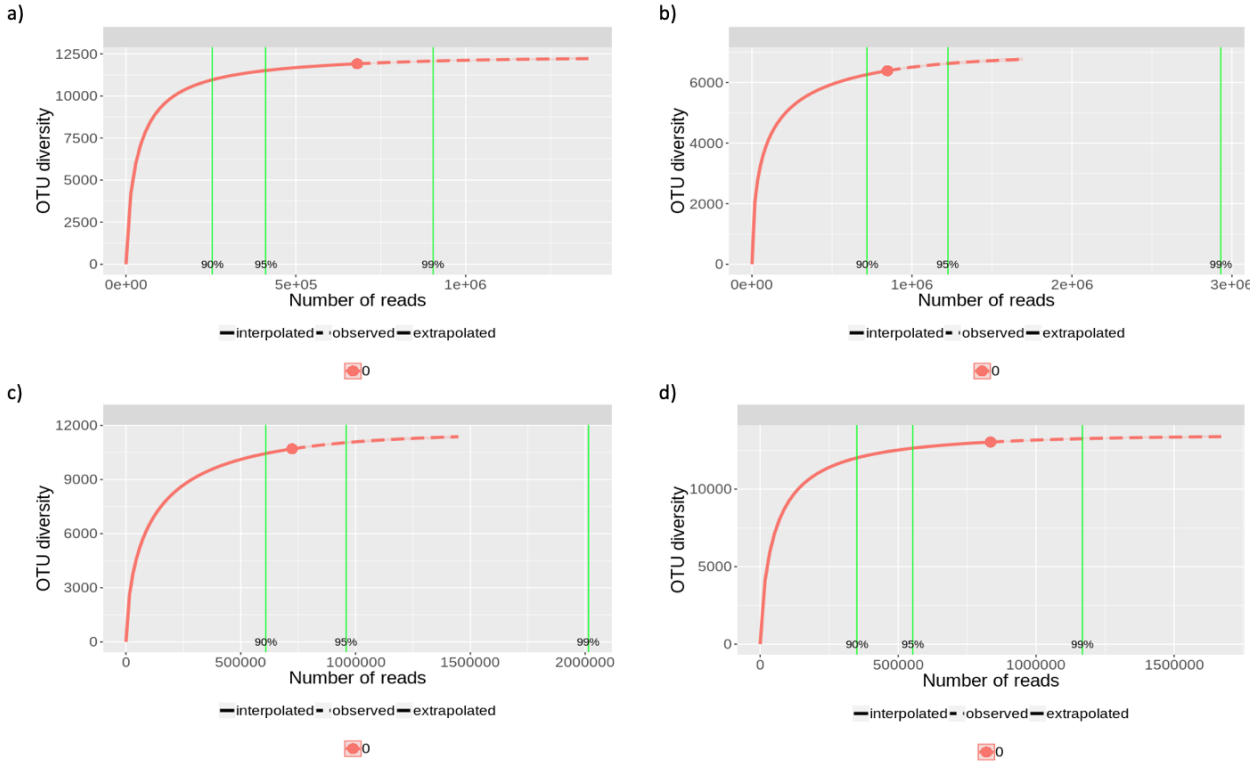
b) Shannon index

		EF	P75	P1000
Upper		**0.001**	**0.021**	0.13
	EF		0.22	0.05
			P75	0.40

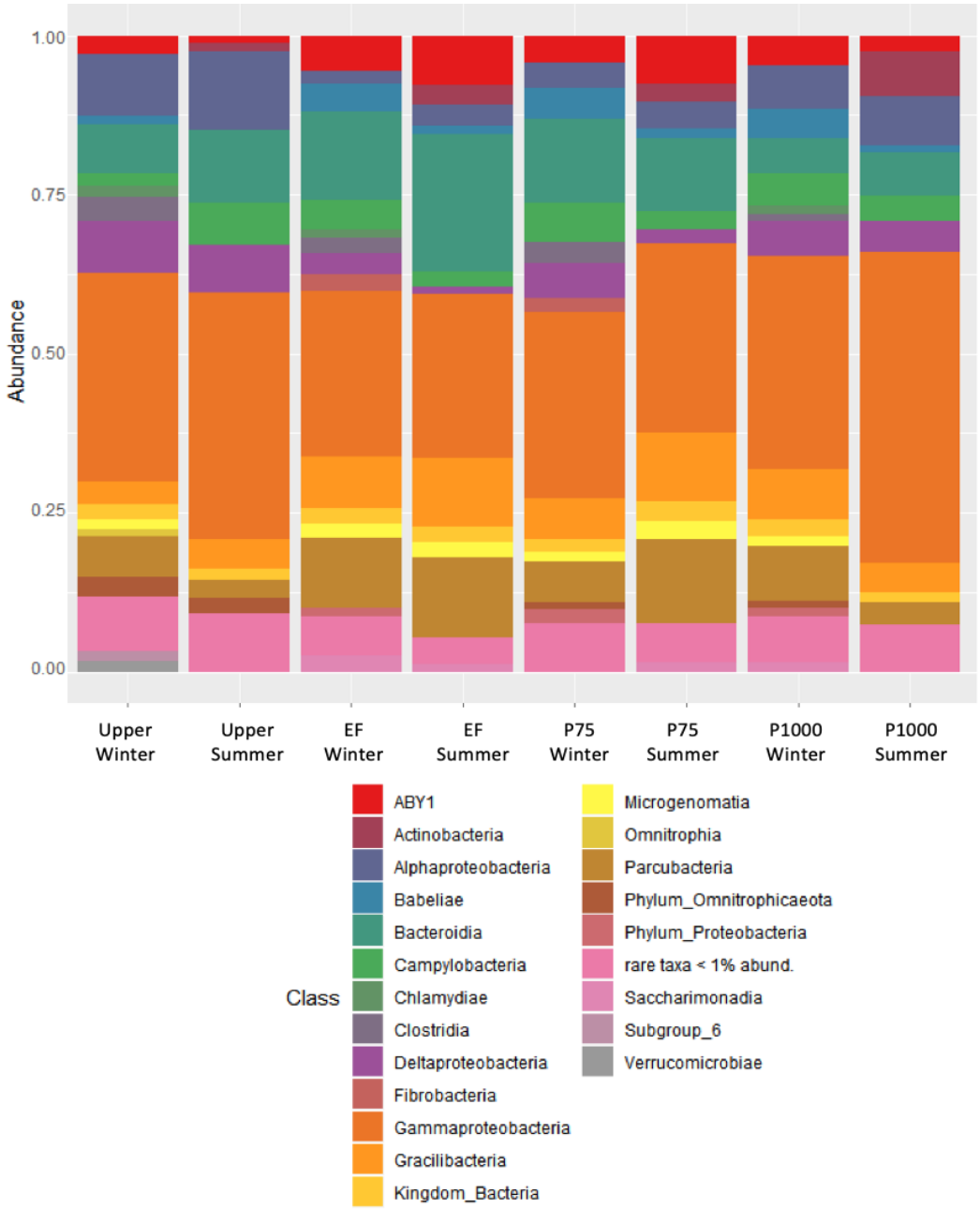
c) Inverse Simpson index

		EF	P75	P1000
Upper		**0.001**	**0.02**	**0.02**
	EF		0.22	0.27
			P75	0.62

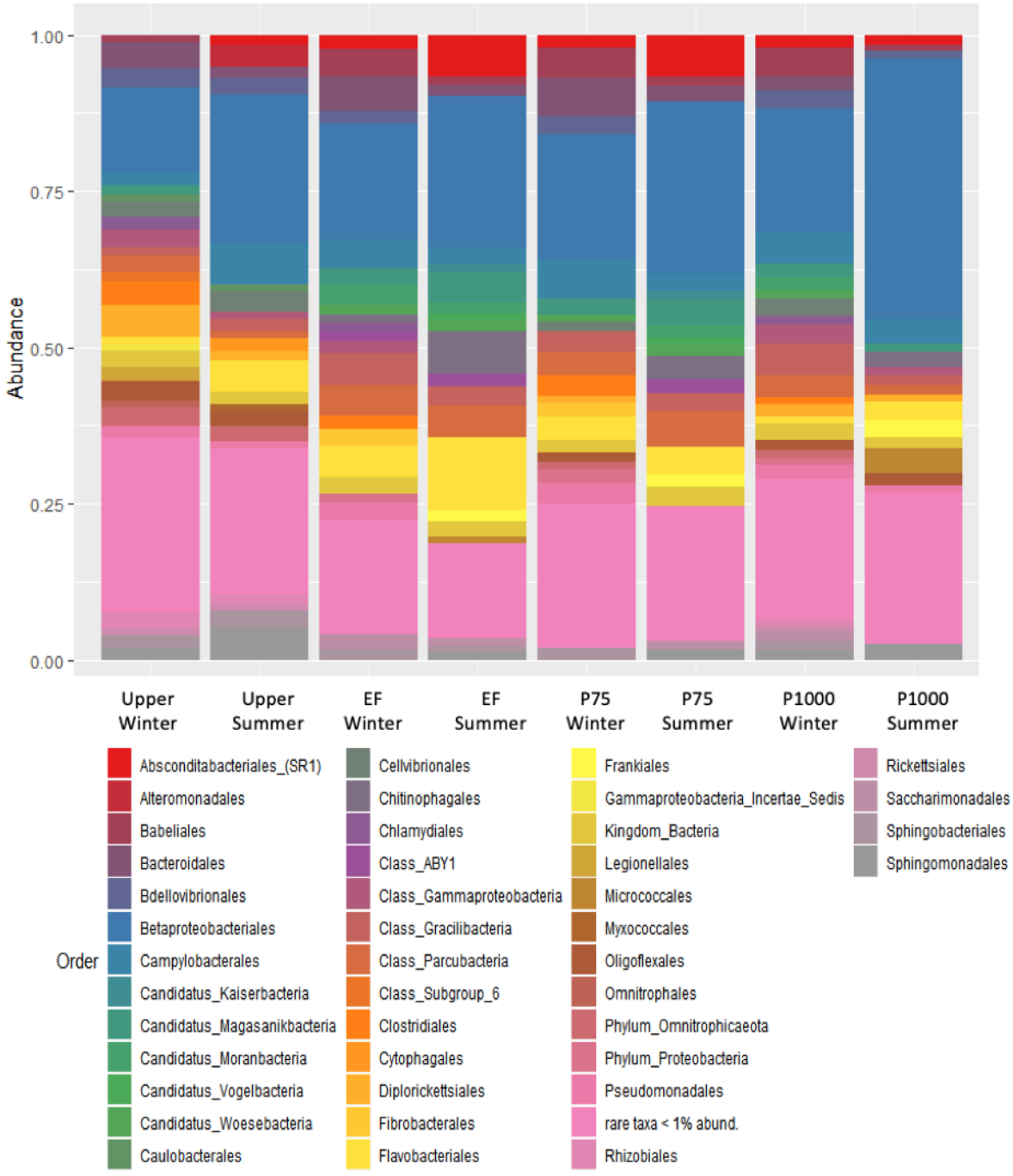
Supplementary Fig. 1: Representability of the sequencing results using rarefaction plots. The graphical representation of the calculated diversity with respect to the expected diversity in relation to the obtained sequencing reads is shown: a) Upper point; b) Effluent point; c) P75 point and d) P1000 point.



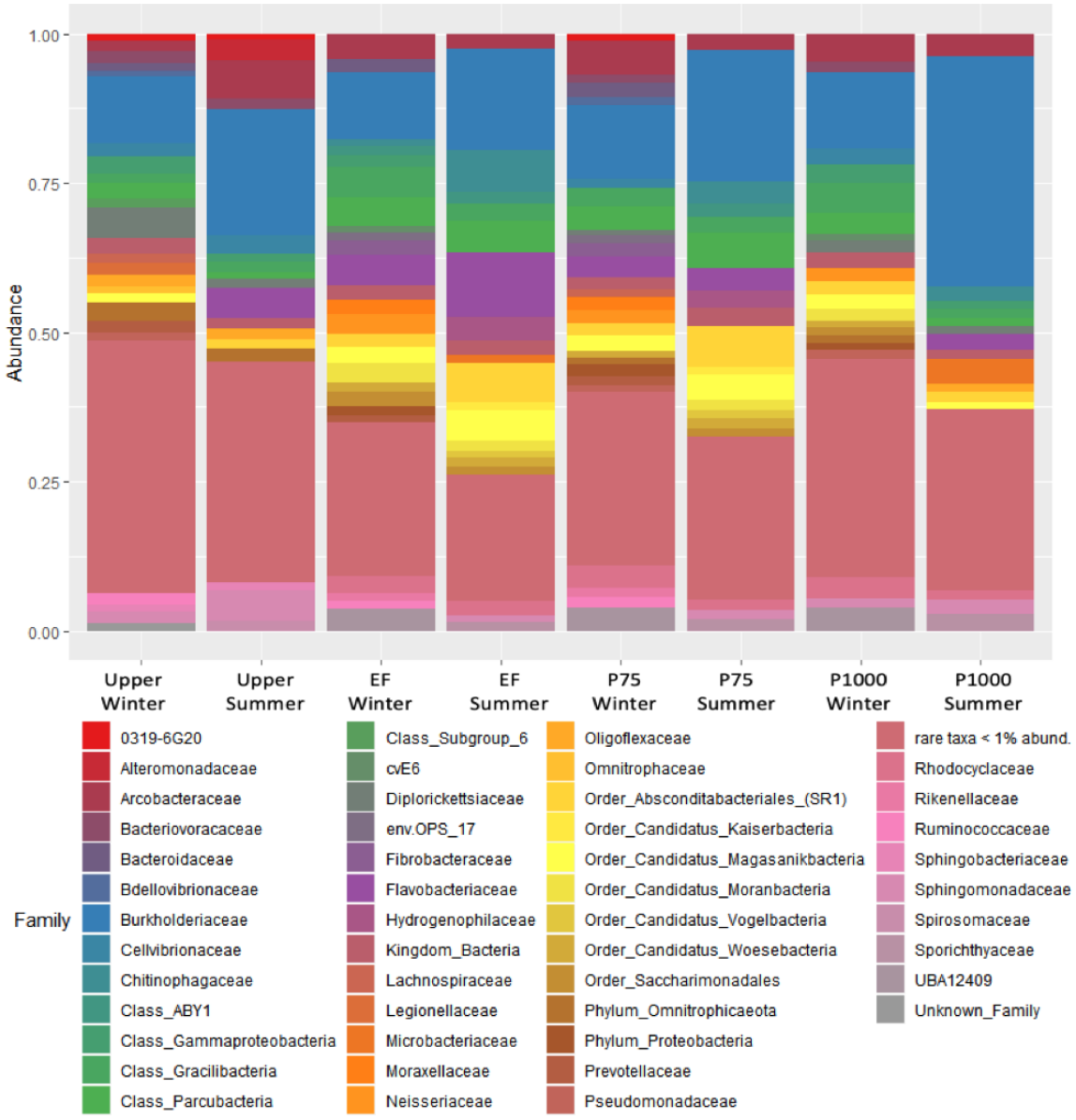
Supplementary Fig. 2: Distribution of classes in each sampling point and separated by season (summer and winter). Rare taxa grouped the genera with proportions <1%.



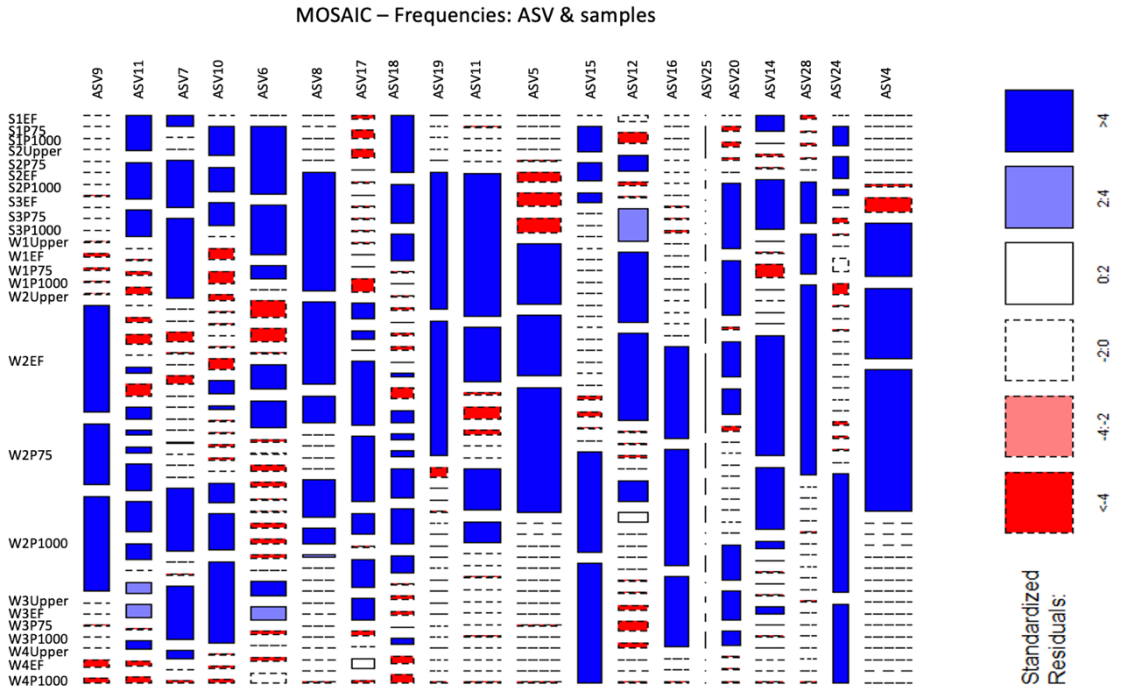
Supplementary Fig. 3: Distribution of orders in each sampling point and separated by season (summer and winter). Rare taxa grouped the genera with proportions <1%.



Supplementary Fig. 4: Distribution of families in each sampling point and separated by season (summer and winter). Rare taxa grouped the genera with proportions <1%.

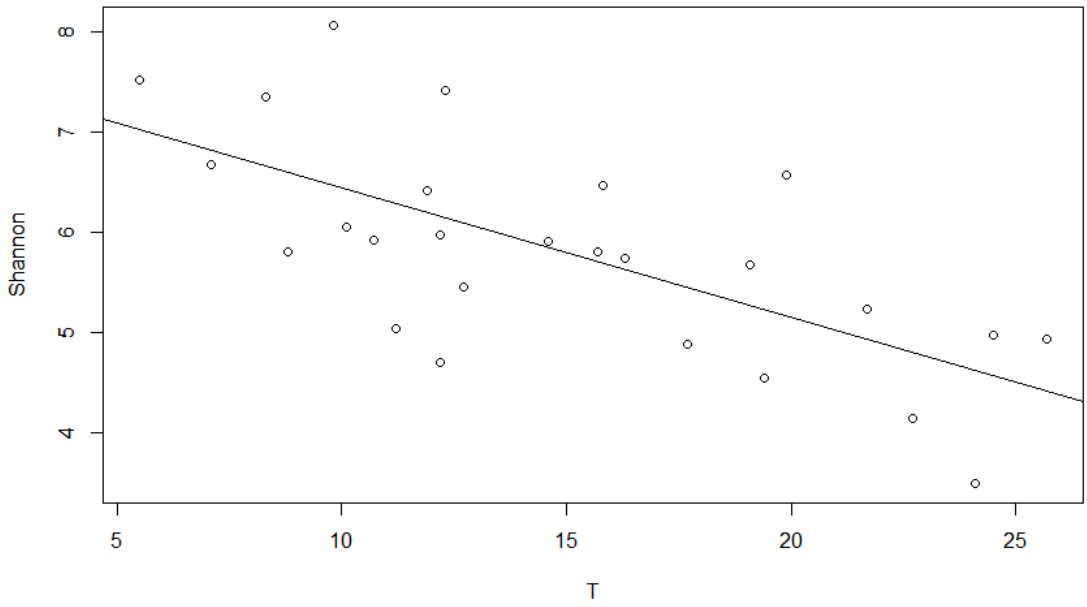


Supplementary Fig. 5: Mosaic plot where all sample groups and the 20 most abundant ASV where used. Standardized residuals are represented as a method to detect sample patterns under the null model (independence) between samples and ASV. Seasonality is indicated in each sample with W or S, corresponding to winter or summer, respectively. Sampling campaign is indicated with a number from 1 to 4 depending on the season, before the sampling site name.

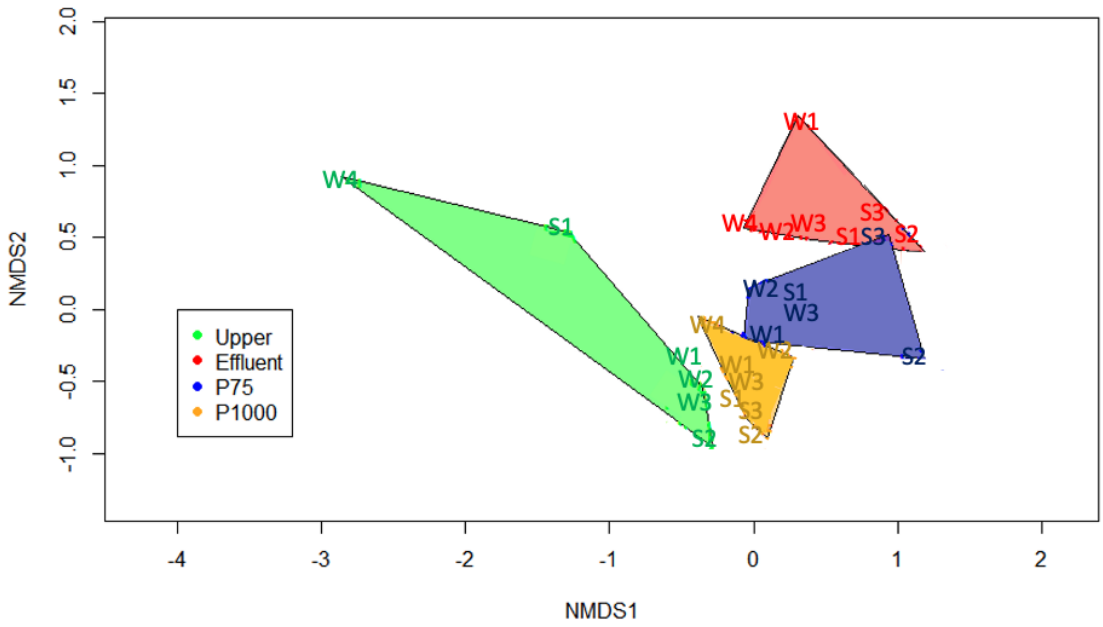


ASV1, *Flavobacterium*; ASV4, *Polynucleobacter*; ASV5, Family_Burkholderiaceae; ASV6, Order_Absconditabacteriales_(SR1); ASV7, *Rhodoferax*; ASV8, *Sediminibacterium*; ASV9, Family_Neisseriaceae; ASV10, Rhodocyclaceae C39; ASV11, *Arcobacter*; ASV12, *Simplicispira*; ASV14, *Flavobacterium*; ASV15, Class_Gracilibacteria; ASV16, Babeliales Family_UBA12409; ASV17, *Aquabacterium*; ASV18, *Arcobacter*; ASV19, Family_Hydrogenophilaceae; ASV20, Order_Candidatus_Magasanikbacteria; ASV24, Order_Candidatus_Moranbacteria; ASV25, Order_Candidatus_Campbellbacteria; ASV26, Aurantimicrobium.

Supplementary Fig. 6: Pearson's correlation of the temperature (T, in °C) and alpha diversity (Shannon index).



867 **Supplementary Fig. 7:** Multidimensional scaling plot of the dissimilarity between the sampling
868 points. Seasonality is indicated in each sample with W or S, corresponding to winter or summer,
869 respectively. Sampling campaign is indicated with a number from 1 to 4 depending on the
870 season.



871