Science of the Total Environment 709 (2020) 135727

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Microbial community shifts in streams receiving treated wastewater effluent

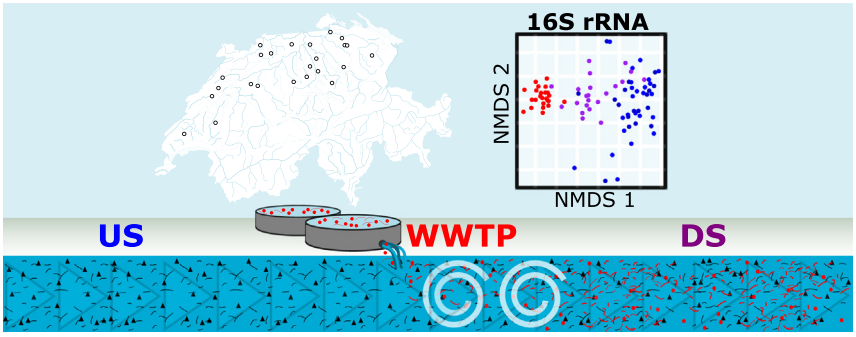
Cresten Mansfeldt [a,1], Kristy Deiner [a,b,*,1], Elvira Mächler [a,c], Kathrin Fenner [a,d,e], Rik I.L. Eggen [a,d], Christian Stamm [a], Urs Schönenberger [a], Jean-Claude Walser [d], Florian Altermatt [a,c]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[b] Natural History Museum London, London, UK
[c] Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland
[d] Department of Environmental Systems Science, ETH, Zürich, Switzerland
[e] Chemistry Department, University of Zürich, Zürich, Switzerland

HIGHLIGHTS

- 23 headwater streams were analyzed to determine bacterial community composition affected by released treated wastewater
- Mixing between the stream and the wastewater effluent predicted downstream community composition for most taxa
- 14 sites showed greater than 50 % of the bacteria taxa were from the wastewater
- Decreases in *phototrophic taxa* could not be explained by mixing alone
- Human-gut related bacteria are indicators of natural streams impacted by wastewater effluent

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 June 2019
Received in revised form 22 October 2019
Accepted 22 November 2019
Available online 11 December 2019

Editor: Patricia Holden

Keywords:
Headwater streams
Wastewater treatment plant effluent
Ruminococcus
Cyanobacteria

ABSTRACT

Wastewater treatment plant (WWTP) effluents release not only chemical constituents in watersheds, but also contain microorganisms. Thus, an understanding of what microorganisms are released and how they change microbial communities within natural streams is needed. To characterize the community shifts in streams receiving WWTP effluent, we sampled water-column microorganisms from upstream, downstream, and the effluent of WWTPs located on 23 headwater streams in which no other effluent was released upstream. We characterized the bacterial community by sequencing the V3-V4 region of the 16S rRNA gene. We hypothesized that the downstream community profile would be a hydraulic mixture between the two sources (i.e., upstream and effluent). In ordination analyses, the downstream bacterial community profile was a mixture between the upstream and effluent. For 14 of the sites, the main contribution (>50%) to the downstream community originated from bacteria in the WWTP effluent and significant shifts in relative abundance of specific sequence variants were detected. These shifts in sequence variants may serve as general bioindicators of wastewater-effluent influenced streams, with a human-gut related *Ruminococcus* genus displaying the highest shift (30-fold higher abundances downstream). However, not all taxa composition changes were predicted based on hydraulic mixing alone. Specifically, the decrease of *Cyanobacteria*/Chloroplast reads was not adequately described by hydraulic mixing. The potential

* Corresponding author: Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
E-mail address: alpinedna@gmail.com (K. Deiner).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.scitotenv.2019.135727

alteration of stream microbial communities via a high inflow of human-gut related bacteria and a decrease in autotrophic functional groups resulting from WWTP effluent creates the potential for general shifts in stream ecosystem function.

1. Introduction

Riverine ecosystems harbor immense diversity of bacteria, archaea, and micro- and macro-eukaryotes (Dudgeon et al., 2006; Heino et al., 2015). This diversity is under pressure from anthropogenic demands placed on these water bodies to provide a range of services including recreation, transport routes, agricultural use, and waste deposition (e.g., treated wastewater, agricultural runoff), inducing stress (Vörösmarty et al., 2010). Within the river ecosystem, the microbial communities participate in biogeochemical cycling (Battin et al., 2016; Zhao et al., 2015) and stress responses (Woodward et al., 2012). However, the understanding and description of the common core river microbial structure (Staley et al., 2013), the demographic shifts of these communities over long river stretches (Savio et al., 2015; Wang et al., 2017), and the responses of these organisms to anthropogenic stresses (Cai et al., 2016; Ibekwe et al., 2016; Li et al., 2018) remain to be fully understood. To ensure that human interactions minimally influence the water quality and ecological roles rivers fulfill, proper management requires a clear understanding of the census, function, and interaction of the resident organisms (Harvey et al., 2017) as well as an adequate and representative monitoring of riverine ecosystems (Jackson et al., 2016).

Riverine ecosystems increase in input complexity, starting out as linear systems but soon form a network of natural and anthropogenic inputs (Altermatt, 2013; Besemer et al., 2013). Natural inputs are confluences of upstream tributaries and non-point terrestrial sources, whereas anthropogenic inputs result from outflows from non-point sources such as those from agricultural fields (Crump et al., 2012) and point sources such as wastewater treatment plants (WWTPs). Anthropogenic sources contain chemicals and microorganisms (Marti and Balcázar, 2014; Price et al., 2018; Rizzo et al., 2013) as well as effluents being sources of antibiotic resistance genes in river ecosystems (Proia et al., 2018; Rodriguez-Mozaz et al., 2015)

Advances in high throughput sequencing have enabled studies of the population structure and responses of microorganisms in rivers (Tan et al., 2015), providing a tool to identify the sources and sinks of the resident organisms. Previous sequencing campaigns have also successfully utilized space-integration to compare microorganisms along the river continuum (Battin et al., 2016; Ruiz-González et al., 2015; Savio et al., 2015; Vannote et al., 1980). The chemical profile of a riverine network often follows dilution principles during the mixing of sources (Alexander et al., 2009). Initial studies exploring bacterial communities suggest the influence of upstream effluents may confound the observed communities downstream (e.g., Marti and Balcázar, 2014; Price et al., 2018; Rizzo et al., 2013). These studies, however, do not precisely address whether aquatic bacterial communities result from the direct hydrological mixing of upstream natural and anthropogenic sources of microorganisms.

To determine the influence of an anthropogenic point source on downstream bacterial communities, shifts in the bacterial community after the contribution of WWTP effluent was determined using high throughput sequencing. Surface water samples were taken from locations upstream of the WWTP effluent (US1 & 2), the discharged effluent (EF), and at the point of complete mixing between the sources downstream (DS; 52–404 m downstream of US1; Table S1) of 24 WWTPs distributed across Switzerland. In total, 23 of these WWTPs discharged into streams that did not receive effluent previously upstream. After obtaining the 16S rRNA sequence profiles at the US, EF, and DS locations, the direct influence of the WWTP effluent on the DS microbial community was determined by uncovering sequence variants that were significantly differentially expressed between sites. Furthermore, the suitability of explaining the resulting DS profile through the physical mixing of the respective water sources was explored in both ordination and community-wide census-based analyses.

2. Methods

2.1. Microbial DNA sampling, filtration, and extraction

Water samples were collected from 22 streams in the Rhine River watershed and 2 streams in the Rhone River watershed (Fig. 1) in Switzerland, representing streams occurring across a wide range of land-use types and geological substrates characteristic of the Swiss Plateau (Table S1; for more details on selected sites see also Figs. 4 and 5 from Stamm et al., 2016). All sites except Val-de-Ruz (*Va*) were streams for which the sampled WWTP was the first on the stream (Fig. 1, indicated in grey). Thus, *Va* was excluded from the location-specific mean taxonomic profiles where mixing effects were tested.

At each of the stream sites, surface water was sampled upstream of the WWTP (US1), the effluent from the WWTP before it flowed into the stream (EF), and at a downstream location (DS) at a point after complete mixing between the stream and the effluent (Fig. S1). The point of complete mixing was determined by onsite measurements of electrical conductivity. Complete mixing was assumed when transects measured across the stream did not reveal any lateral conductivity gradients. As a control for distance on the same stream for selected locations, we sampled a second upstream location (US2) that was an equal distance upstream from US1 as the US1 was from the DS (Table S1). Water samples were transported in a cooler on ice with a maximum transport time of six hours and were stored at −20 °C until further processing.

The method for the filtration and extraction of DNA from water samples followed that of Deiner et al. (2015), in which the water was first thawed in a room temperature water bath, filtered through a glass fiber filter (GF/F, nominal pore size of 0.7 μm, 25 mm, Whatman International Ltd., England), and extracted with a Phenol-Chloroform Isoamyl method succeeded by an ethanol precipitation. Strict adherence to contamination control was ensured using a designated clean lab in which only DNA isolation and pre-polymerase chain reaction (PCR) preparations are performed (Deiner et al., 2015). Between two and eight independent extractions from filters were performed for each sample location. The total volume of water filtered for each extraction replicate ranged from 65 to 350 mL and depended on the filter clogging as a result of suspended solids in the sample. A total 500 to 700 mL of filtered water was used to screen bacterial communities per sample location in the PCR by creating a 50-μL pool of extracted DNA in which equal volumes from each independent extraction replicate were combined. All pooled DNA extractions were cleaned using a OneStep™ PCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA) according to the manufacturer's protocol to minimize PCR inhibition of DNA sourced from riverine samples (McKee et al., 2015). Total DNA in the pooled extraction was then quantified using a Qubit (1.0) fluorometer following the recommended protocol for the dsDNA HS Assay (Life Technologies, Carlsbad, CA, USA) (Table S1).

Filter negative controls were created on each day of filtration (Table S2). A filter negative control was established by decontaminating 250-mL Nanopur filtered water under UVC light for 30 min and

subsequently filtering the water in accordance with the same protocol used for samples. DNA extraction controls were processed side-by-side with samples and used to monitor contamination. A negative extraction control was processed with each batch of extractions; the batches consisted of 18 and 22 filters (Table S2). A subset of controls were sequenced to confirm no major contamination occurred during laboratory processing.

2.2. Library construction and sequencing

Library construction for each pooled DNA extraction followed a three-step PCR process. The first PCR amplified the V3-V4 region of the 16S rRNA gene. Four independent PCRs were performed in 15-μL volumes with final concentrations of 1× supplied buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland), 1000 ng/μL BSA (New England Biolabs, Ipswich, MA, USA), 0.18 mMol dNTPs, 2.0 mMol $MgCl_2$, 0.05 units per μL Taq DNA polymerase (Faststart TAQ, Roche, Inc.), and 0.5 μMol of each forward and reverse primer (Table S3; Microsynth, AG, Balgach, Switzerland). In total, 2-μL of extracted DNA was added, ranging in concentration from 0.03 to 54.0 ng/μL (Table S1). The thermal-cycling regime was 95 °C for 4 min, followed by 25 cycles consisting of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min. A final extension of 72 °C for 5 min was performed, and the PCR was cooled to 4 °C and stored at −20 °C until further processing. Each PCR replicate was cleaned with Exo I Nuclease (EXO I) and Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific Inc., Waltham, Maryland USA). The master mix consisted of 1.6 U/μL Exo I and 0.15 U/μL SAP in a total volume of 1.1-μL which was then added to 7.5-μL of the PCR product. Products were heated to 37 °C for 15 min, followed by 15 min at 80 °C, and was then cooled to 4 °C and stored at −20 °C until further processing.

The second PCR was conducted with the same conditions as the first PCR except the forward and reverse primers were modified to include the Nextera® transposase sequences (Microsynth, AG, Balgach, Switzerland) and only 1-μL of cleaned PCR product was used in the reaction. Between the forward and reverse primer sequence and the transposase sequence a different number of bases were inserted to create more heterogeneity on the flow cell (Table S3). The thermal-cycling regime was the same except only 8 cycles were used. PCR products from the four independent reactions for each sample were then pooled and cleaned. Each of the pooled reactions (a total of 60-μL) were cleaned using a 0.8× bead:sample volume ratio of Agencourt Ampure XP Beads (Beckman Coulter, Webster, TX) and separated with a DynaMag-2 magnet rack (Life Technologies, Grand Island, NY) following the manufacturer's recommended protocol (Beckman Coulter, Brea, CA, USA).

The third PCR was employed to index each sample prior to pooling all samples for sequencing. We dual-indexed samples using the Nextera® Index Kit A (Illumina, Inc., San Diego, CA, USA). The index PCR was performed in a 50-μL volume. Cleaned amplicons from the previous step were added in volumes of either 2-, 5-, or 15-μL for DNA concentrations of >2.5 ng/μL, between 2.5 and 0.12 ng/μL, and <0.12 ng/μL, respectively. We used the KAPA Library Amplification Kit following the manufacturer's recommended protocol (KAPA Biosystems, Wilmington, MA). Each reaction was cleaned using a 1.0× bead:sample volume (50 μL) ratio of Agencourt Ampure XP Beads (Beckman Coulter, Webster, TX) and separated with a DynaMag-2 magnet rack (Life Technologies, Grand Island, NY) following the manufacturer's recommended protocol (Beckman Coulter, Brea, CA, USA).

The DNA concentration of the cleaned and indexed libraries was then determined using a Qubit (1.0) fluorometer following recommended protocols for the dsDNA HS Assay (Table S1). The libraries were then normalized and pooled at a 4-nM concentration. PHiX control was added at a 1% concentration. Paired end (2 × 300 nt) sequencing was performed on an Illumina MiSeq (MiSeq Reagent kit v3, 300 cycles) at the Genomic Diversity Centre (GDC) at the ETH, Zurich, Switzerland following the manufacture's run protocols (Illumina, Inc.). On a different Miseq instrument at the Functional Genomics Center (FCGZ), Zurich, Switzerland, a second run of the same pooled libraries was conducted to increase the sequence depth and test for run effects. The MiSeq Control Software Version 2.2 including MiSeq Reporter 2.2 was used for the primary analysis and the de-multiplexing of the raw reads. All raw sequences are available at the European Nucleotide Archive (ENA) under the accession number PRJEB26649.

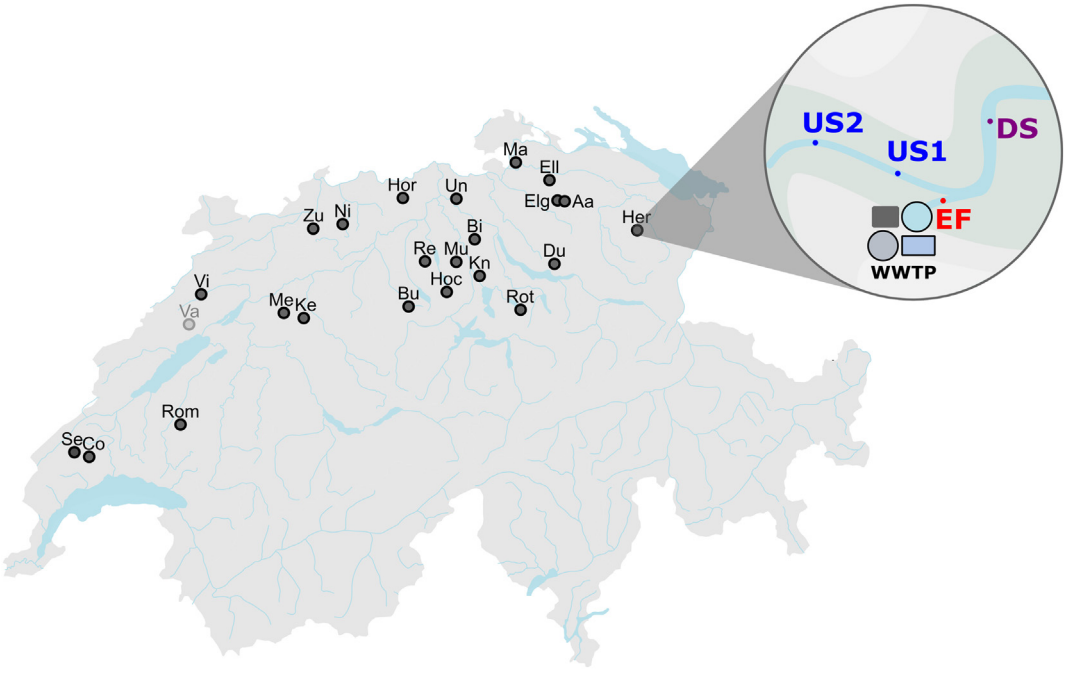


Fig. 1. Sampling sites on streams throughout Switzerland. The inset diagram shows the locations that were sampled at each site: upstream 2 (US2) in blue, upstream 1 (US1) in blue, wastewater treatment plant effluent (EF) in red, and downstream (DS) in purple. Val-de-Ruz (*Va*, colored light-grey) was not the first WWTP on the stream and thus excluded from location-specific averages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. 16S data processing, exact sequence variants binning, and taxonomic assignment

The reads were checked for quality using FastQC v0.11.2 (Andrews, 2010), end-trimmed using seqtk (with 30 and 70 bp removed for read pairs R1 and R2, respectively; minimum length 150 bp; minimum mean quality 15) [https://github.com/lh3/seqtk], and merged using FLASH v1.2.11 (minimum and maximum overlap of 15 and 300 bp, respectively; maximum mismatch density of 0.25) (Magoč and Salzberg, 2011). The primers were trimmed using cutadapt v1.12 (wildcards allowed; full-length overlap; error rate 0.01). Quality filtering was performed with PRINSEQ-lite (minimum quality mean 20; no ambiguous nucleotides; dust low-complexity filter with a threshold of 30) with a subsequent size and GC selection step (size selection range 100–500 bp; GC selection range 40–70%).

The reads were processed using an Exact Sequencing Variants (ESV) analysis (Callahan et al., 2017) to follow current best practices in 16S rRNA gene copy processing (Knight et al., 2018). The sample reads were first denoised into ESVs (i.e., zero OTUs; zOTUs) using UNOISE3 in the USEARCH software v.10.0.240. The final predicted taxonomic assignments were performed using SINTAX in the USEARCH software v.10.0.240 (both strands considered, cutoff of 0.7) (Edgar, 2016). Phyla level annotations were manually replaced for ESVs assigned to the chloroplast class and mitochondria family with *Cyanobacteria*/Chloroplast and *Proteobacteria*/Mitochondria, respectively, to correctly identify true 16S rRNA gene sequences that are of eukaryotic origin and separate them from the bacterial signal. Both chloroplast and mitochondrial sequences were maintained in the analysis because these signals, although not of bacterial origin, remain informative to an environmental-DNA based screening campaign. Additionally, the majority of the *Cyanobacteria*/Chloroplast reads originated from aquatic photosynthetic organisms and were not a result of only leaf litter or other terrestrial species origin (Table S4). A complete comparative analysis with these sequences removed is included in Fig. S2. Additionally, the total reads obtained at each step of bioinformatic filtration are reported in Table S5 from run at GDC and Table S6 from run at FCGZ. All intermediate data files required for the described analyses are provided in a data package.

2.4. Sample and ESV quality control

The data was analyzed in R (v. 3.5.1) using the phyloseq package (v. 3.0.0) (McMurdie and Holmes, 2013) using File S1. The raw reads were first checked to determine whether any samples performed poorly in either the amplification, pooling, or sequencing steps, resulting in too few reads from sampling loss (location EkUS2) or too many reads from possible contamination (location EkDS); these two samples were removed from the analysis. Additionally, the locations CoDS, MeEF, and NiDS resulted in low read counts (however, exceeding the negative controls), potentially influencing any analyses incorporating rarefaction or assuming similar number of reads per sample.

Negative controls were also used to identify ESVs that represent likely sample handling contamination events or errors in tag jumping. In total, 12 ESVs (ESV9, 30, 63, 133, 241, 281, 308, 322, 363, 388, 411, and 487) were detected in high abundance within the controls (each comprising over 2% of the total detected reads across all negative controls). Only ESV9, 30, 63, and 133 were detected within the field samples at >0.1% of the total sample read abundance. To be conservative, all 12 high-abundance ESVs in the controls were removed from all samples before analyses.

2.5. Merging of pseudo-technical replicates

Because the library was split into two prior to the MiSeq run (but after pooling; labelled as Run 01.06.2015 and 16.06.2015) to overcome potential instrument errors at the time of sequencing, the two runs are not true technical replicates. To test whether it was appropriate to combine the two sequencing runs, a correlation between the raw data of the two runs was used and resulted in a highly significant correlation ($r = 1.00$, p-value $= 2.14\times 10^{-187}$). A permutational multivariate analysis of variance (permanova) test was then performed on the Bray-Curtis dissimilarity of the scaled (to an even sampling depth) data using the adonis function in vegan v.2.5.2 in R. The permanova considered the amount of variation determined by grouping the samples into either run, location, or run crossed with location (Table S7). The betadisper and permutest functions in vegan v.2.5.2 were applied to both groups to determine whether artifacts resulting from heterogeneous dispersions were present; no artifacts were found. The grouping by run displayed no significant variation, whereas the location grouping displayed highly significant differences. Therefore, the split runs were merged in all further analyses to not overweight the power of downstream statistical tests with pseudo-technical replicates. The type of analyses dictates the merging method employed and is described below.

2.6. Alpha- and beta-diversity estimation

The split runs were merged using the merge_sample command in phyloseq. All samples passed a 10,000 reads per library cutoff. The resulting number of reads per sample ranged between 14,517–399,377 with a median of 149,643 (Fig. S3). For the alpha diversity calculations for each location category (US2, US1, EF, and DS), richness and Shannon-diversity were calculated using the merged data after and before rarefaction, respectively, following the recommendations of Knight et al. (2018). To determine whether this lower cutoff influenced further analyses by being away from the saturation of the rarefaction curves (Fig. S4), the analysis was repeated for a higher cutoff of 100,000 reads and no major difference were observed (Fig. S5). We therefore present the results after rarefaction in the main text. The subsequent alpha diversities were compared between locations to test for significant differences using the non-parametric Kruskal-Wallis test.

The beta-diversity was determined by calculating the UniFrac distance (both weighted to account for abundance and unweighted to overemphasize the low abundance ESVs; (Hamady et al., 2010) and represented in a non-metric multidimensional scaling (NMDS) plot using the ordinate function within phyloseq.

2.7. Determination of community mixing between stream and WWTP sources

To determine whether hydrological mixing can explain the fraction of EF DNA mixing into the DS, we first calculated the fraction of EF flow (Q_{EF}) on total downstream discharge Q_{tot}. This fraction can be obtained without flow measurements based on either the measured water chemistry parameters (D_{chem}) or the DNA (D_{DNA}) concentrations under the assumption of complete physical mixing (Eq. (1), see Supplemental File S2 for the derivation of the equation illustrating how the flow cancels out):

$$D_{chem\ or\ DNA} = \frac{Q_{EF}}{Q_{up} + Q_{EF}} = \frac{C_{DS} - C_{US}}{C_{EF} - C_{US}} \quad (1)$$

where C_{US1}, C_{DS}, and C_{EF} are the concentrations from chemical parameters (Burdon et al., 2019; Stamm et al., 2016) or the total extracted DNA concentration at the US, DS, and EF, respectively.

The values of D_{chem} were calculated in three steps. First, for each water quality parameter from the two groups general water chemistry (20 parameters) and micropollutants (225 parameters in 2013, 57 in 2014) (see Supplemental File S2 for the full list of parameters) D_{chem} was calculated based on Eq. (1). In 2014, heavy metals (20 parameters) were additionally included as a third group of water quality parameters. Subsequently, the median was derived for each of the two (or three) groups, and finally the average was taken across the groups. This

procedure allowed to consistently include all chemical data and to obtain robust D_{chem} values.

When both US1 and US2 was available for the 20 measurements, the mean was used as the US. Only US1 was used for the DNA concentration to be consistent with subsequent analyses. A linear fit and Pearson's correlation was employed to test for a relationship between the mixing predicted from the calculated D_{chem} and the mixing predicted from the calculated D_{DNA}. The Reinach, Severy, and Zullwil samples are excluded from the above analysis. All three display higher extracted DNA in the DS than either the US1 or EF sources. Thus, these sites did not follow mass conservation and were likely influenced by an additional source of DNA or sample loss during laboratory preparation.

To overcome potential non-mixing mediated shifts in the DS community (e.g., nutrient promoted growth, toxic-substance mediated death) that potentially influence the DS DNA concentration values two further approaches were explored to quantify the mixing of the sources without the need to consider bulk DNA concentrations measured in the DS. First, we determined the fraction F_{DNA} of the DNA mass (M_{DNA}^{EF}) contributed by the WW to the total DNA mass downstream (M_{DNA}^{down}) under the assumption of complete mixing. For the fraction of EF (D_{chem}), F_{DNA} can be derived without flow measurements (see Supplemental File S2 for the derivation of the equation) as shown by Eq. (2):

$$F_{D_{chem}\ corrected\ DNA} = \frac{M_{DNA}^{EF}}{M_{DNA}^{EF} + M_{DNA}^{up}} = \frac{C_{EF} * D_{chem}}{C_{EF} * D_{chem} + C_{US1} * (1 - D_{chem})} \quad (2)$$

where C_{EF} and C_{US1} were the measured DNA concentrations in the EF and US1 samples, respectively.

Second, the contribution of the EF to the DS community profile was predicted using SourceTracker2 (Knights et al., 2011), a program that considers the identity of the DNA sequences (not bulk DNA concentrations) observed. SourceTracker2 uses the ESV data, that when binned at the phyla level exceed 100 reads across all sites) to determine the contributions of the two sources (US1 and EF) to the DS. The rarefaction in SourceTracker2 was disabled because rarefied data from Section 2.6 was converted into biom format and used in the analysis. A linear fit and Pearson's correlation was employed to test for a 1:1 relationship between these two approaches.

The accuracy of the prediction resulting from Eq. (2) was further explored by taking the ratio of the predicted (using the denominator of Eq. (2) and substituting in phyla abundances) to measured DS phyla abundances. Failing predictions were defined as an absolute value of the log-ratio exceeding 2. To test for over- or under-representation of failures within a phylum, a geometric analysis was run using the phyper function in R. The assigned *p*-values were multiple hypotheses corrected with the p.adjust function in R using the false discovery rate (fdr). Corrected values below 0.05 are considered significant (File S3).

2.8. Differential ESV abundance analysis

The runs were first merged using the collapseReplicates function in DESeq2 version 1.20.0 (Love et al., 2014). Samples were not rarefied for these pairwise comparisons because of the limitations in the analyses detailed by McMurdie and Holmes (2015). The analysis followed the standard workflow for DESeq2 analysis, and the scripts to run the test are provided in File S2. To determine whether certain ESVs were characteristic of a location, pairwise differential expression tests were performed between the EF and DS, US1 and DS, EF and US1, and US1 and US2 locations. The analysis employed a Wald test with a parametric fit type. Adjusted p-values below 0.05 for ESVs exceeding a fold-change of two are considered significant.

2.9. Assigning ESVs to environmental descriptors

To explore whether the measured community profiles of bacteria are indicative of a site influenced by treated wastewater effluent, we used the SeqENV v1.2.3 text-mining pipeline (Sinclair et al., 2016) to link the detected ESVs at a given location to habitat descriptors (File S3). In brief, this method searches for 97% sequence identity between previously reported 16S rRNA gene sequences that have an associated isolation source and the top 1000 abundant ESVs found in the NCBI nt database compiled on January 7, 2015 to obtain the now deprecated gi numbers. The isolation sources are parsed to determine words that match to the European Bioinformatics Institute's Environmental Ontology database (EnvO). The observed read counts of the ESVs are assigned to these habitat descriptors. A heatmap was plotted with a cluster dendrogram to visualize habitat names associated across all sites and locations within each site. Wordclouds depicting habitat descriptors varying in size according to abundance of ESV associated with that descriptor in sample were plotted using wordcloud v.2.6 to illustrate habitat descriptor diversity associated with locations at each site as implemented in SeqENV v1.2.3.

3. Results

3.1. Mean taxonomic profiles for sampling locations

The averaged site-specific 16S rRNA gene taxonomic profiles for the four types of sampling locations (US1, US2, EF, and DS) revealed characteristic profiles for each location type across all sites (Fig. 2a; profiles for individual sites are presented in Fig. S6). The ESV alpha diversity metrics of the locations showed that the EF samples had significantly lower richness and diversity (Fig. 2b and c). Upstream locations did not differ from each other in richness or diversity nor did they differ substantially from their downstream locations (Fig. 2b and c).

Considering the identity and phylogenetic relatedness of bacteria ESVs sampled at each location, the Unifrac beta-diversity of the samples plotted on an ordination space showed that the downstream is a mixture of the US and EF communities by intermediate placement between the US and EF (Fig. 3). Additionally, when comparing the weighted (Fig. 3a, abundance considered) and unweighted (Fig. 3b, presence-absence) Unifrac diversity, both the US and EF communities display a shift in their range (i.e., a decrease; likely driven by changes in rarity of taxa across the study locations) within the ordination space, with the DS showing a rather similar profile (i.e., the change in ellipse size when abundance is considered).

3.2. Mixing of community profiles

When comparing the bulk DNA concentration predicted mixing to the chemical mixing, both calculated using Eq. (1), a low correlation ($r = 0.18$) resulted and did not trend with the 1:1 line (Fig. 4a). With the alternative analysis based on SourceTracker2, the majority of the DS 16S rRNA gene profile could be explained by the combination of the US1 and EF for each site (Fig. S7). The percentage contribution between the two sources varied, but most sites (14) had >50% of the 16S rRNA gene profile explained by contributions from the EF. Notably, results of the SourceTracker2 analysis for the fraction of ES in DS displayed a high correlation with the EF fraction derived according to Eq. (2), i.e., predicted using bulk DNA in US and EF only ($r = 0.75$) (Fig. 4b). When exploring individual phyla instead of the bulk community composition (Fig. 4c), the geometric analysis highlighted the relative abundance change for the Chloroplast associated 16S rRNA sequences as containing significantly more outliers not well-explained by mixing (fdr corrected p-value <.05).

3.3. Differential ESVs abundance observed among locations

When US1 was compared to US2, order or genus relative abundances did not change, signifying that distance alone is not explanatory for community shifts (Table S8). When comparing US1 to the DS and EF sites, phototrophs such as *Cyanobacteria,* Chloroplasts associated with

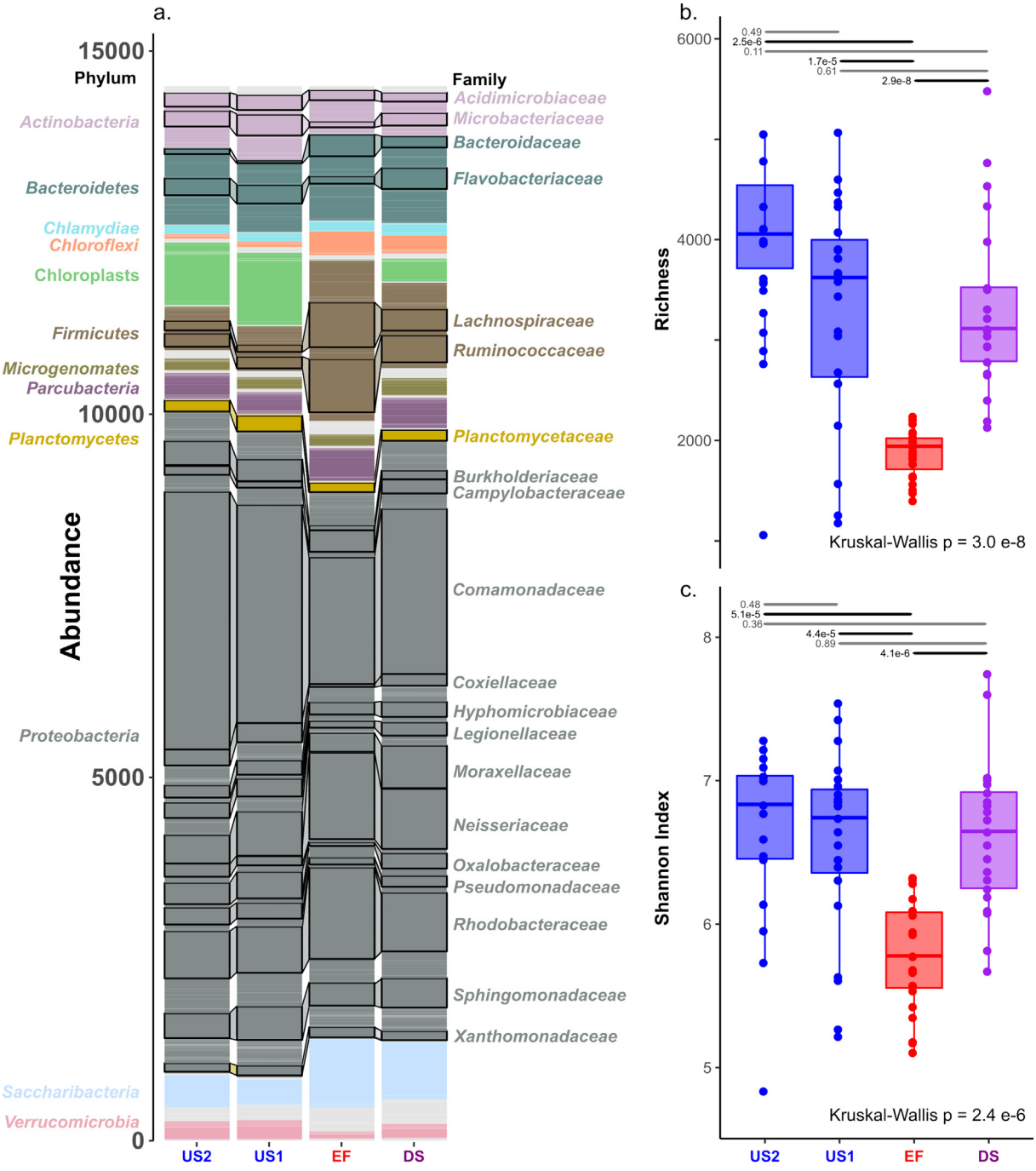


Fig. 2. Summary of the microbial composition measured by 16S rRNA gene sequence profiles for all locations. a) The abundancies averaged across all sites with respect to US2, US1, EF, and DS locations. Phyla accounting for >1% of all reads in at least one sample are colour coded and labelled to the left of the figure. The top twenty families by summed abundances across the four sites (and with a taxonomic descriptor other than uncultured) are represented by the boxes and labelled to the right of the figure. b) Observed richness and c) Shannon diversity indices for each sampling location, for which each individual point represents a single WWTP site. The Kruskal-Wallis *p*-value is also presented for comparisons between locations. For the individual location comparisons, those p-values below the Bonferoni corrected value of 8.3 e^{-3} are considered significant and are bolded. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

algae and members of the *Proteobacteria-α* such as *Sphingomonadales*-related ESVs were significantly lower in relative abundance in DS locations (Fig. 5). Conversly, relative abundances of heterotrophic *Proteobacteria-β* groups such as *Neisseriaceae* and *Comamonadaceae* and the group *Firmicutes* were significantly higher in the DS relative to US1 locations (Fig. 5). Whereas Fig. 5 highlights groups listed that shifted between US1, DS, and EF, many taxonomic groups did not show shifts (Table S7).

3.4. ESVs qualitatively identify outliers

Using the SeqENV pipeline, EF samples were generally characterized by descriptors such as “waste water treatment plant”, “activated sludge”, and “bioreactor”, whereas the US samples are assigned labels such as “river”, “freshwater”, and “glacier” (the two lists of words are colored red and blue in Fig. 6, respectively). US1 and US2 locations generally clustered (Fig. 6; middle two clusters) and were separate from

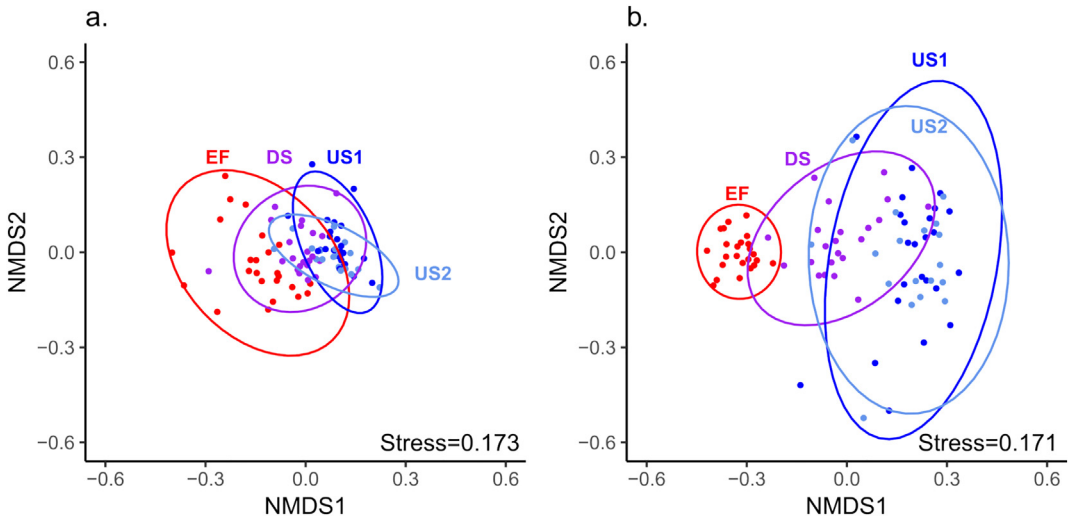


Fig. 3. Ordination plot of the non-metric multidimensional scaling analysis for the individual locations based on the Unifrac (a) weighted and (b) unweighted distance metric. Ellipses denote the 95% confidence interval considering the standard error and a normal distribution. (For interpretation of the plots in colour, the reader is referred to the web version of this article.)

their EF and DS locations (Fig. 6; leftmost two clusters), with a few exceptions. A fifth cluster contained a mix of DS and US locations. DS locations associated with the third cluster tended to be sites for which SourceTracker2 (Fig. S7) predicted a large percent of their composition to be explained by the US rather than the EF community composition.

When investigating shifts in relative phyla abundance, locations appear to be reflected in the descriptors as well. For example, the dominance of *Chloroflexi* in the Villeret (*Vi*) EF and its DS location causes the DS and EF at this site to cluster together and to map well with the "sludge" habitat descriptor. Most US sites are linked to descriptors associated with natural systems (Fig. 7; e.g., "river", "lake"; all wordclouds for each location are depicted in Fig. S8) and are segregated from DS and EF (e.g., Aadorf (*Aa*); Fig. 7a).

Zullwil US1 (*ZuUS1*) and Kernenried US2 (*KeUS2*) are the only two US locations that strongly cluster with EF locations, highlighting that these streams may already be contaminated as indicated by the associated terms of "activated sludge" and "sewage" for the US location (Fig. 6). There are no identified EF sources upstream, but based on near-stream land use, both sites may be influenced by livestock grazing (e.g., Zullwil; Fig. 7b).

4. Discussion

Based on a highly replicated study design of sampling 23 natural streams above and below the first WWTP on that stream, we found that WWTP effluents are acting as additional sources of bacteria in the stream and that the effluent significantly alters downstream water column bacterial communities. At all sites, the hydrological mixing of the EF into the stream is a driver of the downstream community composition indicating that mixing of the two sources can generally explain planktonic community shifts after wastewater enters a stream. For many sites >50% of the bacterial community DNA is likely from the WWTP effluent and the greatest shifts in the downstream community consisted of increased relative abundance of bacteria from human gut

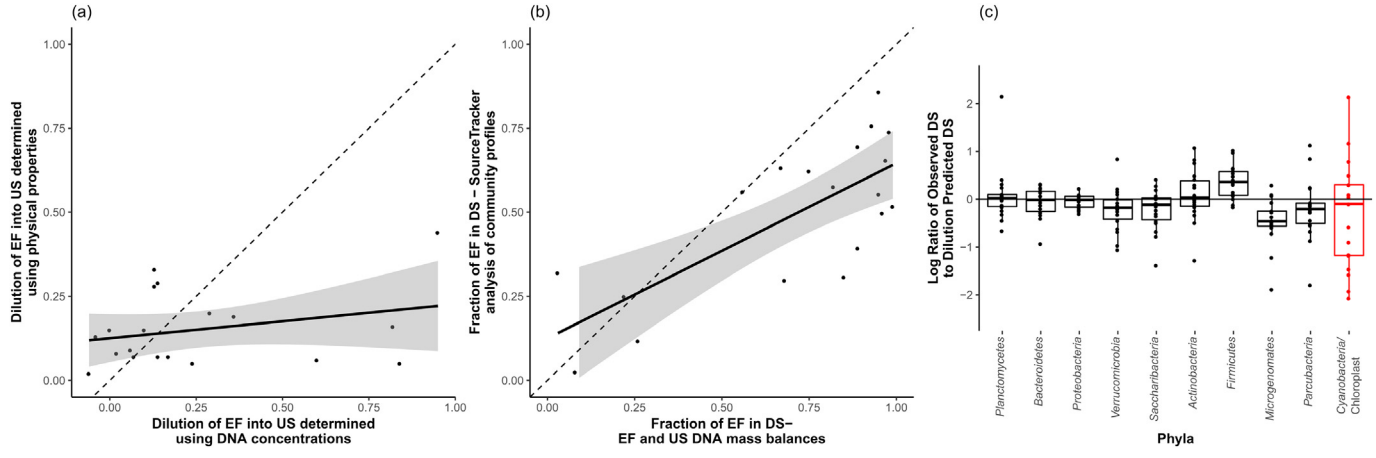


Fig. 4. Comparison of mixing analyses used to predict the composition of the DS 16S rRNA gene profile for (a) the mixing of EF discharged into a stream determined using two DS sample-dependent methods, the stream water chemistry characteristics (D_{chem}) and the measured DNA concentrations (D_{DNA}; Eq. (1)), and (b) the fraction of EF in DS determined using two methods that avoid the bulk DNA concentrations measured DS, SourceTracker2 ($F_{SourceTracker2}$) taking identity of the sequences into account and the D_{chem} corrected DNA mass balances ($F_{Dchem\ corrected\ DNA}$; Eq. (2)). The dashed line indicates the 1:1 line, and the grey shadow indicates the 95% confidence interval of the linear model fit. (c) Box-and-whisker plot identifying the ratio of the observed to the F_{chem} derived mixing-predicted DS abundances of phyla whose median observed value exceeds 100 reads. The box indicates the 75%, median, and 25% interval. Dots appearing above or below the dashed line produced non-finite values and were excluded from the geometric analysis. One phylum was significantly identified as being poorly described (after a false-discovery multiple hypothesis correction) by the mixing predicted DS and is highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

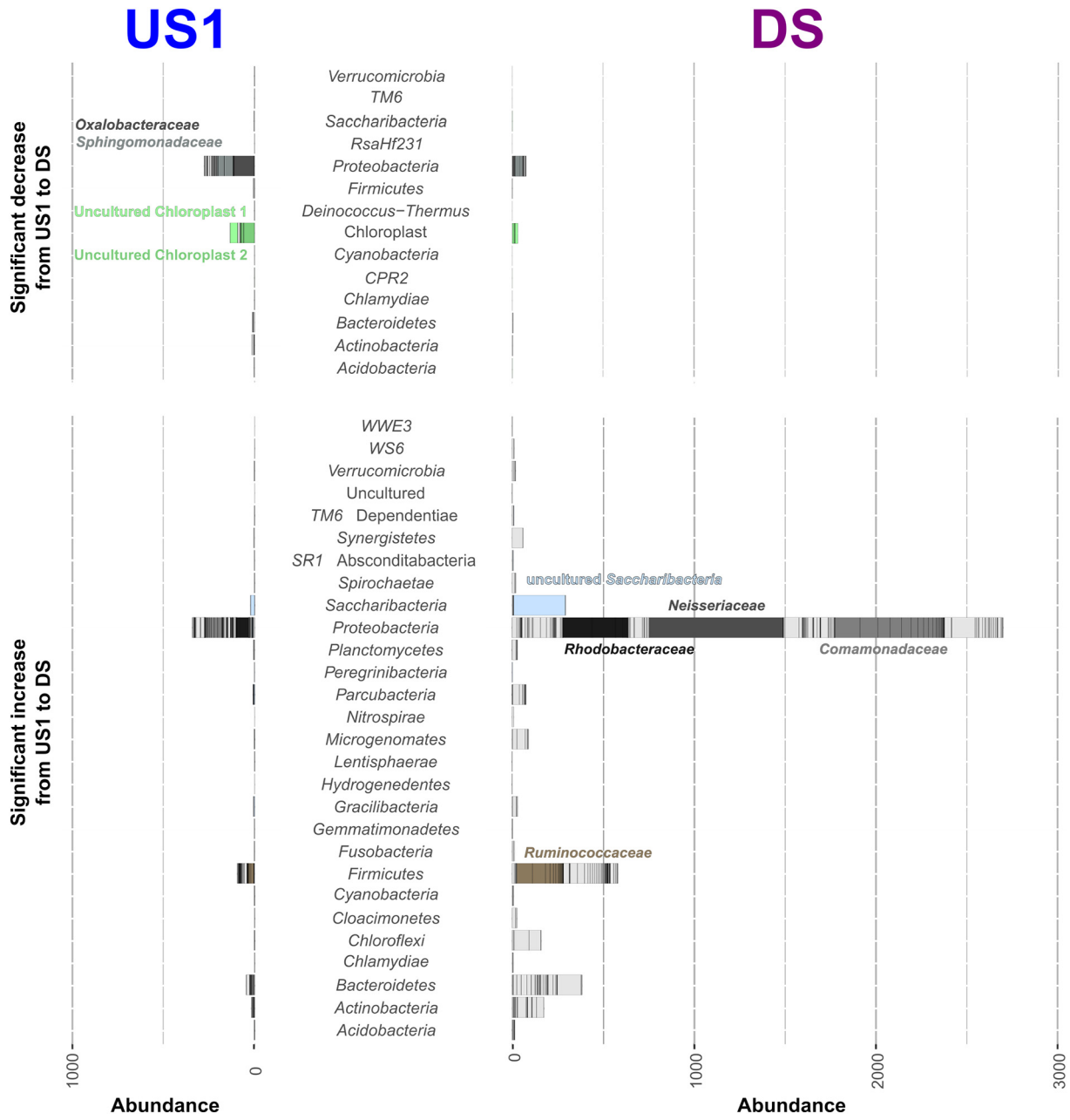


Fig. 5. Stacked bar chart showing the phylum and family composition of groups with significantly decreased or increased downstream (DS) abundance (adjusted p-value <.05, fold-change >2) when compared to upstream (US1) in the DeSeq2 analysis. Taxonomic groups that significantly decreased (top) and increased (bottom) in abundance DS relative to US1 are colour coded to represent families exceeding 5% of the total within increasing or decreasing total abundance. The results of all DeSeq2 comparisons between all locations are provided in Table S7.

microbial taxa (i.e., *Ruminococcus)* and those associated with activated sludge (e.g., *Comamonadaceae*). However, hydrological mixing alone cannot explain the change in relative abundance or occurrence of some naturally occurring 16S rRNA gene sequences. Specifically, we found declines in the naturally occurring 16S rRNA gene sequences associated with Chloroplasts of mainly phototrophic algae (Table S4) and hence the potential of WWTP effluents to alter primary producing microbial communities in stream ecosystems. Bacteria showing an increase in relative abundance downstream include human gut related *Ruminococcus* and are potential indicator groups of surface waters receiving treated wastewater from humans. Thus, changes observed for downstream microbial communities serve as both general bioindicators of anthropogenically influenced streams and indicate a decrease in chloroplasts from primary producing microorganisms, which may alter ecosystem function.

4.1. Taxonomic profiles are indicative of location on stream

The type and read abundance of the phyla among the four locations sampled on each stream are, in general, reflective of a previous survey of the bacterial community composition observed in the water column of a single stream (Zeglin, 2015). A notable distinction observed in this study was the low relative abundance of the *Cyanobacteria* phylum (inclusive of Chloroplasts) in the effluent from the majority of the observed WWTPs (Fig. 2a). The decreased DS relative abundance of chloroplast 16S rRNA gene sequences and that it was not significantly predicted when using the mixing equations, suggests that treated wastewater may shift the relative role of primary production in microbial communities and potentially cause a shift in organelle abundances. This finding potentially extends to all microbial phototrophs because a previous study that monitored the influence of effluent from four wastewater

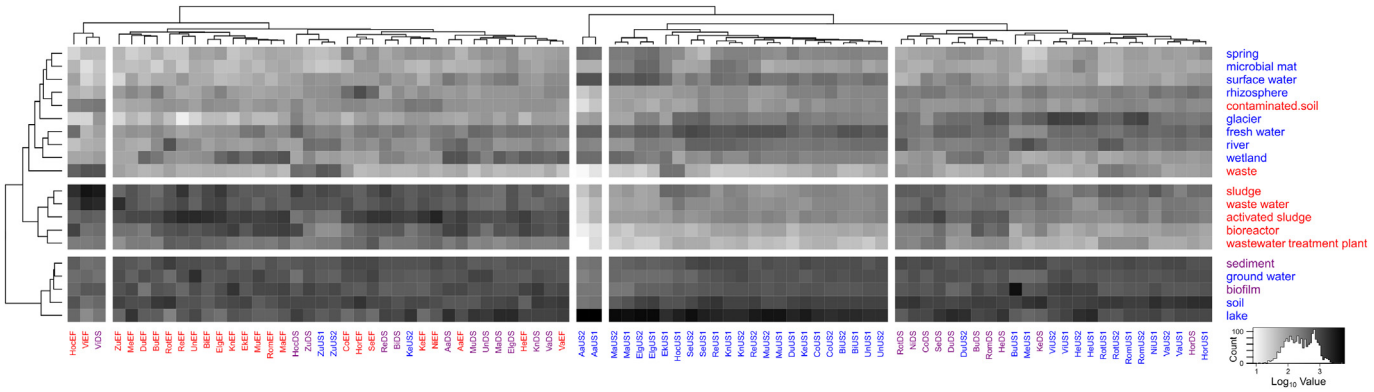


Fig. 6. Clustering of locations and heatmap of read abundance for habitat descriptors identified through SeqENV analysis. Darker squares indicate higher abundance of ESVs associated with habitat descriptor listed on right-hand side with the location of sample indicated below heatmap. Colors of location names follow the same code as the inset of Fig. 1 with US, EF, and DS sites colored blue, red, and purple, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

treatment plants along a river network also reported that the most upstream locations displayed the highest relative abundance of *Cyanobacteria* (Price et al., 2018).

Conversely, the relative abundance of the *Firmicutes* phyla, on average, were highest in the EF and lowest in the US samples, with a mixture displayed in DS samples (Fig. 2a). The *Firmicutes* phylum was characteristic of EF, likely resulting from the dominance of this phylum, along with *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria*, in WWTP activated sludge (Zhang et al., 2011). Additionally, this increase reflects that Firmicutes contain multiple genera that serve as human fecal indicator organisms (e.g., *Enterococcus, Intestinibacter,* and *Roseburia*). Another phylum characteristic of the EF and DS, *Chloroflexi*, was most notable in the EF and DS locations at Villeret (*Vi*) (Fig. S5). Comparatively, the obligate intracellular *Chlamydiae* were more variable, appearing in higher relative abundances in certain US rather than EF locations (i.e., Hornussen (*Hor*), Kernenried (*Ke*), and Rothenthurm (*Rot*)).

When the differences in taxonomic profiles are compared across all sites in the study, the NMDS plots of UniFrac distances show that

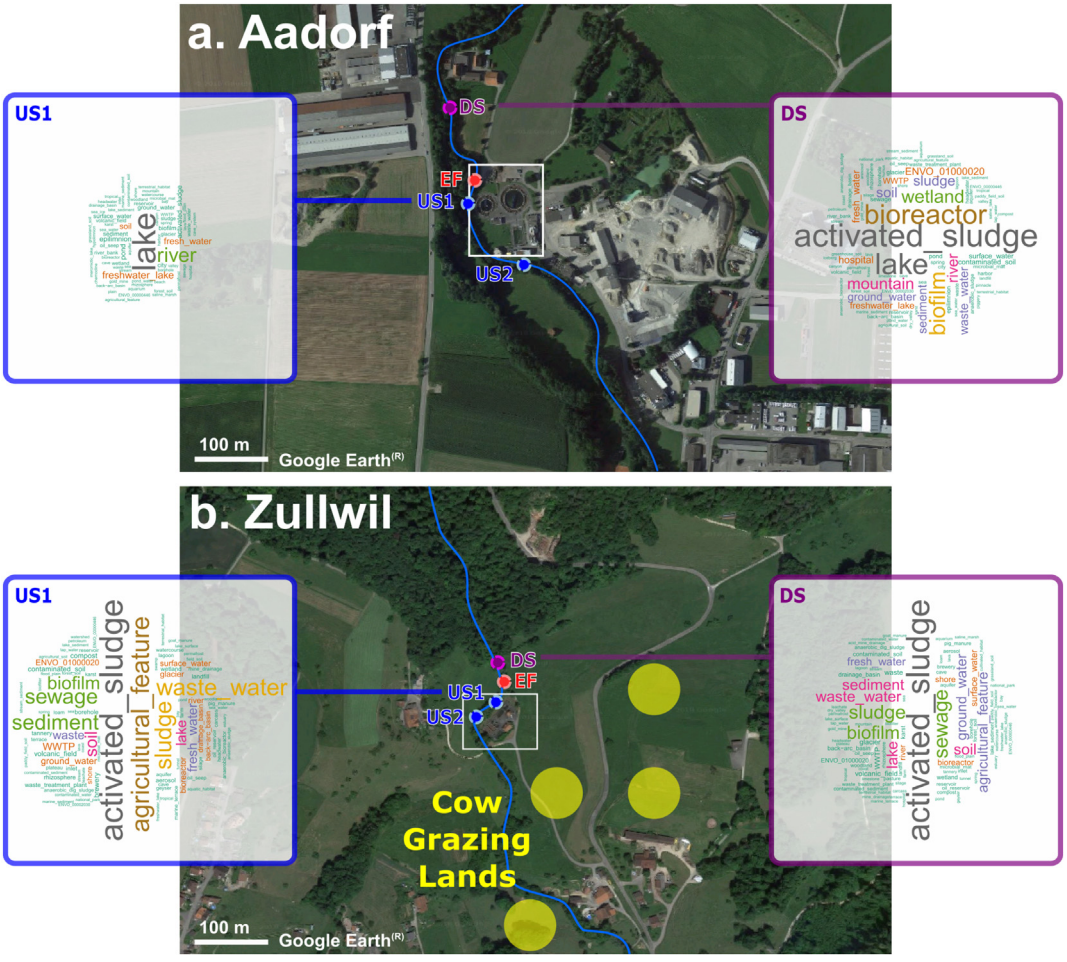


Fig. 7. EnvO term maps (wordclouds) where words are weighted by abundance (i.e., the larger the word, represents more reads were associated with that source) for the isolation sources of the ESVs detected at (a) Aadorf and (b) Zullwil for the US1 and DS locations with aerial images of the two sites (Google Earth).

generally the downstream locations are intermediate between the EF and US communities (Fig. 3). However, abundances of ESVs did shift community similarities in opposite directions for US versus EF communities, indicating that rare taxa are responsible for community dissimilarity in upstream sites, whereas in EF communities shifts in their relative abundance. Notably, the US1 and US2 locations did not exhibit a significant difference in alpha or beta diversity, indicating that the physical distance between locations is not a cause of the shifts observed in the community.

US and DS shifts have also been observed in benthic communities related to effluent (Drury et al., 2013). Monitoring benthic communities allows for describing the diversity profile of bacteria that actively colonize or change in the streams' biofilm, whereas water column bacterial DNA can be from live cells or can be from dead or particle-bound DNA. Thus, by sampling bacteria present in the water column, as is the case in our study, it is not clear which of the effluent taxa may colonize or invade downstream communities. To determine the extent and drivers of community invasion, monitoring the water column community using the analytical and computational approach described in this study could be combined with simultaneous measurements of the biofilm using methods such as glass plate settling (Mußmann et al., 2013) and a larger sampling transect downstream of the effluent release point.

4.2. Hydrological mixing of bacterial communities

Across all sites, the US and EF sources of bacteria accurately predict >95% of the DS bacterial community, supporting the concept that mixing is a dominant process determining downstream bacterial community composition. However, substantial variation in the impact from the EF between sites was noted and was not explained by mixing calculated from chemical parameters. Exceptions such as Zullwil show only minor or no differences in the community profile US vs DS. At these sites, other bacterial sources specific to the surrounding environmental context of animal grazing likely determined the US community profile and in turn directly influence the DS profile, thus minimizing the impact of the EF. Notably, this observation is only possible when the other bacterial sources are similar to the EF and because we had broadly surveyed effluent from 22 additional sites, we could confirm this finding.

A previous study monitoring the community profiles along stretches of a river also displayed profiles in which organisms increased in relative abundance or even disappeared after mixing (Hladilek et al., 2016), suggesting that the underlying hydraulic-mixing properties (e.g., distance and discharge) do not account for all changes in bacterial communities. Along the stream, selected planktonic microorganisms released from the WWTP are known to colonize sediments (Chu et al., 2018) and biofilms (Mußmann et al., 2013), where they can then proliferate and exceed the source contributions, thus further obscuring the mixing profile of the communities. Additionally, the bacterial community structure is also shaped by chemical and physical components of the effluent and their mixing, such as the nitrogen concentration and pH (Ibekwe et al., 2016), adding further controlling variables other than hydraulic mixing. The observed change specific to the phototrophic group of *Cyanobacteria*/Chloroplasts suggest a functional response to the effluent and needs further investigation to understand the consequences of this change for stream ecosystem function.

Alternatively, and not mutually exclusive, is that compositional changes result from technical constraints or biases of methods used in the laboratory or field, producing estimates of community profiles that are not representative of true values in river networks. Specifically, we were constrained in the use of a larger filter pore size (0.7 μm vs 0.2 μm) due to clogging issues when filtering the effluent. To control for clogging issues, we used the same filter pore size for all water samples. Thus, due to this constraint we may have underestimated the total abundance of smaller bacterial cells due to not capturing them on the larger pore size filter. The misrepresentation of community profiles estimated using specific methodologies has been previously highlighted for in silico designed DNA-tracer probes (Abbott et al., 2018; Foppen et al., 2013; Sharma et al., 2012). Although substantial effort was expended to collect informative negative controls within this study and was used to remove any potential contaminations, future studies should expand on site replication and control design by incorporating both temporal sample replicates at each location and apparatus-specific or spike-in positive controls to understand and account for biases introduced through sampling and laboratory methods.

4.3. Community indicators of wastewater and land-use impacted streams

In addition to the richness and relative abundance profiles discussed above, qualitative compositional differences suggested by the habitat descriptors associated with abundant ESVs allowed for detecting wastewater sources in streams (e.g., Fig. 6). Our data demonstrate that associating habitat names with bacterial ESVs allows identifying sources of contamination, highlighting the importance of site-specific influences that can originate from both WWTP and context-dependent land use (Burdon et al., 2019; Burdon et al., 2016). For example, because the sampling locations at the Zullwil site border grazing lands (Fig. 7), agricultural sources of fecal contamination should be further examined, demonstrating a potential local effect. In fact, Zullwil lacks the characteristic Chloroplast associated 16S rRNA sequence abundance typically observed in other upstream locations (Fig. S5). Rather, a species of the genus *Prevotella*, a common indicator of fecal bacteria contamination (Okabe et al., 2007), is uniquely present in high relative abundance upstream. Thus, assessment of bacterial communities using the SeqENV pipeline is a potential method to obtain a first understanding of anthropogenic bacterial contaminations in streams for which no other land-use or detailed information on point source pollution are available.

5. Conclusions

Releasing treated wastewater into natural streams introduces a point source of bacterial pollution in the environment. We demonstrate through a highly replicated study design of sampling 23 natural streams above and below the first WWTP on that stream that bacterial community shifts are predicted based on mixing of effluent bacteria and US sources, where over half of the total 16S rRNA DS community profile originates from the wastewater EF in many of our study sites. Combining our results with that from past studies on antibiotic resistance genes being introduced from WWTPs (Czekalski et al., 2016; Proia et al., 2018; Rodriguez-Mozaz et al., 2015), it is apparent the effluent is altering within-stream microbial communities. The composition of effluent bacteria can be used as a bioindicator of stream influenced by treated wastewater and the functional effects of such alterations observed here is yet unknown. Specifically, we observed that *Cyanobacteria* and Chloroplast associated 16S rRNA sequences decreased in relative abundance downstream from where effluent was released into natural streams and this change was not consistently predicted by mixing. This finding should be complemented with future studies of chromosomal eukaryotic 18S rRNA signatures to determine the eukaryotic community compositions in addition to the organelle associated signal. Overall, two major effects were observed, i.e., the significant increase in abundance of taxa related to human gut bacteria are being introduced into natural streams and signatures of phototrophic microorganisms are decreasing in relative abundance and indicate a potential functional response in microbial communities. The magnitude and mechanisms of functional responses in stream ecosystems to such alterations from wastewater effluent is still unknown and warrants further investigation.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.135727.

Acknowledgements

We thank all people involved in the fieldwork, especially Frank Burdon and Marta Reyes. We additionally thank two anonymous reviewers whom greatly improved the manuscript. Data produced and analyzed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. This work was financially supported by the Swiss National Science Foundation (grants nr. PP00P3_150698, PP00P3_179089, and 31003A_173074 to FA), by EAWAG discretionary funding awarded to KD and FA, and by the European Research Council under the European Union's Seventh Framework Programme (ERC grant agreement no. 614768, PROduCTS for CM).

Declaration of competing interest

The authors declare no conflicts of interest.

Appendix A. List of supplemental files

File S1. Processing the 16S data using Phyloseq.

File S3. Processing the 16S data using Phyloseq for the amplicon sequencing variants version.

File S4. SeqENV analysis.

Compressed File S1. Processed data required for the associated scripts.

Files are available at https://doi.org/10.6084/m9.figshare.c.4540304.v1.

References

Abbott, B.W., Gruau, G., Zarnetske, J.P., Moatar, F., Barbe, L., Thomas, Z., et al., 2018. Unexpected spatial stability of water chemistry in headwater stream networks. Ecol. Lett. 21, 296–308.

Alexander, R.B., Böhlke, J.K., Boyer, E.W., David, M.B., Harvey, J.W., Mulholland, P.J., et al., 2009. Dynamic modeling of nitrogen losses in river networks unravels the coupled effects of hydrological and biogeochemical processes. Biogeochemistry 93, 91–116.

Altermatt, F., 2013. Diversity in riverine metacommunities: a network perspective. Aquat. Ecol. 47, 365–377.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Avilable online at. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., Packmann, A.I., 2016. The ecology and biogeochemistry of stream biofilms. Nat. Rev. Microbiol. 14, 251.

Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., Battin, T.J., 2013. Headwaters are critical reservoirs of microbial diversity for fluvial networks. Proc. R. Soc. B Biol. Sci. 280, 20131760.

Burdon, F.J., Reyes, M., Alder, A.C., Joss, A., Ort, C., Räsänen, K., et al., 2016. Environmental context and magnitude of disturbance influence trait-mediated community responses to wastewater in streams. Ecology and evolution 6, 3923–3939.

Burdon, F., Munz, N., Reyes, M., Focks, A., Joss, A., Räsänen, K., et al., 2019. Agriculture versus wastewater pollution as drivers of macroinvertebrate community structure in streams. Sci. Total Environ. 659, 1256–1265.

Cai, W., Li, Y., Wang, P., Niu, L., Zhang, W., Wang, C., 2016. Revealing the relationship between microbial community structure in natural biofilms and the pollution level in urban rivers: a case study in the Qinhuai River basin, Yangtze River Delta. Water Sci. Technol. 74, 1163–1176.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal 11, 2639.

Chu, B.T., Petrovich, M.L., Chaudhary, A., Wright, D., Murphy, B., Wells, G., et al., 2018. Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. Appl. Environ. Microbiol. 84 (e02168-17).

Crump, B.C., Amaral-Zettler, L.A., Kling, G.W., 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. The Isme Journal 6, 1629.

Czekalski, N., von Gunten, U., Bürgmann, H., 2016. Antibiotikaresistenzen im Wasserkreislauf. Ein Überblick über die Situation in der Schweiz. Aqua & Gas 96, 72–80.

Deiner, K., Walser, J.-C., Mächler, E., Altermatt, F., 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biol. Conserv. 183, 53–63.

Drury, B., Rosi-Marshall, E., Kelly, J.J., 2013. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl. Environ. Microbiol. 79, 1897–1905.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., et al., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. 81, 163–182.

Edgar, R., 2016. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. BioRxiv, 074161 https://doi.org/10.1101/074161.

Foppen, J.W., Seopa, J., Bakobie, N., Bogaard, T., 2013. Development of a methodology for the application of synthetic DNA in stream tracer injection experiments. Water Resour. Res. 49, 5369–5380.

Hamady, M., Lozupone, C., Knight, R., 2010. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. The ISME journal 4, 17.

Harvey, E., Gounand, I., Ward, C.L., Altermatt, F., 2017. Bridging ecology and conservation: from ecological networks to ecosystem function. J. Appl. Ecol. 54, 371–379.

Heino, J., Melo, A.S., Bini, L.M., Altermatt, F., Al-Shami, S.A., Angeler, D.G., et al., 2015. A comparative analysis reveals weak relationships between ecological factors and beta diversity of stream insect metacommunities at two spatial levels. Ecology and Evolution 5, 1235–1248.

Hladilek, M.D., Gaines, K.F., Novak, J.M., Collard, D.A., Johnson, D.B., Canam, T., 2016. Microbial community structure of a freshwater system receiving wastewater effluent. Environ. Monit. Assess. 188, 626.

Ibekwe, A.M., Ma, J., Murinda, S.E., 2016. Bacterial community composition and structure in an Urban River impacted by different pollutant sources. Sci. Total Environ. 566, 1176–1185.

Jackson, M.C., Weyl, O., Altermatt, F., Durance, I., Friberg, N., Dumbrell, A., et al., 2016. Recommendations for the next generation of global freshwater biological monitoring tools. Adv. Ecol. Res. 55, 615–636 Elsevier.

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., et al., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., et al., 2011. Bayesian community-wide culture-independent microbial source tracking. Nat. Methods 8, 761.

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., et al., 2018. Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers. Environmental science & technology 52, 11708–11719.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Magoč, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963.

Marti, E., Balcázar, J.L., 2014. Use of pyrosequencing to explore the benthic bacterial community structure in a river impacted by wastewater treatment plant discharges. Res. Microbiol. 165, 468–471.

McKee, A.M., Spear, S.F., Pierson, T.W., 2015. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biol. Conserv. 183, 70–76.

McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.

McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput. Biol. 10 (4), e1003531.

Mußmann, M., Ribot, M., von Schiller, D., Merbt, S.N., Augspurger, C., Karwautz, C., et al., 2013. Colonization of freshwater biofilms by nitrifying bacteria from activated sludge. FEMS Microbiol. Ecol. 85, 104–115.

Okabe, S., Okayama, N., Savichtcheva, O., Ito, T., 2007. Quantification of host-specific Bacteroides–Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl. Microbiol. Biotechnol. 74, 890–901.

Price, J.R., Ledford, S.H., Ryan, M.O., Toran, L., Sales, C.M., 2018. Wastewater treatment plant effluent introduces recoverable shifts in microbial community composition in receiving streams. Sci. Total Environ. 613, 1104–1116.

Proia, L., Anzil, A., Subirats, J., Borrego, C., Farrè, M., Llorca, M., et al., 2018. Antibiotic resistance along an urban river impacted by treated wastewaters. Sci. Total Environ. 628, 453–466.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., et al., 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. Sci. Total Environ. 447, 345–360.

Rodriguez-Mozaz, S., Chamorro, S., Marti, E., Huerta, B., Gros, M., Sànchez-Melsió, A., et al., 2015. Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. Water Res. 69, 234–242.

Ruiz-González, C., Niño-García, J.P., del Giorgio, P.A., 2015. Terrestrial origin of bacterial communities in complex boreal freshwater networks. Ecol. Lett. 18, 1198–1206.

Savio, D., Sinclair, L., Ijaz, U.Z., Parajka, J., Reischer, G.H., Stadler, P., et al., 2015. Bacterial diversity along a 2600 km river continuum. Environ. Microbiol. 17, 4994–5007.

Sharma, A.N., Luo, D., Walter, M.T., 2012. Hydrological tracers using nanobiotechnology: proof of concept. Environmental science & technology 46, 8928–8936.

Sinclair, L., Ijaz, U.Z., Jensen, L.J., Coolen, M.J., Gubry-Rangin, C., Chroňáková, A., et al., 2016. Seqenv: linking sequences to environments through text mining. PeerJ 4, e2690.

Staley, C., Unno, T., Gould, T., Jarvis, B., Phillips, J., Cotner, J., et al., 2013. Application of Illumina next-generation sequencing to characterize the bacterial community of the upper Mississippi River. J. Appl. Microbiol. 115, 1147–1158.

Stamm, C., Räsänen, K., Burdon, F.J., Altermatt, F., Jokela, J., Joss, A., et al., 2016. Unravelling the impacts of micropollutants in aquatic ecosystems: interdisciplinary studies at the interface of large-scale ecology. Adv. Ecol. Res. 55, 183–223 Elsevier.

Tan, B., Ng, C., Nshimyimana, J., Loh, L.-L., Gin, K., Thompson, J., 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Front. Microbiol. 6.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37, 130–137.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., et al., 2010. Global threats to human water security and river biodiversity. Nature 467, 555.

Wang, P., Wang, X., Wang, C., Miao, L., Hou, J., Yuan, Q., 2017. Shift in bacterioplankton diversity and structure: influence of anthropogenic disturbances along the Yarlung Tsangpo River on the Tibetan Plateau, China. Sci. Rep. 7, 12529.

Woodward, G., Gessner, M.O., Giller, P.S., Gulis, V., Hladyz, S., Lecerf, A., et al., 2012. Continental-scale effects of nutrient pollution on stream ecosystem functioning. Science 336, 1438.
Zeglin, L.H., 2015. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. Front. Microbiol. 6.
Zhang, T., Shao, M.-F., Ye, L., 2011. 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. The Isme Journal 6, 1137.
Zhao, Y., Xia, Y., Ti, C., Shan, J., Li, B., Xia, L., et al., 2015. Nitrogen removal capacity of the river network in a high nitrogen loading region. Environmental Science & Technology 49, 1427–1435.