



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



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

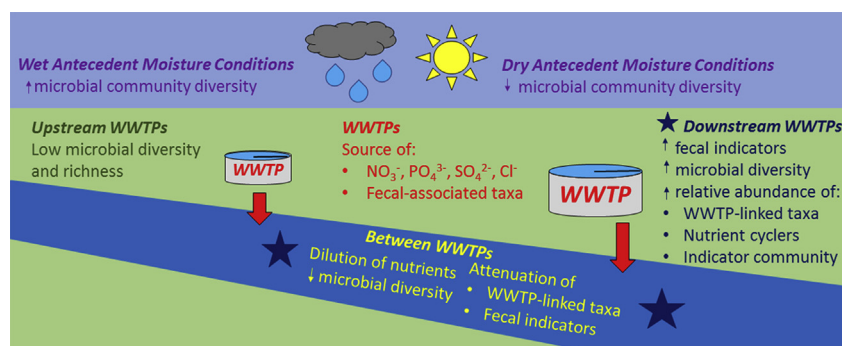
^a Civil, Architectural, and Environmental Engineering, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, United States

^b Earth and Environmental Science, Temple University, 1901 N. 13th St, Philadelphia, PA 19122, United States

HIGHLIGHTS

- Effluent affected diversity and structure of community downstream of WWTPs.
- Effluent-impacts on community composition changed with AMC.
- WWTP-associated taxa significantly decreased with distance from source.
- Major nutrients (N and P) did not control shifts in community structure.
- Efficacy of using a microbial indicator subset was verified.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 28 June 2017

Received in revised form 30 August 2017

Accepted 16 September 2017

Available online xxx

Editor: D. Barcelo

Keywords:

Amplicon sequencing
Microbial community analysis
Water chemistry
Urban stream
Nutrients

ABSTRACT

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

© 2017 Elsevier B.V. All rights reserved.

Abbreviations: AMC, antecedent moisture conditions; MGD, million gallons per day; OTU, operational taxonomic unit; qPCR, real-time polymerase chain reaction; RSV, ribosomal sequence variant; VST, variance stabilizing transformation; WWTP, wastewater treatment plant.

* Corresponding author at: Civil, Architectural, and Environmental Engineering, Drexel University, United States.

E-mail address: chris.sales@drexel.edu (C.M. Sales).

1. Introduction

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

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

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

Most previous studies that use environmental DNA (eDNA) to investigate the impact of WWTP effluent on streams have focused specifically on ammonia-oxidizing assemblages (Merbt et al., 2015; Sonthiphand et al., 2013), denitrifying bacterial communities (Rahm et al., 2016), sediment bacterial communities (Drury et al., 2013), or focus on different pollutant sources (Ibekwe et al., 2016). In this study, we aim to integrate information from multiple sources including traditional stream water chemistry, real-time polymerase chain reaction (qPCR),

and targeted amplicon sequencing to get a more complete, though qualitative due to limited sample size, microbial overview of the impact of WWTP discharge on urban streams. To do this we targeted an urban stream, Wissahickon Creek, outside of Philadelphia, PA, in an urban and suburban setting with four WWTPs discharging effluent. Water quality in the Wissahickon is important to the City of Philadelphia as one of the city's drinking water intake pipes is located approximately half of a mile downstream of the confluence of the Schuylkill River and Wissahickon Creek (PWD, 2007). Samples were collected above and below the plants on two dates and analyzed for water chemistry, qPCR, and targeted amplicon sequencing.

2. Materials and methods

2.1. Site description

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

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

2.2. Field sampling and water quality analysis

Samples were collected from each site on the mornings of May 10 and May 17, 2016. While all samples were collected at baseflow, antecedent moisture conditions (AMC) were wetter on the May 10, 2016 (Fig. S1). The antecedent precipitation index (API) can be used to contrast the two dates, where $API = \sum_{t=-1}^{-i} P_t k^{-t}$ (Ali et al., 2010). Taking t as the 10 days prior, P as the amount of precipitation on each day, and using a recession constant, k , of 0.9, May 10 has an API of 110 and May 17 has an API of only 12. Water samples for chemical analysis were collected from the thalweg in two acid-washed HDPE bottles,

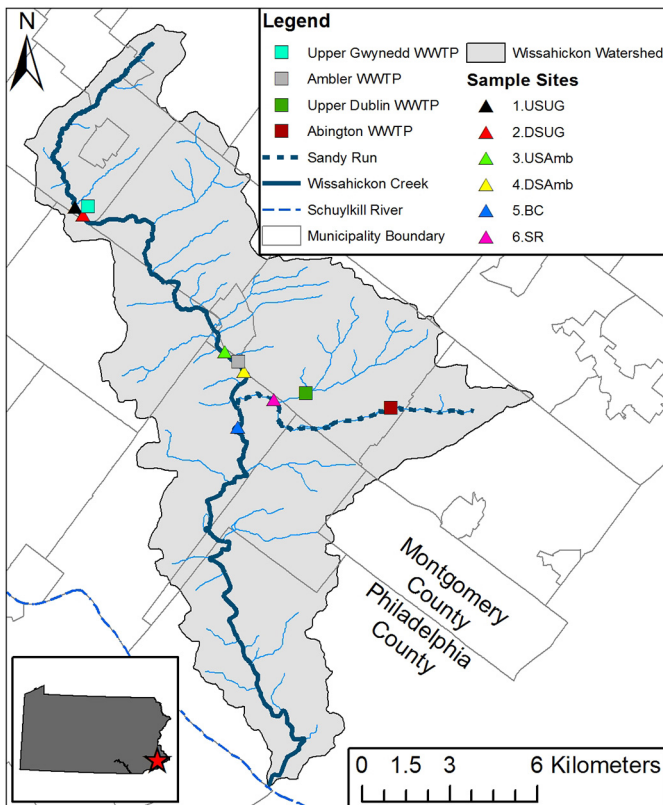


Fig. 1. Wissahickon Creek is a third-order stream that discharges into the Schuylkill River in Montgomery and Philadelphia counties, Pennsylvania. There are four wastewater treatment plants (WWTPs) on the stream, and samples were taken above and below the main stem plants (1.USUG, 2.DSUG, 3.USAmb, and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

after being filtered through 0.45 μm Millipore cellulose filters. One bottle was acidified to approximately $\text{pH} = 2$ with nitric acid and neither bottle had head space. Samples were stored at 4 °C and analyzed within one week. Ions and elements were chosen for analysis due to their common occurrence in natural waters (SO_4^{2-} , Cl^- , NO_3^- , Ca, Mg, Na, and K), potential to be tracers of wastewater (F^- , Sr, Cu, and Mn), or known to be necessary for microbial growth (Fe, NO_2^- , TDP, and Si). Fluoride (F^-), chloride (Cl^-), bromide (Br^-), nitrite (NO_2^- -N), nitrate (NO_3^- -N), and sulfate (SO_4^{2-}) in the non-acidified samples were analyzed on a Dionex ICS-1000 ion chromatograph using three standards, while calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), total iron (Fe), total dissolved phosphorus (TDP), silicon (Si), copper (Cu), manganese (Mn), and strontium (Sr) were measured in elemental quantities in the acidified samples on a Thermo Scientific iCAP 7000 ICP-OES using three standards. Previous work in the watershed indicated that at baseflow orthophosphate comprised all of the TDP and there was minimal particulate phosphorus in the stream water (data not shown). National Pollutant Discharge Elimination System (NPDES) permit reporting data was available from Upper Gwynedd WWTP for orthophosphate on both sample days. NPDES nutrient data for the other plants were not available on the sampling days so an average of WWTP effluent collected and analyzed as described above on three different dates (August 28, 2016; October 15, 2016; February 25, 2017) was reported. A USGS gage, station number 01473900, existed at the 5.BC site and was used for discharge information (Fig. S1).

Water samples were also collected to characterize the microbial community in the streams. Three 1 L replicate samples were collected at each site in autoclaved polypropylene bottles. Samples were collected from the thalweg, facing upstream, and while wearing gloves. The

bottles were rinsed with stream water three times prior to sample collection to remove any carryover that may have occurred during cleaning or in transit. Samples were transported to the laboratory within 6 h for DNA extraction.

2.3. DNA extraction and quantification

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

2.4. Real-time PCR (qPCR)

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

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

2.5. Amplicon library preparation and sequencing

Amplicon libraries, targeting 16S rDNA, were created for each of the samples following the Earth Microbiome Project's (Gilbert et al., 2010) 16S rRNA amplification protocol (Caporaso et al., 2012). Primers 515F-Y (Parada et al., 2015; Quince et al., 2011) and 926R (Parada et al., 2015; Quince et al., 2011) were selected to provide coverage of two hypervariable regions (V4 and V5) within the 16s gene. PCR reactions for amplicon library preparations were carried out using a single PCR reaction for each extracted DNA aliquot. Each 25 μL PCR reaction volume was comprised of 12.5 μL HotStarTaq Plus Master Mix, 1.5 μL of primer pre-mix containing both forward and reverse primers at a concentration of 5 μM , 1 μL of molecular biology-grade water, and 10 μL of gDNA. The thermocycler program entailed a 5 min at 95 °C heat-activation step; 25 replication cycles of 94 °C for 45 s, 50 °C for 45 s,

and 72 °C for 90 s; a final extension was carried out at 72 °C for 5 min. PCR triplicates, three aliquots per sample, were pooled after the amplification step, then 4 µL from each of the 12 pools were then run on a 1% agarose gel to test for amplification and correct amplicon product size. The pools were purified and concentrated with a Qiagen QIAquick Purification kit and the product quantified via a QuBit 2.0 fluorometer. Equal product masses for each sample were combined into a single tube and submitted for paired-end sequencing (2 × 300 bp) on an Illumina MiSeq Sequencer. Raw read files were uploaded to the National Center for Biotechnology Information (NCBI) Sequencing Read Archive (SRA) and are accessible under the SRA Study Accession ID SRP103534 and BioProject Accession ID PRJNA382371; Table S4 contains a list of sample names and their corresponding BioSample Accession IDs.

2.6. Bioinformatics pipeline

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

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

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

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

2.7. Statistical analysis

Preprocessing and filtering of the phyloseq object was carried out to reduce error and noise in the dataset by removing taxa that met any of the following three criteria: 1) taxa with zero counts; 2) taxa with ambiguous taxonomic assignment at the Kingdom or Phylum level; 3) taxa that were seen <2 times in 2 or more samples.

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

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

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

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

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

3. Results

3.1. Downstream dilution in water chemistry

Nitrate concentrations in the surface water ranged from 0.9 at 1.USUG on May 17, 2016 to 10.2 mg N/L at 2.DSUG on May 17, 2016, while average WWTP effluent concentrations were much higher, averaging 17.1 mg N/L at Abington to 25.6 mg N/L at Upper Gwynedd

(Fig. 2A). WWTP discharge on the main stem resulted in an increase in nitrate in surface waters, but the longer distance between the treatment plants on Sandy Run and the sampling site allowed for retention or removal of nitrate before 6.SR. The decrease in concentration between 2.DSUG and 3.USAmb is indicative of a combination of dilution from incoming tributaries and groundwater, assimilatory nitrate uptake, and denitrification along this reach. The decrease with distance continued below the confluence of Sandy Run with the main stem on both dates, but is primarily controlled by mixing between the two sources. May 17, 2016 had higher surface water nitrate concentrations, most likely due to lower discharge and drier AMC (Fig. S1). Although there is some variation in nutrient concentrations coming from plant effluent, this system has been in equilibrium for a while (none of the infrastructure is new or has changed in the recent past). Thus, the system is likely at relative temporal equilibrium.

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

Chloride concentrations in the stream increased below each WWTP outfall (Fig. S2). The cities contributing to the treatment plants do not fluoridate their drinking water, explaining the lack of increase in fluoride below effluent outfalls. Silicon increases between 2.DSUG and 3.USAmb, where small tributaries flowing into the main stem have been observed to have silicon concentrations as high as 8.7 mg/L. Ambler is the only WWTP that is a source of silicon to the stream. Silicon concentrations decrease with drier antecedent moisture conditions. In

the upstream reach, strontium concentrations are controlled by Upper Gwynedd, which has high effluent concentrations. Below this plant, tributaries have varying strontium concentrations, from 0.18 to 0.51 mg/L, resulting in a small decrease in concentration with distance. Strontium concentrations increase with dry antecedent moisture conditions due to concentrative effects. Sulfate concentrations are controlled by Upper Gwynedd, which has very high (>400 mg/L) effluent concentrations. This causes a large increase in sulfate at 2.DSUG, and the rest of the distance along the stream is dilution of this signal as discharge increases.

3.2. Human indicators in real-time PCR

As anticipated, the real-time PCR (qPCR) results revealed elevated concentrations of *B. dorei* at sites located immediately downstream of wastewater effluent outflows (2.DSUG, 4.DSAmb) in comparison to their upstream counterparts (1.USUG, 3.USAmb) (Fig. 3A). *Bacteroides* spp. have been frequently used in culture-independent, 16S rRNA gene based methods for microbial source tracking (MST) (Ahmed et al., 2008; Bae and Wuertz, 2012; Bernhard and Field, 2000; Shanks et al., 2007). Ryan et al. (2013) showed the *B. dorei* specie assay to be highly specific to human host pollution. A universal/non-human-specific *Bacteroides* qPCR protocol was also used to determine total enteric pollution. The non-human-specific assay mirrors the general trend shown by the human specific assay (Fig. 3B). Welch two sample *t*-test's (with unequal variances) confirmed that the abundance of *B. dorei* ($p < 0.001$) and *Bacteroides* spp. ($p < 0.001$) were significantly higher in the downstream sites (2.DSUG, 4.DSAmb in comparison to 1.USUG, 3.USAmb). The 5.BC site shows the influence of the 6.SR input in that there are higher concentrations during dry weather. However, during wet weather, there was little observed influence from the 6.SR input to the mainstem. While both the Upper Gwynedd and Ambler outflows impacted *Bacteroides* spp. abundances, Upper Gwynedd had a larger effect. These results are most easily explained by contribution of wastewater effluent to the stream flows due to the differences in volumetric flow rates at each of these sites; both sites have similar average daily discharge (Upper Gwynedd = 2.31 MGD and Ambler = 3.64 MGD, Table S1), but Upper Gwynedd is located in the headwaters of the stream, where effluent makes up a larger percentage of total flow at baseflow. The Wissahickon has somewhat higher discharge at Ambler due to approximately six small tributaries that flow into the main

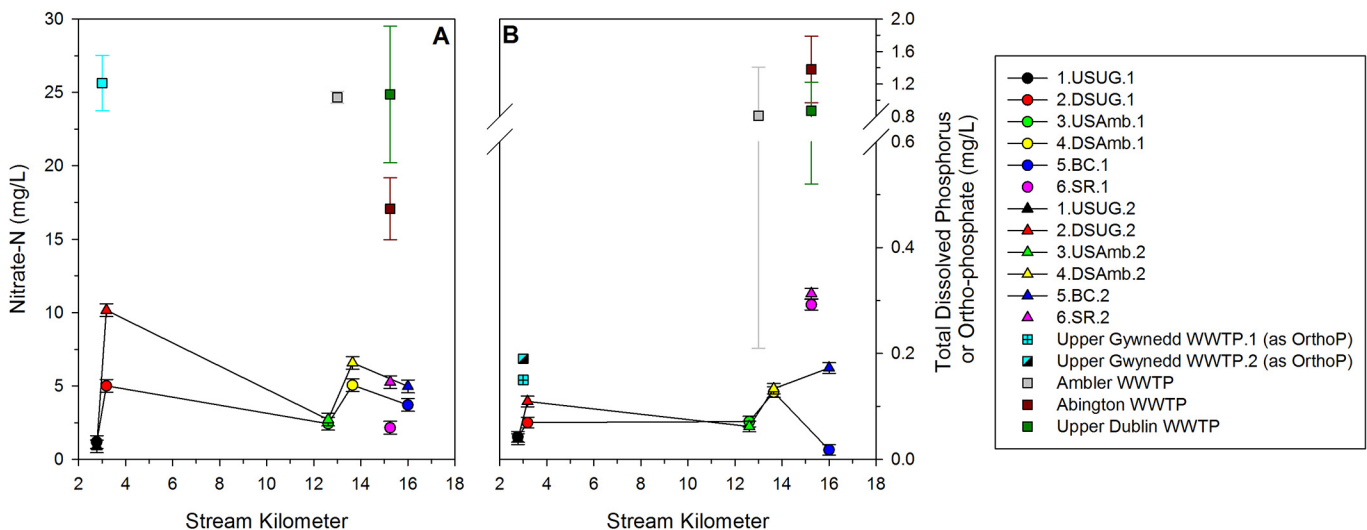


Fig. 2. (A) Nitrate-N concentrations and (B) total dissolved phosphorus and orthophosphate concentrations with distance along the stream relative to the stream's headwaters. Upper Gwynedd WWTP samples from NPDES report on the two sampling dates are reported as orthophosphate and all other samples are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

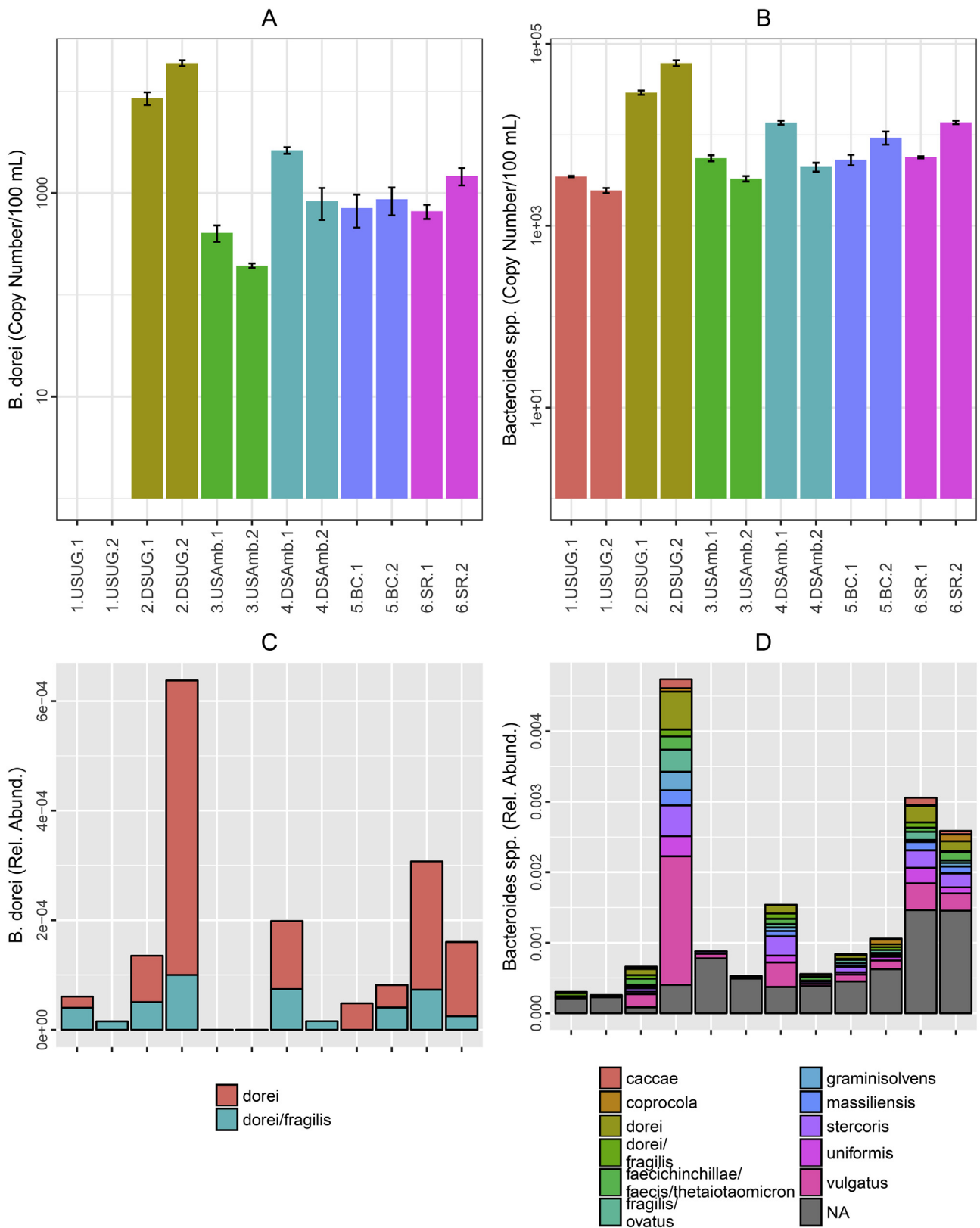


Fig. 3. Observed abundances of human-source fecal indicators (*B. dorei*) and non-specific fecal indicators (*Bacteroides* spp.): The top panels present the Copy Number (CN) concentration of *B. dorei* (A) and *Bacteroides* spp. (B), as determined by qPCR. The bottom panels present the relative abundances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

stem during baseflow between Upper Gwynedd and Ambler, plus an unquantified addition of groundwater. Furthermore, the 2.DSUG site was closer to the Upper Gwynedd WWTP than the other downstream sites (200 m downstream in contrast to 600 m and 1 to 5 km downstream for the 4.DSAmb and 5.SR sites). Antecedent moisture conditions seem to have the strongest impact at 2.DSUG, with higher copy numbers during drier conditions, *Bacteroides* spp. shows the same temporal increase at 5.BC and 6.SR, although to a lesser degree.

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

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

Comparing upstream/downstream pairs (1.USUG and 2.DSUG; 3.USAmb and 4.DSAmb), effluent sources appear to increase diversity within the stream (Fig. 4, Table S6, Chao1 diversity values). Averaging both days, the Upper Gwynedd WWTP appears to have a larger impact on diversity increasing the stream's alpha diversity by 628 taxa, while the Ambler WWTP caused an increase of 113 taxa. A second pattern emerges when comparing the alpha diversity between the two WWTPs (between 2.DSUG 3.USAmb), where diversity decreases by approximately 854 taxa between the two effluent sources. This reduction in diversity, 854 taxa between 2.DSUG and 3.USAmb, exceeds the additional 628 taxa contributed to the stream by the Upper Gwynedd WWTP, indicating that, even with the additional taxa being contributed to the stream by the wastewater effluent, diversity was lost as a result of the effluent source. Samples from Sandy Run (6.SR) had some of the

highest diversities for all three measures. This high diversity appears to account for the increase in diversity between 4.Amb and 5.BC.

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

A total of 5539 RSVs were present in the 12 samples prior to any pre-processing or filtering steps, including 44 singletons. Un-annotated taxa, to some degree, are not unexpected in environmental samples, but the absence of taxonomic assignment makes it difficult to interpret results. To avoid this, RSVs that were not annotated at the kingdom or phylum level (481 RSVs) were removed from the dataset. In an effort to reduce noise and error, only taxa seen more than twice in two independent samples were kept in the final dataset; 3925 RSV's remained in the final set.

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

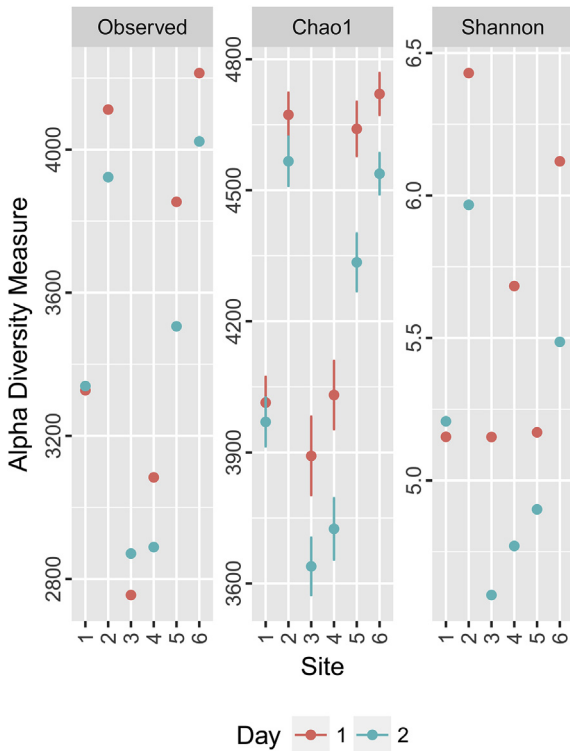


Fig. 4. Alpha Diversity Measures for the full microbial community. Samples are grouped by their site number, found in the first character in the sample name in Fig. 1. Error bars for Chao1 represent standard error (SE).

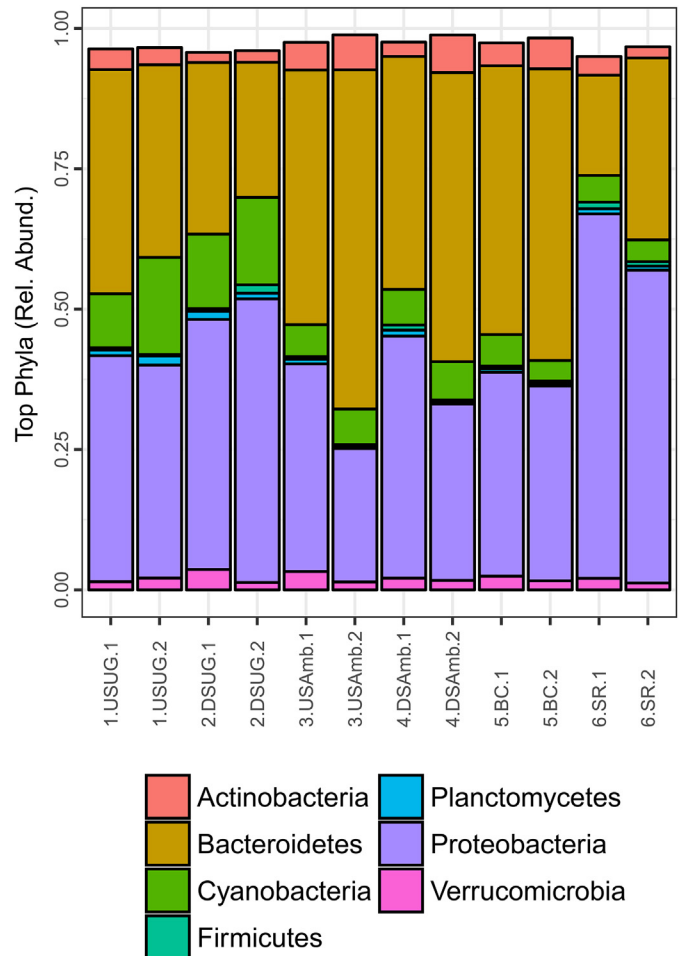


Fig. 5. Barcharts representing the relative abundance of the 7 most abundant Phyla in the microbial community.

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

3.4. Contrasts between wet and dry weather samples using exploratory ordination

Shifts in sample placement are observed in the PCA plot, indicating relative influence of WWTP inputs (Fig. 6). Within the PCA plot, the influence of WWTP effluent outflows appeared to shift the sample placement in the same direction (to the left and upwards). The wet and dry conditions shifted about the same at Upper Gwynedd but the shift was lower under dry conditions for Ambler. The influence of Sandy Run also changes with antecedent moisture conditions. The 5.BC.2 taxa fell between 6.SR.2 and 4.DSAmb.2 in dry conditions, suggesting a mixing relationship. In wet conditions, there was no such mixing relationship and the 5.BC.1 sample falls to the right of 4.DSAmb.1. The 4.DSAmb sample drives this shift and it differs between the two sampling dates although they are only 7 days apart. Microbial cohorts within the three most abundant phyla drove the distinctions between samples, as shown in double principal coordinate analysis (DPCoA) (Fig. S4), and, as in Fig. 6, sample placement appears to be at least partly determined by the influence of WWTP effluent outflows.

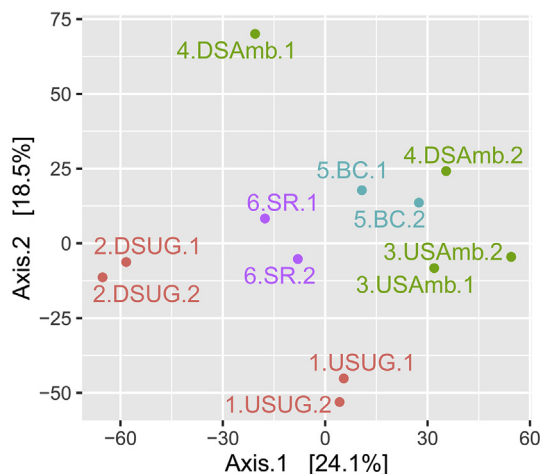


Fig. 6. A plot of the PCA results for the complete microbial community.

3.5. Differential abundance testing

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

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

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

Some of the taxa with the largest increases in abundance are related to wastewater treatment. For example *Zoogloea ramigera* has been observed to comprise up to 10% of the bacterial community within activated sludge systems (Rossello-Mora et al., 1995), and *Zoogloea caeni* was

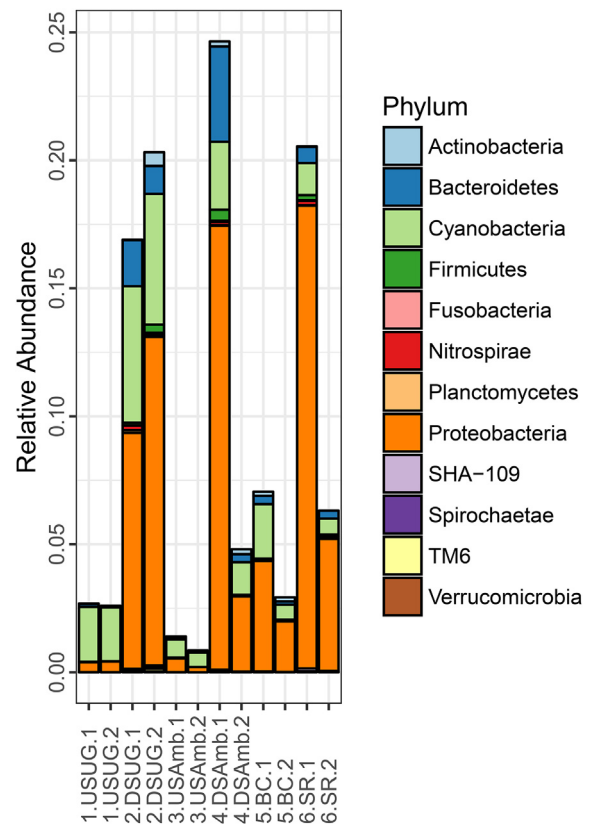


Fig. 7. Relative abundance of differentially abundant taxa, observed during the analysis of the whole microbial community. Samples from sampling sites 5.BC and 6.SR are included only for comparison and were not included in the differential abundance testing.

first isolated from activated sludge systems (Shao et al., 2009). *Malikia granosa* was also first isolated from activated sludge, and has been shown to be a phosphate-accumulating bacterium (Spring et al., 2005).

Wastewater effluent is known to contain excess nutrients, which can influence the health and community structure of receiving streams (Carey and Migliaccio, 2009; Ibeke et al., 2016; Merseburger et al., 2005). In this study, this manifested itself in the form of increased abundance of nutrient cycling organisms within the stream. Perhaps most notably, abundances of *Nitrosomonas* spp., an ammonia oxidizing bacteria (AOB), had an observed log₂-fold change of 4.92 in downstream of effluent sources, while nitrite oxidizing bacteria (NOB), including *Nitrospira defluvii* and another *Nitrospira* spp., were observed to have log₂-fold changes of approximately 3–4. Both *Nitrosomonas* and *Nitrospira* can be found in WWTPs carrying out nitrification/denitrification steps to remove excess nitrogen (Saunders et al., 2016). Several of the taxa have been confirmed to be active denitrifiers within WWTPs, including *Dechloromonas* sp. (1 taxa), *Haliangium* sp. (5 taxa), *Thermomonas brevis* (1 taxa) (McIlroy et al., 2016). For the current study, it is unclear whether the increase in nitrifiers is due to their addition by WWTP effluent (as active or dead biomass). Lastly, 8 taxa falling within Cyanobacteria were found to be differentially abundant; 7 of the 8 were found to have an average log₂-fold change of 3.4, while one was found to decrease in abundance. An increase in Cyanobacteria could be attributed to the influx of nutrients from the wastewater effluent (Vaquer-Sunyer et al., 2016).

In addition to WWTP and nutrient cycling bacteria, 11 of the 137 differentially abundant RSVs fall into two of the three Orders designated as indicators: 5 from Bacteroidales and 6 from Clostridiales. Two of the differentially abundant taxa fall within the *Bacteroides* genus. Differentially abundant Bacteroidales taxa had an average log₂-fold change of 4.32, while Clostridiales taxa had an average log₂-fold change of 4.67.

3.5.2. Differential abundance within the microbial community with distance below Upper Gwynedd WWTP

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

The phyla of Actinobacteria (21), Bacteroidetes (55), Cyanobacteria (34), and Proteobacteria (77) account for the majority of these differentially abundant taxa. Roughly half of the differentially abundant taxa had higher abundance at 3.USAmb (102, positive log₂-fold change in Fig. S6), and the remaining taxa were lower in abundance (103, negative log₂-fold change in Fig. S6). The majority of Actinobacteria (19 of 21) increased in abundance, while taxa falling within Bacteroidetes had mixed responses with 36 (of 55 differentially abundant) taxa having higher abundance and the remaining 19 with lower abundance. The Actinobacteria and Bacteroidetes clusters occur on opposing ends of the first DPCoA axis (Fig. S4). The mixed signals in Bacteroidetes response may explain why the transitions along the first DPCoA axis between the two sampling days (2.DSUG.1 to 3.USAmb.1, and 2.DSUG.2 to 3.USAmb.2) are inconsistent. Taxa falling within Cyanobacteria also showed mixed signals with 20 taxa (of 34) having increased abundance and the remaining 14 differentially abundant Cyanobacteria taxa having decreased abundance. Proteobacteria make up the majority of taxa (56 of the 103) with decreased abundance at 3.USAmb (although an additional 21 Proteobacteria were observed to increase in abundance), and may explain the shifts along the second DPCoA axis in Fig. S4. The abundance of the WWTP-linked taxa *M. granosa*, and three *Zoogloea* species were found to be significantly lower at 3.USAmb than at 2.DSUG, as

were fecal indicators *B. dorei* and *B. vulgatus*, two nitrite oxidizing bacteria belonging to the *Nitrospira* genus, as well as 5 *Haliangium* species.

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

3.6. Environmental influences on the composition of the microbial community

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

Stream water chemistry composition was related to community structure using distance-based redundancy analysis (db-RDA) (Legendre and Anderson, 1999; McArdle and Anderson, 2001) with Bray-Curtis distances (Bray and Curtis, 1957). The final db-RDA model included the effects of Cl⁻, Si, Sr, and F⁻ (all log-scaled), had an adjusted r² value of 0.3351 (unadjusted r² value of 0.5769), and was found to be statistically significant at the 5% significance level ($p = 0.001$); each of the individual terms were also found to be significant when analyzed separately (Cl⁻ $p = 0.001$, Si $p = 0.001$, Sr $p = 0.002$, F⁻ $p = 0.026$). As in previous ordinations, displacement between upstream/downstream sample pairs occurred over both axes (Fig. S7), and the db-RDA model indicates that this displacement is correlated with the concentration of Cl⁻. The relationship between chemical dilution and change in microbial community indicate what portion of the variation can be related to dilution.

3.6.2. BIOENV procedure

The BIOENV procedure (Clarke and Ainsworth, 1993) was applied to identify environmental variables or combination of variables as an alternative method to forward selection db-RDA. The optimal correlational model identified for the full microbial community included (log-transformed) Cl⁻ and SO₄²⁻ and had a BIOENV correlation of 0.6003 (Table S8). Cl⁻, was determined to be the best single parameter for explaining variation in OTU abundance, and appeared in all of the investigated models (Table S8).

3.7. Indicator community results

In addition to investigating the overall microbial community, RSVs within the orders of Bacteroidales, Bifidobacteriales, and Clostridiales were selected for additional study due to their association with human fecal contamination (Halliday et al., 2014; McLellan et al., 2010). The results of this round of analysis mirror those found for the full microbial community. In brief, alpha diversity within these three orders were observed to be higher in samples collected immediately downstream of effluent sources than those collected upstream (Fig. S8) and that the relative abundance of these taxa also increased up to downstream (Fig. S9), indicating that effluent outflows are sources of new species found within these three orders. Roughly 15% of these taxa were found to be significantly higher in abundance at downstream sites (Fig. S12). These differentially abundant taxa showed similar, but more variable, patterns in changes in relative abundance (Fig. S13, Table S7) to those observed for the whole microbial community (Fig. 7, Table S7). In this study, we used forward variable selection to build db-RDA models and applied the BIOENV procedure to identify the chemical parameters that have the most significant correlation with the microbial community structure. The db-RDA models had adjusted r² values of 33–35% (unadjusted r² of 57.69% for the full microbial community and 52.52% for the indicator community). BIOENV produced models that resulted in unadjusted r² values of 60.03% and 68.04% for

the full and indicator communities respectively (Table S8 and Table S9). Full details of the results from this round of analysis are presented in Section 1 of the Supplemental Material.

4. Discussion

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

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

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

These microbial community alterations may be affected by antecedent moisture conditions (AMC). Analysis of storm events in this watershed show that nutrient concentrations recover within one to two days of precipitation events, with overland flow contributing nutrients to tributaries with less riparian buffering than the main stem. In contrast, water level remains high for three to four days after rain, potentially from storage in saturated and unsaturated storage zones slowly draining to the stream, resulting in higher water levels without high nutrient concentrations. We expected a similar response in microbial communities, with lower diversities a few days after precipitation;

however, samples collected after wet AMC showed an increase in microbial community diversity, while those collected after dry conditions showed a decrease in the relative abundance of microbes. The effect of AMC on the microbial communities, however, does seem to be impacted by the discharge at the site. Both 1.USUG and 2.DSUG, where discharge is low, are less impacted by the shift in AMC, while the other sites, which have higher discharge, have larger shifts between sampling dates and AMC. Although a contrast in AMC was observed on the two sample dates only one week apart and collected at the same time of day, the future collection and analysis of samples representing a range of AMC and stream discharge values may enable these trends to be better explored.

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

4.2. Attenuation with distance below WWTPs

Two distinct patterns emerged when investigating microbial community composition with distance downstream of the Upper Gwynedd WWTP. The first was that, despite the instantaneous increase in diversity below effluent outflows, most likely the result of new taxa being introduced into the stream by the effluent, dramatic reductions in diversity were observed between 2.DSUG and 3.USAmb on both sampling dates. The second pattern that emerged was that roughly 40% (53 of 137) of the taxa found at higher abundances downstream of effluent sites decreased in abundance over the course of the 10 km that separated 2.DSUG and 3.USAmb (Fig. 1). Therefore, while these effluent outflows have been demonstrated to be sources of diversity, these effluent-sourced taxa do not appear to proliferate downstream of the effluent source. Moreover, the loss of additional diversity between 2.DSUG and 3.USAmb (on the order of ~850 taxa) is much larger in magnitude than what was determined through differential abundance testing; this disparity is most likely the result of a small sample size and the conservative values selected for differential abundance detection. Loss of abundance and diversity in effluent impacted streams and their beds have been attributed to excess nutrients, sediment organic matter within the water column, or the presence of biologically active compounds (Drury et al., 2013). A variety of these bioactive compounds, including pharmaceuticals, hormones and organic wastewater contaminants, have been found in WWTP effluent (Kolpin et al., 2002). Wastewater treatment methods such as the activated sludge process were not initially designed for the removal of such compounds and, as a result, WWTPs can serve as sources for antibiotics (Akiyama and Savin, 2010) and other pharmaceutical compounds (Bartelt-Hunt et al., 2009).

Another observation that may partly explain these signals comes from the nature of the source material and analysis methods. Culture independent methods such as qPCR and amplicon sequencing do not distinguish between viable live cells, and those that are dead. Rapid

degradation of eDNA has been observed in aquatic systems (Barnes et al., 2014; Strickler et al., 2015; Thomsen et al., 2012a; Thomsen et al., 2012b). Sampling immediately downstream of an effluent source may inadvertently cause the detection of DNA from dead bacteria in the effluent. After traveling approximately 10 km further downstream (3.USAmb), these cells will have had the opportunity to lyse and their components, including DNA, to degrade, and thus would not be detected or detected in lower quantities. If the effluent contains live bacteria, the disappearance of those taxa could be attributed to those cells settling, sorbing to particles in the water column (aiding settling), sorbing to the streambed, or the environment in the stream is not suitable for their survival (such as those taxa requiring hosts or specific nutrients). Other potential controls on the decrease in diversity include shifts in hydrologic characteristics, including small tributary contributions and changes in discharge, velocity, channel morphology (i.e. ripple vs. run vs. pool), and hyporheic flow. Regardless of cause, the removal of these taxa represents a partial attenuation of the microbial community within the stream. Including additional sample sites between 2.DSUG and 3.USAmb in future studies may provide an opportunity to explain, or at least better describe, the interesting changes in community structure including the non-proliferation of effluent-sourced bacteria and the loss of diversity between the two WWTPs.

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

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

Beyond subjective impressions, utilizing only the indicator subset allowed the identification of 26 differentially abundant taxa, in comparison to the 11 taxa falling within these Orders found when using the entire microbial community dataset. The detection of 15 additional taxa may have been facilitated by the removal of potential noise induced by the non-indicator taxa. In terms of relative abundance, the magnitude of the differentially abundant taxa within the indicator subset displayed similar, albeit more variable, increases in magnitude (Fig. S13) in comparison to the taxa identified as differentially abundant in the full community (Fig. 7).

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

5. Conclusions

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

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

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge Shana McDevitt and Dylan Smith for their technical assistance during MiSeq Sequencing. This work was supported under NSF grant number 1245632 to Drexel University and by the William Penn Foundation as part of the Delaware River Watershed Initiative grant to Temple University., under grant number 8-16. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. The authors would like to acknowledge their use of the Drexel University Research Computing Facility. The authors thank the editor and anonymous reviewers for their comments and feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2017.09.162>.

References

- Ahmed, W., Stewart, J., Gardner, T., Powell, D., 2008. A real-time polymerase chain reaction assay for quantitative detection of the human-specific enterococci surface protein marker in sewage and environmental waters. *Environ. Microbiol.* 10, 3255–3264.
- Ahmed, W., Goonetilleke, A., Powell, D., Chauhan, K., Gardner, T., 2009. Comparison of molecular markers to detect fresh sewage in environmental waters. *Water Res.* 43, 4908–4917.
- Akiyama, T., Savin, M.C., 2010. Populations of antibiotic-resistant coliform bacteria change rapidly in a wastewater effluent dominated stream. *Sci. Total Environ.* 408, 6192–6201.
- Ali, S., Ghosh, N.C., Singh, R., 2010. Rainfall-runoff simulation using a normalized antecedent precipitation index. *Hydrol. Sci. J.* 55, 266–274.

- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Aristi, I., von Schiller, D., Arroita, M., Barcelo, D., Ponsati, L., Garcia-Galan, M.J., et al., 2015. Mixed effects of effluents from a wastewater treatment plant on river ecosystem metabolism: subsidy or stress? *Freshw. Biol.* 60, 1398–1410.
- Bae, S., Wuertz, S., 2012. Survival of host-associated bacteroidales cells and their relationship with *Enterococcus* spp., *Campylobacter jejuni*, *Salmonella enterica* serovar Typhimurium, and adenovirus in freshwater microcosms as measured by propidium monoazide-quantitative PCR. *Appl. Environ. Microbiol.* 78, 922–932.
- Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L., Lodge, D.M., 2014. Environmental conditions influence eDNA persistence in aquatic systems. *Environ. Sci. Technol.* 48, 1819–1827.
- Bartelt-Hunt, S.L., Snow, D.D., Damon, T., Shockley, J., Hoagland, K., 2009. The occurrence of illicit and therapeutic pharmaceuticals in wastewater effluent and surface waters in Nebraska. *Environ. Pollut.* 157, 786–791.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Method.* 57, 289–300.
- Bernhard, A.E., Field, K.G., 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 66, 4571–4574.
- Blanchet, F.G., Legendre, P., Borcard, D., 2008. Forward selection of explanatory variables. *Ecology* 89, 2623–2632.
- Bray, R.J., Curtis, J.T., 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P., 2016a. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.
- Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016b. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Res* 5, 1492.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624.
- Carey, R.O., Migliaccio, K.W., 2009. Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems: a review. *Environ. Manag.* 44, 205–217.
- Carpenter, S.R., Caraco, N.F., Correll, D.L., 1998. Nonpoint Pollution of Surface Waters With Phosphorus and Nitrogen.
- Chao, A., 1984. Nonparametric-estimation of the number of classes in a population. *Scand. J. Stat.* 11, 265–270.
- Clarke, K.R., Ainsworth, M., 1993. A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92, 205–219.
- Deiner K, Fronhofer EA, Meachler E, Walsler J-C, Altermatt F. Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *bioRxiv* 2016: 20800.
- 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.
- Gilbert, J.A., Meyer, F., Jansson, J., Gordon, J., Pace, N., Tiedje, J., et al., 2010. The Earth Microbiome Project: meeting report of the “1 EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6 2010. *Stand. Genomic Sci.* 3, 249–253.
- Gücker, B., Brauns, M., Pusch, M.T., 2006. Effects of wastewater treatment plant discharge on ecosystem structure and function of lowland streams. *J. N. Am. Benthol.*
- Haggard, B.E., Storm, D.E., Stanley, E.H., 2001. Effect of a point source input on stream nutrient retention. *American Water Resources Association*—>*J. Am. Water Resour. Assoc.* 37, 1291–1299.
- Haggard, B.E., Stanley, E.H., Storm, D.E., 2005. Nutrient retention in a point-source-enriched stream. *American Benthological Society*—>*J. N. Am. Benthol. Soc.* 24, 29–47.
- Halliday, E., McLellan, S.L., Amaral-Zettler, L.A., Sogin, M.L., Gast, R.J., 2014. Comparison of bacterial communities in sands and water at beaches with bacterial water quality violations. *PLoS One* 9, e90815.
- Harry, I.S., Ameh, E., Coulon, F., Nocker, A., 2016. Impact of treated sewage effluent on the microbiology of a small brook using flow cytometry as a diagnostic tool. *Water Air Soil Pollut.* 227, 57.
- 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.
- Homer, C., Dewitz, J., Yang, L.M., Jin, S., Danielson, P., Xian, G., et al., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* 81, 345–354.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441.
- 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.* 566–567, 1176–1185.
- 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, 761–763.
- Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., et al., 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. *Environ. Sci. Technol.* 36, 1202–1211.
- Lee, Z.M., Poret-Peterson, A.T., Siefert, J.L., Kaul, D., Moustafa, A., Allen, A.E., et al., 2017. Nutrient stoichiometry shapes microbial community structure in an evaporitic shallow pond. *Front. Microbiol.* 8, 949.
- Legendre, P., Anderson, M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.*
- 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.
- Marti, E., Balcazar, 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.
- Marti, E., Aumatell, J., Gode, L., Poch, M., Sabater, F., 2004. Nutrient retention efficiency in streams receiving inputs from wastewater treatment plants. *J. Environ. Qual.* 33, 285–293.
- McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–297.
- McIlroy, S.J., Starnawska, A., Starnawski, P., Saunders, A.M., Nierychlo, M., Nielsen, P.H., et al., 2016. Identification of active denitrifiers in full-scale nutrient removal wastewater treatment systems. *Environ. Microbiol.* 18, 50–64.
- McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N., Sogin, M.L., 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* 12, 378–392.
- McMurdie, P.J., Holmes, S., 2012. PhyloSeq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. *Pac. Symp. Biocomput.* 235–246.
- McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* 10, e1003531.
- Mello, B.L., Alessi, A.M., McQueen-Mason, S., Bruce, N.C., Polikarpov, I., 2016. Nutrient availability shapes the microbial community structure in sugarcane bagasse compost-derived consortia. *Sci. Rep.* 6, 38781.
- Merbt, S.N., Auguet, J.C., Blesa, A., Marti, E., Casamayor, E.O., 2015. Wastewater treatment plant effluents change abundance and composition of ammonia-oxidizing microorganisms in Mediterranean urban stream biofilms. *Microb. Ecol.* 69, 66–74.
- Merseburger, G.C., Marti, E., Sabater, F., 2005. Net changes in nutrient concentrations below a point source input in two streams draining catchments with contrasting land uses. *Sci. Total Environ.* 347, 217–229.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al., 2016. *vegan: Community Ecology Package.*
- Palmer, M.W., 1993. Putting things in even better order – the advantages of canonical correspondence-analysis. *Ecology* 74, 2215–2230.
- Parada, A.E., Needham, D.M., Fuhrman, J.A., 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.*
- Pavoine, S., Dufour, A.B., Chessel, D., 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *J. Theor. Biol. Philos. Mag. Ser. 6* 2, 559–572.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., et al., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196.
- Purdom, E., 2011. Analysis of a data matrix and a graph: metagenomic data and the phylogenetic tree. *Ann. Appl. Stat.* 5, 2326–2358.
- PWD, 2007. *Wissahickon Creek Watershed Comprehensive Characterization Report.* Philadelphia Water Department.
- Quince, C., Lanzen, A., Davenport, R.J., Turnbaugh, P.J., 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinf.* 12, 38.
- R Development Core Team, 2015. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.
- Rahm, B.G., Hill, N.B., Shaw, S.B., Riha, S.J., 2016. Nitrate dynamics in two streams impacted by wastewater treatment plant discharge: point sources or sinks? *American Water Resources Association*—>*J. Am. Water Resour. Assoc.* 52, 592–604.
- Ribot, M., Marti, E., von Schiller, D., Sabater, F., Daims, H., Battin, T.J., 2012. Nitrogen processing and the role of epilithic biofilms downstream of a wastewater treatment plant. *Freshw. Sci.* 31, 1057–1069.
- Rossello-Mora, R.A., Wagner, M., Amann, R., Schleifer, K.H., 1995. The abundance of *Zoogloea ramigera* in sewage treatment plants. *Appl. Environ. Microbiol.* 61, 702–707.
- Ryan, M.O., 2012. Microbial source tracking of human and animal waste pollution of diverse watersheds and of urban drainage systems using molecular methods. *Civil, Architectural, and Environmental Engineering.* Drexel University, Philadelphia, PA.
- Ryan, M.O., Gurian, P.L., Haas, C.N., Rose, J.B., Duzinski, P.J., 2013. Acceptable microbial risk: cost-benefit analysis of a boil water order for *Cryptosporidium*. *J. Am. Water Works Assoc.* 105, E189–E194.
- Saunders, A.M., Albertsen, M., Vollertsen, J., Nielsen, P.H., 2016. The activated sludge ecosystem contains a core community of abundant organisms. *ISME J.* 10, 11–20.
- Savitchcheva, O., Okabe, S., 2006. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. *Water Res.* 40, 2463–2476.
- Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593.
- Shanks, O.C., Domingo, J.W., Lu, J., Kely, C.A., Graham, J.E., 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Appl. Environ. Microbiol.* 73, 2416–2422.
- Shanks, O.C., White, K., Kely, C.A., Hayes, S., Sivaganesan, M., Jenkins, M., et al., 2010. Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 76, 1359–1366.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.
- Shao, Y., Chung, B.S., Lee, S.S., Park, W., Lee, S.S., Jeon, C.O., 2009. *Zoogloea caeni* sp. nov., a floc-forming bacterium isolated from activated sludge. *J. Syst. Evol. Microbiol.*—>*Int. J. Syst. Evol. Microbiol.* 59, 526–530.

- Siefring, S., Varma, M., Atikovic, E., Wymer, L., Haugland, R.A., 2008. Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. *J. Water Health* 6, 225–237.
- Sonthiphand, P., Cejudo, E., Schiff, S.L., Neufeld, J.D., 2013. Wastewater effluent impacts ammonia-oxidizing prokaryotes of the Grand River, Canada. *Appl. Environ. Microbiol.* 79, 7454–7465.
- Spring, S., Wagner, M., Schumann, P., Kampfer, P., 2005. *Malikia granosa* gen. Nov., sp. nov., a novel polyhydroxyalkanoate- and polyphosphate-accumulating bacterium isolated from activated sludge, and reclassification of *Pseudomonas spinosa* as *Malikia spinosa* comb. nov. *J. Syst. Evol. Microbiol.* 55, 621–629.
- Strickler, K.M., Fremier, A.K., Goldberg, C.S., 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biol. Conserv.* 183, 85–92.
- Thomsen, P.F., Kielgast, J., Iversen, L.L., Moller, P.R., Rasmussen, M., Willerslev, E., 2012a. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One* 7, e41732.
- Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T., et al., 2012b. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* 21, 2565–2573.
- Van der Gucht, K., Vandekerckhove, T., Vloemans, N., Cousin, S., Muylaert, K., Sabbe, K., et al., 2005. Characterization of bacterial communities in four freshwater lakes differing in nutrient load and food web structure. *FEMS Microbiol. Ecol.* 53, 205–220.
- Vaquer-Sunyer, R., Reader, H.E., Muthusamy, S., Lindh, M.V., Pinhassi, J., Conley, D.J., et al., 2016. Effects of wastewater treatment plant effluent inputs on planktonic metabolic rates and microbial community composition in the Baltic Sea. *Biogeosciences* 13, 4751–4765.
- Wakelin, S.A., Colloff, M.J., Kookana, R.S., 2008. Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. *Appl. Environ. Microbiol.* 74, 2659–2668.
- Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267.
- Wang, P., Chen, B., Yuan, R., Li, C., Li, Y., 2016. Characteristics of aquatic bacterial community and the influencing factors in an urban river. *Sci. Total Environ.* 569–570, 382–389.
- Wright, E.S., 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinf.* 16, 322.
- Wu, C.H., Sercu, B., Van de Werfhorst, L.C., Wong, J., TZ, DeSantis, Brodie, E.L., et al., 2010. Characterization of coastal urban watershed bacterial communities leads to alternative community-based indicators. *PLoS One* 5, e11285.
- Zwart, G., Crump, B.C., Agterveld, M.P.K.V., Hagen, F., Han, S.K., 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* 28, 141–155.