Science of the Total Environment 566–567 (2016) 1176–1185

Contents lists available at ScienceDirect

Science of the Total Environment

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

Bacterial community composition and structure in an Urban River impacted by different pollutant sources

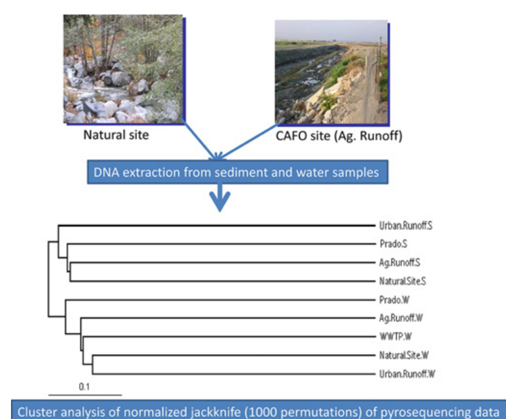
A. Mark Ibekwe [a,*], Jincai Ma [a,b], Shelton E. Murinda [c]

[a] USDA-ARS, U.S. Salinity Laboratory, Riverside, CA 92507, USA
[b] Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, China
[c] Animal and Veterinary Sciences Department, California State Polytechnic University, Pomona, CA 91768, USA

HIGHLIGHTS

- pH, N, and P were the main nutrient sources impacting fecal indicator bacteria.
- Bacteria responded differently to chemical and physical parameters.
- Low flow and contaminants from urban environments decreased microbial composition.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 March 2016
Received in revised form 23 May 2016
Accepted 24 May 2016
Available online 3 June 2016

Editor: D. Barcelo

Keywords:
Pyrosequencing
Microbial diversity
Watershed
Non-point source pollution
Sediment
Surface water

ABSTRACT

Microbial communities in terrestrial fresh water are diverse and dynamic in composition due to different environmental factors. The goal of this study was to undertake a comprehensive analysis of bacterial composition along different rivers and creeks and correlate these to land-use practices and pollutant sources. Here we used 454 pyrosequencing to determine the total bacterial community composition, and bacterial communities that are potentially of fecal origin, and of relevance to water quality assessment. The results were analyzed using UniFrac coupled with principal coordinate analysis (PCoA) to compare diversity, abundance, and community composition. Detrended correspondence analysis (DCA) and canonical correspondence analysis (CCA) were used to correlate bacterial composition in streams and creeks to different environmental parameters impacting bacterial communities in the sediment and surface water within the watershed. Bacteria were dominated by the phyla *Proteobacteria*, *Bacteroidetes*, *Acidobacteria*, and *Actinobacteria*, with *Bacteroidetes* significantly ($P < 0.001$) higher in all water samples than sediment, where as *Acidobacteria* and *Actinobacteria* where significantly higher ($P < 0.05$) in all the sediment samples than surface water. Overall results, using the β diversity measures, coupled with PCoA and DCA showed that bacterial composition in sediment and surface water was significantly different ($P < 0.001$). Also, there were differences in bacterial community composition between agricultural runoff and urban runoff based on parsimony tests using 454 pyrosequencing data. Fecal indicator bacteria in surface water along different creeks and channels were significantly correlated with pH ($P < 0.01$), NO_2

* Corresponding author at: USDA-ARS, U. S. Salinity Laboratory, 450 W. Big Springs Rd, Riverside, CA 92507, USA.
E-mail address: Mark.Ibekwe@ars.usda.gov (A.M. Ibekwe).

http://dx.doi.org/10.1016/j.scitotenv.2016.05.168
0048-9697/Published by Elsevier B.V.

($P < 0.03$), and NH_4N ($P < 0.005$); and in the sediment with NO_3 ($P < 0.015$). Our results suggest that microbial community compositions were influenced by several environmental factors, and pH, NO_2, and NH_4 were the major environmental factors driving FIB in surface water based on CCA analysis, while NO_3 was the only factor in sediment.
Published by Elsevier B.V.

1. Introduction

Terrestrial fresh water sediment bacterial communities are a major component of microbial food webs, biogeochemical cycles and energy flows in the sediment-water interface of rivers and streams (Ruiz-González et al., 2015). Rivers are the chief source of renewable water for humans and freshwater ecosystems (Vörösmarty et al., 2010), yet microbial diversity in flowing freshwater (lotic) is less commonly studied than in marine or lake ecosystems (Zinger et al., 2012). Their diversity and structure are determined by the temporal and spatial variability of physicochemical and biotic parameters and thus, can reflect local environmental conditions (Urakawa et al., 1999; Zhang et al., 2008a, 2008b). Any shift in nutrient, environmental and pollution profiles in the benthic-pelagic ecosystems will directly influence bacterial communities and that in turn will further affects nutrient cycle and other related communities (Hale et al., 2015). Therefore, microbial communities in rivers are diverse and dynamic in composition due to environmental stresses and nutrient compositions (Nogales et al., 2007). Furthermore, the composition of a microbial community in a river has been suggested as an indicator for pollution (Atlas, 1984).

Microbial communities in polluted rivers may contain not only broad functional diversities, but also bacteria, which may be pathogenic for humans and livestock. *Escherichia coli* and enterococci are widely used to monitor fecal contamination in drinking and recreational water. The occurrence of pathogenic bacteria in river water may increase near large urban populations following failure in sewage treatment processes. This is very common in developing countries with inefficient sewage treatment, low income, fast-growing populations and severe water stress where infection rates by water-borne pathogens are high (Abraham, 2011). The current study was conducted in a large urban watershed with varying land use in southern California with low flowing rivers and creeks. Pollutants in the watershed mainly consist of pathogens and nutrients due to the densely populated areas, some agricultural activities, and urban and storm-water runoff in the region. Different federal, state, and private agencies have monitored fecal bacterial composition from the surface water (Izbicki et al., 2004; Rice, 2005), but no studies have been done to uncover bacterial composition within the water bodies and sediment within the watershed using deep sequencing. Recently, terminal restriction fragment length polymorphism (T-RFLP) and Sanger sequencing were used to describe microbial community structure and composition in low flowing surface water within a small section of the watershed with different sources of pollutants (Ibekwe et al., 2012). During that study, bacterial community structure using T-RFLP showed that bacterial contamination of the low flowing river was not significantly different between concentrated animal feeding operations (CAFOs) and urban runoff. The current study expands on that study by using pyrosequencing to assess bacterial composition in the watershed, and correlate this data with some water chemistry parameters collected during the sampling period. River pollution is not limited to this watershed in southern California according to the United States Environmental Protection Agency (US-EPA), 45% of streams and rivers, and 32% of bays and estuaries in the United States are impacted by pathogens and sewage discharge (USEPA, 1986). Therefore, due to the presence of bacterial pollutants in some rivers in the middle Santa Ana River watershed, there is a need for the comprehensive analysis of bacterial composition in water bodies in order to understand their fate and transport.

In a recent study to better understand ecosystem health, and obtain foundational data to answer fundamental questions about microbial communities in flowing river affected by land use, it was found that sequence composition and average genome size vary with sampling site, environmental conditions, and water chemistry (Van Rossum et al., 2015). Other studies have shown that using taxonomic characterization of riverine bacterial microbiomes, bacterioplankton communities vary by location and nutrient concentrations (Hu et al., 2014; Jackson et al., 2014; Read et al., 2015; Ruiz-González et al., 2015; Savio et al., 2015; Wang et al., 2015), and in a 3-year study focusing on spring and summer, (Fortunato et al., 2013). In this study, we examined the microbial community composition for a year in sediment and surface water with inputs from different rivers and creeks impacted by agricultural activities such as concentrated animal feeding operations (CAFOs), urban runoff, waste water treatment plants (WWTPs), and a combination of these pollutant sources (recreational park) using pyrosequencing. We examined both sediment and surface flow microbial diversity, since sediment community may act as a major source to surface flow microbial community. We also monitored fecal indicator bacteria (FIB) for water quality assessments. Our goal was to identify species richness at the local scales, and differences in diversity across space and time, with an attempt to identifying both extrinsic and intrinsic factors that may explain differences in taxon composition among different communities separated by different sources of pollutants within the watershed. We hypothesized that bacterial diversity decreases with increasing pollutants, reflecting specific environmental factors along different rivers and creeks and land-use practices such as agriculture, urban runoff, and WWTPs. However, due to low water flow volume and concrete lined channels along urban runoff creeks; we expect to find low diversity compared with recreational water outlets. Finally, other measured chemical and biological variables were correlated to understand the influence of these variables on the fate of indicator bacteria in the watershed.

2. Materials and methods

2.1. Study area and sample collection

We conducted this study in the middle Santa Ana River (MSAR) watershed in the southwestern corner of San Bernardino County and the northwestern corner of Riverside County (Fig. 1) in southern California, USA. The current population of the watershed is ~1.9 million people, based upon the 2010 census data. Approximately 32% of land use in the watershed is residential, commercial or industrial, and with the rest, urban and agriculture. The area contained approximately 385,000 cows in 1995 (RWQCB, 2005), and as of January 2009, this number was down to about 138,500 (SAWPA, 2013).

In this study, we expanded our sampling sites from 14 to 20 to include another creek that originates from a lower elevation (N2; 207 m) as compared to N1 that originates from 1447 m elevation (Ibekwe et al., 2012). The Prado recreational area was also included because surface flow from the upper section of the watershed empties into this subsection. Sampling locations with site names, descriptions, and global positioning system (GPS) coordinates are listed in Table S1; Fig. 1. Sampling locations were based upon historical data obtained for the Total Maximum Daily Loads (TMDL) for Bacterial Indicators for MSAR watershed (Rice, 2005), and divided into five zones (Table S1; Fig. S1). Zone 1 consist of seven sites (UR2, UR3, UR4, UR5, UR9, and UR12, and UR13) along the Chino Creek and Cucamonga creek representing urban runoff; zone 2 has four sites along Cypress channel (CAFO6, CAFO7, CAFO8, and CAFO10) representing agricultural activities; zone 3 consist of four sites (PRADO2, PRADO3, PRADO4, AND PRADO5) around the Prado Recreational Park; zone 4 consist of samples from three waste-water treatment facilities (WW11, WW13, and WW14); and finally zone 5 consist of two natural sites (N1 and N2). Table S1 identifies the “zones” and

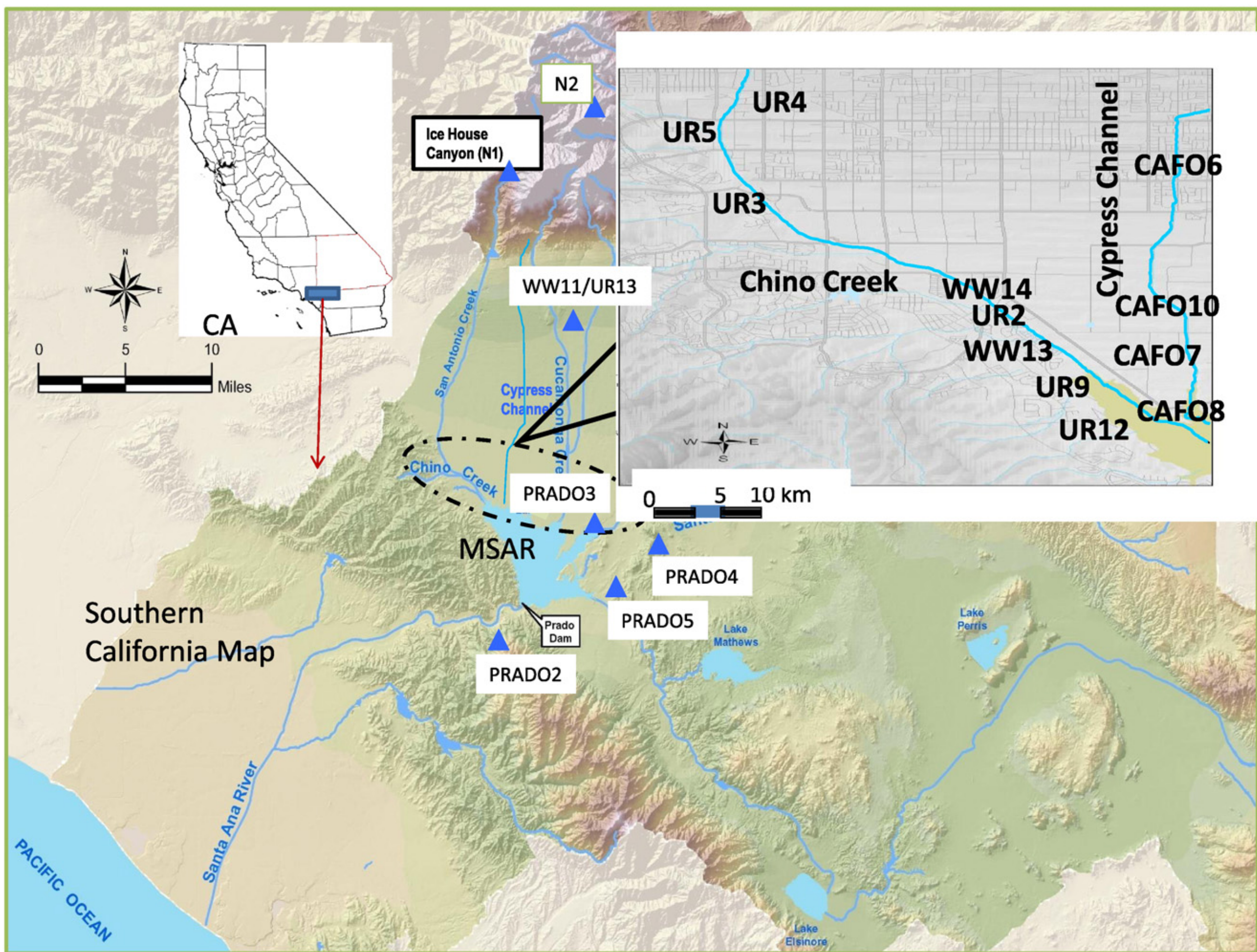


Fig. 1. Various sampling points along Chino creek and Cypress channel within the middle Santa Ana River (MSAR) watershed. Water flow from the natural site at Ice house Canyon (S1) to the San Antonio creek and into Chino creek. This flows into the Prado basin and into Santa Ana River and finally empties into the Pacific Ocean. The Santa Ana River is critical for the replenishment of Orange County's Groundwater Basin since over 2 million residents in Orange County depend on groundwater for 75% of their water supply.

corresponding sites, along with the number of samples acquired at each site. Water samples were collected from the three waste-water treatment plants (WWTPs) at sampling ports at the end of the plant where it discharges tertiary-level-treated water into Chino Creek. The sites along Chino creek had low flow on concrete lined channels resulting in low bacterial concentrations. Agricultural runoff from CAFOs was the main source of pollutant into Cypress channel while Chino Creek was affected more by WWTPs and urban runoff. The natural sites on Ice House Canyon (N1) and N2 on the lower elevation were used mainly as the control sites, and historical data for fecal coliforms had averaged 9 CFU 100 mL^{-1} over a five-year time period, from 2000 to 2005 (Rice, 2005). The sites were divided into zones based on pollutant sources. Analysis of covariance (ANOCOVA) model was used to evaluate our primary factors of interest which were pollutant sources or sites and other environmental parameters as additional covariate factors to determine the effects of zonation on our sampling strategy (Ibekwe et al., 2011). During this study, it was shown that total bacterial counts (API) did not significantly differ between sites and seasons. It was also shown that the summary statistics for the $\log_{10}$ transformed bacteria concentration models for total coliform (TC), fecal coliform (FC), *Escherichia coli*, and enterococci based on Shapiro-Wilk (SW) test scores for residual normality were not significantly different between sites along each zone (Ibekwe et al., 2011). The primary goal of this study was to understand and quantify the spatial-temporal trends in various bacteria measurements along the Chino Creek samples UR2, UR3, UR4, UR5, UR9, and UR12, and UR13 and Cypress Channel samples CAFO6, CAFO7, CAFO8, and CAFO10, and Prado recreational area PRADO2, PRADO3, PRADO4, AND PRADO5, and the natural sites N1 and N2 as baseline control. Towards this goal, these sites were sampled quarterly for a year.

Rainfall in the watershed is predominantly between December and April, with a mean annual rainfall of ≤800 mm, resulting in a variable base stream flow between seasons as previously described (Ibekwe et al., 2011). Briefly, the United States Geological Survey (USGS) gauged data shows the mean annual stream flow from Chino Creek (UR 3-Chino Creek) was 133.6 $m^3\ s^{-1}$) representing urban runoff and at Cypress channel, representing agricultural runoff (CAFO 6 – Cypress channel) was 96.8 $m^3\ s^{-1}$. Furthermore, four water quality parameters were also consistently acquired at each sample point: water pH, salinity, turbidity, and temperature.

Water samples were collected using sterile Nalgene sampling bottles and field parameters such as salinity, pH, temperature, turbidity, and dissolved oxygen were taken at each sample location using standard methods (APHA, 1995). All samples were collected in duplicate. For sites that were deep enough to obtain samples, grab samples were collected about 10–15 cm below the surface of the water. Sites with a shallow flow were sampled using a sterile stainless-steel sampling device. Sample turbidity was determined using a Hach model 2100P Portable Turbidimeter (Loveland, CO) according to manufacturer's instructions and calibrated each day of use. Salinity and pH were recorded with a standard conductivity/pH meter. Concentrations of fecal indicator bacteria were determined within 6 h of sample collection according to Standard Methods 9222B for total coliform, EPA Methods 1600 for *Enterococcus* and 1603 for *E. coli* (USEPA, 1986), based on membrane filtration.

Sediment samples from the 0–15-cm depth were taken from the creek or river banks using ethanol-disinfected core tubes and stored in

Whirl-Pak bags at 4 °C until processed; usually within 24 h. Samples were transported to the laboratory for analysis in coolers maintained between 2°–10 °C using ice packs. Total nitrogen (TN) and organic N were determined using Flash 2000 NC Analyzers (Thermo Scientific, MA). Salinity (EC, dS m^{-1}) of each soil was obtained by determining the conductivity of soil water extract (30 min extraction in horizontal shaker with water to soil ratio of 1:1, vol:wt) using a conductivity meter (Okaton, IL). Concentrations (mM) of Na^+, K^+, Ca^{2+}, Mg^{2+}, P, and S in soil water extracts was determined using an Optima-3300 DV ICP-OES spectrometer (Perkin-Elmer, MA) that was calibrated with certified standards prior to sample analysis. Fecal indicator bacteria (*Enterococcus* and *E. coli*) in sediment samples (10 g) were determined using serial dilution method in a 1:9 sediment PBS ratio. In brief, 90 mL of phosphate buffered saline (PBS) water (0.0425 g/L KH_2PO_4 and 0.4055 g/L $MgCl_2$) was added to sediment and shaken for 15 min. Ten milliliters of the suspension was added to Colilert or Enterolert vessel, diluted 1:10 and mixed. One milliliter from the 1:10 dilution was transferred to another vessel and was further diluted 1:1000, and an aliquot was added to the media, mixed, then sealed in QuantiTrays and incubated at 37 °C for 24 h. Samples were processed following the manufacturer's protocol in accordance with method 9223 (Eaton et al., 1998).

2.2. DNA extraction and purification from sediment and water samples

Total bacterial DNA was extracted from 500 mg of sediment samples and from water sample from the different sites. DNA was extracted using Power Soil and Water DNA kits (MO BIO, Inc., Solana Beach, CA), according to the manufacturer's protocol. Extracted DNA (2 μL) was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington DE), and run on 1.0% agarose gel before used for pyrosequencing.

2.3. Pyrosequencing

Sediment and water DNA samples were submitted to Core for Applied Genomics and Ecology at the University of Nebraska Lincoln for PCR optimization and pyrosequencing analysis. The V1–V2 region of the 16S rRNA gene was amplified using bar-coded fusion primers with the Roche-454 A or B titanium sequencing adapters (in italics), followed by a unique 8-base barcode sequence (B) and finally the 5′ ends of primer A-8FM (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGBBBBBBBBAGAG TTTGATCMTGGCTCAG) and of primer B-357R (5′-CCTATCCCCTGTGTG CCTT-GGCAGTCTCA GBBBBBBBB CTGCTGCCTYCCGTA-3′ (Benson et al., 2010). All PCR reactions were quality-controlled for amplicon saturation by gel electrophoresis; band intensity was quantified against standards using GeneTools software (Syngene). The resulting products were quantified using PicoGreen (Invitrogen, Carlsbad, CA) and a Qubit fluorometer (Invitrogen) before sequencing using Roche-454 GS FLX titanium chemistry (Benson et al., 2010; Wu et al., 2010). Raw data were treated with the Pyrosequencing Pipeline Initial Process (Cole et al., 2009) of the Ribosomal Database Project (RDP) to match barcodes and to trim off the adapters, barcodes and primers using the default parameters, and to remove sequences containing ambiguous 'N' or shorter than 200 bps (Claesson et al., 2009). These raw reads were further denoised to remove sequences that are likely due to pyrosequencing errors (Huse et al., 2010; Roeselers et al., 2011), and chimeras were filtered out using Chimera Slayer (Haas et al., 2011). Bacterial pyrosequencing population data were further analyzed by performing multiple sequence alignment techniques using the dist.seqs function in MOTHUR, version 1.9.1(Schloss et al., 2009). MOTHUR was also used to assign sequences to operational taxonomic units (OTUs, 97% similarity) and calculate both Shannon's diversity index values (H′), and Chao richness estimates. Taxonomic classification of the bacterial sequences of each sample was carried out individually, using the RDP Classifier. A bootstrap cutoff of 80% suggested by the RDP was applied to assign the sequences to different taxonomic levels. The sequence data sets of potential human pathogen (461 sequences) from our previously study using the same sites had been deposited in Sequence read Archive under the project name SRP028870 (Ibekwe et al., 2013), with accession numbers SRX335804 to SRX335812 (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP028870). The new dataset used for this study was based on 12,959 16S rRNA sequence tags generated through 454 pyrosequencing.

2.4. Statistics and analysis of pyrosequencing data

Analysis of variance (ANOVA) model using both site and field parameter classifications were used as the statistical model for analyzing the (log-transformed) density of indicator bacteria from the watershed (Ibekwe et al., 2011). Along with total coliforms, *E. coli*, and enterococci densities, four additional field parameters, i.e., pH, salinity (EC), turbidity, and surface water temperature for each sampling point were measured. Variations in each of these water quality parameters are known to affect bacterial concentrations. Principal coordinates analysis (PCoA) and hierarchical clustering in UniFrac were carried out using MOTHUR. PCoA was conducted based on RDP Classifier results from MOTHUR, OTUs, and weighted UniFrac (Hamady et al., 2010). The relaxed neighbor-joining algorithm in Clearcut (version 1.0.9) (Sheneman et al., 2006), was used to construct phylogenetic trees for between-site comparisons as previously described (Ibekwe et al., 2011) and parsimony tests in Treeclimber (Schloss and Handelsman, 2006). Pairwise comparisons with Bonferonni correction (Neter, 1996; Hollister et al., 2010) was used to test significant differences ($P \leq 0.001$). Additional statistics were performed on pyrosequencing data on the relative percentage of each operational taxonomic unit (OTU), or the sum of OTUs at a specific taxonomic (phylum, class, order, or family) level. The microbial diversity indices were analyzed using the vegan package of R software version 3.1.0 (The R foundation for Statistical Computing: http://www.r-project.org/). Detrended correspondence analysis (DCA) was employed to determine the overall structural changes in the microbial communities by R software version v3.1.0. DCA is an ordination technique that uses detrending to remove the arch effect, where the data points are organized in a horseshoe-like shape, in correspondence analysis (Zhou et al., 2012). Canonical correspondence analysis (CCA) was performed to determine the most significant sediment and water variables shaping microbial community composition and structure. The Mantel test ($p < 0.001$) was used to select the most significant environmental variables and to examine the correlation between community structure or fecal indicator bacteria and these variables (Zhou et al., 2012). Using automatic forward selection in CCA, 11 environmental variables were selected for analysis in sediment and 10 in water, and these include EC, pH, NO_2, NO_3, NH_4N, CA, Mg, Na, K, P, S. CCA was performed using the vegan package in R. During the analysis matrices of pollutant sources or sample units (the five zones) versus bacterial OTUs were paired with matrices of samples units versus environmental variables from sediments and surface water for CCA.

3. Results

3.1. Environmental variables

Physical and chemical characteristics, such as temperature, turbidity, salinity, dissolved oxygen (DO), and pH were determined in various surface water and sediment samples retrieved from the different sites. Salinity (EC), pH, temperature, turbidity, DO, NO_3, NO_2, NH_4, Ca, Na, Mg, P, S, and K are summarized in Table S2. The temperature values for surface water samples and wastewater effluent stream ranged from 6 °C at Ice House Canyon (S1) to 23 °C at WWTP (S14). Average turbidity values were below 1 nephelometric turbidity units (NTU) for the natural site at Ice House Canyon (N1). The basic univariate summary statistics for the three indicator bacteria measurements of interest; i.e., TC, *E. coli*, and enterococci counts are shown in Fig. 2. The statistics summarized the natural log transformed counts for each of these response variables. Indicator bacterial counts were not significantly different between sites in sediment samples (Fig. 2A) and in water samples (Fig. 2B) within

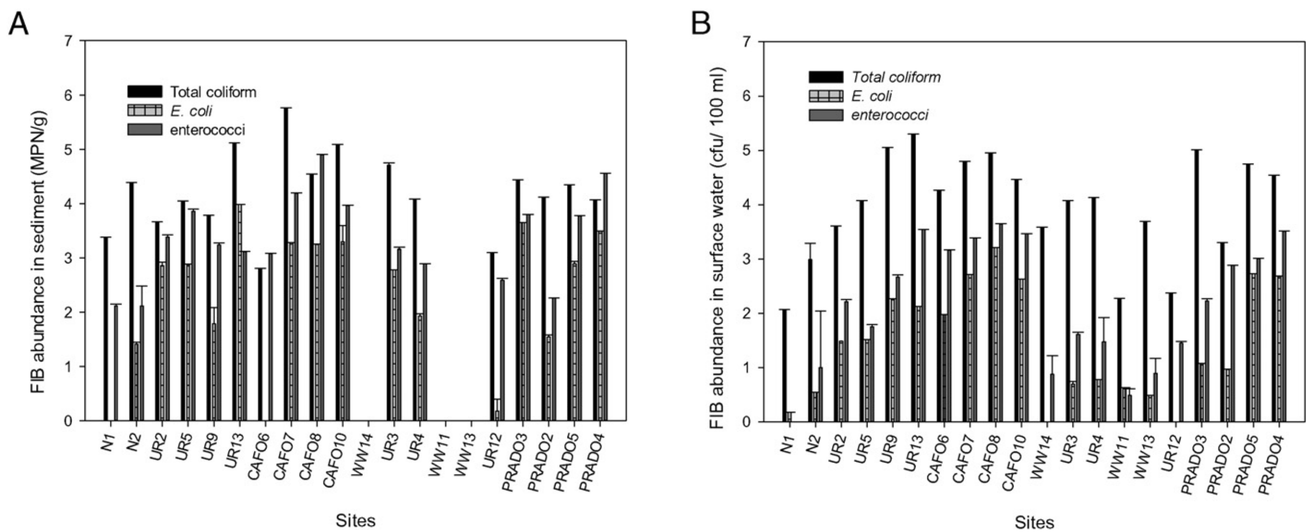


Fig. 2. Univariate summary statistics for bacterial counts using Shipiro-Wilk test for residual normality for total coliforms, *E. coli*, and enterococci. (A) Sediment samples (B) water samples. Error bars represent standard errors of duplicate samples pooled throughout each source. Zone 1 consist of urban runoff (UR) and represents seven sites (UR2, UR3, UR4, UR5, UR9, and UR12, and UR13) along the Chino Creek and Cucamonga creek representing; zone 2 has four sites along Cypress channel (CAFO6, CAFO7, CAFO8, and CAFO10) representing agricultural activities; zone 3 consist of four sites (PRADO2, PRADO3, PRADO4, AND PRADO5) around the Prado Recreational Park; zone 4 consist of samples from three waste-water treatment facilities (WW11, WW13, and WW14); and finally zone 5 consist of two natural sites (N1 and N2). Table S1 identifies the "zones" and corresponding sites, along with the number of samples acquired at each site.

each zone. However, FIB counts were significantly different ($P < 0.023$) between zones. Examination of each site throughout the watershed indicated that fecal bacterial concentrations along urban runoff creeks and CAFO channel routinely exceeded the applicable water quality objectives for *E. coli* (235 CFU/100 mL) except in the natural site (N1) and WWTPs.

3.2. Community composition, diversity, and estimated richness

A total of 12,959 16S rRNA sequence tags were generated through 454 pyrosequencing, with an average read length of about 200 bp (Table S3). The 454 sequence libraries ranged in size from 142 sequence tags (smallest) at water samples from urban runoff to 5666 sequences at the natural site surface water (largest) and contained between 68 OTUs in water samples to 1770 OTUs in natural site sediment (Table S3; Fig. S2A&B). All data were normalized to the smallest sequence tag, and the data reanalyzed to show normal distribution of variances (Table 1). Shannon diversity index values (H′) suggest that diversity varied along the study zones, with the lowest diversity associated with urban runoff water. Diversity values decreased significantly ($P < 0.05$) between water and sediment materials collected from urban runoff, agricultural runoff, and control sites. Although we determined hundreds and thousands of tags per sample, rarefaction curves of OTUs were far from the plateau, indicating that there were more undetermined tags either from real rare species or artificial sequences produced by PCR and sequencing errors (Fig S2A&B). Chao richness estimates suggest that higher sequences were captured in sediment than in surface water. Both the Shannon's diversity index value at a sequencing depth of 142 (Table 1) and its rarefaction curves showed that the sediment had the highest diversity, ranging from 5.82 (natural site) to 5.32 (Prado park), while the water samples had the lowest, ranging from 4.93 (Prado park) to 4.17 (urban runoff), and there were significant differences ($P < 0.05$) among the three groups. Summary statistics of the non-normalized data showed that higher OTUs, Chao richness, and Shannon diversity index values (H′) were also higher in sediment than in surface water (Table S3). Therefore, all analyses were done on both datasets to see if there are differences in the grouping pattern of bacterial community structures. However, most of the data presented are from the normalized dataset, with few exceptions involving sequence library. Furthermore, Fig. S1 shows the physical appearances of sampling locations from the different zones, including zone 1 from urban runoff samples, obtained mainly from concrete lined channels. These channels have very low volume of water flowing with continuous exposure to ultraviolet radiation, resulting in low sequence tag (142). The 454 libraries detected 24 bacterial phyla from the non-normalized dataset with *Proteobacteria* (34.8%) and *Bacteroidetes* (18.70%) encountered most frequently (Table 2; Fig. S3). The community in the sediment was the most evenly distributed, while surface water samples were the most skewed with a dominance of *Proteobacteria* (Table 2). Two phyla (*Proteobacteria* and *Bacteroidetes*) dominated (>50%) bacterial communities of sediment and surface water from the nine potential sources of contaminants classified using RDP Classifier at a confidence threshold of 80%. The two phyla were followed by a few other major (average abundance >1% in at least two sources) phyla, including *Actinobacteria* (3.79%), *Acidobacteria* (2.99%), *Verrucomicrobia* (4.02%), *Chloroflexi* (1.45%), *Planctomycetes* (2.54%), *Firmicutes* (2.41%), and *Gemmatimonadetes* (2.54%). A few phyla with minor abundance (<1%) in one of the 9 sources were also determined (Table 2). Within *Proteobacteria*, *Epsilonproteobacteria* only occurred at very low levels (0.00–1.67%, averaging 0.19%) (Table S4). The *Beta*-subdivision on the average was the most dominant *Proteobacteria*, followed by *Gamma*, *Delta*, and *Alphaproteobacteria*. However, *Gammaproteobacteria* was the dominant class of *Proteobacteria* in surface water from all the sources, and *Delta* and *Alphaproteobacteria* were in the sediment except in the natural sites for *Alphaproteobacteria*.

3.3. Phylogenetic structure of bacterial community from different zones

Phylogenetic structures of microbial communities, as indicated by detrended correspondence analysis (DCA)-based ordination revealed that the phylogenetic community structures were significantly different between water and sediment samples (Fig. 3A). Further analysis showed pollutant sources correlated with factors influencing bacterial assemblages ($R^2 = 0.32$, $P < 0.001$) since bacteria grouped according to the sources of pollutants such as CAFOs or urban runoff. Samples from the different sources on the distribution of bacterial similarity were sorted into different groups such as CAFO or urban runoff by applying principal coordinates analysis (PCoA), and the UPGMA hierarchical clustering analysis to a matrix of UniFrac distances using the UniFrac web interface in MOTHUR. The PCoA (Fig. 3B) showed that microbial community structure were significantly different ($P = 0.001$) between water samples from urban runoff (zone 1) and CAFOs (Cypress channel-zone 2). These samples clustered to the middle portion of Fig. 3B.

Table 1
Normalized summary of sequence library, OTUs, and diversity and richness estimates at 97% level.

Group[a]	Nseqs[b]	OTUs	Chao	Invsimpson	Npshannon	Simpson	Coverage
CAFO S	142	113	641.11	154.01	5.73	0.006	0.309
CAFO W	142	73	167.23	42.24	4.39	0.023	0.647
UR S	142	104	289.88	161.46	5.38	0.006	0.435
UR W	142	68	162.09	19.59	4.17	0.051	0.676
N S	142	115	440.71	270.56	5.82	0.003	0.323
N W	142	75	139.47	30.71	4.44	0.032	0.647
Prado S	142	102	278.64	130.01	5.32	0.007	0.450
Prado W	142	90	220.20	103.21	4.93	0.009	0.556
WWTPs	142	76	186.50	30.06	4.45	0.033	0.634

[a] CAFO S = agricultural runoff sediment, CAFO W = agricultural runoff water, UR S = urban runoff sediment, UR W = urban runoff water, N S = natural site sediment, N W = natural site water, Prado S = Prado sediment, Prado W. = Prado water, WWTPs = waste water treatment plant.
[b] Nseqs = number of sequence tags, OTUs = operational taxonomic units.

This result was different from our previous study using T-RFLP that suggested no significant differences between CAFO runoff and urban runoff (Ibekwe et al., 2012). Furthermore, samples from WWTPs clustered to the bottom right, while sample from Prado region (zone 3) clustered to the far left, and those from the natural sites clustered at the top (zone 5), and these were significantly different ($P < 0.05$). Microbial community structure in sediment samples were also analyzed using the same procedure (Fig. 3C). Parsimony tests confirmed that community structures of the four sources were significantly different from each other ($P < 0.001$), with urban, CAFO, and natural sites clustering on PCoA 1 and Prado samples on PCoA 2. Hierarchical clustering analysis (Fig. 3D) with Jackknife supporting values using both water and sediment samples showed that microbial community structure from the five zones were significantly different from each other when sediment and water samples were analyzed together.

Canonical correspondence analysis (CCA) was performed to determine the most significant water and sediment variables shaping microbial community composition and structure. CCA showed that microbial community structure was significantly shaped by several key physical and chemical variables for water (Fig. 4A), such as NO_2, pH, and NO_3. The biplot score showed 30.5% of the variations were explained by CCA1 while 28.6 were explained by CCA2. In the sediment (Fig. 4B), NO_3 ($P < 0.015$) was the strongest environmental variable contributing to the biplot. In the sediment 35.9% of the variations were explained by CCA1, while by 32.0% of the variations were explained by CCA2.

3.4. Correlation of environmental variables with fecal indicator bacterial populations

CCA was performed to determine the most significant water and sediment variables shaping fecal indicator bacterial population in the watershed. The Mantel test was used to examine the correlation between fecal indicator bacteria population and each variable (Fig. 5). CCA of fecal indicator bacterial data and water variables showed that indicator bacterial population were significantly shaped by several key environmental factors (Fig. 5A): pH ($F = 4.69$, $p < 0.01$), NO_2 ($F = 2.85$, $p < 0.03$), and NH_3N ($F = 8.77$, $p < 0.005$), with 13.6% of the constrained variance explained by CCA1and 4.8% by CCA2. However, in the sediment the effects of environmental factors were not as strong as in the water column (Fig. 5B), with 8.73% of the constrained variance explained by CCA1and 5.46% by CCA2. Fecal indicator bacterial population were significantly shaped by NO_3 ($F = 3.05$, $p < 0.015$).

Table 2
Santa Ana River percent sample counts by phylum. The abundance is presented in terms of percentage in total bacterial sequences in a sample, classified using RDP classifier at a confidence threshold of 50%.

Phyla	Total	CAFO S[*]	CAFO W	UR. S	UR W	Natural. Site. S	Natural Site. W	PradoS	PradoW	WWTP
Acidobacteria	2.99	4.05	0.45	4.00	0.00	4.41	0.68	2.24	0.35	0.76
Actinobacteria	3.79	9.54	1.35	4.80	0.00	4.81	1.90	2.75	4.59	1.53
Armatimonadetes	0.55	0.58	0.00	0.00	0.00	0.52	0.41	0.52	1.06	0.00
Bacteroidetes	18.7	11.2	32.43	6.40	51.67	14.55	28.86	13.94	27.21	42.75
BRC1	0.03	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
Chlorobi	0.26	0.00	0.00	2.40	0.00	0.23	0.00	0.17	0.35	0.00
Chloroflexi	1.45	1.73	0.00	0.00	0.00	1.39	0.27	4.65	0.00	0.76
Deinococcus-Thermus	0.16	0.00	0.00	0.80	1.67	0.12	0.14	0.00	0.00	0.00
Firmicutes	2.41	3.47	1.35	0.80	0.00	2.32	1.90	3.96	1.06	1.91
Fusobacteria	0.10	0.00	0.00	0.80	0.00	0.12	0.14	0.34	0.35	0.00
Gemmatimonadetes	0.35	0.87	0.00	0.00	0.00	0.52	0.54	0.34	0.00	0.00
Lentisphaerae	0.06	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00
Nitrospira	0.16	0.87	0.45	0.80	0.00	0.23	0.14	0.34	0.00	0.00
OD1	0.55	0.58	0.00	0.80	1.67	0.35	0.81%	0.52	1.41	1.15
OP11	0.03	0.29	0.00	0.00	0.00	0.00	0.00%	0.00	0.00	0.00
Planctomycetes	2.54	1.45	0.45	0.80	0.00	3.25	1.08	2.07	0.71	0.38
Proteobacteria	34.8	39.3	44.59	50.40	41.67	36.17	42.28	41.14	42.05	30.92
Spirochaetes	0.42	1.16	0.45	0.00	0.00	0.41	0.41	0.86	0.35	0.38
SR1	0.10	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.38
Synergistetes	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00
Tenericutes	0.06	0.00	0.00	0.00	0.00	0.06	0.14	0.00	0.00	0.00
TM7	0.51	0.29	0.00	0.80	0.00	0.87	0.41	0.52	0.00	0.00
unclassified	25.5	20.2	16.67	23.20	3.33	23.88	16.26	21.69	15.55	16.03
Verrucomicrobia	4.02	4.34	1.80	2.40	0.00	5.39	3.12	3.44	4.59	3.05
WS3	0.29	0.00	0.00	0.80	0.00	0.35	0.00	0.34	0.35	0.00

[*] CAFO S = agricultural runoff sediment, CAFO W = agricultural runoff water, UR S = urban runoff sediment, UR W = urban runoff water, N S = natural site sediment, N W = natural site water, Prado S = Prado sediment, Prado W. = Prado water, WWTPs = waste water treatment plant.

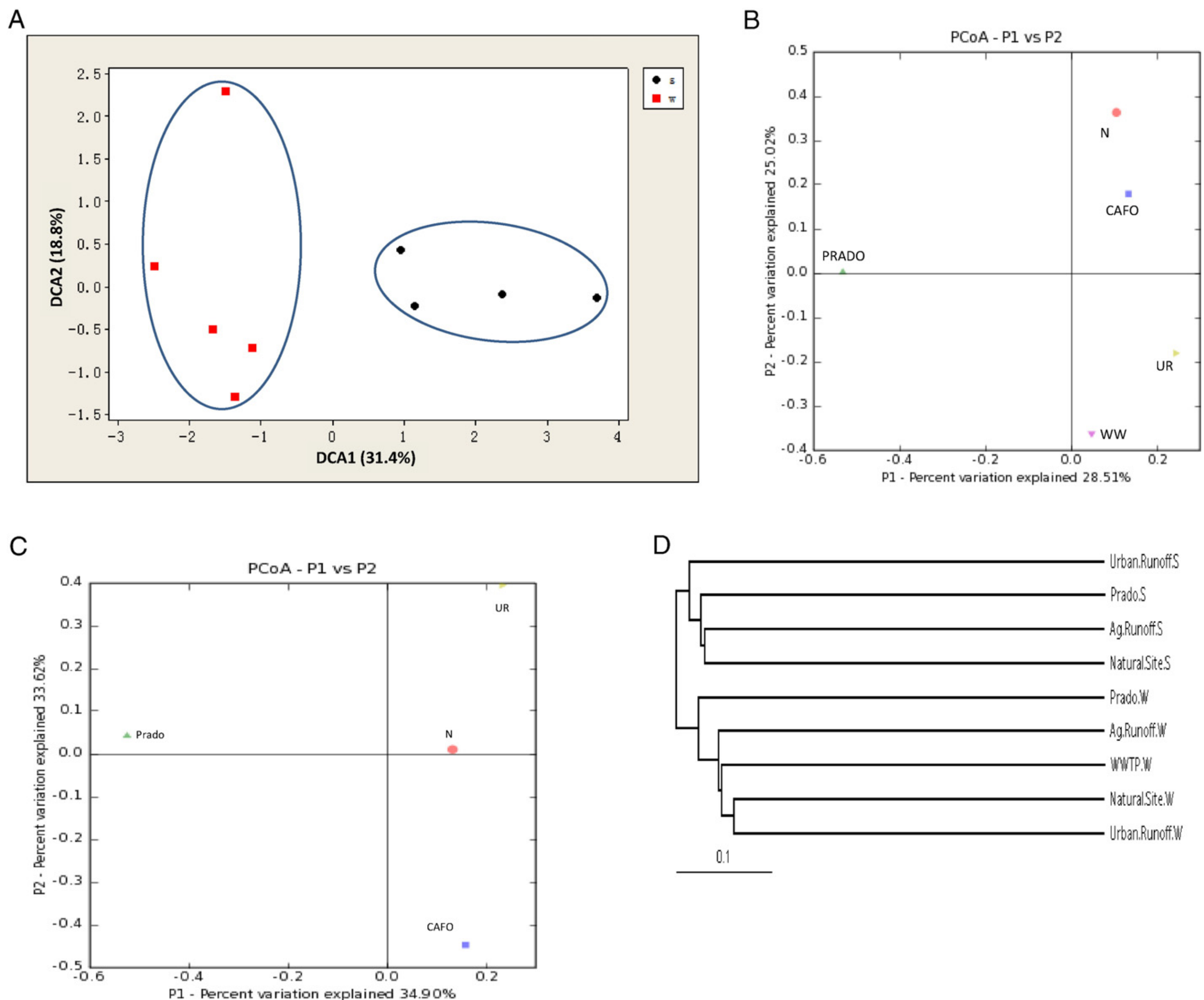


Fig. 3. Detrended correspondence analysis (DCA) of bacterial community structure data showing that sediment and river water samples were significantly separated on DCA1 based on bacterial community composition (3 A). Principal coordinate analysis (PCoA) obtained using the UniFrac distance matrix comparing the five sources of pollutants. Principal coordinate 1 (P1) vs principal coordinate 2 (P2) are represented. Samples are from pyrosequencing to a matrix of UniFrac distances using the UniFrac web interface in MOTHUR. The PCoA (Fig. 3B) showed samples from urban runoff (Chino creek) and CAFO runoff (Cypress channel) clustering to the middle portion while samples from WWTPs clustered to the bottom right and Prado clustered to the far left and the natural site to the top. Microbial community structure in sediment samples were also analyzed using the same procedure (Fig. 3C). Cluster analysis of normalized jackknife (1000 permutations) of pyrosequencing data (3D) from both sediment and water samples from all the sampling points throughout the watershed.

3.5. Potential bacterial community sequence richness for fecal pollution assessment

The assessment of some potential bacterial community sequence relative richness for FIB across the different pollutant sources are presented in Table 3. The top phyla with the highest degree of relative richness associated with fecal pollution were *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, and *Spirochaetes*. *Proteobacterial* sequences dominated our sample with more than 40% OTUs on the average from each source (Table 3). *Proteobacteria* contributed the highest proportion of bacterial richness that may be associated with fecal sources that could cause diseases in humans (Table 3). Some of the potential sequence tags were associated with sequences from *Arcobacter*, *Citrobacter*, *Shigella*, *Aeromonas*, and *Legionella* and were detected in all the sources except urban runoff. The next dominant phylum was the *Actinobacteria*, and the sequences at the genus level were dominated by *Nocardia*, *Corynebacterium* and *Mycobacterium*. Another phylum with a major contribution of bacterial communities that may be associated with fecal sources was *Bacteroidetes*. *Bacteroides* was the dominant genus from this phylum (Table 3), with the highest percent richness in sediment from Prado recreational water. The Phylum *Fermicutes* produced a very diverse group of bacteria of fecal origin that may be very important for water quality assessment. These include *Bacillus*, *Staphylococcus and Clostridium* (Table 3), with *Bacillus* dominating CAFO runoff sediment, natural site sediment and WWTP, and with *Clostridium* in Prado park recreational area sediment and natural site sediment.

4. Discussion

In this study, we employed high-throughput 454 pyrosequencing techniques to quantify bacterial community structures in a large urban watershed impacted by many pollutant sources such as 3 WWTPs, large developing urban population of about 1.9 million, and a dwindling cow population. Within our samples, the natural sites showed the highest bacterial richness and diversity, while urban runoff was the lowest both in the sediment and surface

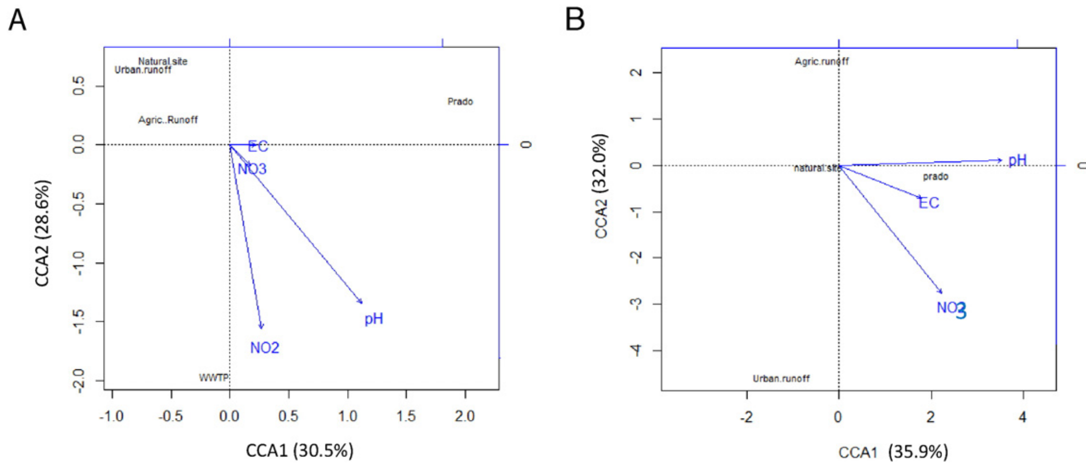


Fig. 4. Canonical correspondence analysis (CCA) of bacterial community structure of the most significant water and sediment variables shaping microbial community composition and structure. Canonical correspondence analysis (CCA) showed that microbial community structure was significantly shaped by several key physical and chemical variables for water (Fig. 4A), such as electrical conductivity (EC), NO_3, NO_2, and pH; and for sediment (Fig. 4B) such as EC, pH, and NO_2.

water. It should be noted that the natural sites were not influenced by pollutants because the water source was dominated mostly by melting snow and FIB from the natural sites as previously shown to be less than 9.0 CFU/mL (Ibekwe et al., 2011). The main reason for the low bacterial diversity in urban runoff could be that the river flows mainly through a concrete channel (Fig. S1), and about 90% of the flow are from tertiary treated waste water. Furthermore, the volume of flow during the dry season (April–August) may be less than 20 cm deep, resulting in continuous exposure to ultra-violet radiation. Also, sediment samples were localized deposits of soil particles on the concrete channel. Therefore, the sediment and water samples from the concrete lines were subjected to extreme conditions during dry weather. During the wet season (November–March) the volume of flow may triple or even higher depending on the amount of rainfall, but are still subjected to the same extreme weather conditions. These data were consistent with recent sediment bacterial community analyses results of pyrosequencing data from urban areas affected by WWTPs showing bacterial communities significantly decreased in population and diversity (Drury et al., 2013) due to the influence of treated waste water. In our results, the Chao index from the natural sites sediment and surface water were significantly higher than all other sites with the non-normalized data (Table S3), although richness estimators may be extremely sensitive to the error-prone sequences (Quince et al., 2009; Reeder and Knight, 2009). However, the Shannon's diversity index which is less sensitive to sequencing and PCR errors (Wang et al., 2012) showed the highest diversity in the natural site sediment using both the normalized and the non-normalized data (Table 1 and Table S3).

Our work in this watershed has produced results that are very similar to results produced from other large/urban rivers in Chicago (Drury et al., 2013). Here we examined the quality of the river water about 75 km from the ocean and observed that FIB were higher than the state and EPA standards (Fig. 2). The high numbers may be associated with nitrogen sources of nutrients as they correlate significantly with FIB from upstream (Fig. 5A). This is not surprising, because urban areas have the potential to pollute water in many ways as this may contain pollutants such as fecal coliform bacteria, nitrates, phosphorus, chemicals, and other bacteria. Therefore, preventing pollution in the middle Santa Ana river watershed may help reduce pollutants entering Pacific Ocean.

Our study showed that microbial communities in sediment were significantly different from microbial communities in surface water (Fig. 3A) as the water samples clustered together and away from sediments. Microbial communities in surface waters are highly responsive to perturbation, dissolved organic matter concentration, and chemical stress among others (Bodtker et al., 2008; Wassel and Mills, 1983; Nelson, 2009; Hirayama et al., 2005). The detection of diverse community

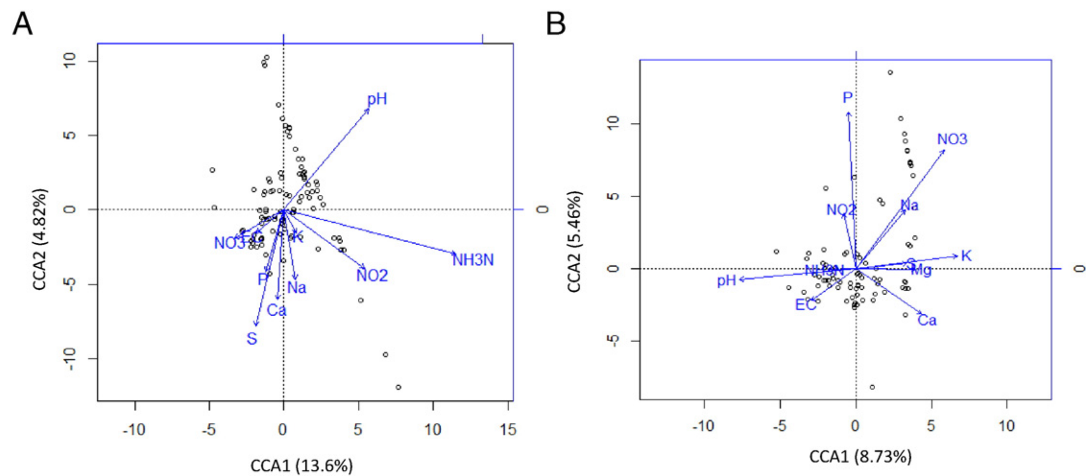


Fig. 5. Canonical correspondence analysis (CCA) was performed to determine the most significant water and sediment variables shaping fecal indicator bacterial composition in the watershed. CCA of fecal indicator bacterial data and water variables showed that indicator bacterial composition were significantly shaped by pH, NO_2, and NH_3N (Fig. 5A): in the sediment the effects of environmental factors were not as strong as in the water column (Fig. 5B). Fecal indicator bacterial composition and structure were significantly shaped by NO_3. The Mantel test was used to examine the correlation between fecal indicator bacteria and each variable.

Table 3
Percent of potential bacterial pathogen sequences at the genus level that may be of concern to the assessment of water quality.

Phyllum	Genus	CAFO S *	CAFO W **	UR. S	UR W	Natural site S	Natural site. W	PradoS	PradoW	WWTP
Actinobacteria	*Nocardia*	0	0	0	0	0.012	0	0	0	0
	Corynebacterium	0	0	0	0	0.012	0.022	0	0	0
	Mycobacterium	0.2144	0	0	0	0.122	0.011	0	0	0
Bacteroidetes	*Bacteroides*	0	0	0	0	0	0	0	0.123	0
Proteobacteria	*Arcobacter*	0.4287	0.360	0	0.4545	0.036	0.123	0.329	0	0.068
	Citrobacter	0	0	0	0	0	0	0.188	0	0
	Shigella	0	0.072	0	0	0.012	0	0	0	0
	Aeromonas	0	0.432	0	0.0136	0.305	0.653	0.047	0.00134	0
	Legionella	0	0	0	0	0.024	0.022	0	0	0
Spirochaetes	*Leptospira*	0	0.144	0	0	0	0.056	0	0	0
Firmicutes	*Bacillus cereus*	1.9293	0	0	0	0.195	0	0	0	0.068
	Staphylococcus	0	0	0	0	0	0.022	0	0	0
	Clostridium	0	0	0	0	0.171	0	0.094	0	0

* CAFO S = agricultural runoff sediment, ** CAFO W = agricultural runoff water, UR S = urban runoff sediment, UR W = urban runoff water, N S = natural site sediment, N W = natural site water, Prado S = Prado sediment, Prado W. = Prado water, WWTPs = waste water treatment plant.

composition and structure within the different zones of the watershed suggests changes in environmental microbial communities in surface waters are highly responsive to changes environmental factors such as nutrients, salinity (EC), and pH (Fig.4 B&C). For instance microbial communities were quite different between the natural sites and urban runoff as well as the other zones. The main drivers may be differences in nutrients and salinity from the urban area that may be influenced by household waste (Table S2), whereas microbial communities from agricultural area may be influenced by excess nutrients impacted by concentrated animal feeding operations (CAFOs). Nutrients from the different zones may enrich different bacterial populations depending on their sources into the Santa Ana River. For instance, fast unifrac analysis showed that the populations of river water were quite different among the five zones (Fig. 3B).

The order *Bacteroidales* is enriched within the gut microbiota of many mammals (Dowd et al., 2008; Cotta and Forster, 2006; Ishii et al., 2006; Fogarty and Voytek, 2005; Juteau et al., 2005), and specific species within this order have been proposed as fecal indicators (Weary et al., 2008). Sequences of the genus *Bacteroides* was commonly identify in the recreational water samples from Prado Park (Table 4). Some species from this genus has been identified as indicators of fecal pollution originating from mammalian sources (Savichtcheva et al., 2007; Savichtcheva and Okabe, 2006). *Bacteroides*, have a high degree of host specificity that reflects differences in the digestive system of the host animal (Bernhard and Field, 2000) and have a small potential to grow in the environment (Kreader, 1998) thus making it a good indicator of recent fecal contamination. The members of this phylum contain diverse bacterial species that have colonized virtually all types of habitats. They are among the major members of the microbiota of animals, especially in the gastrointestinal tract, and are frequently used for microbial source tracking (Dubinsky et al., 2012; Li et al., 2015; Gomi et al., 2014). The presence of DNA sequences in surface water in Prado recreational area may suggest some recent fecal contamination close to sampling. Another group of bacteria that are commonly used as indicator bacteria are the *Enterococcus*. Although we were able to grow pure cultures from the genus (Fig. 2A and B), the pyrosequencing method could not detect their DNA sequences suggesting that this technique may have some limitations when it comes to detecting less abundant members of the community (Table S4). It was also interesting to note that other fecal indicator bacterial DNA sequences, such as *Clostridia*, were found mainly in sediment samples associated with recreational water sediment at Prado recreational area and in the sediment of natural sites. This indicates the wide distribution of *Clostridia* DNA in sediment samples more than surface water and suggests that the sediment may be a major source of this class of bacteria (Weary et al., 2008). The high prevalence of *Enterobacteriales* DNA sequences in Table S4 falls in line with the high FIB counts observed in Fig. 2A&B. Again, DNA sequences belonging to the genus *Escherichia* and *Salmonella* were not detected with this method, but *Shigella* DNA sequences were identified in surface water from agricultural runoff and in the sediment from natural sites.

5. Conclusions

Our data showed that pH, NO_2, and NH_4 were the major environmental factors correlating significantly with FIB sequences in urban river runoff water based on CCA analysis, while NO_3 was the only factor in sediment. Using pyrosequencing, we were able to understand some of the relationships between microbial community structures, nutrient, and environmental factors in a large watershed as they relate to changes in microbial communities from different sources. It shows that microbial communities in the sediment-water interface in terrestrial fresh water with low flow were significant reduced in composition due to different environmental factors. The reduction was the highest from urban sources in comparison to CAFO and recreational sources. The higher level of bacterial composition in the CAFO and recreational sources was due to higher sediment nutrient availability to surface water in these zones.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2016.05.168.

Acknowledgements

This research was supported by the 214 Manure and Byproduct Utilization Project of the USDA-ARS. We thank Damon Baptista for technical assistance. Mention of trademark or proprietary products in this manuscript does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

References

Abraham, W.-R., 2011. Megacities as sources for pathogenic bacteria in rivers and their fate downstream. Int. J. Microbiol. 798292.
American Public Health Association, 1995. Standard Methods for the Examination of the Examination of Water and Wastewater. 19th ed. Am. Public Health Asso, Washington, DC.
Atlas, R., 1984. Use of microbial diversity measurements to assess environmental stress. In: Klug, M.J., Reddy, C.A. (Eds.), Current Perspectives in Microbial Ecology. American Society for Microbiology, Washington, DC, USA, pp. 540–545.
Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S., et al., 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. PNAS 107, 18933–18938.
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. http://dx.doi.org/10.1128/AEM.66.10.4571-4574.2000.
Bodtker, G., Thorstenson, T., Lillebo, B.L., Thorbjornsen, B.E., Ulvoen, R.H., et al., 2008. The effect of long-term nitrate treatment on SRB activity, corrosion rate and bacterial community composition in offshore water injection systems. J. Ind. Microbiol. Biotechnol. 35, 1625–1636. http://dx.doi.org/10.1007/s10295-008-0406-x.
Claesson, M., O'Sullivan, O., Wang, Q., Nikkila, J., Marchesi, J., Smidt, H., et al., 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PLoS One 4, e6669.
Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., et al., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37, D141–D145.

Cotta, M., Forster, R., 2006. The family Lachnospiraceae, including the genera Butyrivibrio, Lachnospira and Roseburia. In: Dworkin, M., Schleifer, K.-H., Rosenberg, E., Falkow, S. (Eds.), The Prokaryotes - A Handbook on the Biology of Bacteria: Bacteria: *Firmicutes*, *Cyanobacteria*, 3 ed. Springer-Verlag New York, LLC., New York, NY, pp. 1002–1021.

Dowd, S.E., Callaway, T.R., Wolcott, R.D., Sun, Y., McKeehan, T., et al., 2008. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol. 8. http://dx.doi.org/10.1007/bf02015128.

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.

Dubinsky, E.A., Esmaili, L., Hulls, J., Cao, Y., Griffith, J., Andersen, G.L., 2012. Application of phylogenetic microarray analysis to discriminate sources of fecal pollution. Environ. Sci. Technol. 46, 4340–4347.

Eaton, A.D., Clesceri, L.S., Greenberg, A.E., Franson, M.A.H., 1998. Standard Method for Examination of Water and Wastewater. American Public Health Association, Washington, DC, USA.

Fogarty, L.R., Voytek, M.A., 2005. Comparison of *Bacteroides-Prevotella* 16S rRNA genetic markers for fecal samples from different animal species. Appl. Environ. Microbiol. 71, 5999–6007.

Fortunato, C.S., Eiler, A., Herfort, L., Needoba, J.A., Peterson, T.D., Crump, B.C., 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia river coastal margin. ISME J. 7, 1899–1911. http://dx.doi.org/10.1038/ismej.2013.79.

Gomi, R., Matsuda, T., Matsui, Y., Yoneda, M., 2014. Fecal source tracking in water by next-generation sequencing technologies using host-specific *Escherichia coli* genetic markers. Environ. Sci. Technol. 48, 9616–9623.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., et al., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504.

Hale, R.L., Turnbull, L., Earl, S.R., Childers, D.L., Grimm, N.B., 2015. Stormwater infrastructure controls runoff and dissolved material export from arid urban watersheds. Ecosystems 18, 62–75. http://dx.doi.org/10.1007/s10021-014-9812-2.

Hamady, M., Lozupone, C., Knight, R., 2010. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J. 4, 17–27.

Hirayama, H., Takai, K., Inagaki, F., Yamato, Y., Suzuki, M., et al., 2005. Bacterial community shift along a subsurface geothermal water stream in a Japanese gold mine. Extremophiles 9, 169–184.

Hollister, E.B., Engledow, A.S., Hammet, A.J., Provin, T.L., Wilkinson, H.H., et al., 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J. 4, 829–838.

Hu, A., Yang, X., Chen, N., Hou, L., Ma, Y., Yu, C.-P., 2014. Response of bacterial communities to environmental changes in a mesoscale subtropical watershed, Southeast China. Sci. Total Environ. 472746–472756 http://dx.doi.org/10.1016/j.scitotenv.2013.11.097.

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L., 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. 12, 1889–1898.

Ibekwe, A.M., Lesch, S.M., Bold, R.M., Leddy, M.B., Graves, A.K., 2011. Variations of indicator bacteria in a large urban watershed. Trans. ASABE 54, 2227–2236.

Ibekwe, A.M., Leddy, M., Bold, R.M., Graves, A.K., 2012. Bacterial community composition in low-flowing river water with different sources of pollutants. FEMS Microbiol. Ecol. 79, 155–166.

Ibekwe, A.M., Leddy, M., Murinda, S.E., 2013. Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. PLoS One 8, e79490. http://dx.doi.org/10.1371/journal.pone.0079490.

Ishii, S., Ksoll, W.B., Hicks, R.E., Sadowsky, M.J., 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from lake superior watersheds. Appl. Environ. Microbiol. 72, 612–621.

Izbicki, J.A., Pimentel, M.I., Leddy, M.B., 2004. Microbial and dissolved organic carbon characterization of stormflow in the Santa Ana River at Imperial Highway, Southern California, 1999-2002. US Geol. Surv. Sci. Investig. Rep. 2002, 2004–5116.

Jackson, C.R., Millar, J.J., Payne, J.T., Ochs, C.A., 2014. Free-living and particle-associated bacterioplankton in large rivers of the Mississippi River Basin demonstrate biogeographic patterns. Appl. Environ. Microbiol. 80, 7186–7195. http://dx.doi.org/10.1128/AEM.01844-14.

Juteau, P., Tremblay, D., Villemur, R., Bisaillon, J.G., Beaudet, R., 2005. Analysis of the bacterial community inhabiting an aerobic thermophilic sequencing batch reactor (AT-SBR) treating swine waste. Appl. Microbiol. Biotechnol. 66, 115–122.

Kreader, C.A., 1998. Persistence of PCR-detectable Bacteroides distasonis from human feces in river water. Appl. Environ. Microbiol. 64, 4103–4105.

Li, X., Harwood, V.J., Nayak, B., Staley, C., Sadowsky, M.J., Weidhaas, J., 2015. A novel microbial source tracking microarray for pathogen detection and fecal source identification in environmental systems. Environ. Sci. Technol. 49, 7319–7329. http://dx.doi.org/10.1021/acs.est.5b00980.

Nelson, C.E., 2009. Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. ISME J. 3, 13–30. http://dx.doi.org/10.1038/ismej.2008.81.

Neter, J., 1996. Applied Linear Regression Models. third ed. Irwin, Chicago, IL (xv720 pp.).

Nogales, B.M., Ma, M., Aguiló-Ferretjans, C., Martín-Cardona, J., Lalucat, J., et al., 2007. Bacterial diversity, composition and dynamics in and around recreational coastal areas. Environ. Microbiol. 9, 1913–1929.

Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., et al., 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat. Methods 6, 639–641.

Read, D.S., Gweon, H.S., Bowes, M.J., Newbold, L.K., Field, D., Bailey, M.J., et al., 2015. Catchment-scale biogeography of riverine bacterioplankton. ISME J. 9, 516–526. http://dx.doi.org/10.1038/ismej.2014.166.

Reeder, J., Knight, R., 2009. The 'rare biosphere': a reality check. Nat. Methods 6, 636–637.

Rice, W., 2005. Staff Report on Bacterial Indicator Total Maximum Daily Loads in the Middle Santa Ana River Watershed. Santa Ana Regional Water Control Board, Santa Ana, CA.

Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K., et al., 2011. Evidence for a core gut microbiota in the zebrafish. ISME J. 5, 1595–1608.

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. http://dx.doi.org/10.1111/ele.12499.

RWQCB, Regional Water Quality Control Board, 2005. Resolution amending the water quality control plans for the Santa Ana River Basin to incorporate bacterial indicator total maximum daily loads (TMDLs) for middle Santa Ana River Watershed Water bodies. Regional Board Resolution R8-2005-0001.

Savichtcheva, O., Okayama, N., Okabe, S., 2007. Relationships between bacteroides 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. Water Res. 41, 3615–3628.

Savichtcheva, 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.

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. http://dx.doi.org/10.1111/1462-2920.12886 (Epub ahead of print).

SAWPA: Santa Ana Watershed Project Authority, 2013. Middle Santa Ana River bacterial indicator TMDL 2013 dry season report. Prepared by CDM Smith on behalf of SAWPA and the Middle Santa Ana River Watershed TMDL Task Force (July).

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., et al., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Schloss, P.D., Handelsman, J., 2006. Introducing treeclimber, a test to compare microbial community structures. Appl. Environ. Microbiol. 72, 2379–2384.

Sheneman, L., Evans, J., Foster, J.A., 2006. Clearcut: a fast implementation of relaxed neighbor joining. Bioinformatics 22, 2823–2824.

USEPA, 1986. Ambient water quality criteria for bacteria – 1986. EPA Office of Water, Washington, DC. EPA 440/5–84-002.

Urakawa, H., Tsukamoto, K.K., Ohwada, K., 1999. Microbial diversity in marine sediment from Sagami Bay and Tokoyo Bay, Japan, as determined by 16S rRNA gene analysis. Microbiology 145, 3305–3315.

Van Rossum, T., Peabody, M.A., Uyaguari-Diaz, M.I., Cronin, K.I., Chan, M., Slobodan, J.R., Nesbitt, M.J., Suttle, C.A., Hsiao, W.W.L., Tang, P.K.C., Prystajecky, N.A., Brinkman, F.S.L., 2015. Year-long metagenomic study of river microbiomes across land use and water quality. Front. Microbiol. 6, 1405. http://dx.doi.org/10.3389/fmicb.2015.01405.

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, 555–561. http://dx.doi.org/10.1038/nature09440.

Wang, Y., Yang, J., Liu, L., Yu, Z., 2015. Quantifying the effects of geographical and environmental factors on distribution of stream bacterioplankton within nature reserves of Fujian China. Environ. Sci. Pollut. Res. 22, 11010–11021. http://dx.doi.org/10.1007/s11356-015-4308-y.

Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., et al., 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl. Environ. Microbiol. 78, 8264–8271.

Wassel, R.A., Mills, A.L., 1983. Changes in water and sediment bacterial community structure in a lake receiving acid-mine drainage. Microb. Ecol. 9, 155–169. http://dx.doi.org/10.1007/bf02015128.

Weary, N., Lhoutellier, C., Ducray, F., Delgenes, J.P., Godon, J.J., 2008. Behaviour of pathogenic and indicator bacteria during urban wastewater treatment and sludge composting, as revealed by quantitative PCR. Water Res. 42, 53–62.

Wu, G.D., Lewis, J.D., Hoffmann, C., Chen, Y.-Y., Knight, R., et al., 2010. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. 10, 206.

Zinger, L., Gobet, A., Pommier, T., 2012. Two decades of describing the unseen majority of aquatic microbial diversity. Mol. Ecol. 21, 1878–1896. http://dx.doi.org/10.1111/j.1365-294X.2011.05362.x.

Zhang, R., Thiyagarajan, V., Qian, P.Y., 2008a. Evaluation of terminal-restriction fragment length polymorphism analysis in contrasting marine environments. FEMS Microbiol. Ecol. 65, 169–178.

Zhang, W., Ki, J.S., Qian, P.Y., 2008b. Microbial diversity in polluted harbor sediments: I. Bacterial community assessment based on four clone libraries of 16 s rDNA. Estuar. Coast. Shelf Sci. 76, 668–681.

Zhou, J., Xue, K., Xie, J., Deng, Y., Wu, L., et al., 2012. Microbial mediation of carbon-cycle feedbacks to climate warming. Nat. Clim. Chang. 2, 106–110. http://dx.doi.org/10.1038/nclimate1331.