Journal of Great Lakes Research 44 (2018) 412–420

Contents lists available at ScienceDirect

Journal of Great Lakes Research

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

Microbial pollution characterization at a TMDL site in Michigan: Source identification

Huiyun Wu [a], Amira Oun [a], Ruth Kline-Robach [b], Irene Xagoraraki [a,*]

[a] Department of Civil and Environmental Engineering, Michigan State University, East Lansing 48824, USA
[b] Department of Community Sustainability, Michigan State University, East Lansing 48824, USA

ARTICLE INFO

Article history:
Received 6 September 2017
Accepted 20 February 2018
Available online 5 March 2018

Communicated by R. Michael McKay

Keywords:
E. coli
Bacteroides
Microbial source tracking
Whole genome shotgun sequencing

ABSTRACT

Communities throughout the Great Lakes basin are developing and implementing watershed management plans to address non-point sources of pollution and meet Total Maximum Daily Load (TMDL) requirements. Investigating sources of microbial contamination in key streams and creeks is critical for the development of effective watershed management plans. This work aims to present an approach that will facilitate source identification. In addition to conventional indicator analysis, the approach includes molecular analysis of species-specific markers and microbial community diversity analysis. We characterized microbial pollution in the Sloan Creek subwatershed in Ingham County MI, an impaired area, located in the Great Lakes Basin. To identify pollution sources (human or animal) and major sites of origin (tributaries with highest pollution loads) water samples were collected from three locations in the subwatershed representing the main creek upstream, main creek downstream, and tributary. A fecal indicator (*E. coli*) and host-specific human and bovine-associated *Bacteroides* genetic markers were quantified in all water samples. Results indicated that 54% of the samples from the three locations exceeded the recreational *E.coli* water quality guidelines. High concentrations of both human and bovine associated-*Bacteroides* indicated influence of multiple sources of fecal contamination. Statistical tests showed significantly different water characteristics between two of the sampling locations. Whole genome shotgun sequencing indicated fecal and sewer signatures, wastewater metagenome, human gut metagenome, and rumen gut metagenome in the water samples. Results suggested that probable sources of contamination were leakage from septic systems and runoff from agriculture activities nearby to Sloan Creek.

Introduction

Numerous water bodies in the Great Lakes Basin are impaired due to pollutants including bacteria (MDEQ, 2017). The Michigan Department of Environmental Quality (MDEQ) estimates that about half of Michigan's river miles were impaired due to elevated *E. coli* concentrations as of 2014 (MDEQ, 2014). Research into microbial contamination of beaches and drinking water intakes has typically been given priority over that of streams and creeks, since these pose a direct risk to human health (Almeida and Soares, 2012; Kistemann et al., 2002; Wong et al., 2009). However, investigating sources of microbial contamination in key streams and creeks is critical for the development of comprehensive watershed management plans.

The goal of watershed planning is to restore and protect water quality (USEPA, 2008). Watershed management plans developed with Clean Water Act Section 319 Nonpoint Source funds must address nine elements, including an identification of sources and causes of pollutants (USEPA 2008). This work aims to present an approach that will allow the identification of sources of microbial pollutants. The approach includes site-specific sampling, conventional indicator analysis, molecular analysis of species-specific markers, and microbial community analysis. The approach was applied in the Red Cedar River Watershed in central, lower Michigan, where *E. coli* was identified as a primary pollutant of concern. Because high bacteria concentrations may impair designated uses of the river, a Total Maximum Daily Load (TMDL) was established by the MDEQ and a watershed planning process was initiated to address the elevated levels of bacteria. This study will help watershed managers better understand the sources of bacteria.

Indicator organisms, such as fecal coliforms and *E. coli*, are typically monitored to indicate the presence or absence of microbial contamination (Simpson et al., 2002; Stoeckel and Harwood, 2007). Using an indicator organism to identify fecal contamination problems in surface waters presents several challenges. Fecal coliforms and *E. coli* are not direct measures of fecal contamination because of the poor correlation with pathogens, and do not differentiate between human and animal pollution sources (Harwood et al., 2005; Lemarchand and Lebaron, 2003; Noble and Fuhrman, 2001; Pusch et al., 2005). Conducting a microbial source tracking (MST) study to identify the potential sources of

* Corresponding author.
E-mail address: xagorara@msu.edu (I. Xagoraraki).

https://doi.org/10.1016/j.jglr.2018.02.006
0380-1330/

pollution is an important component in the watershed planning process. After the pollutant sources and causes are identified appropriate best management practices can be selected to address the impairments.

An appropriate rapid MST method to distinguish human and non-human sources of contamination may incorporate the use of host-specific *Bacteroides* molecular markers. Recently *Bacteroides* species have been used to isolate specific markers and investigate land use and water quality impairments (Peed et al., 2011; Verhougstraete et al., 2015). A study conducted by Furtula et al. (2012) confirmed ruminant, pig, and dog fecal contamination in an agricultural-dominated watershed (Canada) using *Bacteroides* markers. Another study by Verhougstraete et al. (2015) provided a water quality assessment for a large number of watersheds in Michigan and found that human fecal contamination was prevalent. Moreover, the quantitative polymerase reaction (qPCR) based method is culture-independent; therefore, the water quality result can be obtain within 4 h, and detection is more sensitive compared to the traditional cultural method (Layton et al., 2006).

High-throughput sequencing metagenomics methods have been shown as promising MST tools because they are able to scrutinize hundreds to thousands of microbes at one time (Field and Samadpour, 2007; Li et al., 2015; Staley et al., 2013; Uyaguari-Diaz et al., 2016; Wang et al., 2016). Previously researchers have studied microbial metagenomes from different environments and some microbial signatures have been proposed to identify the source of microbial contamination (Fisher et al., 2015; Newton et al., 2013, 2015). Most of these sequencing studies applied to MST investigate microbial communities in surface water by analyzing 16S rRNA amplicons. Whole genome shotgun sequencing (WGS) was able to provide bacterial metagenome profile information to identify microbial sources in water samples from different watersheds with higher sensitivity as compared to 16S rRNA amplicon sequencing (Venter et al., 2004). Currently, there are only a few surface water metagenome studies from the Great Lakes basin region (Fisher et al., 2015; Shanks et al., 2013). This paper aims to characterize bacterial contamination in the Sloan Creek sub-watershed, a TMDL site located in the Great Lakes basin; track microbial contamination by studying water quality in upstream tributaries; and identify the dominant source of microbial contamination, animal vs. human by studying host specific molecular markers and by analyzing the water-borne metagenomic profile.

Material and methods

Site description

The Red Cedar River flows about 50 miles through rural and agricultural land in the south-central lower peninsula of Michigan. The Red Cedar drains into the Grand River and subsequently Lake Michigan. For this study, the Sloan Creek sub-watershed of the Red Cedar River watershed in Ingham County was selected for investigation due to elevated *E. coli* concentrations that exceed the Michigan water quality standards for total and partial body contact (Ingham Conservation District, 2012; MDEQ 2014). Sloan Creek is a tributary of Red Cedar River, which receives drainage from 19 mi^2 of the Sloan Creek sub-watershed (agricultural, rural, and suburban land use). The MDEQ ranked this sub-watershed as a top priority subgroup in the TMDL area based on their stressor analysis. There are two main streams, Sloan Creek and Button Drain, within the sub-watershed and the two streams drain agricultural and residential areas into the Red Cedar River.

Fig. 1 shows the Red Cedar River watershed and Sloan Creek sub-watershed. Button Drain is a tributary of Sloan Creek. Water samples were collected at three sites: Sloan Creek upstream, Sloan Creek downstream, and Button Drain. In Fig. 1, the Red Cedar River watershed and Sloan Creek sub-watershed are defined and characterized with Esri ArcMap GIS (10.3 version). The National Hydrography Dataset (NHD) from USGS was used for the channel and stream network. The USGS National Elevation Dataset (NED), with 30-m resolution was used for the Digital Elevation Model (DEM) for slope and surface runoff direction estimation.

Land use data percentages are calculated based on 30-meter resolution National Land Cover Database (NLCD 2011; http://www.mrlc.gov/nlcd11_data.php). The Land Cover NLCD Classification System includes 16 thematic classes and these were reclassified using the Anderson Land Use/Land Cover Classification system, into five land cover categories (Table 1). Agriculture, shrub land and forest comprise the majority of the studied area; therefore, the Sloan Creek sub-watershed was classified as a rural and agriculturally dominated area.

According to the Red Cedar River Watershed Management Plan (MSU Institute of Water Research, 2015), the Sloan Creek sub-watershed included a human population of 2127, living at a density of 112 people per square mile. About 393 homes were estimated to be using septic systems. This sub-watershed has an estimated 3080 large animals, including 3000 cows, 40 horses and 40 pigs, sheep, goats and alpacas. Most of the cows are housed at a Concentrated Animal Feeding Operation (CAFO), although smaller farms were also present. Large animal density was estimated to be 174 animals per square mile, the highest of any of the Red Cedar River sub-watersheds. Excluding the CAFO, there were an average of 10 animals per farm, and 12 animals per square mile. Suspected sources of bacteria in the sub-watershed include human, agricultural, and wildlife inputs. There were no known point-source sewage inputs to Sloan Creek or Button Drain, but both streams were reported to have animal and human nonpoint sources.

Water sample collection and processing

A comprehensive six-month sampling scheme was designed to collect samples at least twice per week during spring and summer 2015, from March 22nd to August 26th. A total of 192 samples (64 from each sampling location) were collected. Compared with a synoptic sampling scheme, this approach can capture the change of water quality under different flow conditions to address pollution sources.

Two tributaries, Sloan Creek and Button Drain, within the sub-watershed were selected for sampling. Three sites within the Sloan Creek sub-watershed were monitored for the presence of *E. coli* to assess microbial water quality. The Button Drain site was located in Button Drain, the Sloan Creek upstream site was located in the upstream of Sloan Creek, and the Sloan Creek downstream site was located after the intersection with Button Drain, before the confluence with the Red Cedar River (Fig. 1). All sampling sites were located at bridge crossings. Sampling sites were determined based on the watershed elevation slope and ease of access.

Grab samples were collected in sterile one-liter bottles, which were autoclaved at the lab and rinsed three times with sampled water before use. Two water samples were collected at each location, one for *E. coli* analysis and one for *Bacteroides* analysis. Samples were stored on ice and processed in the Water Quality Laboratory at Michigan State University (MSU) within 2 to 4 h of collection.

E. coli analysis

Water samples were analyzed for *E. coli* concentration using the defined substrate method Colilert-18™ Quanti-Tray 2000 (IDEXX Laboratories, Inc.) within 4 h of collection. *E. coli* were measured in duplicate directly or diluted with phosphate-buffered saline solution (PBS) (pH = 7.2) to three serial dilutions 10^0, 10^{-1}, and 10^{-2}. The mean value was taken for the final concentration from the three dilutions. Each sample was mixed with reagent, shaken 10 times, and poured into the tray. Samples were incubated at 35 °C (±0.5 °C) for 24 h (±2 h). Microbial enumeration was conducted following the manufacturer's protocol; fluorescent wells were reported positive for *E. coli* as Most Probable Number (MPN) per 100 mL. Sterile PBS was used as negative control to verify method integrity. The detection limit was 1 MPN/100 mL.

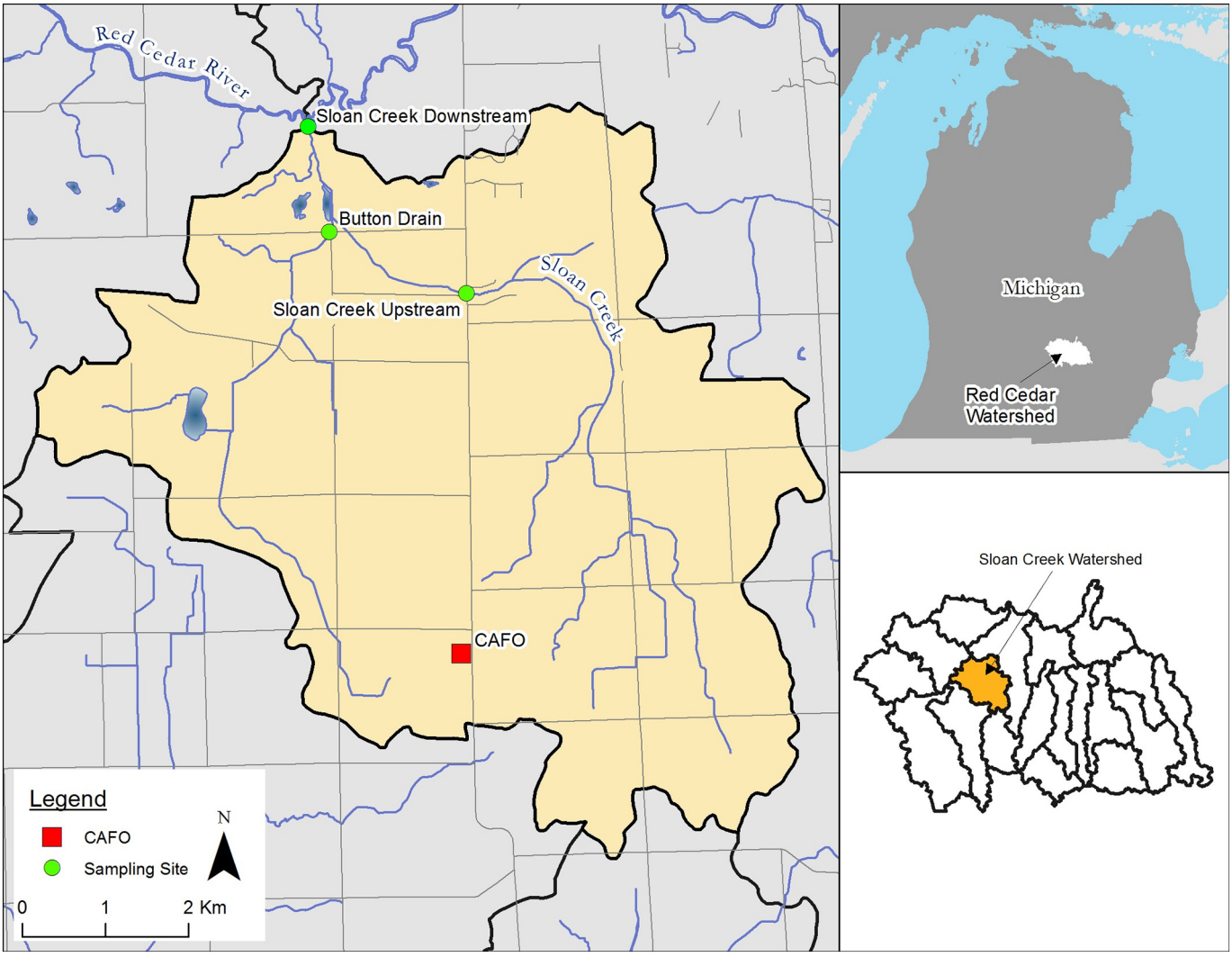

Fig. 1. Red Cedar River watershed, Sloan Creek sub-watershed, and sampling sites. Sloan Creek sub-watershed has a relatively low relief with the maximum elevation recorded as 326 m and a minimum of 249 m.

Molecular analysis

All water samples were tested for human and bovine-associated *Bacteroides* molecular markers quantitatively using qPCR (Layton et al., 2006; Yampara-Iquise et al., 2008). A total of 500 mL of water sample were filtered through 0.45 μm hydrophilic mixed cellulose esters filter (Pall Corporation 66278) under partial vacuum. The filter was placed into a 50 mL sterile disposable centrifuge tube containing 45 mL of sterile phosphate buffered saline PBS, vortexed on high for 10 min, and then centrifuged (30 min; 4500×*g*; 20 °C) to pellet the cells. Samples were concentrated down to 2 mL by decanting 43 mL from the tube and the remaining pellets were stored at −80 °C until DNA could be extracted. After thawing samples, 100 μL of DNA was extracted from 400 μL of pellet using MagNa Pure Compact System automatic machine (Roche Applied Sciences, Indianapolis, IN) with the corresponding kit (MagNA Pure Compact Nucleic Acid Isolation Kit I). Two host-specific qPCR methods were utilized to identify and quantify sources of fecal pollution within the subwatershed. The primers and probes used are listed in Table 2.

All qPCR quantification analyses were carried out with LightCycler® 1.5 Instrument (Roche Applied Sciences, Indianapolis, IN) and LightCycler 480 Probes Master kit with a total reaction volume of 20 μL. Analysis for bovine-associated *Bacteroides* (BoBac) marker was performed according to Layton et al. (2006). Each BoBac assay was carried out with 10 μL of LightCycler 480 probe Mastermix (Roche Applied Sciences), 1 μL forward and reverse primers, 0.4 μL probe, 2.6 μL nuclease-free water, and 5 μL of extracted DNA and processed in triplicate. The qPCR analyses included a 2 min, 50 °C, and 10 min 95 °C pre-incubation cycle, followed by 50 amplification cycles (30 s, 95 °C and 45 s, 57 °C), and a 0.5 min 40 °C cooling cycle. Analysis for human-associated *Bacteroides* (HuBac) marker was performed according to Yampara-Iquise et al. (2008). Each HuBac assay was performed with 10 μL of LightCycler 480 probe Mastermix (Roche Applied Sciences), 0.4 μL forward and reverse primers, 0.2 μL probe, 4 μL nuclease-free water, and 5 μL of extracted DNA for template and processed in triplicate. The qPCR analyses consisted of a 10 min, 95 °C pre-incubation cycle, followed by 45 amplification cycles (15 s, 95 °C; 60 s, 60 °C; and 5 s, 72 °C), and a 0.5 min 40 °C cooling cycle. A diluted plasmid standard

Table 1
Land use in the study area.

Watershed	Watershed land use percentage (NLCD 2011)				
	Agriculture (%)	Shrub (%)	Developed (%)	Forest (%)	Water and wetland (%)
Red Cedar River	35	23	18	10	14
Sloan Creek	45	27	9	11	8

Table 2
Primer and probes used in real-time PCR assays to detect *Bacteroides* genetic markers.

	Forward	Reverse	Probe	Reference
Human-associated *Bacteroides* HuBaciotaomicron α-1–6 mannanase (HuBac)	TCGTTCGTCAGCAGT-AACA	AAGAAAAAGGGACAGTGG	6FAM-ACCTGCTG-NFQ	Yampara-Iquise et al. (2008)
Bovine-associated *Bacteroides* 16srRNA (BoBac)	BoBac367f (GAAG(G/A) CTGAACCAGCCAAGTA)	BoBac467r (GCTTATTCATACGGT-ACATACAAG)	BoBac402Bhqf (TGAAGGATGAAGGTTCTATGGATTGTA-AACTT)	Layton et al. (2006)

was included during each qPCR run as a positive control and molecular grade water was used in place of DNA template for negative controls (Oun et al., 2017). One copy of the targeted *Bacteroides* gene is assumed to be present per cell and thus one gene copy number corresponded to one equivalent cell. The crossing point (Cp) value for each qPCR reaction was automatically determined by the LightCycler® Software 4.0. Gene copies were then converted to and reported as copies/100 mL.

In order to prepare the standard curves to quantify the gene numbers, the DNA was extracted from ATCC (number 29148D-5) genomic DNA for Human-specific *Bacteroides* genetic marker (HuBac), and from bovine feces obtained from Michigan State University dairy farm for Bovine-specific *Bacteroides* genetic marker (BoBac). The amplified PCR products for the target genes were cloned into one shot chemically competent *E. coli* using TOPO TA Cloning kit for Sequencing (Invitrogen Inc., Carlsbad, CA, USA) according to the protocol provided by the manufacturer. Plasmids were extracted with QIAprep Spin MiniPrep kit (Valencia, CA, USA) and were sequenced at the Research Technology Support Facility (RTSF) at Michigan State University to confirm the insertion of the target inside the vector. The DNA concentration in plasmids was quantified using Qubit Fluorometric Quantitation (Thermo Fisher Scientific) and then serially diluted ten-fold to construct qPCR standard curves. Triplicates of dilutions ranging from 10^8 to 10^0 were used for the standard curve.

Statistical analysis

Student's *t*-tests were used to determine the differences in mean concentrations of target organisms, to compare the difference between sampling sites (Figs. 2 and 4) and to compare the difference between BoBac and HuBac concentration in each sampling site (Fig. 3). Descriptive statistical analyses were performed using SPSS Statistics software (Version 22) with a significance threshold $\alpha = 0.05$. *E. coli* concentration were analyzed based on the original data; whereas BoBac and HuBac concentrations were $\log_{10}$-transformed to achieve normality and zero copy/100 mL was converted to 0 $\log_{10}$ copy/100 mL manually. The *t*-test was two-tailed, and the probability of rejecting the null hypothesis when it is true, set at $p < 0.05$.

Microbial community analysis

Three samples, specifically August 15th, 18th, 19th of 2015 from the Sloan Creek downstream site were processed for Whole Genome Shotgun Sequencing (WGS). These samples were chosen since there was a spike in *E. coli* concentration on August 15th, BoBac was detected on August 18th, and HuBac was detected on August 19th. The DNA extracts of the three samples were purified and sequenced on an Illumina platform (Illumina Miseq, Roche Technologies) at the Research Technology Support Facility (RTSF) at MSU. DNA-Seq libraries were prepared using the Rubicon Genomics ThruPLEX DNA-seq Kit. After preparation, libraries underwent quality control and were quantified using Qubit double-stranded DNA (dsDNA), Caliper LabChipGX and Kapa Biosystems Library Quantification qPCR kit. The libraries were pooled together, which was loaded on an Illumina MiSeq v2 standard flow cell. Sequencing was done in a 2 × 250 base pairs (bp) format with a v2 500 cycles reagent cartridge. Base calling was performed by Illumina Real Time Analysis (RTA) v1.18.54 and output of RTA was demultiplexed and

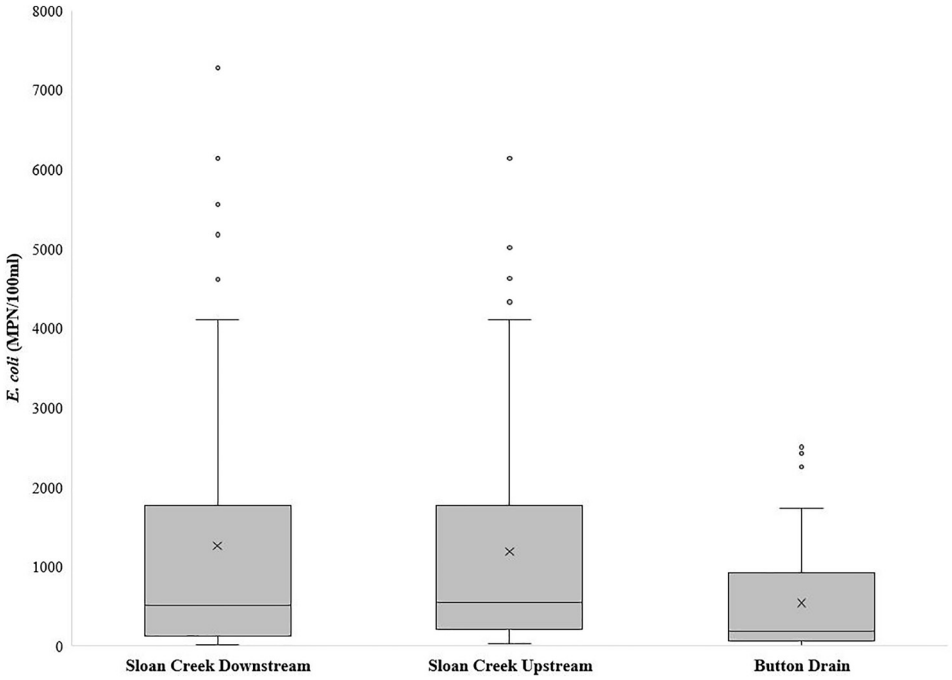

Fig. 2. *E. coli* distribution in the three sampling sites. The box plots displayed the range, quartiles, mean and outliers of the concentration in the sampling sites, $n = 64$ per site.

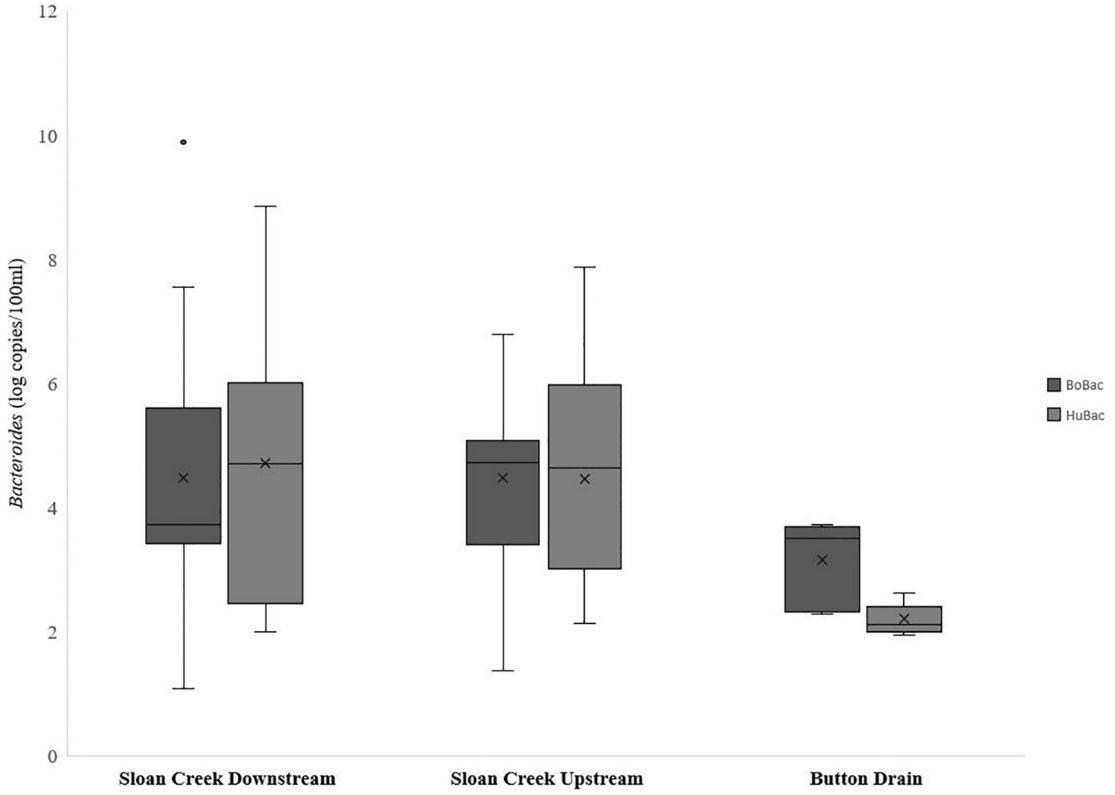

Fig. 3. Boxplots of bovine-associated *Bacteroides* (BoBac) and human-associated *Bacteroides* (HuBac) in three sampling sites, classified by sampling locations. The *Bacteroides* concentration-positive data were shown in the boxplots, and they were log-transformed.

converted to FastQ format with Illumina Bcl2fastq v1.8.4 to produce raw sequencing data.

The raw sequencing data from Illumina sequencer were initially processed using the flexible read-trimming tool (Trimmomatic) to trim the low quality reads and remove adapters (Bolger et al., 2014). Sequences <30 bp were discarded. The trimmed reads were assembled into contiguous datasets (contigs) so as to reduce the chances of false positive detection using an iterative de novo assembler IDBA-UD (Peng et al., 2012). The assembled contig datasets were aligned with NCBI RefSeq release bacteria database (downloaded in February of 2017) and env-nt database (downloaded in May of 2017) using the Basic Local Alignment Search Tool (BLAST) on MSU High Performance Computing Cluster (HPCC) platform for microbial taxonomy annotation. The BLAST outcome data sets were displayed using the software Metagenome Analyzer (MEGAN) with an evaluate cutoff of $1e^{-5}$ for microbial community analysis (Huson et al., 2007).

Results and discussion

E. coli levels

During the study period, Sloan Creek continuously delivered water with high concentrations of *E. coli* to the Red Cedar River. *E. coli* was detected in 100% of all samples at all sampling sites, indicating the high risk of microbial contamination in the watershed. *E. coli* concentrations across the three sampling sites are shown in Fig. 2. The concentrations ranged widely during the sampling events. The single sample highest concentration of *E. coli* (7270 MPN/100 mL) was observed at the Sloan Creek downstream site and it occurred on June 18th 2015. *E. coli* concentrations also peaked on the same day at the Sloan Creek upstream site (6131 MPN/100 mL) and the Button Drain site (2500 MPN/100 mL).

The geometric mean concentrations of *E. coli* during the entire six-month sampling period were 461, 516, and 189 MPN/100 mL for Sloan Creek downstream, Sloan Creek upstream, and Button Drain site. If Michigan's Total Body Contact daily maximum geometric mean of 300 MPN/100 mL is used, then 54% of the total number of collected samples in all three sites (192) exceeded the state of Michigan water quality guidelines for presence of *E. coli* (MDEQ, 2016). In particular, 59% (38 of 64) of Sloan Creek downstream samples, 67% (43 of 64) of Sloan Creek Upstream samples, and 36% (23 of 64) of Button Drain samples exceeded the water quality guidelines for the state of Michigan.

E. coli was also detected in the three sequenced samples analyzed with Illumina. The number of *Escherichia* hits from metagenomic analysis was low, ranging from 8 to 16 hits. The *E. coli* concentrations measured using the Colilert method, in the same three samples, were relatively low ranging from 280 to 914 MPN/100 mL. This low detection in the water samples may be because they were collected at the end of the sampling period (August) when many of the microbial pollutants have been flushed out from the land surface during the high June and July storm events according to rainfall data (not shown here).

Human and bovine-associated Bacteroides levels

To distinguish between the contribution of human and animal sources, human-associated *Bacteroides* and bovine-associated *Bacteroides* genetic markers were quantified in each site (Fig. 3). The occurrence rate for BoBac was 27% in samples from Sloan Creek downstream site, 22%

Table 3
Occurrence rate of microbial contamination indicators in three sampling sites. BoBac is bovine-associated *Bacteroides*, HuBac is human-associated *Bacteroides*.

Sampling locations	*E. coli*	BoBac	HuBac
Sloan Creek downstream	100%	27%	25%
Sloan Creek upstream	100%	22%	27%
Button drain	100%	11%	14%

from Sloan Creek upstream site, and 11% from Button Drain site (Table 4). HuBac were present in 25%, 27% and 14% of the samples from Sloan Creek downstream site, Sloan upstream site, and Button Drain site respectively (Table 3). The lowest occurrence rates of HuBac and BoBac were observed in Button Drain.

Sloan Creek downstream and upstream sites shared similar range and mean *Bacteroides* concentration patterns (Fig. 4). The average concentration of BoBac marker was the highest at the Sloan Creek downstream site ($10^{4.5}$ genomic copies/100 mL), and the lowest at the Button Drain site ($10^{3.15}$ copies/100 mL). At the Sloan Creek upstream site, the average of the BoBac concentration was $10^{4.47}$ copies/100 mL. Similarly, the mean concentrations of HuBac marker were the highest ($10^{4.71}$ copies/100 mL) at the Sloan Creek downstream site, and lowest at the Button Drain site ($10^{2.2}$ copies/100 mL). At the Sloan Creek upstream site, the average of the HuBac concentration was $10^{4.46}$ copies/100 mL.

Statistical comparisons between sampling locations

In the Sloan creek sub-watershed, water flows from Sloan Creek upstream site and Button Drain site into the Sloan Creek downstream and confluence with the Red Cedar River. In general, the impact of the Sloan Creek upstream site was heavier than Button Drain to the Sloan Creek downstream site. *t*-Test results unveiled the relationship of microbial concentration levels between sampling sites. There was no significant difference between Sloan Creek downstream site or Sloan Creek upstream site in either *E. coli* concentrations (*t*-test $t = 0.258$, df = 126, $p = 0.797$), BoBac concentrations (*t*-test $t = 0.039$, df = 29, $p = 0.970$), or HuBac concentrations (*t*-test $t = 324$, df = 33, $p = 0.748$) between Sloan Creek downstream and Sloan Creek Upstream sites (Figs. 2 and 4). However, Sloan Creek downstream site had significantly higher concentrations than Button Drain site in *E. coli* concentrations (*t*-test $t = 3.072$, df = 126, $p = 0.002$), BoBac (*t*-test $t = 1.618$, df = 22, $p = 0.028$) and HuBac (*t*-test $t = 3.005$, df = 24, $p = 0.000$) (Figs. 2 & 4). Therefore, Sloan Creek upstream had more impact on Sloan Creek downstream microbial levels than Button Drain did.

Statistical comparisons between human and bovine-associated Bacteroides levels

t-Test results showed that there was no statistically significant difference between BoBac and HuBac concentrations in Sloan Creek downstream site (*t*-test, $t = -0.3$, df = 31, $p = 0.766$) and Sloan Creek upstream sampling sites (*t*-test, $t = 0.024$, df = 29, $p = 0.981$) (Fig. 3). However BoBac concentrations were higher than HuBac concentrations in the Button Drain site, and the difference was statistically significant (*t*-test, $t = 4.23$, df = 14, $p = 0.006$). Therefore, animal and human feces both affected Sloan Creek upstream and Sloan Creek downstream, whereas bovine feces were a major source of pollution in the Button Drain. The Sloan Creek upstream site receives surface runoff from agricultural activities, a concentrated animal feeding operation, and suspected leaking septic tanks, which may lead to microbial contamination from both animal source and human source, as confirmed by our results.

The level of microbial contaminant concentrations may be predicted according to the types of land use in a watershed. Understanding the influences of land use on surface water quality, especially on sources of fecal contaminants, is critical for effective watershed planning and management efforts. A study showed that bacteria concentrations in storm water run-off were highest in recreational, agricultural and urban areas, while open-space land use types had the lowest bacteria concentrations (Tiefenthaler et al. 2008). Our study showed that a small watershed with low urban land use (9% in Sloan Creek sub-watershed) was sufficient to generate high microbial pollutant concentrations and degrade water quality. Researchers have shown human fecal contamination was prevalent in a large number of Michigan watersheds by

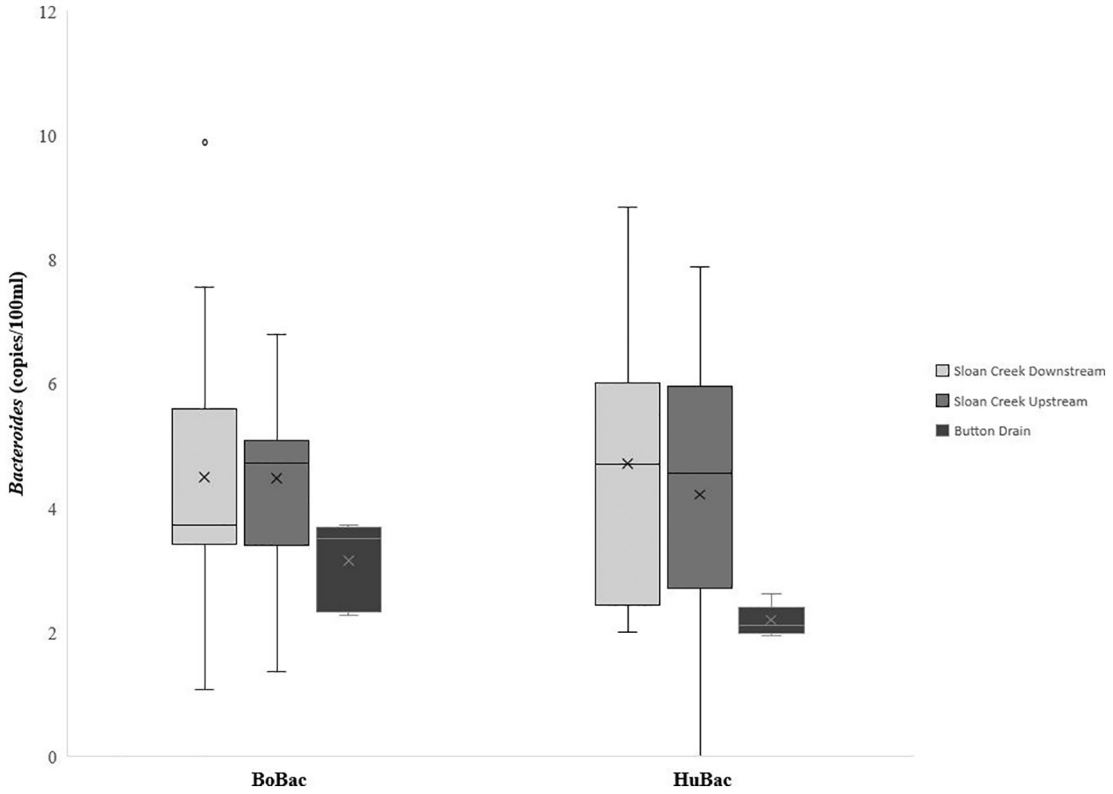

Fig. 4. Human-specific *Bacteroides* (HuBac) and Bovine-specific *Bacteroides* (BoBac) concentrations in the three sampling sites, classified by host-specific *Bacteroides*. The *Bacteroides* concentration-positive data were shown in the boxplots, and they were log-transformed.

Table 4
Metagenome analysis statistics (MEGAN), using NCBI RefSeq bacteria database. The environmental sequences that cannot be found in the reference database were assigned to "not hits". The environmental sequences that cannot meet the algorithm threshold were assigned to "not affiliated."

	Number of contigs	Bacteria	%	Not affiliated	%	Not hits	%
August 15th	105,590	48,633	46.06%	1595	1.51%	55,304	52.38%
August 18th	117,559	68,744	58.48%	1054	0.90%	47,708	40.58%
August 19th	81,874	36,169	44.18%	151	0.18%	45,552	55.64%

detecting human-specific *Bacteroides* (Verhougstraete et al., 2015). High levels of HuBac were detected in all three sampling sites in our study, which indicates the potential for leaking septic tanks in the sub-watershed. In addition, the studied watershed had a significant agricultural land use (45%) and a CAFO in the upstream of Sloan Creek. Indeed, high levels of BoBac were detected in the three sampling sites originating from surface runoff from agricultural land.

Microbial community diversity analysis and relationships with pollution sources

Microbial community analysis can help in the identification of pollution sources. In this study, shotgun sequencing in Illumina Miseq platform was performed and the results were recovered and analyzed with BLAST and MEGAN using the NCBI RefSeq bacteria database. The Guanine-Cytosine (GC) content of the trimmed sequences was about 48% and the average length was 40–250 base pairs. The number of total contigs recovered from August 15th, 18th, and 19th 2015 water samples were 105,590, 117,559, and 81,874, and about 50% of the contigs can be annotated as bacteria (Table 4). Contigs were annotated to >13 phyla, which were dominated by members of *Proteobacteria*, *Actinobacteria*, *Bacteroides*, and *Cyanobacteria* (Fig. 5).

Characterization of the overall water microbial community can be used directly in MST for discerning waste and fecal source (Unno et al., 2010; McLellan et al., 2010). Human-fecal related bacteria such as *Prevotellaceae*, *Porphyromonadaceae*, *Coriobacteriaceae*, *Lachnospiraceae*, and *Ruminococcaceae* are commonly and abundantly present in most surveyed wastewater treatment plants (sewage) across the US (Li et al., 2015; McLellan et al., 2013; Newton et al., 2013), and they were present in sequenced water samples in our study. *Acinetobacter*, *Arcobacter*, and *Trichococcus* have been suggested as signatures of sewer contamination (Newton et al., 2013), and they were present in the sequenced water samples from Sloan Creek. *Betaproteobacteria*, *Gammaproteobacteria*, *Clostridiales* and *Verrucomicrobia* present abundantly in human feces (Dubinsky et al., 2012), and they have high abundance in the sequenced water samples in this study. Overall, these results indicate the presence of sewage and fecal signatures in the Sloan Creek sub-watershed.

In addition, *Clostridia*, *Bacilli* and *Bacteroidetes* are related to ruminant gut (Dubinsky et al., 2012; Unno et al., 2010). Grazer identifier included a variety of Clostridia from cattle rumen such as *Clostridium* and *Ruminococcus* (Dubinsky et al., 2012). These identifiers presented in the sequenced samples from Sloan Creek, indicating the sub-watershed was affected by fecal contamination from ruminant animals.

The microbial community may shift when the land usage changes. To access shifts in microbial community multiple samples need to be collected and analyzed. In this study only limited samples were sequenced, therefore the results can only indicate the potential presence of pollution and not shifts in microbial diversity. The Sloan Creek sub-watershed is an agricultural-urban mixed watershed. *Polynucleobacter*, *Arcobacter*, *Methylotenera*, *Flavobacterium*, *Pseudomonas*, and *Bacteriodes* were ubiquitous in the sequenced water samples, with abundance higher than that of *E. coli*. This pattern was similar in a urban and agricultural watershed (Uyaguari-Diaz et al., 2016).

In addition to the NCBI RefSeq bacteria database, the results were also analyzed using the NCBI_nt database. The number of contigs that could be assigned to environmental metagenomes were 52,348, 69,186, and 33,821, for August 15th, 18th, and 19th 2015 water samples

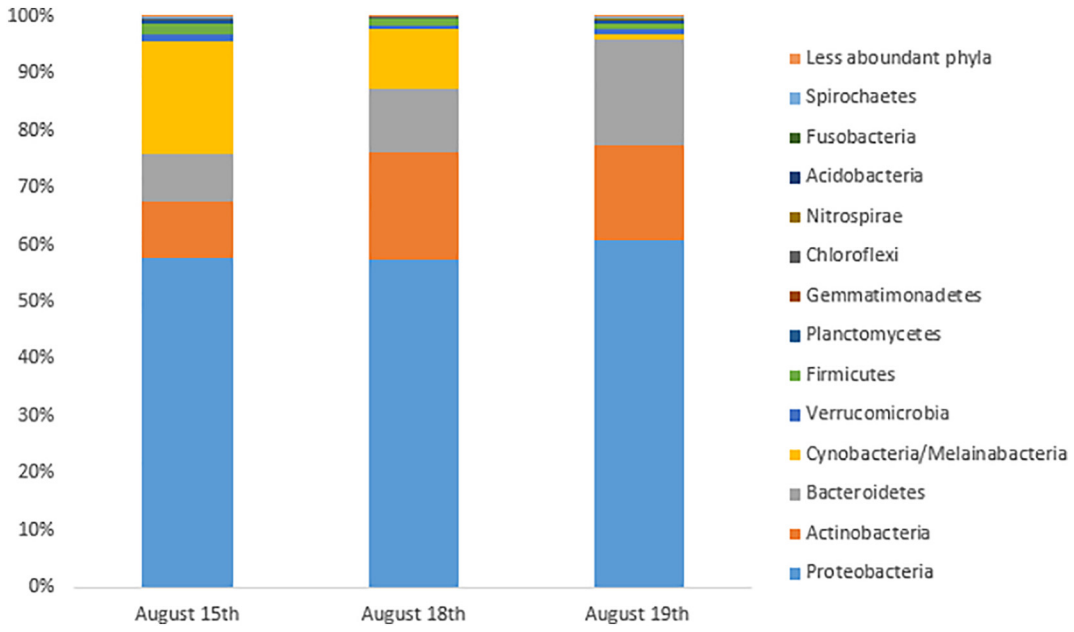

Fig. 5. Microbial community distribution in water samples from Sloan creek (using NCBI RefSeq bacteria database). Samples were from the Sloan Creek downstream site and were obtained on August 15th, 18th, and 19th 2015. Samples were analyzed with whole genome sequencing, results were analyzed with BLAST and MEGAN. Approximately 50% of the total contigs can be annotated.

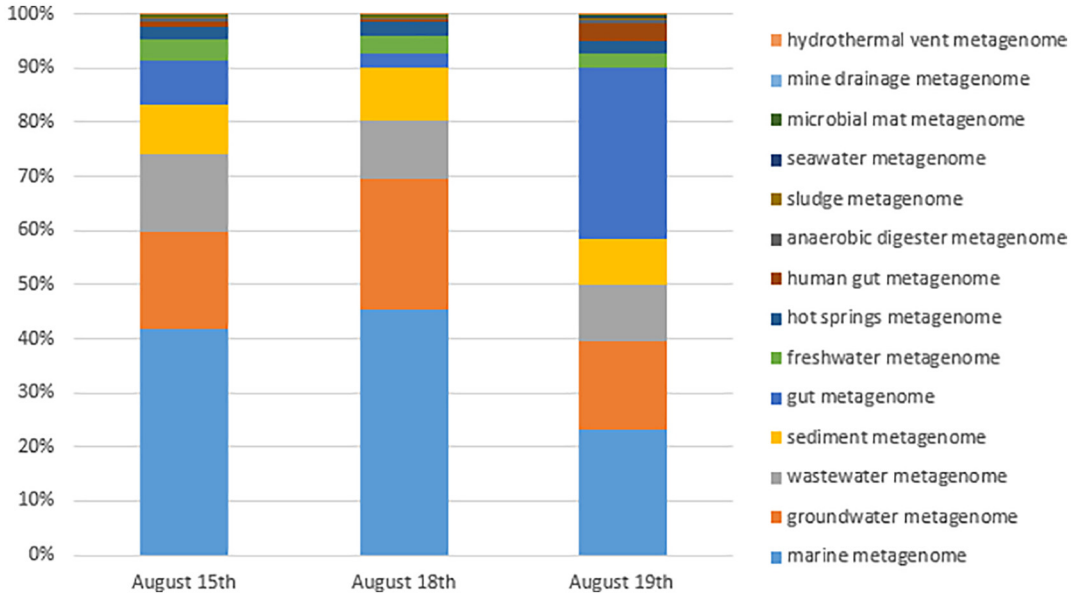

Fig. 6. Characterized metagenomes for the water samples from Sloan creek (using NCBI Env_nt database). The displayed data were retrieved from the annotated contigs, and it was about 30% of the total annotated contigs for the sequenced samples. Samples were from the Sloan Creek downstream site and were obtained on August 15th, 18th, and 19th 2015. Samples were analyzed with whole genome sequencing, results were analyzed with BLAST and MEGAN. Approximately 50% of the total contigs can be annotated, and 30% of the annotated contigs can be further characterized to specific environmental metagenomes.

respectively. Approximately 30% of the assigned contigs could be further annotated to 14 specific metagenomes, which can help to identify the source of microbial contamination (Fig. 6). In particular, 1728, 1533, and 1259 contigs were annotated as wastewater or sewage metagenome. A total of 989, 347, and 3873 were annotated as gut metagenome, mainly rumen gut metagenome (Yutin et al., 2015) respectively for the three samples. The results indicate that human gut and sludge metagenomes were present in the water samples.

To the best of our knowledge, this is the first study in the Great Lakes Basin that analyzed the freshwater microbiome by whole genome sequencing. The detailed microbial community information in the impacted watershed may help to develop new fecal indicators and source signatures in surface water for future studies in the Great Lakes region. Even though the microbial community in the water samples can be annotated using metagenomic analysis, there was a significant part of the contigs that could not be assigned based on NCBI reference databases (about 50%). Based on the classification of the NCBI ENV-NT database, only about 30% of the assigned contigs could be annotated. Moreover, in this study only three water samples were sequenced; therefore, there is a limitation to directly comparing the microbial community of the water samples to the suspected fecal sources in the sub-watershed. Consequently, the metagenomic analysis results can only serve as supportive evidence in addition to qPCR and traditional culture methods.

Conclusions

The high exceedance rate of *E. coli* standards confirmed that the Sloan Creek sub-watershed was under a high risk of microbial contamination. Microbial pollutant concentrations in the upstream Sloan Creek samples and downstream Sloan Creek samples were significantly correlated. Statistical analysis indicates that the main source of microbial contamination originated from the upstream of Sloan Creek, and the impact from the Button Drain tributary was weak. Pollutants are not equally distributed in a watershed; therefore, identifying critical locations in space where the majority of pollution is released, such as tributaries with high impact, is important for remediation purposes.

The detection of both bovine and human-associated *Bacteroides* markers revealed that there were contaminant inputs from animal and human sources. Moreover, microbial community analysis detected fecal and sewer signatures and environmental metagenomes such as wastewater, sludge, human gut, and rumen gut in the water samples. The metagenomic analysis suggested that the major sources of contaminants originated from human and animal feces and sewage. The results indicate that microbial contaminants in the watershed may originate from agricultural activities, concentrated animal feeding operations and leaking septic tanks. However, the metagenomic analysis results are only based on a limited number of samples and are not confirmed since wastewater, fecal and manure samples were not analyzed for comparison purposes.

This mixed-use watershed study provided multiple lines of evidence to identify the sources and location of fecal pollution in Sloan Creek using host-specific markers and whole genome sequencing microbial community analysis. This study offered a proposed MST methodology path to assess water quality at a sub-watershed level that will facilitate watershed management planning efforts for a TMDL site.

Acknowledgments

This work was funded by USGS project 2015MI234B. We thank Dr. Shi-han Shui, Department of Plant Biology, Michigan State University, for help with metagenomics analysis.

References

Almeida, C., Soares, F., 2012. Microbiological monitoring of bivalves from the Ria Formosa Lagoon (south coast of Portugal): a 20 years of sanitary survey. Mar. Pollut. Bull. 64, 252–262.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Dubinsky, E.A., Esmaili, L., Hulls, J.R., Cao, Y., Griffith, J.F., Andersen, G.L., 2012. Application of phylogenetic microarray analysis to discriminate sources of fecal pollution. Environ. Sci. Technol. 46:4340–4347. https://doi.org/10.1021/es2040366.

Field, K.G., Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 41:3517–3538. https://doi.org/10.1016/j.watres.2007.06.056.

Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elem. Wash. DC 3.

Furtula, V., Osachoff, H., Derksen, G., Juahir, H., Colodey, A., Chambers, P., 2012. Inorganic nitrogen, sterols and bacterial source tracking as tools to characterize water quality and possible contamination sources in surface water. Water Res. 46, 1079–1092.

Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R., Rose, J.B., 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl. Environ. Microbiol. 71, 3163–3170.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386.

Ingham Conservation District, 2012. 2012 Natural Resource Assessment.

Kistemann, T., Claben, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J., Vacata, V., Exner, M., 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl. Environ. Microbiol. 68, 2188–2197.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., Sayler, G., 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl. Environ. Microbiol. 72, 4214–4224.

Lemarchand, K., Lebaron, P., 2003. Occurrence of *Salmonella* spp. and *Cryptosporidium* spp. in a French coastal watershed: relationship with fecal indicators. FEMS MicrobiolLett. 218, 203–209.

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. https://doi.org/10.1021/acs.est.5b00980.

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.

McLellan, S.L., Newton, R.J., Vandewalle, J.L., Shanks, O.C., Huse, S.M., Eren, A.M., Sogin, M. L., 2013. Sewage reflects the distribution of human faecal *Lachnospiraceae*: structure of *Lachnospiraceae* in sewage. Environ. Microbiol. 15:2213–2227. https://doi.org/10.1111/1462-2920.12092.

MDEQ, 2014. Water quality and pollution control in Michigan. (2014 Sections 303 (d), 305 (b), and 314 Integrated Report. MDEQ Report# MI/DEQ/WRD-14/001. http://www. michigan. gov/deq/0, 4561, 7-135-3313_3686_3728-12711–, 00. html.)

MDEQ, 2016, *E. coli* in surface waters ([WWW document], n.d. URL: http://www.michigan.gov/deq/0,4561,7-135-3313_3681_3686_3728-383659–,00.html, accessed 8.20.17) .

MDEQ, 2017 Water quality and pollution control in Michigan, 2016 (Sections 303(d), 305 (b), and 314 Integrated Report [WWW Document]. URL: http://www.michigan.gov/documents/deq/wrd-swas-ir2016-report_541402_7.pdf, accessed 12.25.17a).

MSU Institute of Water Research, 2015. Red Cedar River watershed management plan (No. MDEQ tracking code: #2011-0014).

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L., McLellan, S.L., 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb. Ecol. 65:1011–1023. https://doi.org/10.1007/s00248-013-0200-9.

Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M., Sogin, M.L., 2015. Sewage Reflects the Microbiomes of Human Populations. mBio 6, e02574-14. https://doi.org/10.1128/mBio.02574-14.

Noble, R.T., Fuhrman, J.A., 2001. Enteroviruses detected by reverse transcriptase polymerase chain reaction from the coastal waters of Santa Monica Bay, California: low correlation to bacterial indicator levels. The Ecology and Etiology of Newly Emerging Marine Diseases. Springer, pp. 175–184.

Oun, A., Yin, Z., Munir, M., Xagoraraki, I., 2017. Microbial pollution characterization of water and sediment at two beaches in Saginaw Bay, Michigan. J. Gt. Lakes Res. https://doi.org/10.1016/j.jglr.2017.01.014.

Peed, L.A., Nietch, C.T., Kelty, C.A., Meckes, M., Mooney, T., Sivaganesan, M., Shanks, O.C., 2011. Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. Environ. Sci. Technol. 45, 5652–5659.

Peng, Y., Leung, H.C., Yiu, S.-M., Chin, F.Y., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428.

Pusch, D., Oh, D.-Y., Wolf, S., Dumke, R., Schröter-Bobsin, U., Höhne, M., Röske, I., Schreier, E., 2005. Detection of enteric viruses and bacterial indicators in German environmental waters. Arch. Virol. 150, 929–947.

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl. Environ. Microbiol. 79, 2906–2913.

Simpson, J.M., Santo Domingo, J.W., Reasoner, D.J., 2002. Microbial source tracking: state of the science. Environ. Sci. Technol. 36, 5279–5288.

Staley, C., Unno, T., Gould, T.j., Jarvis, B., Phillips, J., Cotner, J.b., Sadowsky, M.j., 2013. Application of Illumina next-generation sequencing to characterize the bacterial community of the upper Mississippi River. J. Appl. Microbiol. 115:1147–1158. https://doi.org/10.1111/jam.12323.

Stoeckel, D.M., Harwood, V.J., 2007. Performance, design, and analysis in microbial source tracking studies. Appl. Environ. Microbiol. 73, 2405–2415.

Tiefenthaler, L.L., Stein, E.D., Schiff, K.C., 2008. Watershed and land use–based sources of trace metals in urban storm water. Environ. Toxicol. Chem. 27 (2), 277–287.

Unno, T., Jang, J., Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.-S., Chun, J., Hur, H.-G., 2010. Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environ. Sci. Technol. 44:7777–7782. https://doi.org/10.1021/es101500z.

USEPA, 2008. Handbook for developing watershed plans to restore and protect our waters. (EPA 841-B-08-002 - Google Search [WWW Document], n.d. URL https://www.google.com/search?q=USEPA%2C+2008.+Handbook+for+Developing+Watershed+Plans+to+Restore+and+Protect+Our+Waters.+EPA+841-B-08-002&oq=USEPA%2C+2008.+Handbook+for+Developing+Watershed+Plans+to+Restore+and+Protect+Our+Waters.+EPA+841-B-08-002&aqs=chrome..69i57j69i60.572j0j4&sourceid=chrome&ie=UTF-8, accessed 12.25.17).

Uyaguari-Diaz, M.I., Chan, M., Chaban, B.L., Croxen, M.A., Finke, J.F., Hill, J.E., Peabody, M.A., Van Rossum, T., Suttle, C.A., Brinkman, F.S.L., Isaac-Renton, J., Prystajecky, N.A., Tang, P., 2016. A Comprehensive Method for Amplicon-based and Metagenomic Characterization of Viruses, Bacteria, and Eukaryotes in Freshwater Samples. Microbiome 4. https://doi.org/10.1186/s40168-016-0166-1.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Verhougstraete, M.P., Martin, S.L., Kendall, A.D., Hyndman, D.W., Rose, J.B., 2015. Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. Proc. Natl. Acad. Sci. 112, 10419–10424.

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. https://doi.org/10.1016/j.scitotenv.2016.06.130.

Wong, M., Kumar, L., Jenkins, T.M., Xagoraraki, I., Phanikumar, M.S., Rose, J.B., 2009. Evaluation of public health risks at recreational beaches in Lake Michigan via detection of enteric viruses and a human-specific bacteriological marker. Water Res. 43, 1137–1149.

Yampara-Iquise, H., Zheng, G., Jones, J.e., Carson, C.A., 2008. Use of a *Bacteroides* thetaiotaomicron-specific α-1-6, mannanase quantitative PCR to detect human faecal pollution in water. J. Appl. Microbiol. 105:1686–1693. https://doi.org/10.1111/j.1365-2672.2008.03895.x.

Yutin, N., Kapitonov, V.V., Koonin, E.V., 2015. A new family of hybrid virophages from an animal gut metagenome. Biol. Direct 10. https://doi.org/10.1186/s13062-015-0054-9.