# MICROBES IN WATER: METHODS TO EVALUATE SOURCES, TEMPORAL VARIABILITY, AND POTENTIAL DISEASE SIGNALS

By

Huiyun Wu

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Environmental Engineering – Doctor of Philosophy

## ABSTRACT

# MICROBES IN WATER: METHODS TO EVALUATE SOURCES, TEMPORAL VARIABILITY, AND POTENTIAL DISEASE SIGNALS

#### By

#### Huiyun Wu

Microbial contamination in waters has been one of the most pressing environmental health concerns in the world and the US. Understanding sources, temporal variability, and transport pathways of microbes in water is critical for the development of watershed remediation and pollution prevention plans. Furthermore, understating the occurrence of pathogens in water and its relationship to human and animal health may facilitate early detection and prevention of disease. This dissertation focuses on the study of impaired water bodies in the Great Lakes basin. In particular, maximum pollutant loading and its relationship to hydrological events is studied, microbial pollution source identification methods are investigated, and methods for analyzing water samples to identify disease signals are explored. The selected sites, located in Michigan, are sites for which Total Maximum Daily Load (TMDL) requirement for bacteria have been identified by the state and watershed management plans are being developed.

#### ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Dr. Irene Xagoraraki for her extraordinary support and guidance throughout my PhD program at Michigan State University. Her foresighted research view has brought me into an interesting and promising field that is worth of lifetime endeavor; her detail-oriented work and research style has shaped me into a better researcher; her warm and kind heart has helped me out through many difficulties in my study and my life.

Furthermore, I would like to thank Dr. Tom Voice, and Dr. David Long, who were also my advisors in my Master's program, for providing me the opportunity to work with them in February 2014. Their expert advice and encouragement in my graduate life are invaluable to me. I would also like to extend the thanks Dr. Alison Cupples for her support in my learning process, as well as Dr. Shuguang Li, Dr. Phanikumar Mantha, and Dr. Yadu Pokhrel for being extraordinary faculty of record when I was a teaching assistant. I have to appreciate the work of staff in the department of Civil and Environmental Engineering, especially Lori Larner, Laura Post, and Joseph Nguyen.

I would also like to thank my lab mates Amira Oun for her selfless instructions in the lab, to Evan O'Brien for completing our research projects together, to Camille McCall and Brijen Miyani for their collaboration and friendship in our current project.

Special thanks goes to Canadian Studies Center, especially the director AnnMarie Schneider for introducing me to the One Health Leadership field, for providing me the graduate staff position in the center, and opportunities to attend conferences and workshops in Canada. These experiences have broadened my eyes to be better prepared into the professional world.

iii

Last but not least, many thanks goes to my dear friends Chelsea Weiskerger, Mengying Sun, Charifa Hejase , Vidhya Ramalingam, and my family for their love and support during my PhD program.

# TABLE OF CONTENTS

LIST OF TABLES	. vii
LIST OF FIGURES	.viii
Introduction	1
Chapter 1:	4
Microbial Pollution Characterization at a TMDL Site in Michigan: Effect of Hydrological	
Conditions on Pollution Loading	4
1.1. Abstract	4
1.2. Introduction	5
1.3 Materials and Methods	7
1.3.1. Site description and sampling	7
1.3.2. E. coli quantification	8
1.3.3. Bacteroides quantification	9
1.3.4. Data analyses	. 11
1.4 Results and Discussion	. 12
1.4.1. E. coli temporal variation	. 12
1.4.2. Host-specific <i>bacteroides</i> temporal variation	. 14
1.4.3. Correlation between microbial indicators, discharge, and rainfall	. 15
1.5. Conclusions	. 21
1.6. Acknowledgments	. 21
REFERENCES	. 22
Chapter 2:	. 26
Microbial Pollution Characterization at a TMDL Site in Michigan: Source Identification	. 26
2.1. Abstract	. 26
2.2. Introduction	. 27
2.3. Material and Methods	. 29
2.3.1. Site description	. 29
2.3.2. Water sample collection and processing	. 32
2.3.3. <i>E. coli</i> analysis	. 33
2.3.4 Molecular analysis	. 34
2.4. Results and Discussion	. 38
2.4.1. <i>E. coli</i> levels	. 38
2.4.2. Human and bovine-associated bacteroides levels	. 39
2.4.3. Statistical comparisons between sampling locations	. 41
2.4.4. Statistical comparisons between human and bovine-associated bacteroides levels	. 42
2.5. Conclusions	. 47
2.6. Acknowledgments	. 48
REFERENCES	. 49

Chapter 5	54
3.1 Abstract	54
3.2 Introduction	55
3.3 Materials and Methods	58
3.3.1 Site description	58
3.3.2. Sample collection	60
3.3.3. <i>E. coli</i> quantification	61
3.3.4. Bovine Bacteroides quantification	61
3.3.5. Microbial community analysis	62
3.3.6. MG-RAST analysis	63
3.3 Results and Discussion	64
3.3.1 E. coli and bovine-associated bacteroides levels	64
3.3.2 WGS results	65
3.3.3. Screening for Livestock Pathogens	66
3.5. Conclusions	70
REFERENCES	71
Chapter 4:	75
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central	
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River	75
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River	75 75
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River	75 75 75
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River	75 75 75 75
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> </ul>	75 75 75 77 77
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification.</li> </ul>	75 75 75 77 77 77
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification</li> <li>4.3.3. <i>Bacteroides</i> quantification</li> </ul>	75 75 75 77 77 77 78 79
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification</li> <li>4.3.3. <i>Bacteroides</i> quantification</li> <li>4.3.4. Nitrate and chloride quantification</li> </ul>	75 75 77 77 77 78 79 80
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification</li> <li>4.3.3. <i>Bacteroides</i> quantification</li> <li>4.3.4. Nitrate and chloride quantification</li> <li>4.3.5. C-q plot generation</li> </ul>	75 75 77 77 77 78 79 80 80
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification</li> <li>4.3.3. <i>Bacteroides</i> quantification</li> <li>4.3.4. Nitrate and chloride quantification</li> <li>4.3.5. C-q plot generation</li> <li>4.4. Results and discussions</li> </ul>	75 75 77 77 77 78 79 80 80 80
<ul> <li>Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River</li> <li>4.1. Abstract</li> <li>4.2. Introduction</li> <li>4.3. Materials and methods</li> <li>4.3.1. Site description and sampling</li> <li>4.3.2. <i>E. coli</i> quantification</li> <li>4.3.3. <i>Bacteroides</i> quantification</li> <li>4.3.4. Nitrate and chloride quantification</li> <li>4.3.5. C-q plot generation</li> <li>4.4. Results and discussions</li> <li>4.5. Conclusions:</li> </ul>	75 75 75 77 77 78 79 80 80 80 81 86
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River 4.1. Abstract 4.2. Introduction 4.3. Materials and methods 4.3.1. Site description and sampling 4.3.2. <i>E. coli</i> quantification 4.3.3. <i>Bacteroides</i> quantification 4.3.4. Nitrate and chloride quantification	75 75 77 77 78 79 80 80 80 81 86 88
Concentration-Discharge Diagrams of <i>E. coli</i> and Bovine-Associated <i>Bacteroides</i> in a Central Michigan River	75 75 75 77 77 78 79 80 80 80 81 86 88

# LIST OF TABLES

Table 1.1: Primers used in this study for human and bovine associated <i>bacteroides</i> genetic         markers
<b>Table 1.2:</b> Correlation analysis (2-tailed) for microbial contaminants and hydrological conditions
<b>Table 1.3:</b> E. coli exceedance based on hydrological events
<b>Table 1.4:</b> Results of non-parametric Wilcoxon rank sum test for <i>E. coli</i> , bovine-associated <i>bacteroides</i> (BoBac), and human-associated <i>bacteroides</i> (HuBac)
Table 2.1: Land use in the study area
<b>Table 2.2:</b> Primer and probes used in real-time PCR assays to detect <i>Bacteroides</i> genetic markers
<b>Table 2.3:</b> Occurrence rate of microbial contamination indicators in three sampling sites40
<b>Table 2.4:</b> Metagenome analysis statistics (MEGAN), using NCBI refseq bacteria database44
<b>Table 3.1:</b> Land use in the Sloan creek sub-watershed (NLCD 2011)60
<b>Table 3.2:</b> E. coli and BoBac levels and hydrological conditions
<b>Table 3.3:</b> Whole Genome Shotgun Sequencing Analysis Statistics
<b>Table 3.4:</b> Domain distribution of the microbial community
Table 3.5: Mycobacterium species observed sequences
<b>Table 3.6:</b> Brucella species observed sequences
<b>Table 4.1:</b> Primer/probe sets for qPCR tested in water samples
<b>Table 4.2:</b> Event division for the sampling period
Table 4.3: Dominant source summary for the three events

# LIST OF FIGURES

<b>Figure 1.1:</b> River discharge and concentrations of a) <i>E. coli</i> , b) bovine-associated <i>Bacteroides</i> ( <i>BoBac</i> ), c) human-associated <i>Bacteroides</i> ( <i>HuBac</i> )
Figure 1.2: Rainfall and hydrograph of Sloan Creek discharge during the study period17
<b>Figure 1.3:</b> Box and whisker plots for the three different hydrological events a) <i>E. coli</i> , b) bovine-associated Bacteroides (BoBac), c) human-associated Bacteroides (HuBac)19
Figure 2.1: Red Cedar River watershed, Sloan Creek sub-watershed, and sampling sites31
<b>Figure 2.2:</b> <i>E. coli</i> distribution in the three sampling sites. The box plots displayed the range, quartiles, mean and outliers of the concentration in the sampling sites
<b>Figure 2.3:</b> Boxplots of bovine-associated <i>bacteroides</i> (BoBac) and human-associated <i>bacteroides</i> (HuBac) in three sampling sites, classified by sampling locations40
<b>Figure 2.4:</b> Human-specific <i>Bacteroides</i> (HuBac) and Bovine-specific <i>Bacteroides</i> (BoBac) concentrations in the three sampling sites, classified by host-specific <i>bacteroides</i> 41
Figure 2.5: Microbial community distribution in water samples from Sloan creek (using NCBI RefSeq bacteria database)
<b>Figure 2.6:</b> Characterized metagenomes for the water samples from Sloan creek (using NCBI Env_nt database)
Figure 3.1: Sloan Creek sub-watershed
Figure 3.2: Phylum distribution of the bacterial community
Figure 4.1: Red Cedar River watershed
Figure 4.2: Division of hydrological events during sampling period
Figure 4.3: Pollutant Concentrations in the Red Cedar River
Figure 4.4: Concentration discharge (C-q) hysteresis loop summary for Events A to C83

# **KEY TO ABBREVIATIONS**

BoBac	Bovine-associated bacteroides
bp	base pair
Ср	Crossing point
CAFO	Concentrated animal feeding operation
CDC	Centers for Disease Control
C <sub>G</sub>	Concentration in Ground Water
C <sub>SO</sub>	Concentration in Soil Water
C <sub>SE</sub>	Concentration in Surface Water
Contigs	Contiguous datasets
C-q	Concentration-discharge
DNA	Deoxyribonucleic acid
DEM	Digital Elevation Model
dsDNA	Doubles-stranded Deoxyribonucleic acid
E. coli	Escherichia coli
EPA	Environmental Protection Agency
GIS	Geographic Information System
HuBac	Human-associated bacteroides
MDEQ	Michigan Department of Environmental Quality
MST	Microbial Source Tracking
MSU	Michigan State University

MPN	Most Probable Number			
NAHSS	National Animal Health Surveillance System			
NGS	Next generation sequencing			
NLCD	National Land Cover Database			
NHD	National Hydrography Dataset			
PBS	Phosphate Buffered Saline			
PCA	Principal component analysis			
PCR	Polymerase chain reaction			
qPCR	Quantitative polymerase chain reaction			
RCR	Red Cedar River			
RNA	Ribonucleic acid			
RT-qPCR	Real-time quantitative polymerase chain reaction			
RTSF	Research Technology Support Facility			
TMDL	Total Maximum Daily Load			
USDA	United States Department of Agriculture			
USGS	United States Geological Survey			
WGS	whole genome shotgun sequencing			
WWTP	Wastewater treatment plant			

# Introduction

Microbial contamination in waters has been one of the most pressing environmental health concerns in the world and the US. Understanding sources, temporal variability, and transport pathways of microbes in water is critical for the development of watershed remediation and pollution prevention plans. Furthermore, understating the occurrence of pathogens in water and its relationship to human and animal health may facilitate early detection and prevention of disease. This dissertation focuses on the study of impaired water bodies in the Great Lakes basin. In particular, maximum pollutant loading and its relationship to hydrological events is studied, microbial pollution source identification methods are investigated, and methods for analyzing water samples to identify disease signals are explored. The selected sites, located in Michigan, are sites for which Total Maximum Daily Load (TMDL) requirement for bacteria have been identified by the state and watershed management plans are being developed. Four studies have been conducted in total.

The objective of the first study was to investigate the effect of hydrological conditions on microbial pollutant levels at a TMDL site during spring and summer storm events. Water samples were collected and analyzed to quantify concentrations of *E. coli*, bovine-associated *bacteroides* (BoBac) gene markers, and human-associated *bacteroides* (HuBac) gene markers. *E. coli* concentrations in water samples had significant strong correlations with precipitation and discharge and BoBac concentrations were positively related to discharge. *E. coli*, BoBac and HuBac patterns suggested first-flush phenomena occurred during summer storms. *E. coli* permit exceedance rates raised from 31% before first-flush, to 100% during first-flush in the summer. The resulting information may help develop a plan for restoring impaired waters and establish

the maximum amount of pollutants that the body of water can receive during different hydrological conditions.

The objective of the second study was to present an approach that will facilitate pollution source identification. In addition to conventional indicator analysis, the approach included molecular analysis of species-specific markers and microbial community diversity analysis. 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 a sub-watershed representing the main creek upstream, main creek downstream, and tributary. Host-specific human and bovine-associated *Bacteroides* genetic markers were quantified in all water samples. High concentrations of both human and bovine associated-*Bacteroides* indicated influence of multiple sources of fecal contamination. 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 of the particular sub-watershed were leakage from septic systems and runoff from agriculture activities.

The objective of the third study was to identify livestock disease signals in water samples from agriculture-dominated sub-watersheds. Early detection and prevention of livestock disease outbreaks is paramount to the animal agriculture industry. In agriculture-dominated watersheds, it is impractical to test every animal for potential disease. Sampling runoff-impacted surface water from agricultural areas represents a community fecal and urine sample of the livestock population in the sub-watershed; therefore, it can serve as a screening tool for the presence of potential disease outbreaks in the corresponding livestock population. We conducted whole genome shotgun sequencing analysis of water samples collected in the mouth of the subwatershed. The analysis of the genomic sequences was focused on the identification of potential

cattle pathogens. We observed genomic sequences related to *Mycobacterium*, *Brucella*, and other species. The information serves as a screening tool for the identification and early detection of signals of potential livestock disease, including bovine tuberculosis. This proposed approach may only serve as a screening tool for the presence for potential disease. When signals of disease of interest are observed, further testing of manure and individual animals is required.

The objective of the fourth study was to investigate the potential origin of microbial contaminants (surface, soil, or groundwater). Concentration-discharge (C-q) relationships can help explain the source/origin and transport of contaminants in water. Water samples were collected from an agriculture-dominated sub-watershed on a daily basis in the spring and early summer and analyzed for *E. coli*, bovine-associated bacteroides and nitrate. Based on the C-q hysteresis loops the dominant water sources of bacteria was studied. *E. coli* and bovine-associated *bacteroides* were associated with different dominant water sources in the early spring. On the contrary, similar water sources were observed in the late spring and early summer. In the early spring, *E. coli* was associated with surface event water sources and soil water, whereas BoBac was associated with groundwater and soil water. Surface event water became the dominant source for both *E. coli* and BoBac in early summer especially after manure application. The behavior of nitrate was also evaluated and it appeared similar to BoBac.

#### Chapter 1:

Microbial Pollution Characterization at a TMDL Site in Michigan: Effect of Hydrological Conditions on Pollution Loading

# 1.1. Abstract

Communities throughout the United States are developing and implementing watershed management plans to address nonpoint sources of pollution and meet Total Maximum Daily Load (TMDL) requirements. Once a TMDL is established, a watershed management plan is developed and implemented to reduce contaminant sources and attain TMDL goals. Developing an effective TMDL and remediation plan should take into account fluctuation of pollution loadings and the timing of first-flush events. The objective of this study is to investigate the effect of hydrological conditions on microbial pollutant levels at a TMDL site during spring and summer storm events. A total of 64 water samples were collected from Sloan Creek in Mid-Michigan in the Spring/Summer of 2015. All samples were analyzed to quantify concentrations of E. coli, bovine-associated bacteroides (BoBac) gene markers, and human-associated bacteroides (HuBac) gene markers. Discharge was the driving force of the microbial contaminant loading in the studied water body. E. coli concentrations had significant strong correlation with precipitation and discharge and BoBac concentrations were positively related to discharge. E. coli, BoBac and HuBac patterns suggested first-flush phenomena occurred during summer storms. E. coli permit exceedance rates raised from 59% before first-flush, to 100% during and after first-flush in the summer. The resulting information may help develop a plan for restoring impaired waters and establish the maximum amount of pollutants that the body of water can receive during different hydrological conditions.

### **1.2. Introduction**

Communities throughout the United States are developing and implementing watershed management plans to address nonpoint sources of pollution and meet total Maximum Daily Load (TMDL) requirements. Once a TMDL is established, a watershed management plan is developed and implemented to reduce contaminant sources and attain TMDL goals ("USEPA, 2008.). In addition to identifying pollution sources, effective watershed management plans should identify the timing of maximum pollutant loading and its relationship to hydrological events.

It is well known that storm water runoff, from urban and agricultural areas, significantly contributes to the microbial pollution loads of surface water. Changes in hydrological conditions have been shown to be associated with microbial contamination levels in surface waters (Almeida and Soares, 2012; Kistemann et al., 2002). Rainfall and discharge have been shown to correlate with *Escherichia coli (E. coli)* concentrations in urban and agricultural watersheds (Bach et al., 2010; Krometis et al., 2007; Reischer et al., 2008). Rainfall events can lead to increasing discharge and first-flush phenomena, which are associated with high contaminant concentrations. Initial runoff may significantly increase contaminant concentrations in surface waters due to compounds accumulating on the ground surface during dry periods. In addition, first flush phenomena may raise significant public health issues, such as the spread of epidemiologic diseases (Kleinheinz et al., 2010). Thus, studying first-flush phenomena can help to understand patterns of contaminant release and to improve land management decisions.

Researchers have primarily focused on chemicals in surface runoff when studying firstflush phenomena. The major chemical contaminants that have been studied were Na<sup>+</sup>, Cl<sup>-</sup>, K<sup>+</sup>, TSS, and DOC (Deletic, 1998; Evans and Davies, 1998; Lee et al., 2004). Limited studies have focused on first-flush assessment of microbial pollutants. Doyle (2008) observed an increased

concentration in total coliform and *E. coli* in rainwater harvesting system in Rwanda following a first-flush event. Similarly, a first-flush of *E. coli* was observed in an urban catchment in Australia (Bach et al., 2010). Hathaway and Hunt, 2011 used *E. coli*, Fecal *coliform*, and *Enterococci* as indicators in an urban watershed with mixed land use conditions, and first-flush phenomena of *E. coli* was observed. Several studies have been conducted in North Carolina. Stumpf et al., (2010) investigated *E. coli* and *Enterococci* in a headwater catchment, but there was no first-flush phenomena observed; Rowny and Stewart, 2012 studied Fecal *coliform* and *E. coli* in an urban watershed, but no first-flush phenomena were observed.

In recent years, Bacteroides genetic markers have been used to identify fecal contamination in water and provide a powerful tool to trace the source of fecal contamination (Field and Samadpour, 2007). Several *Bacteroides* quantitative polymerase chain reaction (qPCR) assays have been developed to detect fecal sources from multiple animal and human sources (Bernhard and Field, 2000; Dick et al., 2005; Field and Samadpour, 2007; Layton et al., 2006; Okabe et al., 2007). However, first-flush studies or studies focusing on the effect of hydrological conditions on microbial source tracking (MST) tools such as bacteroides concentrations have not been reported yet. Such studies are important in understanding the loading patterns of pollution originating from different sources, and can provide valuable information for addressing TMDL sites. In this study, the concentrations of the traditional fecal contamination indicator, E. coli, and the concentrations of the genetic markers for Humanassociated Bacteroides (HuBac) and Bovine-associated Bacteroides (BoBac) were quantified in water samples collected from Sloan Creek, a TMDL site in Mid-Michigan, in the Great Lakes Basin. The purpose of this study was to assess the impact of spring and early summer rainfall events on microbial contaminant loadings. This research provides an understanding of microbial contaminant sources during different types of hydrological events (storm events and dry events) in a mixed watershed, and helps with developing land use management plans to protect water quality.

The sampling scheme and study design should be given special attention when conducting a microbial contamination loading study, taking into consideration all regional hydrological conditions and pollution dynamics (Kay et al., 2007; Reischer et al., 2008). Therefore, this study established and evaluated a comprehensive sampling plan for the Sloan Creek with consideration given to spring and summer seasonal hydrological conditions. The sampling scheme was designed to characterize and compare spring hydrological conditions (snow-melt) with the summer hydrological events (rainfall)

# **1.3 Materials and Methods**

# 1.3.1. Site description and sampling

Sloan Creek (Ingham County, MI) is a tributary of the Red Cedar River, which 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 to Lake Michigan. In the state of Michigan, the daily maximum geometric mean is 300 *E. coli* /100 mL for total body contact recreation, and 1000 *E. coli* /100 ml for partial body contact recreation (MDEQ, 2016). The Sloan Creek sub-watershed is a small sub-watershed within the Red Cedar River watershed, and was selected for this study since it previously exceeded State of Michigan *E. coli* water quality standards for total and partial body contact recreation (ICD Red Cedar Monitoring Project 2013, MDEQ 2014). The Sloan Creek flows through both agricultural and residential areas. Suspected sources of bacteria in the sub-watershed include human, agricultural and wildlife fecal inputs. South-central Michigan has hot summers with frequent precipitation. The average maximum temperature in the summer is 27.8 °C and the average annual precipitation is 962 mm. 59% of yearly precipitation falls during the cropping season (the months of May-October), when fertilizers are applied on the land surface in agricultural areas (Ingham Conservation District, 2012). 45% of the Sloan Creek sub-watershed is agricultural land. Precipitation data were obtained from the Michigan Automated Weather Network (MAWN) East Lansing Michigan State University (MSU) Hort, Michigan station, (42.6734, -84.4870). Discharge data in Sloan Creek was collected from a United States Geological Survey gauging station (USGS 04112000) located at Sloan Creek near the City of Williamston, Michigan (42.6758, -84.3638). The base flow in Sloan Creek is 0.0283 m<sup>3</sup>/s.

A six-month sampling scheme was designed to collect samples (N=64) at least twice per week, in addition to sampling during rain events in spring and summer 2015, from March  $22^{nd}$  to August  $26^{th}$ . The sampling site was located in Legg Park, Michigan, at the mouth of Sloan Creek where it drains into the Red Cedar River. One-liter autoclaved sampling bottles were rinsed three times with water samples prior to collection. Two liters of water samples were collected for *E. coli* and *bacteroides* measurements. Grab water samples were collected, stored on ice, and analyzed in the laboratory within 3 (± 1) hours.

#### 1.3.2. E. coli quantification

*E. coli* enumeration was performed using Colilert-18®, which has a detection limit of 1 Most Probable Number per 100 mL (MPN/100mL). Samples were diluted (1:10 and 1:100) with deionized water to make a 100 mL solution. Colilert-18® was added to each sample, dissolved by shaking, poured into a Quanti–Tray/2000 tray, and the trays were incubated overnight at 35°C for 18 hours (Colilert-18 procedure). The wells of the trays were counted and the most probable

number (MPN) per 100 mL of sample was calculated according to the manufacturer's instructions. The measurement was performed twice on each sample and their corresponding dilutions and average MPN values were calculated.

# 1.3.3. Bacteroides quantification

All water samples were tested for Hubac and BoBac molecular markers quantitatively using qPCR. 500 mL of water sample were filtered through 0.45  $\mu$ m 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 the samples, 100  $\mu$ L of DNA was extracted from 400  $\mu$ L 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), according to the manufacturer instructions. Two hostspecific qPCR methods were utilized to identify and quantify potential sources of fecal pollution within the sub-watershed. Primers used are listed in Table 1.1.

Table 1	1.1:	Primers	used	in th	nis stu	dy fe	or	human	and	bovine	associ	ated	bacteroid	des	geneti	С
marker	S															

	Forward	Reverse	Probe	References
Human-associated				
Bacteroides B.	TCGTTCGTC	AAGAAAAA	6FAM-	Yampara-
thetaiotaomicron	AGCAGT -	GGGACAGT	ACCTGCTG-	Iquise et al.,
α-1–6 mannanase	AACA	GG	NFQ	2008
(HuBac)				
			BoBac402Bhq	
Devine esserieted	BoBac367f	BoBac467r	f	
Dovine-associated	(GAAG(G/A)C	(GCTTATTC	(TGAAGGAT	Layton et
16grDNA (DoDao)	TGAACCAGC	ATACGGTA	GAAGGTTCT	al., 2006
TOSIKINA (DODac)	CAAGTA)	CATACAAG)	ATGGATTGT	
			A AACTT)	

All qPCR quantification analysis was 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. DNA extracted from samples was analyzed in triplicate with 5 µL of extract used for template. The crossing point (Cp) value for each qPCR reaction was automatically determined by the LightCycler® Software 4.0. One copy of the targeted gene is assumed present per cell; thus, one gene copy number corresponded to one cell equivalent. Gene copies were then converted and reported using the unit "copies/100mL".

In order to prepare the standard curves to quantify the gene numbers, the DNA was extracted from American Type Culture Collection (ATCC), number 29148D-5, genomic DNA for *B. thetaiotaomicron*, because of its high host-specificity (Xu et al., 2003). Bovine feces obtained from the Michigan State University dairy farm was used for DNA extraction for 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<sup>8</sup> to 10<sup>0</sup> were used for the standard curves. One plasmid standard was included during each qPCR run as a positive control and molecular-grade water was used in place of DNA template for negative controls.

#### **1.3.4.** Data analyses

The association of all measured microbial contaminants with river discharge and rainfall was investigated. Statistical analyses were performed using SPSS Statistics software (Version 22) with a significance  $\alpha$ =0.05. *Bacteroides* concentrations were log-transformed to achieve normality and meet the assumptions of a parametric test. Simple t-tests were used to determine the differences in mean concentrations of target organisms among each other and with the precipitation and the discharge. The t-test was two-tailed, with alpha levels, or the probability of rejecting the null hypothesis when it is true, set at *p* < 0.05. Pearson's correlation coefficient was used to test the relationship between *E. coli, Bacteroides* markers, precipitation, and discharge.

To define first-flush phenomena quantitatively the total river discharge during the sampling period was accumulated and divided into three events of equal amounts. Our goal was to evaluate the impact of a set amount of cumulative discharge on pollutant concentrations. The cumulative discharge of the sampling period was 30.28 cms and the event size of 10 cms cumulative discharge was chosen. The purpose of dividing into equal amounts of cumulative discharge was to capture the impact of event size on the first flush. The distribution of mean concentration of microbial contaminants *E. coli*, BoBac, and HuBac in each event was characterized by using the non-parametric Wilcoxon rank sum test. According to Bach et al. 2010, first-flush phenomenon is confirmed if there is a statistically significant difference of pollutant concentrations between the events. In case the differences are not statistically significance alternative amount of equal cumulative discharge should be selected. The 5% significance level was used for event grouping (Bach et al., 2010). First-flush events were confirmed with Mass/Volume (M/V) method (plots not shown). The M/V method assesses and quantifies first-flush phenomena by using dimensionless cumulative pollutant mass load vs.

cumulative runoff volume curves (Bertrand-Krajewski et al., 1998). This theory sets the criteria for first flush phenomenon when over 80% of the total pollutant mass load is transported within the first 30% of total discharge volume (Bertrand-Krajewski et al., 1998).

#### **1.4 Results and Discussion**

### 1.4.1. E. coli temporal variation

Sloan Creek was monitored for the presence of the microbial fecal indicator *E. coli*. The detection rate of *E. coli* was 100%. Monitoring results for *E. coli* concentrations are shown in Figure 1.1. High concentrations of *E. coli* were found within 24 to 72 hours following each rain event. 59% (38 of 64) of the measured water samples exceeded the water quality standards since the single sample limit of *E. coli* is 300 MPN/100 mL for total body contact recreation (MDEQ, 2016).



**Figure 1.1:** River discharge and concentrations of a) *E. coli*, b) bovine-associated *Bacteroides* (*BoBac*), c) human-associated *Bacteroides* (*HuBac*)

Figure 1.1 (cont'd):



In our study, the highest concentration of *E. coli* was measured on day 169 (June 18<sup>th</sup>, 2015), four days after the first large rain event of the season on day 165 (June 14<sup>th</sup>, 2015, 42.42

mm). A sharp increase was observed on June  $18^{th}$ , 2015 (7270 MPN/100mL) compared with day 163 (June  $12^{th}$ , 2015, 272.3 MPN/100mL), then *E. coli* continued to decrease gradually until the end of the season and fell back within the base flow conditions. It is worthwhile to mention that on day 222 (August  $10^{th}$ , 2015) the largest rain event of the season was recorded (57.66 mm) but *E. coli* levels did not raise significantly (Figure 1.1 a). This pattern shows that *E. coli* was depleted from the land surface and soil after the series of rainfall events.

## 1.4.2. Host-specific bacteroides temporal variation

Human and bovine *bacteroides* gene markers shared similar patterns as *E. coli* (Figure 1.1 b and 1c). The detection rate for bovine *bacteroides* was similar to human *bacteroides*, and they were 26.6% and 25% respectively. The highest concentration of bovine and human *bacteroides* occurred on day 178 (June 27<sup>th</sup>, 2015) after the highest discharge peak with the concentration of  $7.6*10^9$  copies/100mL and  $6.99*10^9$  copies/100 mL respectively.

*Bacteroides* markers underwent some delay on the time series. The highest hits for the *Bacteroides* markers were not tied with *E. coli*. Both *Bacteroides* markers had a sharp increase on day 166 (June 15<sup>th</sup>, 2015) within 24 hours of the first large rain event. HuBac was raised from undetected on day 165 (June 14<sup>th</sup>, 2015) to 1.35x10<sup>6</sup> copies/100mL on day 166 (June 15<sup>th</sup>), and reached its highest concentration on June 27<sup>th</sup>, 2015 (6.99x10<sup>8</sup> copies/100mL). BoBac behaved similarly to HuBac; it rose from undetected on day 178 (June 27<sup>th</sup>, 2015). Later in the season on day 222 (August 10<sup>th</sup>) the largest rain event was recorded (2.27 in) and *E. coli* had a sharp increase from 357.5 MPN/100 mL on day 220 (August 8<sup>th</sup>, 2015) to 1046.2 MPN/100mL on day 222 (August 10<sup>th</sup>, 2015). Similarly, the BoBac concentration was raised from undetected to 6.97x10<sup>4</sup> copies/100mL. No human marker was detected during this event. This pattern indicates

that there was a strong loading of fecal contamination originating from bovine sources to the stream during the largest rainfall event.

#### 1.4.3. Correlation between microbial indicators, discharge, and rainfall

Our results showed that E. coli concentrations were significantly elevated and strongly correlated with precipitation (p =0.001, r=0.422) and discharge (p=0.001, r=0.414). Similar to these findings, bacterial loading rates in rivers and beaches have been shown to increase during hydrologic events (Daly et al., 2013; Kistemann et al., 2002; Kleinheinz et al., 2010; Krometis et al., 2007; Rowny and Stewart, 2012; Stumpf et al., 2010). A strong relationship between fecal indicator bacteria and wet weather has been documented previously. For example, Reeves et al., 2004 carried out a series of field studies to identify the spatial distribution of fecal indicator bacteria in dry and wet weather run-off, and indicated that stormwater runoff was an important factor correlating with nonpoint source pollution. A long-term water quality study (1994-2010) estimated the loads from streams and drains supplying an estuary in Australia and it suggested that stormwater was a significant source of *E. coli* during wet-weather flow (Daly et al., 2013). In the Hoosic River watershed in Massachusetts, bacterial levels were found to be higher in summer than in winter and higher during storms than during base flow conditions (Traister and Anisfeld, 2006). Similarly, during winter the lowest microorganisms concentrations in Navesink River watershed in New Jersey were observed (Selvakumar and Borst, 2006). Additionally, the importance of precipitation and streamflow in the transport of protozoan and bacterial pathogens and fecal indicator bacteria has been frequently reported (Dorner et al., 2006; Ferguson et al., 2003; Wu et al., 2011). Our study also illustrated that late season storms had a greater frequency of water quality exceedances compared to early season storm events, possibly because in early spring the soil temperature is still low, which makes the deposited pollution hard to release into

the stream. Another reason could be the effect of cold temperature on E. coli.

In addition to *E. coli*, our study demonstrated that bovine-associated *bacteroides* concentrations were significantly correlated with discharge (p=0.005, r=0.642). There was no significant relationship with discharge and precipitation for HuBac (Table 1.2).

**Table 1.2:** Correlation analysis (2-tailed) for microbial contaminants and hydrological conditions.

Parameters		E. coli	BoBac	HuBac	
Precipitation	Pearson	.422**	.407	.216	
(mm)	Correlation				
	Sig.	.001	.105	.422	
Discharge	Pearson	.414**	.642**	.383	
(cms)	Correlation				
	Sig.	.001	.005	.144	

\*\*. Correlation is significant at the 0.01 level

\*. Correlation is significant at the 0.05 level

Note: E. coli is Escherichia coli, BoBac is bovine-associated bacteroides, and HuBac is human-associated bacteroides. The number of samples that were involved in the statistical analysis for E. coli, BoBac, and Hubac were 64, 17, 16 respectively.

Our study investigated the effect of selected hydrometeorology factors (precipitation and discharge) on the increased concentration of *E. coli* and host-specific *bacteroides*. However, *E. coli* concentrations were also reported to be positively correlated with antecedent climate, rainfall intensity, stream water temperature, and sediment (Crabill et al., 1999; Liao et al., 2014; McCarthy et al., 2012; Oun et al., 2017). These factors might affect water quality along with precipitation and river discharge, leading to the elevated *E. coli* and *bacteroides* concentrations. Furthermore, the persistence and fate of microbes after they are released to the natural environment should be considered when studying the variation of the concentrations.

Effect of hydrological events on pollution loadings

Both discharge and rainfall fluctuated during the sampling period (Figure 1.2); In general, two main storms were observed during the sampling period which caused drastic increases to the discharge in the creek. The first storm was on day 165 (June 14<sup>th</sup>, 2015) with precipitation of 42.42 mm and the second on day 173 (June 22<sup>nd</sup>, 2015) with precipitation of 43.94 mm.



**Figure 1.2.** Rainfall and hydrograph of Sloan Creek discharge during the study period. *Note: River discharge levels are expressed as daily mean values. Event 1 ranges from Julian day* 79-164; Event 2 ranges from day 165-175; Event 3 ranges from day 176-238. Each hydrological event was of equal total volume of water.

The cumulative discharge of the sampling period was 30.28 m<sup>3</sup>/s (cms), and the event size 10 cms was chosen for the merits of calculation. Therefore, there were three events in this study (Figure 1.2). As explained in the methods section, the purpose of diving into equal volume events was to capture the impact of first flush. The base flow condition for Sloan Creek is 0.0283 cms according to the USGS gauging station 04112000. When the increment of surface runoff (event size) 10 cms was used, day 79-164 were grouped into Event 1; day 165-175 were grouped into Event 2; and day 176-238 were grouped into Event 3. The *E. coli* exceedance rates for each

event are shown in Table 1.3. Our results showed that *E. coli* exceedance rates raised from 59% before first flush to 100% during first flush in the summer.

Event	Day range	Date	Exceedance rate
Event 1:	Day 79-164	March 20 <sup>th</sup> to June 13 <sup>th</sup>	31%
Event 2:	Day 165-175	June 14 <sup>th</sup> to June 24 <sup>th</sup>	100%
Event 3:	Day 176-238	June 25 <sup>th</sup> to August 26 <sup>th</sup>	95%

 Table 1.3: E. coli exceedance based on hydrological events.

Note: The total river discharge during the sampling period was accumulated and divided into three hydrological events of equal volume, as explained in the methods section. The overall E. coli exceedance rate was 59% (based on 300 MPM/100 mL limit)

To examine the change of concentrations over the absolute cumulative surface runoff and to characterize the distribution of the events, box-and-whisker plots were constructed for each event (Figure 1.3) and the non-parametric Wilcoxon Rank Sum test was used to group events of statistically similar concentrations. In addition, the non-parametric Wilcoxon rank sum test was conducted for the three microbial contamination indicators at the three different hydrological events (Table 1.4). The microbial contamination indicators *E. coli*, BoBac, HuBac all showed significant difference in Event 2 when compared to Event 1 and Event 3 ( $p \le 0.05$ ). First-flush phenomena is believed to have occurred during Event 2 as the mean contaminant concentration in this period was significantly greater than Event 1 and Event 3 (Bach et al., 2010).



**Figure 1.3:** Box and whisker plots for the three different hydrological events a) *E. coli* b) bovine-associated *Bacteroides (BoBac)*, c) human-associated *Bacteroides (HuBac)* 



Figure 1.3 (cont'd):



The state of Michigan receives an average 700 to 1000 mm precipitation throughout the year, and it has a typical moist continental mid-latitude climate. In 2015, the annual precipitation of the sampling site was 741.3 mm. It started to snow from November of 2014 to March of 2015 and 69% of yearly precipitation fell during March to August as rainfall events. This climate creates a long period of pollutant build-up deposited on surfaces during dry weather (November-March), which is washed away in the spring when the snow starts to melt into surface waters. The initial storms of the spring season are usually expected to be associated with higher pollutant concentrations, which create first-flush phenomena. This is not the case in mid-Michigan's climate because the first-flush runoff may begin in early March when the snow starts to melt, but the soil is still frozen. In this case pollutants will not be flushed until the soil temperature starts to increase during the summer. Indeed, our results show that the first-flush phenomena for *E. coli* and *bacteroides* occurred in early summer, rather than spring. To the best of our knowledge, this

study is the first work to adopt *bacteroides* genetic markers in the first-flush analysis.

Microbial	Comparison of average	Ν	
contamination	concentration between	(Number of p-valu	
indicator	events	sample)	
E. coli	Event 1 and Event 2	43	0.000*
E. coli	Event 2 and Event 3	28	0.002*
BoBac	Event 1 and Event 2	43	0.004*
BoBac	Event 2 and Event 3	28	0.048*
HuBac	Event 1 and Event 2	43	0.000*
HuBac	Event 2 and Event 3	28	0.008*

**Table 1.4:** Results of non-parametric Wilcoxon rank sum test for *E. coli*, bovine-associated *bacteroides* (BoBac), and human-associated *bacteroides* (HuBac). *Note: Event 2 had higher concentrations than Event 1 and Event 3.* 

Significance indicated by \* at 0.05 level

#### **1.5.** Conclusions

This study examined the influence of hydrological conditions on *E. coli* and *bacteroides* concentrations in Sloan Creek, located in mid- Michigan in the Great Lakes Basin. *E. coli* and bovine-associated *bacteroides* concentrations were strongly influenced by precipitation and stream discharge. The study also identified the timing of first-flush phenomena. High levels of microbial contamination were observed and first-flush phenomena of fecal contaminants occurred during summer rainfall events. The study revealed that the majority of pollution loading was contributed to the river over a short period of time in early summer. Developing an effective TMDL and remediation plan should take into account fluctuation of pollutant loadings and the timing of first-flush events.

# 1.6. Acknowledgments

This work was funded by USGS project 2015MI234B.

REFERENCES

# REFERENCES

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

Bach, P.M., McCarthy, D.T., Deletic, A., 2010. Redefining the stormwater first flush phenomenon. Water Res. 44, 2487–2498. https://doi.org/10.1016/j.watres.2010.01.022

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. https://doi.org/10.1128/AEM.66.10.4571-4574.2000

Bertrand-Krajewski, J.-L., Chebbo, G., Saget, A., 1998. Distrubution of Pollutant Mass Vs Volume in Stromwater Discharges and the First Flush Phenomenon. Water Res. 32, 2341–2356. Colilert-18 [WWW Document], n.d. URL https://www.idexx.com/water/products/colilert-18.html (accessed 3.9.17).

Crabill, C., Donald, R., Snelling, J., Foust, R., Southam, G., 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. Water Res. 33, 2163–2171.

Daly, E., Kolotelo, P., Schang, C., Osborne, C.A., Coleman, R., Deletic, A., McCarthy, D.T., 2013. Escherichia coli concentrations and loads in an urbanised catchment: The Yarra River, Australia. J. Hydrol. 497, 51–61. https://doi.org/10.1016/j.jhydrol.2013.05.024

Deletic, A., 1998. The first flush load of urban surface runoff. Water Res. 32, 2462–2470. https://doi.org/10.1016/S0043-1354(97)00470-3

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Santo Domingo, J.W., Simpson, J.M., Walters, S.P., Field, K.G., 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. Appl. Environ. Microbiol. 71, 3184–3191.

Dorner, S.M., Anderson, W.B., Slawson, R.M., Kouwen, N., Huck, P.M., 2006. Hydrologic modeling of pathogen fate and transport. Environ. Sci. Technol. 40, 4746–4753.

Doyle, K.C., 2008. Sizing the first flush and its effect on the storage-reliability-yield behavior of rainwater harvesting in Rwanda. Citeseer.

Ferguson, C., Husman, A.M. de R., Altavilla, N., Deere, D., Ashbolt, N., 2003. Fate and Transport of Surface Water Pathogens in Watersheds. Crit. Rev. Environ. Sci. Technol. 33, 299– 361. https://doi.org/10.1080/10643380390814497

Field, K.G., Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res., Identifying Sources of Fecal Pollution 41, 3517–3538.

https://doi.org/10.1016/j.watres.2007.06.056

Hathaway, J.M., Hunt, W.F., 2011. Evaluation of First Flush for Indicator Bacteria and Total Suspended Solids in Urban Stormwater Runoff. Water. Air. Soil Pollut. 217, 135–147. https://doi.org/10.1007/s11270-010-0574-y

Ingham Conservation District, 2012. 2012 Natural Resource Assessment.

Kay, E.R., Leigh, D.A., Zerbetto, F., 2007. Synthetic molecular motors and mechanical machines. Angew. Chem. Int. Ed. 46, 72–191.

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.

Kleinheinz, G.T., McDermott, C.M., Hughes, S., Brown, A., 2010. Effects of rainfall on E. coli concentrations at Door County, Wisconsin beaches. Int. J. Microbiol. 2009.

Krometis, L.-A.H., Characklis, G.W., Simmons, O.D., Dilts, M.J., Likirdopulos, C.A., Sobsey, M.D., 2007. Intra-storm variability in microbial partitioning and microbial loading rates. Water Res. 41, 506–516.

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.

Lee, H., Lau, S.-L., Kayhanian, M., Stenstrom, M.K., 2004. Seasonal first flush phenomenon of urban stormwater discharges. Water Res. 38, 4153–4163. https://doi.org/10.1016/j.watres.2004.07.012

Liao, H., Krometis, L.-A.H., Hession, W.C., House, L.L., Kline, K., Badgley, B.D., 2014. Hydrometeorological and Physicochemical Drivers of Fecal Indicator Bacteria in Urban Stream Bottom Sediments. J. Environ. Qual. 43, 2034. https://doi.org/10.2134/jeq2014.06.0255

McCarthy, D.T., Hathaway, J.M., Hunt, W.F., Deletic, A., 2012. Intra-event variability of Escherichia coli and total suspended solids in urban stormwater runoff. Water Res. 46, 6661–670. https://doi.org/10.1016/j.watres.2012.01.006

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).

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

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. Reeves, R.L., Grant, S.B., Mrse, R.D., Copil Oancea, C.M., Sanders, B.F., Boehm, A.B., 2004. Scaling and management of fecal indicator bacteria in runoff from a coastal urban watershed in southern California. Environ. Sci. Technol. 38, 2637–2648.

Reischer, G.H., Haider, J.M., Sommer, R., Stadler, H., Keiblinger, K.M., Hornek, R., Zerobin, W., Mach, R.L., Farnleitner, A.H., 2008. Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. Environ. Microbiol. 10, 2598–2608.

Rowny, J.G., Stewart, J.R., 2012. Characterization of nonpoint source microbial contamination in an urbanizing watershed serving as a municipal water supply. Water Res. 46, 6143–6153. Selvakumar, A., Borst, M., 2006. Variation of microorganism concentrations in urban stormwater runoff with land use and seasons. J. Water Health 4, 109–124.

Stumpf, C.H., Piehler, M.F., Thompson, S., Noble, R.T., 2010. Loading of fecal indicator bacteria in North Carolina tidal creek headwaters: hydrographic patterns and terrestrial runoff relationships. Water Res. 44, 4704–4715.

Traister, E., Anisfeld, S.C., 2006. Variability of indicator bacteria at different time scales in the upper Hoosic River watershed. Environ. Sci. Technol. 40, 4990–4995.

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).

Wu, J., Rees, P., Dorner, S., 2011. Variability of E. coli density and sources in an urban watershed. J. Water Health 9, 94–106.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., Gordon, J.I., 2003. A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis Science 299, 2074–2076. https://doi.org/10.1126/science.1080029

Weather state: www.agweather.geo.msu.edu/mawn

### Chapter 2:

Microbial Pollution Characterization at a TMDL Site in Michigan: Source Identification

# 2.1. 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, one of the impaired areas 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 Sloan Creek.
Keywords: E. coli, bacteroides, microbial source tracking, whole genome shotgun sequencing

### 2.2. Introduction

Multiple water bodies in the Great Lakes Basin are impaired due to a number of 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 given priority over that in 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 effective 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 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., 2014). 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. (2014) 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 cultural-independent, therefore the water quality result can be obtain within 4 hours, and detection is more sensitive compared to the traditional cultural method (Layton et al., 2006).

Additionally, high-throughput sequencing metagenomics methods have been shown as promising MST tools since 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 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 metagenomic profile.

### 2.3. Material and Methods

### **2.3.1. 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 square miles 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.

Figure 2.1 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 Figure 2.1, 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.



**Figure 2.1. Red Cedar River watershed, Sloan Creek sub-watershed, and sampling sites.** *Note: Button Drain is a tributary of Sloan Creek. Button Drain sampling site and Sloan Creek upstream site were located in the bridge crossings and road crossings. Sloan Creek downstream was located before the confluence with Red Cedar River. Sloan Creek sub-watershed has a relatively low relief with the maximum elevation recorded as 326 m and a minimum of 249 m.* 

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 2.1). Agriculture, shrub land and forest comprises the majority of the studied area; therefore, the Sloan Creek sub-watershed was classified as a rural and agriculturally dominant area.

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

Table 2.1: Land use in the study area.

According to the Red Cedar River Watershed Management Plan (MSU Institue of Water Research, 2015), the Sloan Creek sub-watershed included a human population of 2,127, 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 3,080 large animals, including 3,000 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.

### 2.3.2. 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 22<sup>nd</sup> to August 26<sup>th</sup>. A total of 192 samples (64 from each sampling location) were collected. Compared with 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 Red Cedar River (Figure 2.1). 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 two to four hours of collection.

### 2.3.3. E. coli analysis

Water samples were analyzed for *E. coli* concentration using the defined substrate method Colilert-18<sup>TM</sup> Quanti-Tray 2000 (IDEXX Laboratories, Inc.) within 4 hours of collection. *E. coli* were measured in duplicate directly or diluted with phosphate-buffered saline solution (PBS) (pH = 7.2) to three serial dilutions  $10^0$ ,  $10^{-1}$ ,  $10^{-2}$ . 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 hr). 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.

## 2.3.4 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  $\mu$ m 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  $\mu$ L of DNA was extracted from 400  $\mu$ L 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 2.2.

	Forward	Reverse	Probe	References
Human-associated Bacteroides HuBaciotaomicron α-1–6 mannanase (HuBac)	TCGTTCGTC AGCAGT- AACA	AAGAAAAAG GGACAGTGG	6FAM- ACCTGCTG-NFQ	Yampara- Iquise et al., 2008
Devine anosisted	BoBac367f	BoBac467r	BoBac402Bhqf	
Bovine-associated	(GAAG(G/A)	(GCTTATTCAT	(TGAAGGATGAA	Layton et al.,
(BoBac)	CTGAACCA	ACGGT-	GGTTCTATGGAT	2006
(Dobac)	GCCAAGTA)	ACATACAAG)	TGTA -AACTT)	

Table 2.2: Primer and probes used in real-time PCR assays to detect *Bacteroides* genetic markers

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. Each BoBac assay was carried out with 10 µl of LightCycler 480 probe Mastermix (Roche Applied Sciences), 1  $\mu$ L forward and reverse primers, 0.4  $\mu$ L 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 min 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. Each HuBac assay was performed with 10 µL of LightCycler 480 probe Mastermix (Roche Applied Sciences), 0.4  $\mu$ L forward and reverse primers, 0.2  $\mu$ L probe, 4  $\mu$ L 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 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<sup>®</sup> Software 4.0. Gene copies were then converted to and reported as copies/100mL.

In order to prepare the standards curve 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 Bovinespecific *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 were 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.

### 2.3.5. 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 (Figure 2.2 and Figure 2.4) and to compare the difference between BoBac and HuBac concentration in each sampling site (Figure 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<sub>10</sub>-transformed to achieve normality; zero copy/100 mL was converted to 0 log<sub>10</sub> copy/100mL manually. The t-test was two-tailed, and the probability of rejecting the null hypothesis when it is true, set at *p* < 0.05.

### 2.3.6 Microbial community analysis

Three samples, specifically August 15<sup>th</sup>, 18<sup>th</sup>, 19<sup>th</sup> 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 15<sup>th</sup>, BoBac was detected

on August 18<sup>th</sup>, and HuBac was detected on August 19<sup>th</sup>. 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 2x250 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 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 less than 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<sup>-5</sup> for microbial community analysis (Huson et al., 2007).

### 2.4. Results and Discussion

### 2.4.1. 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 Figure 2.2. 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 18<sup>th</sup> 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).



**Figure 2.2:** *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. *Note:* n=64 *per site* 

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. When 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 exceed 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).

### 2.4.2. Human and bovine-associated bacteroides levels

To distinguish between the contribution of human and animal sources, human-associated *bacteroides* and bovine-associate *bacteroides* genetic markers have been quantified in each site (Figure 2.3). The occurrence rate for BoBac was 27% in samples from Sloan Creek downstream site, 22% from Sloan Creek upstream site, and 11% from Button Drain site (Table 2.4). 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 2.3). The lowest occurrence rates of HuBac and BoBac were observed in Button Drain.



**Figure 2.3**: Boxplots of bovine-associated *bacteroides* (BoBac) and human-associated *bacteroides* (HuBac) in three sampling sites, classified by sampling locations. *Note: The bacteroides concentration-positive data were shown in the boxplots, and they were log-transformed.* 

Table 2.3: Occurrence rate of microbial contamination indicators in three sampling sites.

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

Note: BoBac is Bovine-associated bacteroides, HuBac is Human-associated bacteroides.

Sloan Creek downstream and upstream sites shared similar range and mean *bacteroides* concentration patterns (Figure 2.4). 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/100mL). 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/100mL.



**Figure 2.4**: Human-specific *Bacteroides* (HuBac) and Bovine-specific *Bacteroides* (BoBac) concentrations in the three sampling sites, classified by host-specific *bacteroides*. *Note: The bacteroides concentration-positive data were shown in the boxplots, and they were log-transformed*.

## 2.4.3. 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 (Figure2&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) (Figure 2.2&2.4). Therefore, Sloan Creek upstream had more impact on Sloan Creek downstream microbial levels than Button Drain did.

## 2.4.4. 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), (Figure 2.3). 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 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 detecting human-specific *bacteroides* (Verhougstraete et al. 2014). 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 15<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> 2015 water samples were 105590, 117559, and 81874, and about 50% of the contigs can be annotated as bacteria (Table 2.4). Contigs were annotated to more than 13 phyla, which were dominated by members of *Proteobacteria, Actinobacteria, Bacteroides*, and *Cyanobacteria* (Figure 2.5).

	Number of contigs	Bacteria	%	Not affiliated	%	Not hits	%
August 15 <sup>th</sup>	105590	48633	46.06%	1595	1.51%	55304	52.38%
August 18 <sup>th</sup>	117559	68744	58.48%	1054	0.90%	47708	40.58%
August 19 <sup>th</sup>	81874	36169	44.18%	151	0.18%	45552	55.64%

Table 2.4: Metagenome analysis statistics (MEGAN), using NCBI refseq bacteria database.

*Note: 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". The contigs that could be annotated were 46.06%, 58.48%, and 44.18% for August 15<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> samples respectively.* 



**Figure 2.5**: Microbial community distribution in water samples from Sloan creek (using NCBI RefSeq bacteria database).

*Note:* Samples were from the Sloan Creek downstream site and were obtained on August 15<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> 2015. Samples were analyzed with whole genome sequencing, results were analyzed with BLAST and MEGAN. Approximately 50% of the total contigs can be annotated.

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*, *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 sample 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 sequence samples from Sloan Creek, indicating the sub-watershed was affected by the fecal contamination from ruminant animals.

The microbial community shifts when the land usage changes. The Sloan Creek subwatershed was an agricultural-urban mixed watershed. *Polynucleobacter, Arcobacter, Methylotenera, Flavobacterium, Psedomonas, and Bacteriodes* were ubiquitous in the sequence water samples and they had more abundance than *E. coli*. The pattern was similar to Uyaguari-Diaz et al., 2016, showing the studied sub-watershed was impaired.

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 52348, 69186, and 33821, for August 15<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> 2015 water samples

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 (Figure 3.6). In particular, 1728, 1533, and 1259 contigs were annotated as wastewater metagenome. Sewage in septic systems has similar microbial community as the influent of wastewater treatment plants, so sewage may be annotated as wastewater treatment plant. 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 metagenome were present in the water samples.



# Figure 2.6. Characterized metagenomes for the water samples from Sloan creek (using NCBI Env\_nt database).

Note: The displayed data was 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 15<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> 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.

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. Therefore, there was a significant amount of contigs that could not be assigned. 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 serve as supportive evidence in addition to qPCR and traditional culture methods.

# **2.5.** Conclusions

The high exceedance rate of *E. coli* standards confirmed that the Sloan Creek subwatershed 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. 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 further proved 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. 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.

# 2.6. 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

# REFERENCES

Almeida, C., Soares, F., 2012. Microbiological monitoring of bivalves from the Ria Formosa Lagoon (south coast of Portugal): A 20years 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., Identifying Sources of Fecal Pollution 41, 3517–3538.

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 Microbiol. Lett. 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., 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/10.1111/1462-920.12092

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, n.d. 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).

MDEQ, n.d. MDEQ, 2012. Total Maximum Daily Load for E. coli in Portions of the Red Cedar River and Grand River Watersheds; including Sycamore, Sullivan, Squaw, and Doan Creeks. Ingham, Eaton, Clinton, Jackson, and Livingston Counties. - Google Search [WWW Document]. URL (accessed 12.25.17b).

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).

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, in: 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.

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.

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  $\alpha$ -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

### Chapter 3:

Watershed Bio-surveillance for Identification of Pollution Sources and Potential Disease Signals

## **3.1 Abstract**

Early detection and prevention of livestock disease outbreaks is paramount to the animal agriculture industry. In agriculture-dominated watersheds it is impractical to test every animal for potential disease. Sampling runoff-impacted surface water from agricultural areas represents a community fecal and urine sample of the livestock population in the sub-watershed; therefore, it can serve as a screening tool for the presence of potential disease outbreaks in the corresponding livestock population. Whole genome shotgun sequencing analysis of the collected samples will provide a wide range of potential pathogens present in the sample. In this paper we characterized bacterial contamination in Sloan Creek sub-watershed, located in the Great Lakes basin. In addition to E. coli, we quantified bovine-associated bacteroides indicating that pollution is partly originating from livestock sources. We conducted whole genome shotgun sequencing analysis of water samples collected in the mouth of the sub-watershed. The analysis of the genomic sequences was focused on the identification of potential cattle pathogens. We observed genomic sequences related to *Mycobacterium*, *Brucella*, and other species. The information serves as a screening tool for the identification and early detection of signals of potential livestock disease, including bovine tuberculosis. This proposed approach may only serve as a screening tool for the presence for potential disease. When signals of disease of interest are observed, further testing of manure and individual animals is required.

**Key words**: Bio-surveillance, Whole-Genome Shotgun Sequencing, Watershed, Livestock disease

Core ideas:

- Sampling runoff from agricultural sub-watersheds represents a community fecal and urine sample of the livestock population in the sub-watershed.
- Whole genome shotgun sequencing analysis of the collected samples will provide a wide range of potential pathogens present in the sample.
- When signals of disease of interest are observed, further testing of manure and individual animals is required for confirmation of potential outbreaks.

## **3.2 Introduction**

Enhancing animal health and well-being is a vital component to the animal agriculture industry and food supply. Even though significant efforts have been made in preparing for and detecting emerging livestock disease, early detection and prevention actions remain to be accomplished. In agriculture-dominated watersheds it is impractical to test every animal for potential disease. Traditional livestock disease detection and management systems are based on diagnostic analyses of clinical samples. However, these systems fail to detect early warnings of public health threats at a wide population level, and fail to predict outbreaks in a timely manner.

Watersheds are the geographical space in which water flows, as well as a subsystem that agriculture, industrial development, population growth, climate change, and governance are of the driving forces; therefore, watersheds offer an ideal context for transdisciplinary research and seeking solutions to provide health for humans, domestic animals, wild animals, and ecosystems (Jenkins et al., 2018).

Increasingly, the rapid advancement of PCR-based nucleic acid detection methods and high throughput sequencing analysis for species and strain identification is providing new

opportunities to understand the pollution sources and potential pathogen signals of livestock and zoonotic diseases. High-resolution genetic typing data allow the characterization and comparison of pathogens enabling development of host-associated genetic markers (Yamahara et al., 2007). Many studies have used the techniques to identify the source of microbial contamination and ensure water quality to protect public health (Bradshaw et al., 2016; Dubinsky et al., 2016; Wu et al., 2018a).

High-throughput sequencing metagenomics methods, such as WGS, have been shown as promising tools since 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 (Wu et al. 2018a; Fisher et al., 2015; Newton et al., 2013, 2015). Also, metagenome analysis has been used to screen for the presence of pathogens, human or livestock. For example, Shaw et al. (2012) and Gomez-Alvarez et al. (2012) have identified bacteria pathogens in water distribution systems using 16s rRNA metagenomics analysis and pyrosequencing respectively. Ye and Zhang (2011) and Bibby et al. (2010) used 16s rRNA analysis to identify bacteria pathogens in wastewater and biosolids. Li et al. 2015 used 16s rRNA and whole genome sequencing to identify bacteria in livestock waste and sewage. Furthermore, whole genome sequencing has been used to identify bacteria pathogens in clinical samples (Anderson et al. 2013; Loman et al. 2013; Seth-Smith et al, 2013; Castellarin et al. 2012; Bertelli and Greub 2013).

The central premise of the proposed screening approach is that community fecal pollution represents a snapshot of the status of public health or livestock health. Sampling runoff-impacted

surface water from agricultural areas represents a community fecal and urine sample of the livestock population in the sub-watershed; therefore, it can serve as a screening tool for the presence of potential disease outbreaks in the corresponding livestock population. Whole genome shotgun sequencing analysis of the collected samples will provide a wide range of potential pathogens present in the sample. The proposed approach can only serve as a screening tool for the presence for potential disease. To effectively use the disease control opportunities empowered by PCR-based techniques such as host-associated qPCR assays and whole genome shotgun sequencing, there is a need to develop novel surveillance approaches that focus on the water microbial community for the targeted watershed.

In this paper we characterize bacterial contamination in Sloan Creek sub-watershed, located in the Great Lakes basin. Previous work demonstrated the high exceedance rate of *E. coli* standards (Wu et al. 2018a; Wu et al. 2018b) confirming that the Sloan Creek sub-watershed was under a high risk of microbial contamination. There are no known point-source sewage inputs to Sloan Creek, but the stream has been reported to have animal and human nonpoint sources. For this paper, in addition to *E.coli*, we detected and quantified bovine-associated *bacteroides* markers confirming that there were contaminant inputs originating from both bovine and human sources. Furthermore, we conducted whole genome shotgun sequencing analysis of samples collected in August of 2015. The analysis of the genomic sequences was focused on the identification of potential cattle pathogens. The resulting information indicates that in addition to human sources, *E. coli* signals may also originate from livestock sources. Most importantly, the information serves as a screening tool for the early detection of potential livestock disease, including bovine tuberculosis and brucellosis.

## **3.3 Materials and Methods**

## 3.3.1 Site description

The Red Cedar River flows about 80 KM through rural and agricultural land in the south-central Lower Peninsula of Michigan. The Red Cedar River drains into the Grand River and subsequently Lake Michigan. For this study, the Sloan Creek sub-watershed (Figure 3.1) 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 the Red Cedar River, which receives drainage from 900 KM<sup>2</sup> of the Sloan Creek sub-watershed (agricultural, rural, and suburban land use).



**Figure 3.1**: Sloan Creek sub-watershed Note: Samples were collected at the mouth of Sloan Creek, right before the confluence with the Red Cedar River

Agriculture, shrub land and forest comprises the majority of the studied area; therefore, the Sloan Creek sub-watershed is classified as a rural and agriculturally dominant area. 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 3.1).

Watershed	% Land Use						
	Agriculture	Shrub	Developed	Forest	Water & wetland		
Red cedar River	35	23	18	10	14		
Sloan Creek	45	27	9	11	8		

**Table 3.1:** Land use in the Sloan creek sub-watershed (NLCD 2011)

According to the Red Cedar River Watershed Management Plan from the Institute of Water Research at Michigan State University (MSU) in 2015, the sub-watershed has an estimated 3,080 large animals. 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 Concentrated Animal Feeding Operation (CAFO), there were an average of 10 animals per farm, and 12 animals per square mile. The Sloan Creek sub-watershed included a human population of 2,127, living at a density of 112 people per square mile. About 393 homes were estimated to be using septic systems. Suspected sources of bacteria in the sub-watershed include agricultural, wildlife, and human inputs. There were no known point-source sewage inputs to Sloan Creek or Button Drain, but both streams have been reported to have animal and human nonpoint sources (Wu et al. 2018a; Wu et al. 2018b).

#### **3.3.2.** Sample collection

Water samples were collected in August 2015 (N=11) at least twice per week, in addition to sampling during rain events in spring and summer 2015, from March  $22^{nd}$  to August  $26^{th}$ . The sampling location was located in Legg Park, Michigan, at the mouth of Sloan Creek where it drains into the Red Cedar River. One-liter autoclaved sampling bottles were rinsed three times with water samples prior to collection. Two liters of water samples were collected for *E. coli*,

Bovine *Bacteroides* and whole genome shotgun measurements. Grab water samples were collected, stored on ice, and analyzed in the laboratory within  $3 (\pm 1)$  hours.

### 3.3.3. E. coli quantification

*E. coli* enumeration was performed using Colilert-18<sup>®</sup>, which has a detection limit of 1 Most Probable Number per 100 mL (MPN/100mL). Samples were diluted (1:10 and 1:100) with deionized water to make a 100 mL solution. Colilert-18<sup>®</sup> was added to each sample, dissolved by shaking, poured into a Quanti–Tray/2000 tray, and the trays were incubated overnight at 35°C for 18 hours ("Colilert-18," n.d.). The wells of the trays were counted and the most probable number (MPN) per 100 mL of sample was calculated according to the manufacturer's instructions. The measurement was performed twice on each sample and their corresponding dilutions and average MPN values were calculated.

### 3.3.4. Bovine *Bacteroides* quantification

Bacteroides gene markers were measured by qPCR assays. A volume of 500 to 1000 ml water sample was filtered onto a 0.45 µm pore size, 47 mm diameter mixed cellulose ester filter ("Pall Canada Ltd, St. Laurent, QC, Canada) under partial vacuum depending on sample's turbidity. The filter was then placed into a 50 mL sterile DNase-free micro-centrifuge tube, containing sterile phosphate buffered saline (PBS), vortexed and centrifuged to pellet the cells. The supernatants were decanted and the remaining pellets (about 2 ml) were stored at -80°C prior to nucleic acid extraction. DNA extraction was performed with automatic MagNa Pure Compact System (Roche Applied Sciences, Indianapolis, IN) with a 100 µl elution for each sample as described in our previous work (Oun et al., 2017). After extraction, qPCR was performed using previously published methods for bovine-associated *bacteroides* genetic markers (Layton et al., 2006). All qPCR reactions were run in sealed glass capillaries (Roche,

Germany). Each reaction was prepared with 20 µl reaction volumes, consisting of 5 µl DNA extraction template and 15 µl of 2x LightCycler 480 Probes Master (Life Science, Roche, Germany) and primers/probe mix. The qPCR amplification was performed using the LightCycler® 1.5 (Roche Applied Sciences, Indianapolis, IN) with LightCycler® Software 4.0. Results were reported as genomic copies/100 ml. Standard curves of *bacteroides* genes were prepared by cloning reactions with DNA templates obtained from manure in dairy farmland. The qPCR standard was run in triplicates with efficiencies >95% and R<sup>2</sup> values >0.999 as described in our previous work (Oun et al., 2017).

### **3.3.5.** Microbial community analysis

Three samples, specifically August 15<sup>th</sup>, 18<sup>th</sup>, 19<sup>th</sup> of 2015 from the Sloan Creek downstream site were processed for whole genome shotgun sequencing. Samples were collected in sterilized, triple-rinsed 1L Nalgene bottles, then placed on ice and processed within 6 hours of collection. A volume of 1000 ml water sample was filtered onto a 0.45 µm pore size, 47 mm diameter mixed cellulose ester filter ("Pall Corporation: Filtration, Purification, Separation and Environmental Technologies," n.d.) under partial vacuum. The filter was then placed into a 50 mL sterile DNase-free micro-centrifuge tube, containing sterile phosphate buffered saline (PBS), vortexed and centrifuged to pellet the cells. The supernatants were decanted and the remaining pellets (about 2ml) were stored at -80°C prior to nucleic acid extraction. DNA extraction was performed with automatic MagNa Pure Compact System (Roche Applied Sciences, Indianapolis, IN) with a 100 µl elution for each sample as described in our previous work (Oun et al., 2017).

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 2x250 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 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 less than 30 bp were discarded. The Guanine-Cytosine (GC) content of the trimmed sequences was about 48% and the average length was 40-250 base pairs. 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).

#### **3.3.6. MG-RAST analysis**

The assembled contig files for all samples were uploaded to Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) web server for analysis. MG-RAST is an online database designed for phylogenetic analysis of metagenomes. The uploaded shotgun metagenome data were performed with annotation mapping. By default, the integrated reference database m5NR, which incorporated with multiple databased such as RefSeq, and IMG is used (Keegan et al., 2016). MG-RAST implemented Best hit, Representative hit, Lowest Common Ancestor to determine the number of hits in the samples. The maximum e-value cutoff is 1e-5, and the minimum % identity cutoff is 60%.

## **3.3 Results and Discussion**

## 3.3.1 E. coli and bovine-associated bacteroides levels

*E. coli* levels are presented in Table 2. The observed concentrations ranged from 281 to 1046 MPN/100mL. In the state of Michigan, the daily maximum geometric mean is 300 *E. coli* /100 mL for total body contact recreation, and 1000 *E. coli* /100 ml for partial body contact recreation (DEQ - E. coli in Surface Waters,). In the majority of samples, *E. coli* concentrations exceeded the total body contact standard.

Bovine-associated *bacteroides* were detected in the August samples. Four out of eleven samples were detected positive with the highest concentration 3560 copies/ 100 ml. The occurrence of bovine-associated *bacteroides* indicates the microbial contamination from bovine sources (Table 3.2). Precipitation events may flush contaminants on the surface or in the soil into streams. Our study showed that the highest concentration of *E. coli* in August (1046 MPN/100 ml) occurred after a heavy rainfall event on 8/10/15. The highest concentration of bovineassociated *bacteroides* in August (3560 copies/100ml) occurred on 8/18/15.

Date		Discharge	precipitation	E. coli	BoBac
		(cms)	( <b>mm</b> )	(MPN/100ml)	(copies/100ml)
	8/2/2015	0.018	17.3	816	12
	8/3/2015	0.023	4.3	613	
	8/7/2015	0.019	1.5	548	
	8/8/2015	0.02	0.3	358	
	8/10/2015	0.481	57.7	1046	
	8/15/2015	0.054	0.8	914	1840
	8/18/2015	0.034	6.1	687	3560
	8/19/2015	0.045	1.8	281	
	8/20/2015	0.105	10.7	411	
	8/23/2015	0.065	16	462	
	8/26/2015	0.062	0	770	531

Table 3.2: E. coli and BoBac levels and hydrological conditions

## 3.3.2 WGS results

Three of the samples described in Table 1 (August 15, 18, and 19) were further processed - DNA was extracted and WGS analysis was conducted. The assembled contigs files for all samples were uploaded to Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) web server for analysis. MG-RAST results are characterized in Table 3 and summarized in Table 3.3 and Figure 3.2.

WGS analysis indicated that the majority of observed *E. coli* species were non-pathogens. Out of a total of 78, 109, and 169 contigs associated to *E. coli* observed in August 15, August 18 and August 19 respectively; 15, 31, and 28 hits were associated with potential pathogenic strains O103, O111, O127, O157, O26 and O55 respectively.

**Table 3.3:** Whole Genome Shotgun Sequencing Analysis Statistics

	<u> </u>		
	8/15/2015	8/18/2015	8/19/2015
Upload: bp Count	79,382,487 bp	95,865,383 bp	46,487,802 bp
Upload: Sequences Count	110190	125,525	85,792
Upload: Mean Sequence Length	$720 \pm 1,530 \text{ bp}$	$764 \pm 1,087$ bp	$542 \pm 410 \text{ bp}$
Upload: Mean GC percent	$48\pm11~\%$	$50\pm10$ %	$50\pm10$ %
Processed: Predicted Protein Features	126331	136,152	88,824
Processed: Predicted rRNA Features	8541	8,035	6,610
Alignment: Identified Protein Features	80104	96,261	45,759
Alignment: Identified rRNA Features	805	588	624
Annotation: Identified Functional Categories	68235	84,741	37,269

**Table 3.4:** Domain distribution of the microbial community

	8/15/2015	8/18/2015	8/19/2015
Bacteria	75282 (97.81%)	96321 (99.07%)	43623 (95.52%)
Eukaryota	1179 (1.53%)	467 (0.48%)	1730 (3.79%)
Viruses	297 (0.39%)	273 (0.28%)	183 (0.40%)
Archaea	115 (0.15%)	72 (0.07%)	95 (0.21%)
Unclassified sequences	95 (0.12%)	88 (0.09%)	36 (0.08%)



Figure 3.2: Phylum distribution of the bacterial community

# 3.3.3. Screening for Livestock Pathogens

The metagenome was searched for genomic sequences related to potential livestock pathogens. The potential pathogenic species were selected based on reportable cattle bacterial disease in Michigan (MDARD, 1988). The results, shown in Tables 5 and 6, are presented as number of observed sequences (number of hits). We observed genomic sequences related to *Mycobacterium and Brucella* species. The presented sequences for potential cattle-associated bacterial pathogens can only be used as a preliminary screening tool. They are only indicators of potential presence of the associated bacteria. For further confirmation, samples should be analyzed using qPCR and culture methods. Additionally, if the proposed methodology were to be applied for early detection of livestock disease, manure samples and clinical samples should also be collected and analyzed for comparison and confirmation purposes.

Table 3.5: Mycobacterium	species	observed sequences
--------------------------	---------	--------------------

	8/15/2015	8/18/2015	8/19/2015
Mycobacterium abscessus	82	122	49
Mycobacterium avium	127	194	117
Mycobacterium bovis	18	16	13
Mycobacterium gilvum	62	145	60

## Table 3.5 (cont'd)

Mycobacterium introcellulare	21	14	14
Mycobacterium kansasii Mycobacterium leprae Mycobacterium marinum	27 52 46	27 93 70	22 24 42
Mycobacterium parascrofulaceum	27	35	30
Mycobacterium smegmatis	230	360	115
Mycobacterium tuberculosis	159	207	103
Mycobacterium ulcerans	65	79	37
Mycobacterium vanbaalenii	168	247	99
Mycobacterium sp.	237	399	171
Total	1269	1916	872

Note: The bolded bacteria species are often observed in cattle.

Our data show the potential presence of *Mycobacterium bovis* in the sub-watershed (Table 5). The reported *Mycobacterium bovis* is the cause of bovine TB. *Mycobacterium avium* may related to both human and bovine sources. Bovine tuberculosis (TB) is a bacterial disease primarily affecting cattle; however, it can be spread between wildlife populations and other mammals, including humans. *Mycobacterium bovis*, the causative agent, is an aerobic bacterium. It is related to *Mycobacterium tuberculosis*, the bacterium which causes tuberculosis in humans, and can jump the species barrier and cause tuberculosis in humans and other mammals (Grange et al. 1996). *Mycobacterium bovis* can also be transmitted from human to human and from human to cattle (Griffith and Munro 1994; Tice, 944). *Mycobacterium bovis* is endemic in Michigan's white tailed deer and has been circulating since 1994 (Wilkins et al, 2008; O'Brien et al, 2002). The strain circulating in deer has remained genotypically consistent and was detected in 2 humans in 2002 and 2004. The focal area of the endemic disease is in northern-lower

Michigan. The Michigan Department of Agriculture and Rural Development (MDARD) reported that in 2017 out of 5428 deer tested in the northern-lower Michigan area, 49 were positive for bovine TB. Additionally, on April 18 2017, MDARD reported that a small beef cattle herd was confirmed as bovine tuberculosis positive. The herd was identified through routine surveillance testing. In 2018 it has been reported by the press that Lansing officials stated that bovine tuberculosis has been detected in two southwestern Michigan cattle herds, In 2015, when the sampling described in this paper was conducted (prior to the reported spread of *mycobacterium bovis* disease in lower western Michigan) we detected the presence of *Mycobacterium* species in Sloan Creek, in the Red Cedar River watershed, Ingham County (data reported in this paper).

	<b>Table 3.6:</b>	Brucella	species	observed	sequences
--	-------------------	----------	---------	----------	-----------

	8/15/2015	8/18/2015	8/19/2015
Brucella abortus	26	39	31
Brucella canis	2	12	4
Brucella ceti	18	16	4
Brucella melitensis	32	36	16
Brucella microti	2	4	3
Brucella neotomae	1	1	0
Brucella ovis	2	4	4
Brucella pinnipedialis	10	15	7
Brucella sp. 83/13	4	8	2
Brucella sp. BO1	3	7	0
Brucella sp. BO2	2	5	2
Brucella sp. F5/99	4	7	0
Brucella sp. NF 2653	2	6	7
Brucella sp. NVSL 07-0026	0	1	0
Brucella suis	28	35	16
Total	136	196	96

Note: The bolded bacteria species are often observed in cattle.

Brucella is a fastidious, aerobic, small, gram-negative coccbaccilus bacteria that causes

brucellosis in both animals and human (ASM Brecella 2016 March report). Brucella abortus

most commonly affects cattle, B. melitensis is most common in goats, as well as in cattle; B.

*Canine* is most common in canines; and *B. Suis* is found in swine (Michigan DNR brucellosis). *B. melitensis* and *B. abortus* are the most common *Brucella* species that may affect humans. Potential source of infection can be milk from infected cattle and goats. Wildlife animals are more likely to acquire *Brucella* from domestic animals, but the incidence of brucellosis in wildlife in very low. In Michigan, all brucellosis species are reportable to MDARD. From 2007, the MDARD received an unprecedented number of reports of *B. canis* infection, and the outbreaks continued till 2016 (Johnson et al., 2018). In 2011, six Michigan counties confirmed canine brucellosis outbreaks, including Ingham county, where the water samples were taken. In the sequenced samples, *B. abortus, B. Suis, B. melitensis, and B. Canine* were detected in 2015 (Table 3.6). The presence of B. *abortus* and B. *melitensis* indicated potential disease in cattle.

The described work is based on that fact that wastewater analysis, manure analysis, or polluted water analysis is equivalent to obtaining and analyzing a community-based urine and fecal sample of the representative sub-watershed. Monitoring temporal changes in pathogen concentration and diversity excreted in a sub-watershed may allow for early detection of outbreaks (critical moments for the onset of an outbreak). In addition, carefully designed spatial sampling may help detect locations where an outbreak may begin to develop and spread (critical locations for the onset of an outbreak). For accurate application of the proposed methodology modeling the fate of pathogens, including shedding rates, transport, growth and inactivation processes in the environmental, will be critical to ensure the effectiveness of the proposed method. When signals of disease of interest are observed, further testing of manure and individual animals is required.

## **3.5.** Conclusions

In this paper we report the presence of livestock associated bacteria sequences in water samples collected from an agriculture impacted sub-watershed in Michigan. Whole genome shotgun sequencing analysis provides data sequences of a wide range of microorganisms present in the sample. Sequences related to potential cattle pathogens such as *Mycobacterium bovis* and *Brucella abortus* were present in the samples. The resulting information may be used as a tool for identification of potential endemic disease signals.

In an agricultural-dominated watershed, it is impractical to test every animal for potential disease. Sampling runoff-impacted surface water from agricultural areas represents a community fecal and urine sample of the livestock population in the sub-watershed; therefore, watershed bio-surveillance may serve as a screening tool for the presence of potential disease outbreaks in the corresponding livestock population. Since the resulting information can only be used and a screening tool of potential presence of disease in a certain area, when signals of disease of interest are observed further testing of manure and individual animals is required. Additionally, for full application of the proposed methodology, watershed hydrological modeling should be included to guide location and timing of appropriate sampling locations.

REFERENCES

# REFERENCES

Anderson, P., et al. "Sequences of multiple bacterial genomes and a Chlamydia trachomatis genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen." Clinical Microbiology and Infection 19.9 (2013): E405-E408.

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 Bertelli, C., and G. Greub. "Rapid bacterial genome sequencing: methods and applications in clinical microbiology." Clinical Microbiology and Infection 19.9 (2013): 803-813.

Bradshaw, J.K., Snyder, B.J., Oladeinde, A., Spidle, D., Berrang, M.E., Meinersmann, R.J., Oakley, B., Sidle, R.C., Sullivan, K., Molina, M., 2016. Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. Water Res. 101, 498–509. https://doi.org/10.1016/j.watres.2016.05.014

Bibby, Kyle, Emily Viau, and Jordan Peccia. "Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids." Water research 44.14 (2010): 4252-4260. Colilert-18 [WWW Document], n.d. URL https://www.idexx.com/water/products/colilert-18.html (accessed 3.9.17).

Castellarin, Mauro, et al. "Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma." Genome research 22.2 (2012): 299-306.

DEQ - 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).

Dubinsky, E.A., Butkus, S.R., Andersen, G.L., 2016. Microbial source tracking in impaired watersheds using PhyloChip and machine-learning classification. Water Res. 105, 56–64.

Field, K.G., Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res., Identifying Sources of Fecal Pollution 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.

Gomez-Alvarez, Vicente, Randy P. Revetta, and Jorge W. Santo Domingo. "Metagenomic analyses of drinking water receiving different disinfection treatments." Applied and environmental microbiology (2012): AEM-01018.

Grange J., Yates M., de Kantor I. (1996). Guidelines for speciation within the Mycobacterium

tuberculosis complex. Second edition". World Health Organization. Retrieved 2007-08-02.

Griffith A. and Munro W (1944). Human pulmonary tuberculosis of bovine origin in Great Britain". J Hyg. 43 (4): 229–40.

Ingham Conservation District, 2012. 2012 Natural Resource Assessment.

Jenkins, A., Capon, A., Negin, J., Marais, B., Sorrell, T., Parkes, M., Horwitz, P., 2018. Watersheds in planetary health research and action. Lancet Planet. Health 2, e510–e511. https://doi.org/10.1016/S2542-5196(18)30228-6

Johnson, C.A., Carter, T.D., Dunn, J.R., Baer, S.R., Schalow, M.M., Bellay, Y.M., Guerra, M.A., Frank, N.A., 2018. Investigation and characterization of Brucella canis infections in pet-quality dogs and associated human exposures during a 2007–2016 outbreak in Michigan. J. Am. Vet. Med. Assoc. 253, 322–336. https://doi.org/10.2460/javma.253.3.322

Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer (2016) "MG-RAST, a metagenomics service for analysis of microbial community structure and function." Microbial Environmental Genomics (MEG). Humana Press, New York, NY, 207-233. 49, 7319–7329. https://doi.org/10.1021/acs.est.5b00980

Loman, Nicholas J., et al. "A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104: H4." Jama309.14 (2013): 1502-1510.

MDARD (1988) Reportable Animal Diseases in Michigan, Act 466 of 1988, MCL 287.709, Animal Industry Division, Michigan Department of Agriculture and Rural Development MDEQ, 2016 - E. coli in Surface Waters [WWW Document], n.d. URL h

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

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

O'Brien D., Schmitt S., Fierke J., Hogle S. Winterstein S., Cooley T., Moritz W., Diegel K., Fitzgerald S., Berry D., Kaneene J. (2002) Epidemiology of Mycobacterium bovis in free-ranging white-tailed deer, Michigan, USA, 1995–2000, Preventive Veterinary Medicine. 54: 47–63

Seth-Smith, Helena MB, et al. "Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture." Genome research (2013). Shaw, Jennifer LA, et al. "Using amplicon sequencing to characterize and monitor bacterial diversity in drinking water distribution systems." Applied and environmental microbiology(2015): AEM-01297.

Pall Corporation: Filtration, Purification, Separation and Environmental Technologies [WWW Document], n.d. URL http://www.pall.com/main/home.page (accessed 3.9.17).

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.

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

Tice F. (1944). "Man, a source of bovine tuberculosis in cattle". Cornell Vet. 34: 363–5.

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

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

Wilkins M., Meyerson J., Bartlett P., Spieldenner S., Berry D., Mosher L., Kaneene J., Robinson-Dunn B., Stobierski M., Boulton M. (2008) Hyman Mycobacterium bovis Infection and Bovine Tuberculosis Outbreak, Michigan, 1994-2007. Emerging Infectious Diseases. 14(4): 657-660

Wu H., Oun A., Kline-Robach R., I Xagoraraki (2018a) Microbial pollution characterization at a TMDL site in Michigan: Source identification. Journal of Great Lakes Research. 44 (3), 412-420

Wu H., Oun A., Kline-Robach R., I Xagoraraki (2018a) Microbial pollution characterization at a TMDL site in Michigan: Source identification. Journal of Great Lakes Research. 44 (3), 412-420

Ye, Lin, and Tong Zhang. "Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing." Environmental science & technology 45.17 (2011): 7173-7179.

Yamahara, K.M., Layton, B.A., Santoro, A.E., Boehm, A.B., 2007. Beach sands along the California coast are diffuse sources of fecal bacteria to coastal waters. Environ. Sci. Technol. 41, 4515–4521.

#### Chapter 4:

Concentration-Discharge Diagrams of *E. coli* and Bovine-Associated *Bacteroides* in a Central Michigan River

## 4.1. Abstract

Concentration-discharge (C-q) relationships can help explain the source and transport of contaminants. The goal of this study is to evaluate the behavior of a commonly used generic microbial pollution indicator (*E.coli*) and a bovine-associated pollution indicator (Bovine-associated *bacteroides*) in a river draining a mixed-use watershed in central Michigan. Water samples were collected on a daily basis in the spring and early summer of 2013 from the Red Cedar River in central Michigan. Based on the C-q hysteresis loops the dominant water sources of *E. coli* and Bovine-associated *bacteroides* (BoBac) were studied. Our results indicate that *E. coli* and BoBac were associated with different dominant water sources in the early spring (April 7<sup>th</sup> to 17<sup>th</sup>). On the contrary, similar water sources were observed in the late spring (April 18<sup>th</sup> to 30<sup>th</sup>) and early summer (May 25<sup>th</sup> to June 16<sup>th</sup>). In the early spring, *E. coli* was associated with surface event water sources and soil water, whereas BoBac was associated with groundwater and soil water. Surface event water became the dominant source for both *E. coli* and BoBac in early summer especially after manure application. The behavior of nitrate was also evaluated and it appears similar to BoBac.

# 4.2. Introduction

Microbial contamination in waters has been one of the most pressing environmental health concerns in the world. Because of increasing demand for safe water caused by urbanization and growing population, researchers have put efforts in water treatment and remediation of contaminated sites. However, the study of the source and transportation of

microbes in water systems is very limited. To monitor water quality, *E. coli* is commonly used as a fecal contamination indicator; bovine associated *Bacteroides* (BoBac) is used to trace contamination originating from cattle livestock.

Stream water can originate from various sources such as precipitation, surface runoff, interflow, and baseflow. To understand water sources Evans and Davis 1998 introduced a three-component system classification: surface event water, soil water, and the ground water. They also modified concentration-discharge (C-q) diagrams, which have distinct patters according to the relative importance of the three components.

Concentration-discharge (C-q) diagrams are useful to infer pathways and sources of pollutants (Creed et al., 2015; Evans and Davies, 1998; Long et al., 2016). The diagrams compare the peak position of the contaminant and the peak position of the hydrograph. The rotation direction of the diagrams is associated with the sources of water in the stream. The clockwise direction indicates the contaminants are from local sources to the stream, whereas the anticlockwise direction shows the contaminants are from distal sources to the stream (Creed et al., 2015; Evans and Davies, 1998). Therefore, the clockwise direction would show that surface water plays an important role, and anticlockwise direction reveals soil water or groundwater is the main source of the water in the stream. The relative importance of surface event water, soil water, and groundwater can be found by generating the event C-q diagram and referring to the chart developed by Evans and Davies (1998).

Most of the published C-q diagrams are applied to chemicals and particles. Nitrate and chloride were commonly studied in salted watershed (Long et al., 2016; Sun et al., 2014, 2012); DOM and DOC were studied in a subtropical river (Yang et al., 2013). Nitrate and potassium were studied in forest or agriculture watershed (Carey et al., 2014; Griffioen, 2001); Total

suspended solid was studied in urban watershed (Gellis, 2013; Williams, 1989). Long et al. 2016 used the C-q concept to describe source of *E.coli* over a sequence of multiple events. In our study, we applied the C-q diagram to demonstrate the potential differences in *E.coli* behavior with the behavior of bovine-associated *bacteroides*. Understanding differences or similarities in the behavior of these two indicators may help facilitate remediation of polluted mixed-use watersheds.

## 4.3. Materials and methods

# 4.3.1. Site description and sampling

This study was conducted in the 1222 km<sup>2</sup> Red Cedar River (RCR) watershed in the central lower-peninsula of Michigan, USA (Fig. 4.1). The RCR watershed is characterized by glacial sediment deposits with elevations ranging from 249 to 324 m above sea level. Soils are primarily well-drained alfisols and mollisols with moderate permeability (Sommers, 1984), which contribute to terrestrial runoff during hydrological events. The majority of the land drains by drain tiles, which are buried conduits that collect and convey water to the river, and the storm sewer drain the urban areas. The Red Cedar River flows through farmlands, concentrated animal feeding operation (CAFO), and receives wastewater effluent from several municipalities. The river flows through the campus of Michigan State University (MSU) and the sampling site is on the MSU campus. The study site experiences long winter, short spring, and humid summer. The bank of river is flooded during spring and summer due to snowmelt and rainfall events. The ArcGIS 10.3 software was used to delineate watershed boundaries on the RCR watershed and to display the streamline. The discharge data in the RCR watershed obtained from a USGS gaging station (Hydrologic Unit 04112500) located along the river on MSU campus.



Figure 4.1. Red Cedar River watershed

Water samples were collected from April 17<sup>th</sup> to June 16<sup>th</sup> in 2013. The site was sampled daily. Samples were collected in sterilized, triple-rinsed 1L Nalgene bottles, then placed on ice and processed within 6 hours of collection. 107 samples were collected to capture a range of hydrological conditions, including base flow and rainfall events. Spring and summer rain events were the dominant hydrologic events in the watershed.

# 4.3.2. E. coli quantification

Colilert-18<sup>®</sup> was used for *E. coli* enumeration and the detection limit of this method was 1 MPN/100 ml (IDEXX Laboratories, Westbrook, ME). Samples were prepared at 1:10 and 1:100 dilutions with deionized water to make 100 ml solutions. Colilert-18<sup>®</sup> was added to each sample, dissolved by shaking, poured into a Quanti–Tray/2000 tray, and the trays were incubated overnight at 37°C for 18 hours (Colilert-18 procedure). The wells of the trays were counted and the most probable number (MPN) per 100 ml of sample was calculated according to the

manufacturer's instructions. Duplicate experiments were performed on each sample and their dilutions, and the average MPN values were calculated based on all the tested values.

#### 4.3.3. Bacteroides quantification

Bacteroides gene markers were measured by qPCR assays with Bovine-associated primers. A volume of 500 to 1000 ml water sample was filtered onto a 0.45 µm pore size, 47 mm diameter mixed cellulose ester filter ("Pall Canada Ltd, St. Laurent, QC, Canada) under partial vacuum depending on sample's turbidity. The filter was then placed into a 50 mL sterile DNasefree micro-centrifuge tube, containing sterile phosphate buffered saline (PBS), vortexed and centrifuged to pellet the cells. The supernatants were decanted and the remaining pellets (about 2ml) were stored at -80°C prior to nucleic acid extraction. DNA extraction was performed with automatic MagNa Pure Compact System (Roche Applied Sciences, Indianapolis, IN) with a 100 µl elution for each sample as described in our previous work (Oun et al., 2017). After extraction, qPCR was performed using previously published methods for bovine-associated Bacteroides genetic markers (Layton et al., 2006). The primers and probes are shown in Table 4.1. All qPCR reactions were run in sealed glass capillaries (Roche, Germany). Each reaction was prepared with 20  $\mu$ l reaction volumes, consisting of 5  $\mu$ l DNA extraction template and 15  $\mu$ l of 2x LightCycler 480 Probes Master (Life Science, Roche, Germany) and primers/probe mix. The cycling conditions of qPCR reactions were described by Layton et al., 2006. The qPCR amplification was performed using the LightCycler<sup>®</sup> 1.5 (Roche Applied Sciences, Indianapolis, IN) with LightCycler® Software 4.0. Results were reported as copy number per 100 ml (Copies/100 ml). Standard curves of Bacteroides 16S rRNA gene for BoBac were prepared by cloning reactions with DNA templates obtained from manure in dairy farmland. The qPCR

standard was run in triplicates with efficiencies >95% and  $R^2$  values >0.999 as described in our previous work (Oun et al., 2017).

Assay	Primer Name	Sequence (5'-3')	Amplicon	References
			Size	
	BtH-P	FAM-ACCTGCTG-NFQ		
BoBac	BoBac367f	GAAG(G/A)CTGAACCAGCCAAGTA	100	Layton et
	BoBac467r	GCTTATTCATACGGTACATACAAG		al., 2006
	BoBac402Bhqf	TGAAGGATGAAGGTTCTATGGAT		
		TGTAAACTT		

**Table 4.1:** Primer/probe sets for qPCR tested in water samples

## **4.3.4.** Nitrate and chloride quantification

Water samples for nitrate and chloride quantification were passed through 0.45  $\mu$ m Millipore filters on the sampling site. Water samples were stored at 4 °C until analysis by ion chromatography using EPA method 300.1.

## **4.3.5.** C-q plot generation

Concentration-discharge plots was generated by using discharge as the x axis and pollutant concentration as the y axis for *E. coli*, BoBac, and nitrate respectively. The hysteresis loops can be viewed when there is a lag between the peak of concentration and the peak of discharge. The loops were separated by hydrological events that were divided by the peak of discharge (Evans and Davies, 1998). According to Evans and Davis (1998), when the C-q plot is clockwise, with a zero or positive slope (C1 or C2), the pollutant is associated with surface event water ( $C_{SE}$ ). When the plot is clock-wise, with a negative slope (C3), the pollutant is associated with groundwater ( $C_G$ ). When the plot counter-clock-wise, with a zero or positive slope (A1 or A2) the pollutant is associated with soil water ( $C_{SO}$ ). Finally, when the plot is counter-clock-wise with a negative slope (A3), the pollutant is associated with groundwater.

# 4.4. Results and discussions

During the sampling period, there were 31 precipitation events, as shown in the hydrograph that includes data from April 7<sup>th</sup> to June 16<sup>th</sup> (Julian days 97-167) (Figure 2). The heavy rainfalls occurred in spring and summer and resulted three major peaks in hydrograph, which was divided into three parts and the sampling time had three major events (Table 4.2 & Figure 4.2). The stream flowrate and contaminants concentration were relatively stable from Day 120 to Day 145, therefore this period was not considered in our study.

**Table 4.2:** Event division for the sampling period

	Julian Day (2013)	Calendar Day (2013)
Event A	97-107	April 7 <sup>th</sup> – April 17 <sup>th</sup>
Event B	108-120	April 18 <sup>th</sup> – April 30 <sup>th</sup>
Event C	145-167	May 25 <sup>th</sup> – June 16 <sup>th</sup>



Figure 4.2: Division of hydrological events during sampling period.





Figure 4.3 shows the pollutant concentrations observed and plotted on the hydrograph. *E. coli*, BoBac, and nitrate concentrations appear to fluctuate over time, but the patterns are not identical.

*E. coli* is an indicator of fecal contamination. The original source of *E. coli* in the study can be wastewater discharge, leaking septic tanks, domestic animals, pets, and wild animals. *E. coli* is capable to proliferate in the natural environment, so the source of *E. coli* in this study can be continuous. BoBac is associated with fecal contamination from cattle livestock. The original source of BoBac can be farms and CAFO in the upstream areas. BoBac is an anaerobic bacterium and it has very limited chances to grow in the water environment. Nitrate is associated with application of fertilizers, such as manure.

Figure 4.4 shows the C-q plots generated for *E. coli*, BoBac, and nitrate for the three hydrological events. The data range of BoBac was  $10^1$  to  $10^5$  copies/100 ml, so the data was log-transformed for event A and event C. The C-q patters and their interpretations (Creed et al. 2015; Evans and Davies 1998) are summarized in Table 4.3.



Figure 4.4: Concentration discharge (C-q) hysteresis loop summary for Events A to C

<b>Table 4.3:</b>	Dominant	source	summary	for	the	three	events
-------------------	----------	--------	---------	-----	-----	-------	--------

		E. coli	BoBac	Nitrate
E	C-q pattern:	Clockwise (C2)	Anticlockwise (A3)	Anticlockwise (A2)
Event A	Dominant source:	Surface event water	Groundwater	Soil water
	C-q pattern:	Anticlockwise (A2)	Anticlockwise (A2)	Anticlockwise (A2)
Event B	Dominant source:	Soil water	Soil water	Soil water
E	C-q pattern:	Anticlockwise to clockwise (A2 to C2)	Clockwise (C1)	Clockwise (C2)
Event C	ent C Dominant source:	Soil water to surface event water	Surface event water	Surface event water

During event A, *E. coli* had a peak before the hydrograph peak, and showed clockwise pattern ( $C_{SE}>C_G>C_{SO}$ ) (Creed et al., 2015; Evans and Davies, 1998) which indicates that surface

event water was the dominant source in this event. We speculate that rainfall events flushed *E*. *coli* to the river and caused the observed clockwise pattern. BoBac showed an anticlockwise pattern A3 ( $C_G > C_{SO} > C_{SE}$ ). Since BoBac (anaerobic bacterium) did not have continuous sources, its behavior was more conservative than *E. coli*. The top few inches of soil that were mixed with manure might be the source of BoBac and NO<sub>3</sub>. We saw that in event A the source of BoBac was ground water and in events B and C it becomes soil and surface event water.

This overtime pattern may be explained by the fact that in early April we observe BoBac originating from manure application that took place last growing season and it was in shallow groundwater (event A) and soil (event B). Groundwater maybe pushed into surface during rainfall/ flooding events (Negrel et. al, 2005), which was what we observed in event C. Additionally, during event C we have new manure application (in May). What we observed in event C maybe a combination effect. BoBac and nitrate displayed A3 pattern,  $C_G > C_{SO} > C_{SE}$ . The distal source may come from the manure application from previous farming season, which infiltrated in shallow groundwater, and the groundwater was pushed into surface during rainfall/ flooding events (Negrel er al, 2005). Nitrate showed anticlockwise pattern A2 ( $C_{so}>C_{SE}>C_G$ ), so the dominant source is soil water. Nitrate had similar source as BoBac but nitrification should be considered as well.

During event B, there was a one-day lag between the second *E. coli* peak and the second hydrograph peak, showing a dilution effect (Figure 5). We observed anticlockwise pattern A2 for *E. coli* ( $C_{SO}>C_{SE}>C_G$ ), so the dominant source was soil water. We speculate that this was because the surface source was depleted in event A. During event A, the soil was probably still frozen (Wu et al., 2018), and in event B soil temperature was elevated, therefore in event B soil water plays a more important role. BoBac exhibited an anticlockwise pattern A2 (Cso>C<sub>SE</sub>>C<sub>G</sub>).

The dominant source for BoBac was soil water. Its concentration was correlated with flowrate, but the peak occurred after the peak of the hydrograph, showing a dilution effect. In this event, *E. coli* and BoBac peaked at the same time and showed the same loop pattern. Soil water and groundwater became dominant source of BoBac and nitrate since the shallow groundwater was pushed into surface and interacted with soil water continuously. Nitrate showed anticlockwise pattern A3 ( $C_G>C_{SO}>C_{SE}$ ) transitioning to A1 ( $C_{SO}>C_G>C_{SE}$ ). Nitrification process needs to be considered as well.

During Event C, E. coli showed the same peak as the hydrograph peak, showing chemostatic stage. Accordingly, the behavior of E. coli in the C-q diagram reveal the flushing/dilution effect (Figure 6). E. coli displayed anticlockwise pattern A2 (Cso>CsE>CG), then twisted to C2 ( $C_{SE}>C_{SO}>C_G$ ). The dominant source was surface event water. This might have happened because manure was applied during this event. BoBac presented a clockwise pattern C1 ( $C_{SE}>C_G>C_{SO}$ ). In Figure 6, the orange arrow represented the first peak and the black arrow was the second peak. The dominant source of BoBac was the surface event water. This might be because manure application took place during this event. Manure was titled with soil (mixed layer was about 6 inches). The surface water, which carried manure contamination, was flushed into the stream, which explaines the first peak, the second peak originated from the bottom part of the tilted soil. Nitrate showed a clockwise pattern C2,  $(C_{SE} > C_{SO} > C_G)$ , showing a flushing pattern. Surface event water became important for BoBac and nitrate in Event C (Figure 6). Nitrate exhibited clockwise pattern C2 ( $C_{SE}>C_{SO}>C_G$ ), and the dominant source was surface event water. Since nitrate had similar source as BoBac, it is speculated that manure application lead to the C2 pattern.

## **4.5. Conclusions:**

In this study *E. coli* was detected in each sample, and three major peaks of *E. coli* were recorded, with the highest concentration in the early summer. The source of *E. coli* changed from surface event water (early spring) to soil water (late spring) and then back to surface event water (early summer). BoBac reached the highest concentration in early summer in the beginning of the farming season. The hysteresis loop analysis showed a change of dominant source associated with BoBac, from shallow ground water (early spring) to soil water (late spring) to soil water (late spring) to surface event water (early summer).

*E. coli* and BoBac showed different transportation patterns between groundwater, soil water and surface water. *E. coli* may originate from continuous sources and was flushed into the stream in the first rainfall event of this study. BoBac was more conservative in transportation compared to *E. coli*, since BoBac are anaerobic bacteria. Shallow groundwater and soil water were its dominant sources in the earlier stage of this study. BoBac and *E. coli* was flushed to the stream by surface event water after manure application during farming season in the early summer. BoBac's behavior was more similar to nitrate than *E. coli*.

Our study revealed that the C-q hysteresis loop was helpful to quantitatively understand the transport pathways of microbial indicators to stream water. This research revealed that we cannot assume microbes as an entity in terms of their transportation in the water pathways, since the behavior is heavily influenced by whether the microbes have continuous sources or not. In other words, the microbes that cannot multiply in the water systems are more conservative (BoBac). Our study showed that BoBac can be monitored by nitrate since they shared more similarity than *E. coli*.

Further study can include more chemicals to explain the C-q pattern of microbes. SUVA<sub>254</sub>, potassium, DOC are parameters to likely facilitate the transportation process since they not only share similar sources as the microbes, but also provide essential elements for microbial proliferation, which would be particularly useful to understand aerobic microbes such as *E. coli*. The result of this study can help to predict the concentration of microbial contamination in streams, as well as to apply the appropriate microbe indicator for the level of microbial contamination in the streams. REFERENCES

# REFERENCES

Carey, R.O., Wollheim, W.M., Mulukutla, G.K., Mineau, M.M., 2014. Characterizing Storm-Event Nitrate Fluxes in a Fifth Order Suburbanizing Watershed Using In Situ Sensors. Environ. Sci. Technol. 48, 7756–7765. https://doi.org/10.1021/es500252j

Colilert-18 [WWW Document], n.d. URL https://www.idexx.com/water/products/colilert-18.html (accessed 3.9.17).

Creed, I.F., McKnight, D.M., Pellerin, B.A., Green, M.B., Bergamaschi, B.A., Aiken, G.R., Burns, D.A., Findlay, S.E.G., Shanley, J.B., Striegl, R.G., Aulenbach, B.T., Clow, D.W., Laudon, H., McGlynn, B.L., McGuire, K.J., Smith, R.A., Stackpoole, S.M., Smith, R., 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. Canadian Journal of Fisheries and Aquatic Sciences 72, 1272–1285. https://doi.org/10.1139/cjfas-2014-0400

Evans, C., Davies, T., 1998. Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry.

Gellis, A.C., 2013. Factors influencing storm-generated suspended-sediment concentrations and loads in four basins of contrasting land use, humid-tropical Puerto Rico. CATENA 104, 39–57. https://doi.org/10.1016/j.catena.2012.10.018

Griffioen, J., 2001. Potassium adsorption ratios as an indicator for the fate of agricultural potassium in groundwater. Journal of Hydrology 254, 244–254. https://doi.org/10.1016/S0022-1694(01)00503-0

LightCycler<sup>®</sup> Capillaries (20 µl) [WWW Document], n.d. URL

Long, D., Voice, T., Chen, A., Wu, H., Eunsang, L., Amira, O., Fangli, X., 2016. Patterns of c-q hysteresis for selected inorganic and organic solutes and E. coli in an urban salted watershed during winter-early spring periods.

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

Pall Corporation: Filtration, Purification, Separation and Environmental Technologies [WWW Document], n.d. URL http://www.pall.com/main/home.page (accessed 3.9.17).

Sommers, L.M., 1984. Michigan: A geography. Westview Press.

Sun, H., Alexander, J., Gove, B., Pezzi, E., Chakowski, N., Husch, J., 2014. Mineralogical and anthropogenic controls of stream water chemistry in salted watersheds. Applied Geochemistry

48, 141-154. https://doi.org/10.1016/j.apgeochem.2014.06.028

Sun, H., Huffine, M., Husch, J., Sinpatanasakul, L., 2012. Na/Cl molar ratio changes during a salting cycle and its application to the estimation of sodium retention in salted watersheds. Journal of Contaminant Hydrology 136–137, 96–105.

Williams, G.P., 1989. Sediment concentration versus water discharge during single hydrologic events in rivers. Journal of Hydrology 111, 89–106. https://doi.org/10.1016/0022-1694(89)90254-0

Yang, L., Guo, W., Chen, N., Hong, H., Huang, J., Xu, J., Huang, S., 2013. Influence of a summer storm event on the flux and composition of dissolved organic matter in a subtropical river, China. Applied Geochemistry 28, 164–171. https://doi.org/10.1016/j.apgeochem.2012.10.004

#### Conclusions

Development and implementation of watershed protection plans are imperative to preserve water quality and protect human and animal health. Understanding the sources and temporal behavior of microbes in water are critical steps in the process. This dissertation presents methodological approaches for understanding sources and fate of microbes in water. Some of these approaches include studying the land-use of the particular watershed of interest, collecting and studying hydrological data, sampling surface water at critical times and locations, analyzing water samples with culture-independent methods such as host-specific gene markers and whole genome shot-gun sequencing of DNA extracts, performing statistical analysis, and others. Furthermore, this dissertation presents an approach for identification of disease signals in water samples.

Four studies have been conducted in total in impaired water bodies in the Great Lakes basin. The first study concluded that an effective watershed remediation plan should take into account fluctuation of pollutant loadings and the timing of first-flush events. This second study offered a proposed microbial pollution source tracking methodology path that involves multiple lines of evidence to identify the sources and location of microbial pollution using host-specific markers and whole genome sequencing microbial community analysis, in association with landuse information. The third study suggested that sampling runoff-impacted surface water from agricultural areas represents a community fecal and urine sample of the livestock population in the sub-watershed; therefore, watershed bio-surveillance may serve as a screening tool for the presence of potential disease outbreaks in the corresponding livestock population. The fourth study proposed the use of bacteria concentration-discharge relationships to explain the source/origin and transport of contaminants in water.