USING DNA-BASED METHODS TO DETECT AND IDENTIFY FECAL CONTAMINATION SOURCE IN GROUNDWATER TO AUGMENT CULTURE-BASED DETECTION OF FECAL POLLUTION

USING DNA-BASED METHODS TO DETECT AND IDENTIFY FECAL CONTAMINATION SOURCE IN GROUNDWATER TO AUGMENT CULTURE-BASED DETECTION OF FECAL POLLUTION

BY PAUL NAPHTALI, B. Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

(Biology) Hamilton, Ontario

TITLE: Using DNA-Based Methods to Detect and Identify Fecal Contamination Sources in Groundwater to Augment Culture-based Detection of Fecal Pollution

AUTHOR: Paul Naphtali, B. Sc. (Redeemer University College, Ancaster, ON)

SUPERVISOR: Dr. Herb E. Schellhorn

Number of Pages: xii, 139.

ABSTRACT

Residents in rural communities across Canada rely on groundwater as their main drinking water source, but the private maintenance of this source may increase the risk of fecal contamination caused by human or animal wastes. Wainfleet, a rural Ontario community, has been under an active boil water advisory for the past decade. The last study to assess groundwater quality, performed in 2007, determined that half of the 586 groundwater wells contained exceedances in total fecal coliform and *E. coli* counts. An examination of fecal contamination levels and its sources is not only necessary for maintaining public health in the township, but is also an opportunity to examine the robustness of culture-independent methods for quantifying and sourcing fecal contamination in groundwater environments across Canada. For this project, culture-based and culture-independent methods were utilized to quantify and source any fecal contaminants in Wainfleet's groundwater. Culture counts of fecal indicator bacteria (FIB) suggested that some of the groundwater wells were receiving more fecal contamination than others, as expected based on a previous study that was conducted 10 years prior. The groundwater wells with higher *E. coli* counts also had higher abundances of sequences annotated to microbes like *Campylobacterales* which could come from septic tanks and higher concentrations of oxidized nitrogen which could also indicate human-based fecal contamination. Finally, fecal contamination in groundwater wells with *E. coli* tested positive for the human *Bacteroidales* marker. Taken together, this study shows that fecal contamination pervades groundwater wells across the boil water advisory zone, much of which originates from leaking septic tanks and poorly-constructed groundwater wells. In

this study, we have shown that a suite of protocols, from physiochemical quantification to targeted sequencing and qPCR, can be used to complement culture-based assays in quantifying and pinpointing fecal contamination in groundwater sources.

ACKNOWLEDGEMENTS

The Masters Degree has been my first notable foray into the world of scientific research. As a student, I received much feedback and advice about my research which will help me as continue my journey into the world of academia. I would like to thank a host of people for helping me through my thesis project and subsequently my exploration in the world of microbial ecology and bioinformatics. I would first like to express my gratitude towards my supervisor Dr. Herb E. Schellhorn for providing me with research suggestions and providing feedback throughout my Masters project. Your mentorship has fostered my love for science further and helped me harness those energies towards best practice and development into an independent researcher. I would also like to extend my sincere thanks to my committee members, Dr. Brian Golding and Dr. Tom Edge, for also providing me with support and guidance for sourcing fecal contamination in Wainfleet, and to Dr. Jian-ping Xu for being the chair for my MSc Defense.

I would also like to extend my thanks to Trevor Imhoff and the Township of Wainfleet for allowing us to collect groundwater well samples for analysis. I am very thankful for the time you have provided to making this project successful, from sampling the wells to acting as a liaison between the lab, the government, and the residents. Thank you also for helping me get out of the chicken poop mound when we collected fecal samples in the dead of winter! I am also thankful to the volunteers who allowed us to sample their groundwater and installed UV treatment systems as part of the study. Your cooperation made this thesis possible, and I hope this study becomes a positive influence for the rural community.

I am also very grateful for the support of former and current lab members. Thank you Farhaan for spending extra hours to help me process the groundwater samples whenever they arrived in the summer. I hope you learned a lot while working with me. I also want to thank Deepinder, Rachelle, Mahi, Cathy, Vivian, Ashish, William, and Kristine for their unending moments of laughter and discussion that gave me reasons to smile and keep plowing through my thesis. You all facilitated my maturation as a person, and I will never forget that. One person from the Schellhorn lab especially receives thanks. So Sakis, thank you so much for answering all of my questions about my research and facilitating scientific discussion in our lab. Thank you again for all your contributions to our lab.

I would also like to thank the All Things Campus Ministry and Dr. Michael Fallon for welcoming me to the group. Michael, the discussions I had with you helped me learn to incorporate my beliefs into my research and everything else I do. I would like to especially thank Adam and the Men's Bible study group, along with Amanda, Gerrit, and Megan. All of you have brought sunshine into my life when I struggled with my research and have helped me become a more holistic person.

Last, but most certainly not least, I would like to thank my family for the countless times they have listened to me ramble about *E. coli* and water quality. Mom and dad, thank you for the emotional and financial support during my first journey into academia. I want to thank Alice for editing my thesis and making sure I was as succinct as possible, and James for working with me the past summer and helping me process water samples while I worked on my thesis. None of this would be possible without my family.

Table of Contents

Table of Figures and Tables

Chapter 1: Microbial Source Tracking: Using Culture-Independent Assays and Metagenomics Tools to Pinpoint Sources of Fecal Contamination in Source Waters

Abstract

Drinking water sources may become contaminated with human and animal fecal matter. Culture-based assays of fecal indicator bacteria are the standard method for detecting fecal contamination, but they, in themselves, cannot be used to conclusively track pollution sources. Molecular assays have been developed to detect genetic markers that can provide additional diagnostic information for identifying contamination sources. These assays generally target a single diagnostic gene marker of interest and may be subject to biases that can compromise the quantification of these markers. In contrast, wholescale surveys of microbial communities in drinking water and potential pollutant sources with targeted DNA sequencing may provide additional insights into fecal contamination sources. The advent of next-generation sequencing can enable the detection of genera-containing waterborne pathogens in environmental samples to provide a preliminary risk assessment of an infectious disease outbreak caused by contaminated drinking water. The increased affordability and continual improvement upon this technology will provide water quality monitors and end-users better tools to detect and source fecal contamination in source waters across Canada and the globe.

1.1 Introduction

Drinking and recreational waters can be contaminated with anthropogenic inputs like sewage and agricultural runoff. Such activities can result in ecological and health costs like source water contamination and ecosystem destruction (Vorosmarty et al. 2010). Human and animal wastes contain waterborne pathogens that can be transferred into aquatic environments. These pathogens can be etiological agents of regional waterborne outbreaks like the Walkerton Tragedy (Hrudey et al. 2002). The identification and quantification of fecal contamination sources for any given water source could help to prevent waterborne disease outbreaks.

This literature review discusses the different methods that can be utilized to detect fecal contamination in source water environments. While standard culture-based methods can provide a preliminary survey of drinking water quality, culture-independent methods can be used to identify and quantify fecal contamination. Reduced sequencing costs have also facilitated the use of targeted sequencing of the 16S gene to detect fecal contamination in source waters. Taken together, water analysts have a plethora of tools available for quantifying and sourcing fecal contamination which can help end-users ensure public health.

1.2 Culture-Based Detection of Fecal Contamination

Pathogen detection in source waters is time-consuming and laborious because there are many potential pathogenic species that could be present at concentrations close to detection limits (Field and Samadpour 2007; Stoeckel and Harwood 2007). While monitoring water quality began with the use of fecal coliform counts (Ashbolt et al. 2001),

E. coli and enterococci, fecal indicator bacteria (FIB), are now considered the gold standard for assessing the safety of drinking water sources (Figueras and Borrego 2010) and are used as surrogates for public health risks (Harwood et al. 2014).

The use of standard methods assumes that FIBs are present together with waterborne pathogens when FIBs are detected. However, FIBs counts can be poorly correlated with waterborne pathogen detection in beach waters (Colford et al. 2007) and freshwater environments (Wilkes et al. 2009). The FIBs that are detected upon culturing may also represent environmental and fecal strains. *E. coli* isolates can be naturalized in source water environments (Walk et al. 2009), and enterococci populations can survive and grow in freshwater environments (Ran et al. 2013). The high FIB diversity in environmental samples compromises the ability to accurately quantify fecal contamination in drinking water sources. Culture-based methods on its own cannot distinguish fecal and environmental FIB colonies on culture plates, which also weakens the correlation between FIB and waterborne pathogen detection. Thus, there is a need to use other diagnostic markers to quantify and source fecal contamination.

1.3 Culture-Independent Markers of Fecal Contamination

Microbial source tracking (MST) is a discipline can be used to determine sources of fecal pollution and pathogens in drinking water sources (Simpson et al. 2002). The use of a DNA marker or indicator that is unambiguously associated with a fecal source in a water sample forms the basis for all MST assays (Field and Samadpour 2007). The two major objectives of MST as a discipline are to pinpoint sources of fecal contamination in drinking water and to detect waterborne pathogens when applicable. To accomplish these

goals, genetic markers can be used to detect host-specific fecal microbes in source waters. 16S genetic markers that belong to host-specific *Bacteroidales* and potential waterborne pathogens will be the focus for this section of the literature review.

1.3.1 Host-specific *Bacteroidales* as Genetic Markers of Fecal Contamination

16S rRNA sequences can act as fingerprints that distinguish one microbial group from another (Woese and Fox 1977). 16S rRNA markers of host-specific fecal microbes were first identified for human-based *Bacteroidales*, designated the HF183 marker (Bernhard and Field 2000). Gut *Bacteroidales* are obligate anaerobes, found in high concentrations in fecal matter, and contain host-dependent, niche-specific features that delineate *Bacteroidales* in the guts of warm-blooded mammals (Coyne and Comstock 2008). Since the identification of the HF183 marker, other target sequences within the *Bacteroidales* order that are found in specific animal hosts (Table 1). All primers were developed by isolating DNA from *Bacteroidales* isolates that originated from fecal samples and sequencing their 16S rRNA gene. During the development of primers targeting the universal and host-specific *Bacteroidales* 16S rRNA sequences, host-specific sequences were identified by determining sequence identity of a given 16S rRNA fragment to a host. Phylogenetic analyses of the 16S rRNA sequences also generates host-specific clades where host-specific sequences can be identified and amplified. Specific conditions in the host's gut can facilitate the establishment of host-specific microbial populations (Ley et al. 2008a). These conditions in turn can result in the formation of host-specific phylogenetic clusters within the same bacterial order, like *Bacteroidales* (Dick et al. 2005).

Bacteroidales markers have been used to quantify the contributions of animal farms and sewage to source water contamination (Okabe et al. 2007; Parker et al. 2010; Marti et al. 2013). In each case, the sites with the most host-specific 16S *Bacteroidales* gene markers were detected at freshwater sites nearest to host-specific farm sites. Furthermore, the detection of *Bacteroidales* markers was also increased with *E. coli* counts during stormwater events, suggesting that wastewater can be transferred at elevated rates during a storm (Parker et al. 2010). These studies show that host-specific *Bacteroidales* markers can be used to track sources of fecal contamination in freshwater environments alongside culturable FIB counts.

Bacteroidales markers can also be used to assess groundwater contamination levels. Like freshwater environments, increased *Bacteroidales* 16S rRNA copy numbers were correlated with increased culturable FIB counts in karst groundwater samples (Johnson et al. 2011). This trend was also recorded for residential groundwater wells, where elevated human *Bacteroidales* 16S rRNA concentrations represented fecal contamination originating from leaking septic tanks (Krolik et al. 2014), with many of the wells located within the Hastings and Prince Edward counties. Levels of human fecal contamination was also assessed in a shallow sandy aquifer, where monitoring wells closest to ponds with *E. coli* exceedances had the highest human *Bacteroidales* concentrations (Knappett et al. 2012).

Bacteroidales bacteria can comprise environmental populations. These microbes can confound the quantification of total fecal contamination loads with a universal fecal *Bacteroidales* PCR assay. In a previous study, pristine soils have been tested positive for

the AllBac and BacUni markers (Vierheilig et al. 2012). The quantification of total *Bacteroidales* markers may be confounded by the presence of environmental *Bacteroidales* populations. Although environmental *Bacteroidales* populations exist, the development of host-specific *Bacteroidales* markers can circumvent this problem as each of the primers listed in Table 1 are specific to fecal matter shed by their respective animals.

Table 1. List of primers targeting general and host-specific *Bacteroidales* populations for qPCR assays.

Size of the gene fragment, along with the gene fragment and primer name and provided within the context of the reference study.

Gene Target Name	Primer Combination	Fecal Target	Size of Amplicon(bp)	Citation
UniBac	BacUni520f-BacUni690r	General *Bacteroidales*	170	(Kildare et al. 2007)
HF183	HF183F-Bac708R	Human *Bacteroidales*	541	(Bernhard and Field 2000; Green et al. 2014)
HuBac	HuBac566f-HuBac692r	Human *Bacteroidales*	116	(Layton et al. 2006)
CF128	CF128F-CF196R	Cow *Bacteroidales*	596	(Bernhard and Field 2000)
Pig-2-Bac	Pig-2-Bac41F-Pig-2-Bac163R	Pig *Bacteroidales*	116	(Mieszkin et al. 2009)
HorseBact	Ho622F-Ho722R	Horse *Bacteroidales*	100	(Silkie and Nelson 2009)
Chicken-Bac	qC160FU-HU-qCD462R-HU	Chicken *Bacteroidales*	102	(Kobayashi et al. 2013)

1.3.2 Waterborne Pathogen Detection with Genetic Markers

Genetic markers specific for waterborne pathogens can be used to help assess the risks of a waterborne outbreak. Many of the qPCR assays that amplify pathogen-specific sequences target virulence genes (Table 2). The three *stx* genes commonly found in pathogenic *E. coli* strains encode Shiga toxins that can cause kidney damage and hemolytic uremia syndrome (HUS) (O'Brien et al. 1984). The *invA* gene, expressed only by *Salmonella typhimurium*, encodes the ability to invade intestinal epithelial cells (Galan et al. 1992).

qPCR assays that quantify these virulence factors in water samples have been used to assess the ability of a municipal wastewater treatment system to remove waterborne pathogens (Shannon et al. 2007). The qPCR-based detection of multiple virulence genes belonging to waterborne pathogens like *Campylobacter spp.* and *Salmonella* was also used to determine the risk of a waterborne outbreak originating from a freshwater lake, where pathogens shed from migratory birds were detected at concentrations that can cause disease risk (Ishii et al. 2014). The virulence factors belonging to waterborne pathogens were also quantified in addition to fecal indicator chemical quantification in a freshwater river that received a manure spill (Haack et al. 2015). FIB counts and virulence genes were still detectable a month after the spill, suggesting that long-term contamination of the river can occur as a result of a manure spill. The detection of genetic markers belonging to waterborne pathogens can be used in place of selective culturing to detect pathogenic microbes in source waters.

Conserved signature indels (CSIs), formed as a result of a genetic change observed in all members of a given taxonomic group (Gupta 1998), could delineate bacterial groups at the species level. The conserved nature of these indels makes CSIs useful markers for waterborne pathogens. Recently, a 9-bp CSI sequence was identified that is specific to *E. coli* O157:H7 pathogen and subsequently used to develop a pathogen-specific test for O157:H7 strains (Wong et al. 2014). Other CSIs may also be present in other waterborne pathogens that can be used for other MST assays. CSIs were identified *in silico* for members of the Epsilonproteobacteria class, containing many waterborne pathogens like *Campylobacter jejuni* and *Helicobacter pylori* (Gupta 2006). CSIs can thus serve as potential markers of waterborne pathogen contamination and public health risks in a drinking water source.

Table 2. List of primers targeting virulence factors in waterborne pathogens.

Size of the gene fragment, along with the gene fragment, primer name, and type of fecal matter the waterborne pathogen is associated with are provided.

Gene Target Name	Organism	Fecal Target	Size of Amplicon (bp)	Citation
glyA	*Campylobacter jejuni*	Cow, chicken	154	(Leblanc-Maridor et al. 2011)
invA	*Salmonella typhimurium*	Variable	244	(Chiu and Ou 1996)
exoT	*Pseudomonas aeruginosa*	Human	285	(Clark et al. 2011)
stx1-stx3	*E. coli O*157:H7	Variable	95	(Jothikumar and Griffiths 2002)
rtxA, epsM, mshA	*Vibrio cholera*	Variable	120, 145, 113	(Gubala 2006)

1.4 Metagenomics: an Introduction

The world microbiome contains over 10^{30} prokaryotes, with over 99% of all microbial species being uncultivatable (Bunge et al. 2014). While the advent of culturomics has increased the resolution for identifying culturable microbes (Lagier et al. 2012), targeted sequencing of the 16S rRNA gene has provided insights into the overall microbial community structure in environments around the world. The 16S rRNA gene encodes a component of the 30S subunit of the prokaryote ribosome which can act as a genetic fingerprint that distinguishes microbes from each other (Woese and Fox 1977). 16S rRNA gene sequencing surveys can help to identify sources of FIB counts and therefore fecal contamination in source waters. Targeted sequencing of the 16S rRNA gene can also be used to identify markers of human-based contamination and waterborne pathogens in drinking water sources. Although culture-based assays only provide a qualitative estimate of fecal contamination levels, sequencing the 16S rRNA gene can profile entire microbial communities to estimate the relative contributions of potential contamination sources to source water contamination.

1.4.1 Sequencing Platforms and Tools for Profiling Microbial Communities

Targeted sequencing of the 16S rRNA gene is a type of metagenomics analysis used for profiling microbial communities. Sequences belonging to the 16S rRNA, encoding the small subunit of the prokaryotic ribosome, are conserved enough to identify a particular microbe, but diverse enough to distinguish different microbial groups from each other (Woese and Fox 1977). There are six hypervariable regions within the 16S gene, and these

regions can be used to differentiate between bacterial groups (Stackebrandt and Goebel 1994). Most studies focus on sequencing the V3-V4 hypervariable regions because it can distinguish the most microbes at the genus level while minimizing sequencing bias (Chakravorty et al. 2007).

Two sequencing platforms have emerged for sequencing 16S fragments from environmental samples in the past decade. The first of these platforms is 454 pyrosequencing (Roche). While pyrosequencing was formerly the standard sequencing pipeline, Illumina's MiSeq and HiSeq platforms have not only increased the amount of sequences that are outputted, but also provide longer, if not equal read lengths as pyrosequencing (van Dijk et al. 2014). As longer read lengths are recommended for ensuring that the sequences are annotated to the correct microorganisms (Wommack et al. 2008), the Illumina platforms have now become the standard sequencing platform for targeted sequencing.

16S rRNA sequences that are obtained after running a sequencing platform must be quality-trimmed and then annotated to database sequences belonging to known microorganisms. Programs like Trimmomatic (Bolger et al. 2014) can be employed to trim low-quality sequences. Once the sequences are trimmed, software pipelines like QIIME (Quantitative Insights into Microbial Ecology) (Caporaso et al. 2010) and Mothur (Schloss et al. 2009) can be utilized to annotate the remaining sequences against known databases like Greengenes (DeSantis et al. 2006) to form reads and then cluster the reads into operational taxonomic units (OTUs). OTUs are microbial identifiers based on sequencing reads that are used to generate community profiles. Preparing an established bioinformatics

pipeline is necessary to robustly profile the microbial communities within drinking water sources.

1.4.2 Sourcing Fecal Contamination with Targeted Sequencing

The guts of warm-blooded animals contain host-specific microbial populations as a result of their unique diets (Ley et al. 2008a). Source water microbiotas are distinct from human and animal gut microbiotas (McLellan et al. 2010). Thus a source water environment whose microbiota is more similar to a specific sewage or fecal microbiota may suggest contamination of the water source by that fecal input (Unno et al. 2010). As a proof-of-concept, amplicon sequencing of the 16S rRNA gene in groundwater samples and surrounding sea water samples revealed that groundwater wells receiving sea water inputs were also receiving microbes found only in seawater samples (Unno et al. 2015). As an evaluation of the Illumina MiSeq platform, pristine water samples were spiked with either one or two types of fecal matter (Cao et al. 2013). The Illumina MiSeq pipeline correctly identified the dominant fecal contaminant in 95% of the 64 challenge samples using a reference dataset of 12 fecal types. The decreasing cost for 16S rRNA amplicon sequencing and the ability to process many water and fecal samples simultaneously enables the use of targeted sequencing to track sources of fecal contamination in source water environments.

Sequencing the 16S rRNA gene can be employed to identify novel fecal markers that can be used to pinpoint fecal contamination sources. 16S rRNA sequencing can be used to identify genera that are associated with sewage, like *Arcobacter* and *Acinetobacter* in wastewater influent samples (Newton et al. 2013). *Lachnospiraceae* are a family within the *Clostridiales* order that could provide huma-specific markers, as determined by 16S

rRNA sequencing (Fisher et al. 2015). At the community-scale, β-diversity analyses can provide insights on the similarity or dissimilarity of microbial communities between fecal and source water samples. Total microbial community diversity decreased when more effluent was dumped into freshwater environments (Drury et al. 2013). Genetic markers at the microbial and community scale can be used to identify and source fecal contamination in source waters.

A bioinformatics program has also been developed to track potential sources of fecal contamination in source waters. Known as SourceTracker, the R-based software can be used to predict the likely sources of reagent contamination within a DNA extract using a Bayesian Method (Knights et al. 2011). Although the program is originally intended for minimizing the effects of DNA contamination in microbial community analyses, SourceTracker has become more prominent for assessing the primary contamination sources in recreational beach sands and correlating rainfall and sand resuspension with the detection of human fecal indicators (Henry et al. 2016).

1.4.3 Detection of Waterborne Pathogens with Metagenomics

MST assays are primarily used to pinpoint sources of fecal contamination. However, the detection of waterborne pathogens in source waters constitutes another MST goal (Simpson et al. 2002). Chromogenic methods are available to detect pathogens like *E. coli* O157:H7 strains on selective plates like Rainbow Agar (Bettelheim 1998). Non-*E. coli* O157:H7 strains may grow on these selective plates, yielding false positive results. (Quilliam et al. 2011). Bacteria can also survive on plates but not form colonies by entering the viable but non-culturable state (VBNC), representing false-negatives (Oliver 2010).

Bacteria enter the VBNC state when they are exposed to environmental stresses like exposure to white light and temperatures outside their optimal range for growth (Oliver 2005). These bacteria can also be resuscitated when exposed to favourable growth conditions (Ramamurthy et al. 2014). The non-detection of bacteria that are in the VBNC state can increase the risk of underestimating the risk of waterborne illness caused by exposure to the contaminated water source.

The use of targeted sequencing has enabled the detection of bacterial taxa which potentially contain waterborne pathogens in drinking water and wastewater environments (Aw and Rose 2012). Although 16S rRNA sequencing cannot always distinguish microbial taxa to the species level, 16S sequencing can still be used to detect genera that are associated with a waterborne pathogen. In wastewater treatment plants, biosolid samples contained elevated abundances of 16S OTUs annotated to *Mycobacterium* and *Clostridium* (Bibby et al. 2010). 16S rRNA sequencing can also be used to detect waterborne pathogens in recreational and freshwater sites (Ibekwe et al. 2013) and activated sludge samples (Kumaraswamy et al. 2014). In both studies, a diverse group of genera containing waterborne pathogens have been detected in human waste samples, including members of the *Clostridium*, *Aeromonas*, *Pseudomonas*, and *Bacillus* genera.

1.4 Conclusions

The provision of safe drinking water is a priority for maintaining public health. For the past half-century, methods for quantifying FIB like *E. coli* and enterococci as a proxy for fecal contamination were standardized. These methods, in themselves, cannot pinpoint sources of fecal contamination. qPCR assays targeting 16S rRNA markers belonging to

host-specific fecal *Bacteroidales* and virulence genes have enabled the quantification of host-specific fecal contamination loads and improve risk assessment caused by pathogens in source waters. Targeted sequencing of the 16S rRNA gene is also increasingly used to characterize microbial community structures in many environments. 16S rRNA gene sequencing can be employed to identify host-specific microbes and conduct wholescale comparisons of microbial communities in potential contaminants and drinking water sources. These two datasets, when used together, can identify the primary fecal contamination sources in drinking water. The detection of waterborne pathogens with 16S rRNA sequencing can also help to determine the risks of an infectious disease outbreak. Water analysts have many tools available to assess the safety of a drinking water source. Standard culture-based methods can now be complemented with genetic-based methods to provide end users with the data necessary ensure the provision of safe drinking water to the general populace.

Chapter 2: Culture Counts and Physiochemical Quantitation in Selected Sites across Wainfleet

2.1 Abstract

Groundwater aquifers are an important source of drinking water for rural communities across Canada. Boil water advisories for drinking water sources are issued by public health officials when fecal contamination of the drinking water source is suspected. Residents of Wainfleet, an Ontario township, rely on groundwater as their primary drinking water source. Wainfleet's homeowners are under the longest active boil water advisory in Canada because of fecal contamination in their groundwater. Standard methods for detecting fecal pollution in drinking water sources use fecal indicator bacteria (FIB) as representative of fecal contamination. For this component of the study, groundwater wells across Wainfleet were sampled for *E. coli* and enterococci, the standard FIB. Nitrate and phosphate concentrations in selected groundwater wells were also measured. *E. coli* counts remained the same over the past decade, whether they previously had high or low *E. coli* counts. FIB contamination pervaded the groundwater wells within the boil water advisory zone. Furthermore, some groundwater wells had far higher FIB counts than the rest of the wells. While molecular assays are needed to delineate fecal contamination sources, the use of culture and physiochemical assays has identified not only the general state of the town's groundwater contamination, but also specific groundwater wells that may have higher contamination levels and must be further examined for pollution sources.

2.2 Introduction

Waterborne pathogens are present in sewage and animal manure. The consumption of drinking water from contaminated drinking water sources can lead to waterborne outbreaks like the Walkerton Tragedy (Hrudey et al. 2002). The direct detection of all potential waterborne pathogens in drinking water is not currently feasible as there are diverse pathogenic species and strains that found in low concentrations (Straub and Chandler 2003). Standard methods for identifying fecal contamination utilize fecal indicator bacteria (FIB) that, when detected, point to fecal pollution in drinking water sources. When FIB from fecal matter are shed into aquatic environments, they can be isolated from water samples and grown on selective plates for colony counts. Fecal coliforms, *E. coli*, and enterococci are the FIBs most commonly used to detect fecal contaminants in drinking water sources and are considered the gold standards for FIB (Figueras and Borrego 2010).

Wainfleet is a rural township located by the shores of Lake Erie. Groundwater is the only source of potable water in the town, and residents use privately-owned groundwater wells to pump the water. Homeowners across Wainfleet are currently under the longest active boil water advisory in Canada, an indicator of the town's deteriorating groundwater quality. A report that was conducted on private groundwater wells across Wainfleet 16 years ago determined that 28% and 54% of the groundwater wells contained *E. coli* and fecal coliform counts that exceeded provincial limits of non-detection respectively. The sewage effluent from malfunctioning septic tanks could diffuse through the thin clay soil, penetrating the aquifer and causing groundwater contamination. An

update on FIB counts, along with genetic assays that evaluate septic tanks as a fecal contamination source, is needed to provide an up-to-date assessment on groundwater quality and to minimize the risks of a waterborne outbreak in Wainfleet.

Physiochemical assays can be performed alongside culture counts to quantify contamination from sewage and animal manure. Two of the most commonly used chemical indicators of sewage and fertilizer contamination are nitrogen and phosphate ions. The cadmium reduction method and stannous chloride method are the standard protocols for quantifying nitrate and organic phosphate ions respectively (Rice et al. 2012). The cadmium reduction method uses two steps. The first step is the reduction of all nitrates in the sample to nitrites with cadmium granules and all the nitrites are then reacted with a colour reagent, which forms a pink solution (Cortas and Wakid 1990). The stannous chloride method converts phosphate ions into molybdophosphoric acid which then reacts with stannous chloride to form a blue-emitting molybdenum-blue complex (Crouch and Malmstadt 1967). Both methods measure the absorbance of the resultant complexes which in turn can be converted into nitrate and phosphate concentrations respectively. The recommended limit for both of these compounds in source waters is 10 mg/L, as determined by Canada Public Health and the US Environmental Protection Agency respectively. Elevated oxidized nitrate concentrations were detected in groundwater wells that were located near cattle and fruit farms, indicating manure and fertilizer contamination (Lockhart et al. 2013). Excess phosphate concentrations can also indicate fertilizer contamination from farms (Heathwaite et al. 2005) and septic tank effluents (Jarvie et al. 2006).

This component of the study provides a preliminary quantitation of fecal contamination in groundwater wells collected across Wainfleet. Here, FIB culture counts, along with nitrate and phosphate concentrations, were determined in privately owned drinking water sources for comparison with data obtained from the previous water quality study done in Wainfleet 16 years prior. With this information, specific groundwater wells were identified for further examination of fecal pollution with molecular assays.

2.3 Methods

2.3.1 Study Site, Water and Fecal Sample Collection

Drinking water samples were collected from the taps that collected water from residential groundwater wells. Willing volunteers in Wainfleet were identified through interviews, and tap water samples were collected from these residences between April and November 2015 (Figure 1). For each month, between four and seven unique groundwater wells were sampled. During the spring and summer months, water samples were collected in single replicates into 500-mL autoclaved Nalgene plastic bottles from residential tap systems that collect water from privately owned groundwater wells. Groundwater samples were collected in duplicate in October and November. UV treatment was also conducted on a subset of groundwater wells. All water samples that were exposed to UV treatment were labelled with the sample ID and then the letter "A". All water samples were transported to the lab inside a cooler box and processed within 8 h upon arrival.

2.3.2 Culture-Based Assays

Escherichia coli and enterococci colonies were counted on all groundwater samples using the membrane filtration assay, following standard procedures (Rice et al. 2012). Briefly, 100-mL of groundwater tap and 100-fold serially-diluted water samples were passed through 0.45 μm membrane filters. All filters were transferred aseptically onto differential coliform (DC, Oxoid) and mEI (BD Difco) agar plates for selective growth of *E. coli* and enterococci counts respectively. All plates were incubated for 24 h at 42^{o}C. All

counts were recorded as CFU/100 mL after multiplying the colony counts on the plate containing 20-100 colonies by its dilution factor.

2.3.3 Cadmium Reduction and Stannous Chloride Methods

The cadmium reduction method was performed on water samples collected from private groundwater wells from June to August 2015, following the standard methods closely (Rice et al. 2012). Oxidized nitrogen concentrations were then calculated against a standard curve that was generated with nitrate standards with a concentration range from 0 ppm to 10 ppm (Figure S1).

All phosphate quantitation was performed with the stannous chloride method, following standard procedures (Rice et al. 2012). Briefly, a drop of phenolphthalein was added to 50 mL of water sample. 2.0 mL and 500-mL of ammonium molybdate reagent and stannous chloride reagent was then added. The solution was mixed well and left to stand for 10 min but not longer than 12 min. The absorbance of the solution at 660 nm was then measured using a spectrophotometer against a previously prepared stannous chloride standard curve (Figure S2).

Figure 1. Map of Wainfleet providing the locations of groundwater samples collected across all sampling months.

Red points on the map represent the streets where the groundwater samples were located. Images were obtained from Google Maps.

2.4 Results

2.4.1 *E. coli* and enterococci counts in private groundwater wells

100-ml of each water sample was used to quantify culturable *E. coli* and enterococci colonies on selective agar plates. No more than six unique groundwater wells were sampled within a single sampling month. A total of 36 samples was collected across the sampling months during the year. On average, over 60% of the groundwater wells tested positive for the *E. coli* marker in all sampling seasons and over 70% of the groundwater wells were positive for *Enterococcus* in the summer and autumn months (Figure 2). Log-transformed *E. coli* counts in four individual groundwater wells were also of similar magnitudes as they were ten years ago (Figure 3).

The average *E. coli* and enterococci counts were substantially higher in two of the selected groundwater wells, one from Burnaby Road and the other from Harbourview Road (Figure 4). On average, *E. coli* and enterococci counts were 1000-fold and 100-fold higher in the groundwater wells labelled 11615 and 11215 than all other groundwater wells, regardless of the sampling month (Figure 4). 11615 contained the highest *E. coli* and enterococci counts with a $\log_{10}$ count of 6.08 and 3.53 CFU/100 mL respectively. In contrast, none of the other groundwater wells exceeded a $\log_{10}$ FIB count of 2.00 CFU/100 mL with the exception of 11431, which was sampled in August 2015 and had a $\log_{10}$ *E. coli* counts of 2.16 CFU/100 mL. A subset of the groundwater wells had a UV treatment system installed. Groundwater samples were collected from these wells before and after the UV treatment system was activated. Treating the drinking water with UV light decreased culturable bacteria to non-detection for all UV-treated water samples (Figure 5).

2.4.2 Nitrate and Phosphate Concentrations in Wainfleet's Drinking Groundwater

Nitrate concentrations were recorded in groundwater wells collected from June to August, while phosphate concentrations were measured in groundwater wells collected in May, July, August, and October. Nitrate and phosphate concentrations were the highest in groundwater wells whose waters were collected in August (Figure 6, Figure 7). On the other hand, groundwater wells that were collected in July had groundwater wells less than 1 mg/L of phosphate ions and non-detectable nitrate concentrations. A linear regression between nitrate concentrations and log-transformed *E. coli* counts resulted an R^2 value of 0.3363. A significant positive correlation was also determined by the Spearman's Rank Correlation Test, with a correlation coefficient of 0.647 (Figure 8, $P < 0.05$).

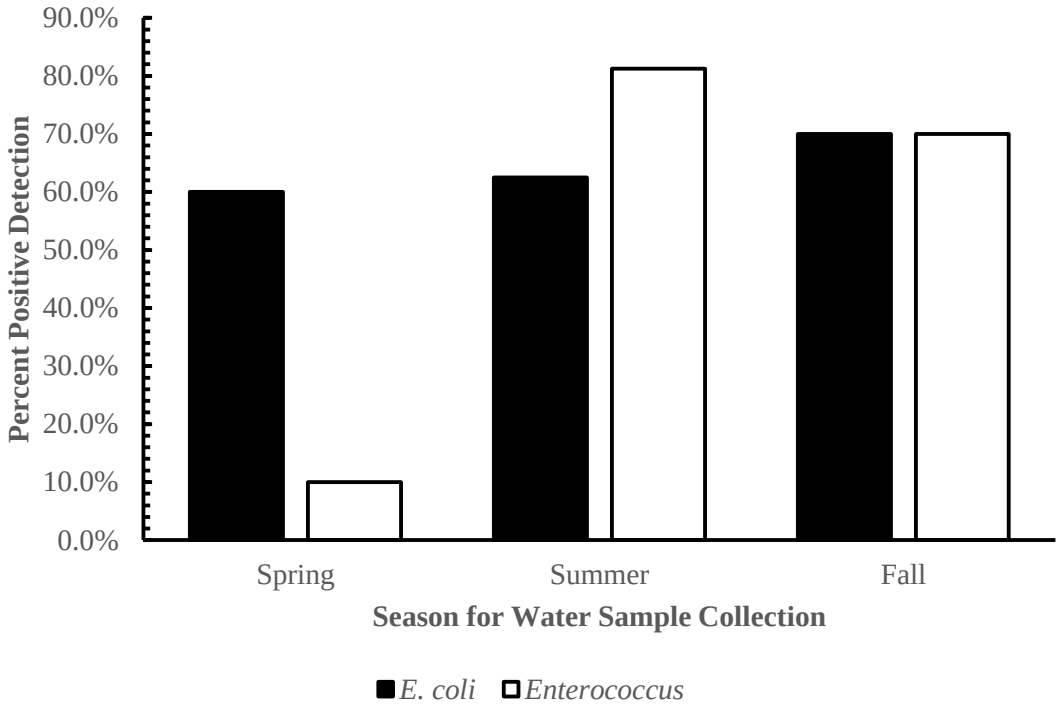

Figure 2. Percent Positive Detection of *E. coli* and *Enterococcus* colonies from groundwater wells collected during the three seasons in 2015.

Groundwater wells were considered positive for the fecal indicator bacteria if at least one CFU/100 mL was detected on agar-selective plates. n(Spring) = 10; n(Summer) = 16; n(Fall) = 10.

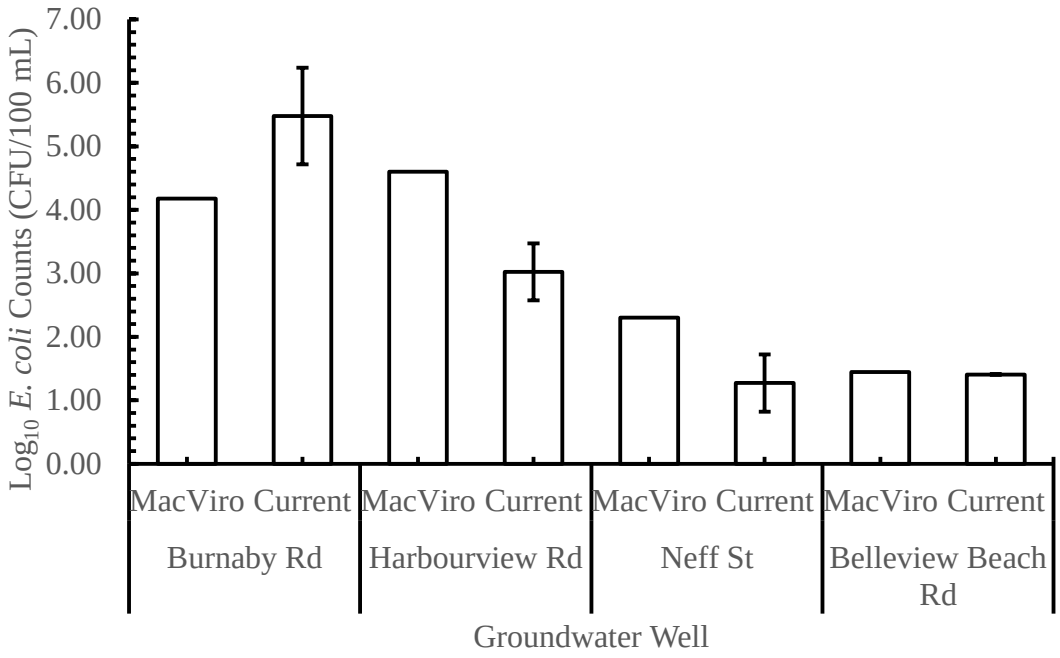

Figure 3. Log-transformed *E. coli* Counts in Four Groundwater Wells In the Previous MacViro Study and the Current Study

The MacViro study was conducted a decade ago as the first study to assess fecal contamination levels in Wainfleet's drinking water. Standard error bars were for the *E. coli* counts that were measured in the current study. n(Burnaby Road) = 4; n(Harbourview Road) = 3; n(Neff Street) = 3; n(Morgan's Point Road) = 2. *E. coli* counts for tap water collected from the groundwater wells across different sampling months were averaged to obtain the mean *E. coli* counts in the current study.

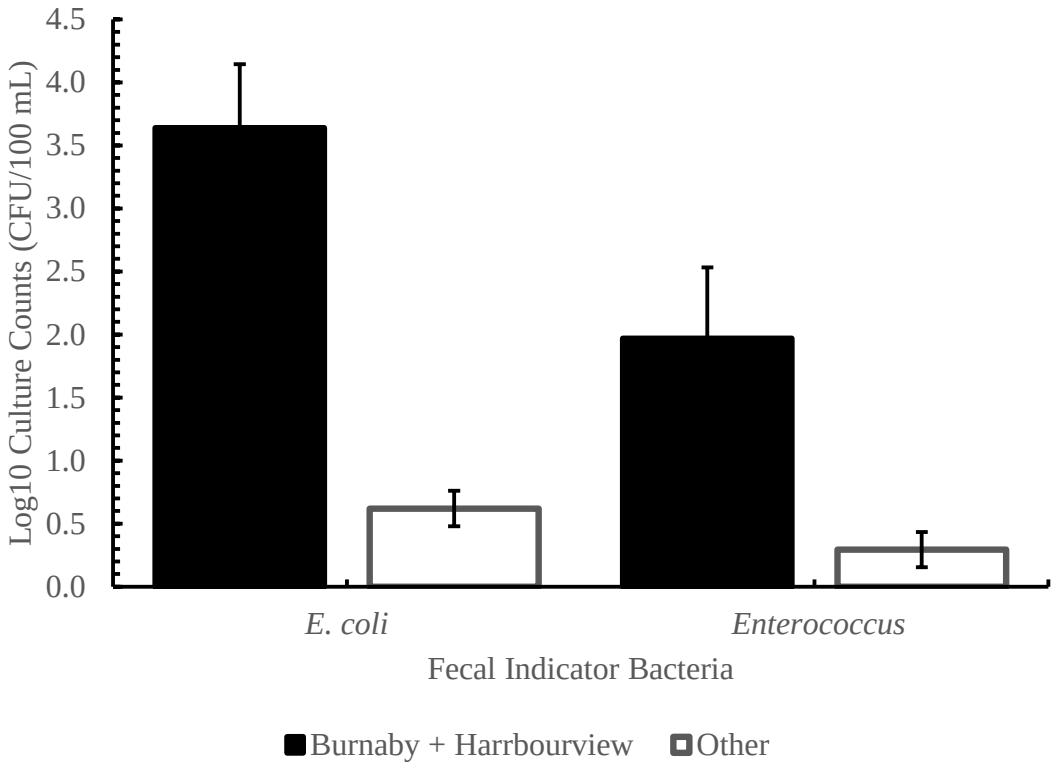

Figure 4. *E. coli* and *Enterococcus* counts in two groundwater wells vs other wells

Bar graph comparing *E. coli* and *Enterococcus* counts between two groundwater wells and all other groundwater wells that were sampled for this study. *E. coli* and *Enterococcus* counts were log-transformed to normalize the data. The population size for the study combines the amount of residences that were sampled and the months that the water samples were collected. n(Burnaby + Harbourview) = 7; n(Other) = 30. Standard error bars are provided.

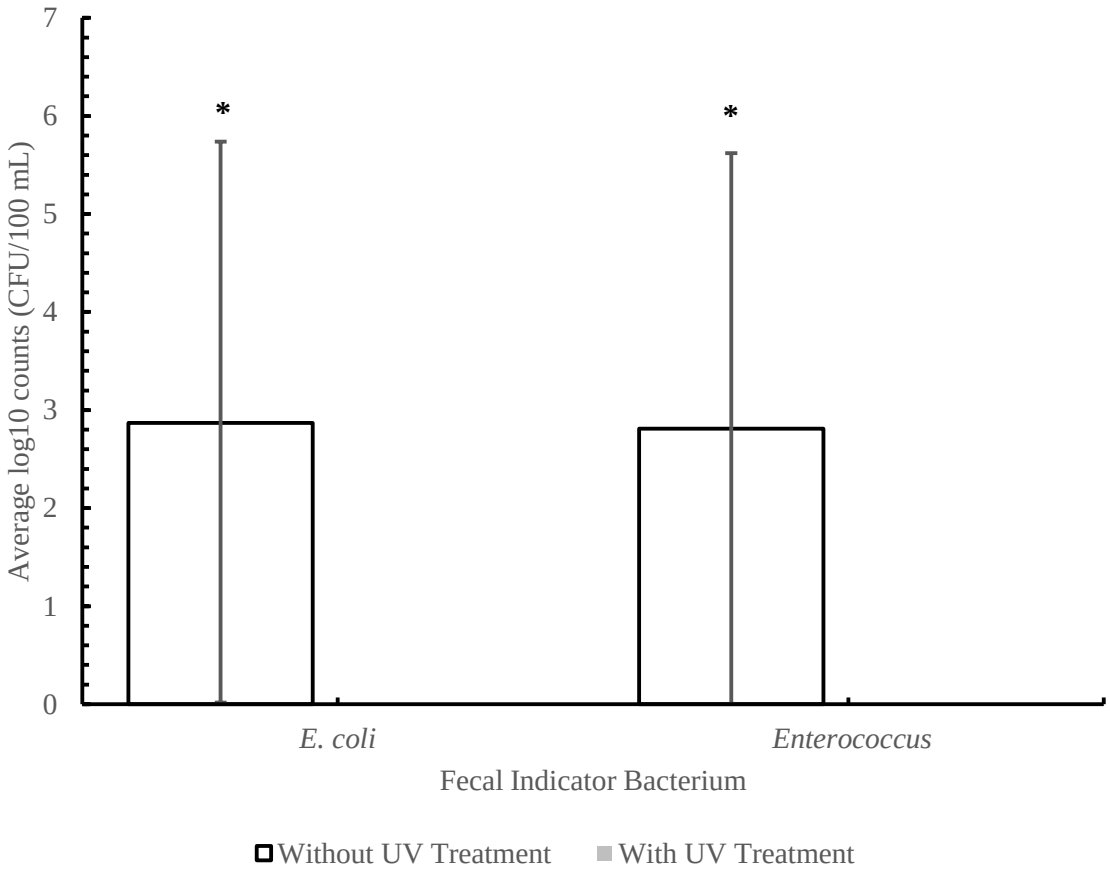

Figure 5. *E. coli* and *Enterococcus* counts with and without UV treatment

Groundwater wells with UV treatment systems installed were kept off to examine *E. coli* and *Enterococcus* counts without UV irradiation in the water. *E. coli* and *Enterococcus* counts were decreased to non-detection when the UV treatment systems were activated before water samples were collected. n(Without UV Treatment) = 6; n(With UV Treatment). Standard error bars are provided.

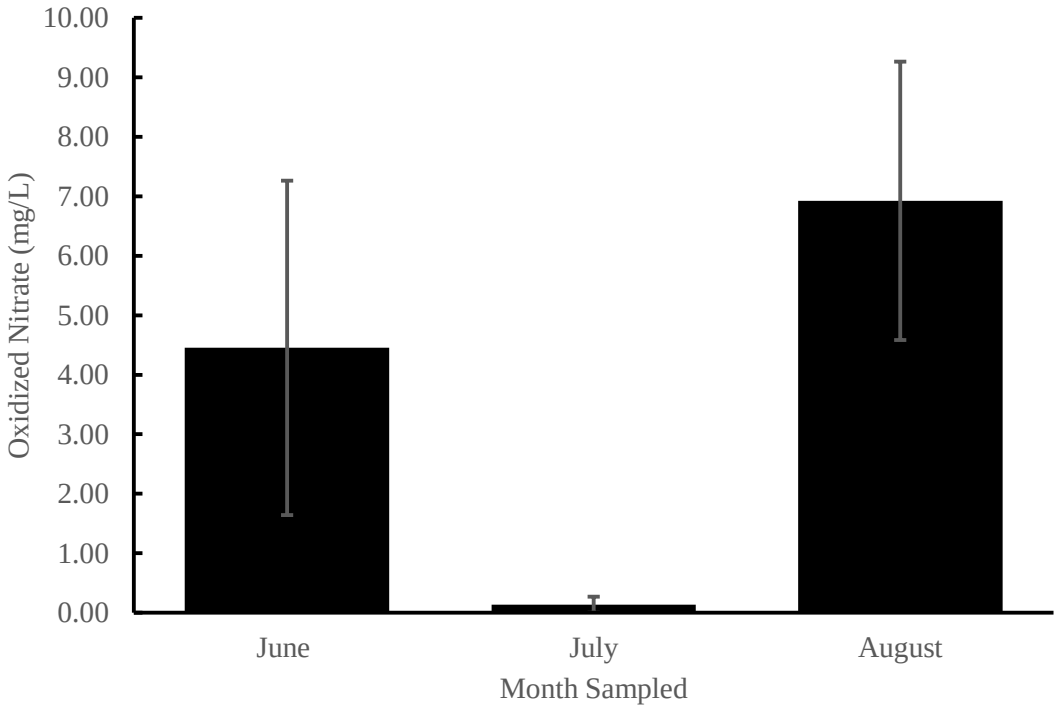

Figure 6. Mean concentration of oxidized nitrates in selected groundwater wells during the summer months

[NO_3^-N] was determined with the cadmium reduction method after comparing absorbances against a standard curve (Figure S1). n(June) = 6, n(July) = 7 n(August) = 5. Standard error bars were prepared.

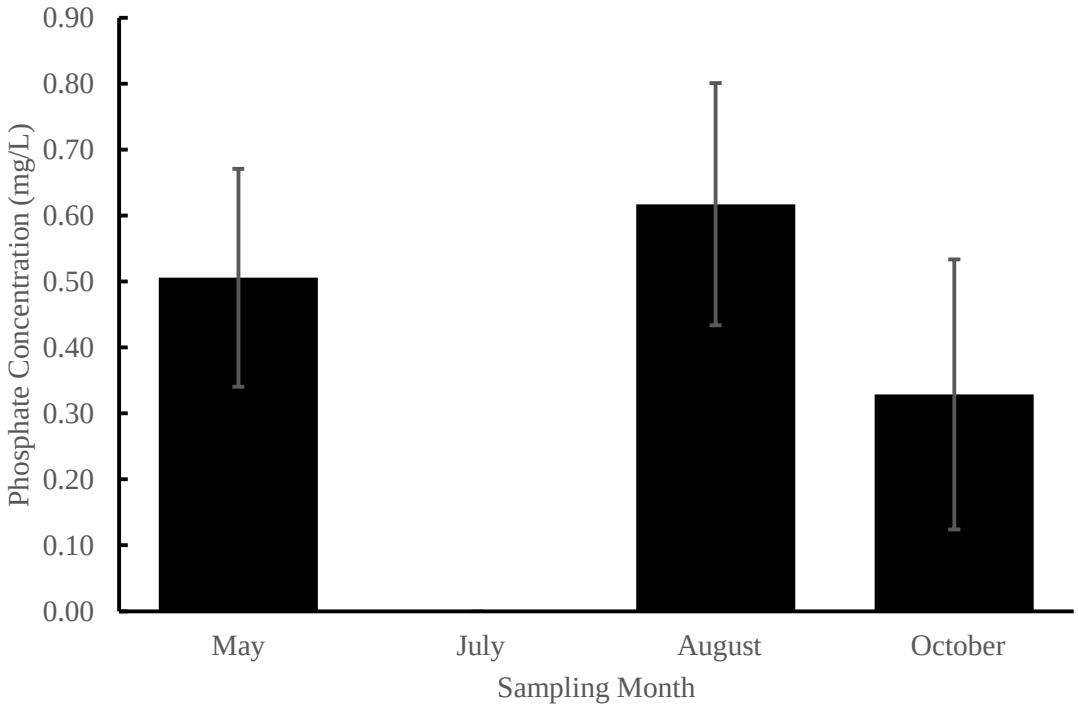

Figure 7. Phosphate concentrations in Wainfleet's groundwater wells

Mean phosphate concentrations were recorded among the groundwater well waters that were sampled within a given month. Phosphate concentrations were determined by comparing the absorbances of the solutions against a standard curve after using the stannous chloride method (Figure S2). All but one groundwater sample collected in July did not have detectable phosphate concentrations. n(May) = 7; n(July) = 7; n(August) = 5; n(October) = 3. Standard error bars were prepared.

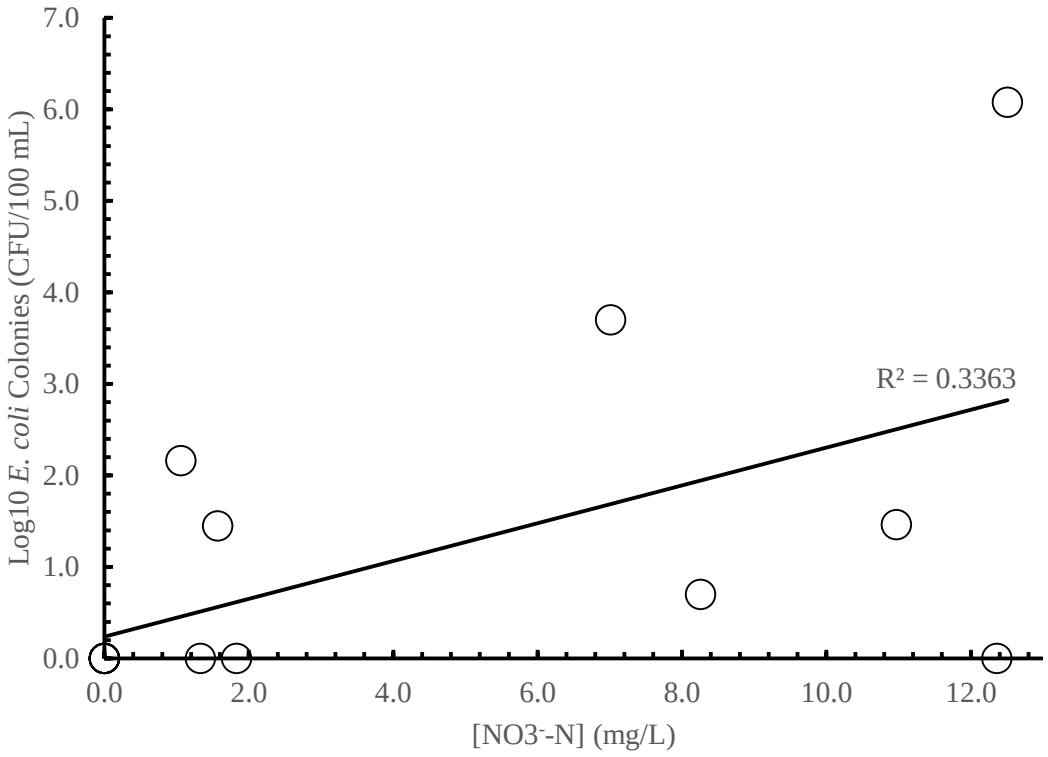

Figure 8. Linear Regression Analysis of Log-transformed *E. coli* Counts and Oxidized Nitrate Concentrations

Groundwater samples that were collected during the months when *E. coli* colony counts and oxidized nitrate concentrations were measured. An R^2 value of 0.3363 was obtained, with a correlation coefficient of 0.647 which was statistically significant ($P < 0.05$).

2.5 Discussion

Residents living in rural communities rely on private groundwater wells as their drinking water source. While guidelines for proper construction and maintenance of groundwater wells and septic tanks exist, they are not enforced as strictly as they are for public drinking water distribution systems and wastewater treatment plants. Groundwater sources are more susceptible to fecal contaminants from sewage and manure that can leech through the soils into the underlying aquifer. Groundwater contamination accounts for 64% of drinking water outbreaks between 1989 and 2002 in the United States (Fong et al. 2007). Monitoring fecal contamination in groundwater aquifers is necessary to prevent a waterborne outbreak from occurring. Over ten years prior, a study conducted on private groundwater wells across Wainfleet detected *E. coli* and fecal coliform counts well above background levels for over half of the 280 groundwater wells. Because the detection of these FIB indicates fecal contamination, an update on the *E. coli* and enterococci counts, along with nitrate and phosphate quantitation, is essential to ensure the safety of Wainfleet's residents.

2.5.1 *E. coli* and Enterococci Counts in Selected Groundwater Wells

FIB counts can provide a preliminary assessment of fecal contamination in a drinking water source. To provide an up-to-date assessment on the safety of Wainfleet's groundwater aquifer, four to seven groundwater wells were sampled monthly. Counts obtained from these wells can also be compared with those obtained from the previous study to determine whether fecal contamination levels decreased. This data can then be

used to provide end-users the means to evaluate the need to continue the boil water advisory for the township.

Over 60% of the tap water that was collected from groundwater wells were positive for *E. coli* or enterococci across all three sampling seasons (Figure 2). The consistent detection of *E. coli* across all sampling months and in many of the wells suggests that most of the wells are contaminated by fecal matter at some level. Fecal contamination is consistently found across wells within the boil water advisory zone. Furthermore, when compared with the previous study, *E. coli* and *Enterococcus* counts in four groundwater wells remained the same as they were a decade prior (Figure 3). Some groundwater wells also have higher FIB counts than others, as seen with the groundwater wells located in Burnaby Road and Harbourview Road (Figure 4). Little success to decrease and prevent the detection of FIB in groundwater wells has been recorded in this study, providing a reason to continue enacting the boil water advisory for the township. More action must be taken to ensure the removal and absence of FIB in drinking water sources to guarantee public health.

2.5.2 UV Treatment and FIB Counts in Wainfleet's Groundwater

UV irradiation is a common water treatment practice used in private groundwater well systems to inactivate enteric bacteria and waterborne pathogens without producing by-products such as disinfectants and chemical wastes (Koivunen and Heinonen-Tanski 2005; Hijnen et al. 2006). UV treatment could become a way to inactivate fecal bacteria, disinfecting the groundwater. The efficient inactivation of FIB in groundwater with UV treatment could encourage homeowners to install UV systems to disinfect their

groundwater wells. To this end, groundwater wells equipped with a functional UV treatment system were sampled before and after the UV systems were activated. Comparison of culture counts of these water samples reveal the inactivation of these bacteria to non-detection on selective plates (Figure 5). The successful elimination of culturable bacteria with UV-radiation could provide a short-term remedy for the deteriorating water quality in the township.

Although UV treatment systems can inactivate FIB, microbes can adopt strategies to survive the UV treatment. When bacteria are exposed to conditions of stress, like UV irradiation, they can enter the viable but nonculturable state (VBNC). Bacteria entering this state cannot be grown on selective culture media, but are still capable of metabolic activity (Oliver 2005). Bacteria enter the VBNC state when exposed to environmental stressors like temperature and pH fluctuations, oxidative stress, and UV light (Ramamurthy et al. 2014). Many waterborne pathogens can enter the VBNC state, and when resuscitated with environmental conditions that favour their growth, they provide an additional health risk (Oliver 2010). Fecal bacteria could adopt this survival strategy and resist UV treatment, returning to its normal state once ingested into the body and retain its virulence if the strain is pathogenic. The ability for bacteria to persist in this manner undermines the robustness of UV treatment to provide safe drinking water for its residents. While UV-treatment systems can be installed in groundwater wells to decrease the risk of contracting waterborne disease, a long-term solution is required to prevent fecal matter from contaminating the groundwater aquifer in the first place.

2.5.3 Nitrate and Phosphate Concentrations in Selected Groundwater Wells

Nitrate concentrations in environmental samples can represent agricultural runoff seeping into groundwater environments (Lockhart et al. 2013). More importantly, the presence of excess nitrate concentrations could indicate inputs of fecal contamination that originate from sewage, animal manure, or nitrogen-based fertilizers (Lundberg et al. 2004). While the Canadian government does not have a maximum nitrate concentration limit in place, the Ontario Public Health authorities place a 10 mg/L for [NO_3^--N] limit for drinking water sources. As a complementary assay to quantify fecal contamination, nitrate concentrations were determined for Wainfleet's private groundwater wells in the summer months.

Nitrate concentrations, as determined by the cadmium reduction method, are the lowest in July, with only one groundwater well containing detectable oxidized nitrate concentrations (Figure 6). In contrast, the groundwater well labelled 11615, sampled in August, contained the highest nitrate concentrations in the study at 12.36 mg/L, which is over the recommended limit. Two other groundwater wells, 11376 in August and 10554 in July, also contained nitrate concentrations above the recommended limit and contributed to the higher mean oxidized nitrate concentrations in the groundwater wells. The detection of nitrate concentrations above recommended limits coincides with the *E. coli* levels in the same time period, where the highest and lowest *E. coli* counts were recorded in August and July respectively. The co-detection of FIB and nitrates raises the possibility that the FIB in the groundwater wells are shed from either sewage or agricultural wastes.

Phosphates are the limiting nutrient for microbial growth in terrestrial environments. In oligotrophic environments like source waters, the presence of phosphate ions, even in minute quantities, can enhance overall microbial growth (Miettinen et al. 1997). Elevated phosphate concentrations in source waters can lead to microbial overgrowth and eutrophication (Holman et al. 2008). Phosphate ions in source waters can also indicate fertilizer and sewage inputs into the groundwater aquifer (Holman et al. 2010). The Canadian government follows the US Environmental Protection Agency's recommended limit of 10 mg/L of phosphate ions in drinking water sources. To determine the potential for microbial growth in groundwater, the stannous chloride method was used to quantify phosphate ions in selected groundwater samples.

Phosphate concentrations were consistently near the detection limit of 0.1 mg/L for all groundwater well samples, regardless of the sampling month. Wastes derived from fertilizers and septic tank systems may not be leaching through the soils and into the groundwater aquifer. The low phosphate concentrations in the groundwater wells could more be the result of mobile phosphorus being adsorbed into the soil as nutrients for plant growth (Holman et al. 2010) than the lack of fecal contamination inputs caused by fertilizers and septic tanks. Phosphate concentrations should not be used as a marker of fecal contamination in groundwater samples as phosphate concentrations are near detection levels for all groundwater wells.

2.5.4 Correlation between *E. coli* Counts and Nitrate Concentrations

Quantification data for culturable *E. coli* and nitrate ions can together act as indicators of fecal contamination in environmental samples. To this end, linear regression

analysis was conducted between oxidized nitrate concentrations and log-transformed *E. coli* counts in groundwater wells. A significant, positive correlation between *E. coli* counts and nitrate concentrations is determined by the Spearman Rank test (Figure 6, $P < 0.05$). However, the significant correlation between *E. coli* and nitrates is not robust, since removing the lone point with high nitrate concentrations and *E. coli* counts reduces the R^2 value to 0.1288 and renders the correlation insignificant under the Spearman Rank test ($P > 0.05$). Most of the tap water collected from separate groundwater wells had low *E. coli* counts and high nitrate concentrations and vice versa. A larger sample size comprised of groundwater wells known to have substantial fecal contamination inputs from STEs or manure from animal farms must be sampled further in the future to more robustly determine whether nitrates and *E. coli* counts can concurrently detect human or animal-based fecal contamination in groundwater.

2.5.5 Conclusions and Future Directions

To analyze drinking water quality in Wainfleet, private groundwater wells were sampled monthly for *E. coli* and *Enterococcus* colonies, two FIB used as standard indicators of fecal contamination. FIB counts in each individual groundwater well remained the same for the past ten years, so little successful action has been taken to improve drinking water quality. Two groundwater wells also received substantially higher FIB counts than the other groundwater samples. These wells may be receiving elevated fecal contamination loads from an unidentified source. A significant ($P < 0.05$), positive correlation between *E. coli* counts and oxidized nitrogen concentrations is also observed. Finally, the groundwater wells that have the highest *E. coli* counts also contain the highest oxidized nitrogen

concentrations. Using culture-based and physiochemical assays has determined that fecal contamination is still being leached into the underlying aquifer, evidenced by the high percentage of wells positive for FIB. Elevated fecal contamination levels were also detected in two groundwater wells, 11615 and 11215. This portion of the project has not only identified overall fecal contamination levels, but also identified wells that should be further examined for pinpointing fecal contamination sources. Genetic-based methods like 16S rRNA sequencing and qPCR of genetic markers belonging to fecal bacteria can be used to identify and confirm potential contamination sources in groundwater wells located within the boil water advisory region.

Chapter 3:

Characterization of Wainfleet's Groundwater, Sewage, and Manure Microbiotas with Deep Amplicon Sequencing of the 16S rRNA Gene

3.1 Abstract

Culture-based assays can provide a preliminary estimation of fecal contamination levels in drinking water. These assays, in themselves, cannot precisely determine sources of fecal contamination. To determine source of FIB counts and fecal contamination in groundwater wells within the boil water advisory region, the 16S rRNA gene was sequenced from DNA extracted from groundwater and septic tank effluent (STE), along with cow, horse, pig, and chicken manure samples. Bacterial taxa specific to STE samples in two groundwater wells, namely *Campylobacterales* and *Bacteroidales*, were identified using targeted sequencing of the 16S gene. The two groundwater wells also contained the highest *E. coli* counts, raising the possibility of STE contamination. Other genetic markers of animal manure contamination were also identified, like the *Turicibacteraceae* and *Ruminococcaceae* families in cow manure. However, these microbial groups were not detected above background levels in any of the groundwater wells, eliminating animal manure as a contamination source. Genera-containing waterborne pathogens, most notably *Arcobacter*, were also detected in STE and groundwater wells that contained the highest *E. coli* counts. The detection of STE microbes and waterborne pathogens associated with sewage further points to leaking septic tanks as the primary fecal contamination source in groundwater wells with elevated FIB levels. These results demonstrate the usefulness of deep amplicon sequencing to not only characterize environmental microbiotas, but to also pinpoint sources of fecal pollution in Wainfleet's groundwater aquifer.

3.2 Introduction

Standard methods for detecting fecal contamination in water samples utilize fecal indicator bacteria (FIB) in drinking water samples. Although FIB counts can provide a survey of fecal contamination levels in groundwater wells, *E. coli* and enterococci can survive for weeks to months outside the gut (Anderson et al. 2005). Furthermore, the conventional FIB are diverse, comprising both fecal and environmental isolates. *E. coli* and enterococci are comprised of environmental populations that are distinct from those that are shed from fecal matter (McLellan 2004; Sidhu et al. 2014). While the detection of *E. coli* does not remove the possibility of fecal contamination in source waters, it does call to question whether the detection of *E. coli* always represents a fecal contamination event. These drawbacks compromise the robustness of culture-based assays to identify and source recent fecal contamination events.

Water samples also contain microbial DNA. 16S rRNA sequences encoding part of the 30S ribosomal subunit in prokaryotes can be used as genetic fingerprints that delineate microbial groups in environmental samples. Targeted sequencing of the 16S rRNA gene can provide a survey of entire microbial communities through 16S rRNA sequence annotation to bacterial groups using a bioinformatics database like Greengenes (DeSantis et al. 2006). Targeted sequencing can be used to compare microbial communities between groundwater and shared fecal samples. The more shared bacterial groups between a potential fecal source and the drinking water source, the more likely the water source is contaminated by that fecal matter (Unno et al. 2010). The Illumina MiSeq platform correctly identified the dominant fecal contaminants in all pristine freshwater samples

spiked with a singleton or doubleton contaminant, reinforcing its ability to detect specific components of a host's microbiota in source waters (Cao et al. 2013).

Targeted sequencing of the 16S rRNA gene can also be used to identify bacteria found exclusively in the feces of a warm blooded animal. Sewage effluents a higher abundance of *Campylobacterales* at the order level (Korajkic et al. 2014), and are distinct from microbiotas in raw human feces (McLellan et al. 2010). Microbiotas in cow feces are typically abundant in cow-specific *Bacteroidales* and *Ruminococcaceae* sequences that facilitates nutrient metabolism in the ruminant gut (Huws et al. 2011). The coevolution of the human and animal host with its enteric microbes contributes to the formation of host-specific phylogenetic clusters, resulting in the proliferation of host-specific fecal microbes, even with the same microbial group (Dick et al. 2005). All the microbes that are specific to certain fecal contaminants could thus be screened with next-generation sequencing techniques and bioinformatics tools to identify potential fecal pollution sources in drinking water environments. In the same way, waterborne pathogens can also be identified in many types of samples, like biosolids (Bibby et al. 2010).

For this component of the study, the microbial communities in private groundwater wells between July and November, two septic tank effluent (STE), and four manure samples were characterized. A small sample size of fecal matter was obtained, but the microbial communities in these environments can still be compared as a pilot study to pinpoint fecal contamination sources. To this end, waste and groundwater microbiota compositions were compared to determine potential sources of FIB contamination in a subset of Wainfleet's residential wells. Manure and STE microbiotas were also examined

for fecal markers that, when detected in groundwater wells, represent a host-specific fecal contaminant. The sequencing data obtained from the groundwater and fecal samples can then be used to focus future efforts on specific DNA extracts for other genetic assays to confirm sources of fecal contamination in Wainfleet's groundwater.

3.3 Methods

3.3.1 Sample Collection and DNA Extraction

500-mL of water samples were collected from private groundwater wells and into autoclaved plastic bottle containers (Nalgene). Water samples were collected in single replicates from May to September and then in duplicate in October and November. All 500-mL of the water samples were passed through a 0.45 μm pore-sized membrane filter with vacuum filtration. Membrane filters were then folded into microfuge tubes and stored in -80^{o}C until DNA extracted was ready to be performed. Manure samples were collected from fecal mounds in animal farms. Chicken, cow, pig, and horse fecal samples were also collected for this study. Fecal samples were stored at -80^{o}C, and DNA was extracted from 0.25 g of each fecal sample after fecal samples were thawed overnight. 500-mL of septic tank samples were collected from two Wainfleet households into autoclaved plastic bottle containers (Nalgene). 50-mL of septic tank sample was then passed through 0.45 μm membrane filters using vacuum filtration. Membrane filters were stored at -20^{o}C overnight before DNA was extracted from the filters.

DNA was extracted from all samples with the Soil DNA Isolation Kit (Norgen Biotek, Thorold), the final elution volume for all DNA extracts being 100 μL. For water and septic tank samples, the folded membrane filters were transferred into tubes containing

zirconium beads. For fecal samples, 0.25 g of the fecal sample was transferred into a tube partially filled with glass beads before DNA extraction. For all samples, one modification was made: instead of a single wash step, two washing steps were conducted with Wash Solution A. DNA concentrations were determined with the Nanodrop 2000 (Thermo Scientific).

3.3.2 PCR of the 16S rRNA Gene with 96-Well Plates

The 16S rRNA PCR assay used in this study amplifies the V3-V4 hypervariable regions of the 16S rRNA gene (Klindworth et al. 2013). This region was chosen because it provided the least bias towards and against different bacterial groups that could be present in the DNA extracts (Chakravorty et al. 2007). 25 μL reaction mixes were prepared into 96 well qPCR plates as follows: 2.5 μL 10x PCR buffer minus Mg^{2+} (Invitrogen), 0.5 μL of 100 mM dNTP solution, 1 μL each of the forward and reverse primer with unique adapter sequences attached to them, 1 μL 10 mg UV-treated BSA solution, 0.75 μL $MgCl_2$ solution (Invitrogen), 0.25 μL Taq DNA Polymerase (Invitrogen), and 2 μL of DNA template. Each well contained a primer pair with a unique combination of adapter sequences (Illumina) that delineated the DNA templates once the sequencing run was completed (Figure S3). Thermal cycling conditions were prepared using the CFX96 Touch Real-Time PCR Detection System (BioRad) and conducted as follows: Initial denaturation at 95^{o}C for 3 min, 40 cycles of denaturation at 94^{o}C for 30 s, annealing at 50^{o}C for 30 s, and elongation at 72^{o}C for 1 min, and a final extension step at 72^{o}C for 10 min.

All PCR products were loaded and run on 0.9% agarose gels. Products from a given row of the 96-well plate were loaded onto the same row of gel wells. After gel

electrophoresis was completed, a row containing these amplicons was cut into a single gel slice which was then transferred into a sterile 50-mL Falcon Tube for gel extraction with the Nucleospin PCR and Gel Cleanup Kit (Macherey-Negel) with modifications. 200 μL of NTI buffer per 100 mg of the gel slice was added into the 50 mL Falcon Tube and mixed. A container filled with water was repeatedly microwaved to 55^{o}C for melting the agarose gels. 700 μL of the gel solutions was then loaded onto a single gel column repeatedly to concentrate the column with the 16S amplicons.

The rest of the gel extraction was followed according to manufacturer's instructions, performing two elution steps of 50 μL each for a total final volume of 100 μL. Gel DNA extracts were concentrated to 25 μL with the concentrator plus vacuum centrifuge (Eppendorf). The DNA solution was then loaded onto a newly prepared 0.9% agarose gel. A second gel extraction was then performed with the Nucleospin PCR and Gel Cleanup Kit, following the manufacturer's instructions. 2 x 50 μL of eluent was then concentrated again to 25 μL with the vacuum centrifuge (Eppendorf) before the gel DNA extracts were pooled into a single 1.7 mL microfuge tube. The samples were then sent to the Farncombe Institute for Digestive Health for sequencing of the 16S gene fragments with the Illumina MiSeq platform. All sequences were outputted into two fastq files containing the forward and reverse reads of the 16S gene fragment. The fastq files were subsequently processed with a custom script prepared by SR Botts that utilized the bioinformatics pipeline prepared by Quantitative Insights into Microbial Ecology (QIIME) (Caporaso et al. 2010).

The custom script contained three steps which acted as quality controls, annotated the remaining sequences to a 16S rRNA sequence data, and visualizing the data. During

the quality controls, sequences containing nucleotides with a quality score less than 25 were removed from downstream analyses. The QIIME pipeline was also used to cluster sequences into operational taxonomic units (OTUs) against the 16S Greengenes Database (DeSantis et al. 2006). From here, the QIIME script was used to perform statistical analyses and visualize microbial communities.

3.3.3 Statistical Analysis and Waterborne Pathogen Screening

A suite of statistical analyses was conducted to quantify sewage and animal manure bacteria in private groundwater wells and source fecal contamination in Wainfleet's primary drinking water source. First, multiple diversity metrics were used to ascertain α-diversity. Chao Diversity and total OTU counts were determined with Python scripts and plotted on bar and line graphs. Rarefaction and rank-abundance curves were also generated for all groundwater and manure samples as part of the α-diversity analyses. Emperor was used to prepare Principal Coordinate Analysis (PCoA) plots for comparing sequences with each other with the weighted Unifrac indice (Lozupone and Knight 2005).

Once the data was obtained from QIIME, a list of genera that contained waterborne pathogens was collected (Taylor et al. 2001). 16S rRNA annotations were screened for these pathogens at the genus level. Their relative abundances were then collected from reference fecal samples and groundwater well samples by month. Only the genera that contained higher than 0.2% relative abundance were considered for characterization in these samples, as only genera-containing waterborne pathogens that were abundant were considered more likely to cause infectious disease.

3.4 Results

3.4.1 Total Microbial Diversity and Identification of Taxa Found Only in Host-specific Fecal Samples

16S rRNA sequences were obtained from two septic tanks and manure from four animal farms. These sequences are annotated into operational taxonomic units (OTUs) to characterize the microbial communities of potential fecal contaminants. From the septic tank samples, an average of 6612 reads were classified into 972 OTUs. A mean of 49035 reads was obtained from animal manure samples which were classified into 2987 OTUs. Rarefaction curves associated with fecal matter rose more sharply than the curves associated with STE samples (Figure 9). While the septic tank rarefaction curve began to plateau, the rarefaction curves for fecal samples did not plateau (Figure 9). α-diversity metrics for total diversity were also measured in sewage and manure samples. All waste samples had a similar Shannon Diversity with each other. Cow fecal samples had the lowest Shannon diversity at 3.97, while pig fecal samples had the highest Shannon Diversity at 4.917. The septic tank samples had an average Shannon Diversity of 4.24 (Figure 11).

At the order level, the 16S microbial community profiles were similar between the two STE samples that were sampled (Figure S4). Members of the *Campylobacterales* and *Bacteroidales* taxa comprised a mean of 43.0% across both STEs (Figure S4). Both sets of STE samples also contained *Rhodocyclales* and *Clostridiales* sequences. Any differences to the septic tank microbiotas were attributed to changes in the relative abundance of the less abundant microbial taxa, like *Pseudomonadales* which comprised 4.20% and 1.70% of 16S sequences in septic tanks from 10554 Lakeshore Road and 11660 Belleview Beach

respectively (Figure S4). Sequences annotated to *Aeromonadales* were also abundant in the 11660 septic tank, but not in the 10554 septic tank.

The microbial communities of animal manure samples were distinct from the microbiotas of STE samples. The average relative abundance of *Campylobacterales* was less than 1% for all four animal manure samples. Whereas septic tank samples were largely comprised of the Proteobacteria and Bacteroidetes phyla, chicken and cow manure samples were dominated by members of the Firmicutes and Actinobacteria phyla. At the order level, members of the *Clostridiales* and *Bacillales* comprised 73.4% of the 16S operational taxonomic units (OTUs) (Figure S5). In contrast, 0.3% and 0.8% of *Clostridiales* and *Bacillales* respectively were annotated to the two sampled septic tanks. Cow fecal samples also contained the highest relative abundance of *Turicibacterales*, comprising 9.4% of 16S OTUs. Members of *Turicibacterales* were also detected in septic tanks, but only comprised 1.4% of their sequences. Chicken fecal samples also contained the highest relative abundance of *Clostridiales* and *Lactobacilles* OTUs at 54.4% and 16.8% respectively, but the lowest abundance of *Bacillales* OTUs at 6.6% (Figure S5). Although chicken and cow fecal samples were largely comprised of Bacteroidetes and Firmicutes bacteria, horse and pig fecal microbiotas contained far more microbial taxa at the order level (Figure S5). The *Burkhoderiales* and *Xanthomonadales* orders, belonging to the Proteobacteria phylum, comprised over 26.6% and 16.1% of the OTUs in pig and horse feces respectively.

At the genus level, *Arcobacter* and *Sulfurospirillum* genera were most abundant in STE samples, accounting for 28-30% of all genera in STEs. These genera were also detected at less than 1% abundance in all other fecal samples. In chicken fecal samples,

members of the *Gallicola* genus covered a mean relative abundance of 42.2% at the genus level. Members of the *Clostridiaceae*, *Planococcaceae*, *Turicibacteraceae*, and *Ruminococcaceae* families made up an average relative abundance of 62.5% in cow feces at the family level. Pig and horse fecal microbiotas were far more diverse in the amount of unique taxa that were detected. No one taxa was substantially more prevalent than the other in pig and horse fecal matter. Furthermore, members of the Proteobacteria phylum were detected in higher relative abundance than the Bacteroidetes and Firmicutes phyla combined.

3.4.2 Wainfleet's Groundwater Microbiota

Deep amplicon sequencing of the 16S rRNA gene was also conducted on DNA that was extracted from private residential groundwater wells. An average of 3101 OTUs was collected from the sequencing data. The rarefaction curve associated with groundwater samples rose without forming a plateau (Figure 9). Proteobacteria were consistently the most abundant phylum in groundwater samples, regardless of sampling month and location of the groundwater well. In July, some groundwater samples had higher relative abundances of *Burkhoderiales* and *Pseudomonadales*, like in Burnaby Road and Neff Street, while *Enterobacteriales* (41.9%) and *Gallionellales* (33.1%) were the most abundant bacterial taxa in 11315 and 10426 respectively (Figure S5). In August, members of *Rickettsiales* taxa were more abundant in 11413, comprising 43.4% of 16S sequences at the order level (Figure S6). *Enterobacteriaceae* were among the predominant members of the groundwater samples collected in September, along with *Cryomorphaceae* and *Pelagibacteraceae* (Figure S7). The relative abundances of members within the

Proteobacteria phylum made substantial contributions to the diversity and stability of the groundwater microbiota.

A PCoA plot was generated to perform wholescale microbial community comparisons between groundwater microbiotas by sampling month and potential fecal contamination sources (Figure 12). Groundwater samples were dispersed along one side of the PCoA graph, with most of the groundwater wells located between -0.1 and 0.1 of the x and y-axes. Two groundwater samples collected in August were located away from the rest of the other summer samples (Figure 12). November groundwater samples were dispersed along the horizontal axis, with one groundwater sample farther away from the other November samples (Figure 12).

3.4.3 Comparing Wainfleet Groundwater and Fecal Microbiotas

PCoA analysis was also employed to compare fecal and groundwater microbiotas in Wainfleet samples. Only three private groundwater samples were clustered with the two septic tank samples by sampling month, two of which were collected in August and the other being collected in November (Figure 12). Two of the groundwater samples were also obtained from wells that contained the highest *E. coli* counts, labelled 11615 and 11215 and collected in August and November respectively. The other groundwater sample that was clustered closely with the septic tank samples was labelled 11376A and collected in August.

To provide evidence towards a potential source of fecal contamination with 16S rRNA sequencing, the microbiotas of groundwater samples were searched for microbial taxa that are found in in STE or animal manure samples in high abundances. *Bacteroides*

and *Campylobacterales*, the two most abundant taxa in STE samples and detected in low abundances in animal manure microbiotas, were searched in groundwater samples taxa to detect STE contamination. *Turicibacteraeceae* and *Ruminococcaceae* were also screened in groundwater wells to determine potential cow manure contamination. Finally, members of *Gallicola* were searched in groundwater wells to determine chicken manure contamination levels as they were absent in other fecal sources but abundant in the chicken manure sample. Groundwater wells with the highest *E. coli* counts contained three times the mean relative abundance of *Bacteroides* and *Campylobacterales* as all the other groundwater wells Figure 13. Despite the increase in the relative abundances of these microbial groups, these differences were not statistically significant as determined by the Mann-Whitney U test (Figure 13). Sequences annotated to the *Ruminococcaceae* family were the highest in the groundwater well 11215 collected in October, at 2.8%. All other groundwater wells had an average relative abundance of less than 2.0% for *Ruminonococcaceae*. Members of the *Gallicola* genus were most abundant in wells 11215 and 11615 collected in October at 2.95%, contrasted with other groundwater wells where less than 1.0% of the sequences were annotated to *Gallicola* at the genus level.

Correlation analysis was also conducted between log-transformed *E. coli* colony forming unit (CFU/100 mL) counts with the relative abundance of *Bacteroides* and *Campylobacterales* taxa. A weak but positive correlation was observed between the two variables across sampling months and locations (Figure 14). However, the relationship was not statistically significant when tested by the Spearman's Rank Correlation test ($n > 0.05$). The relative abundances of *Bacteroides* and *Campylobacterales* taxa were also correlated

with oxidized nitrate concentrations for groundwater samples collected in July and August (Figure 15). The correlation coefficient, R^2, was double that which was obtained when the abundances of the two taxa was correlated with *E. coli* counts. The moderately positive correlation was further examined with the Spearman's Rank Correlation test, which yielded a significant correlation ($n < 0.05$).

3.4.4 Detection of Genera Containing Potential Waterborne Pathogens in Fecal Isolates

The sequencing data obtained from the QIIME pipeline was also used to identify any genera containing potential waterborne pathogens. At the genus level, both STE samples contained members of the *Arcobacter*, *Acinetobacter*, and *Bacillus* taxa (Figure 16). The same types of potentially pathogenic microbes were found in both septic tank samples despite 10554 containing almost three times the relative abundance of these microbes as 11660, at 23.1% and 8.1% respectively. 14.8% of the microbes that were annotated at the genus level could contain pathogenic species in chicken feces, with 70% of these sequences belonging to members of the *Corynebacterium* taxa (Figure 17). In cow manure samples, 8.6% of the 16S sequences were annotated to waterborne-pathogen containing genera. Half of these sequences were annotated to members of the *Bacillus* genus (Figure 17). Members of the *Pseudomonas* genus comprised 5.93% of all 16S sequences that were annotated in horse manure samples at the genus level (Figure 17). In contrast with the rest of the farm animals, a single dominant microbial genus containing waterborne pathogens was not detected in pig fecal samples (Figure 17).

3.4.5 Detection of Genera Containing Potential Waterborne Pathogens in Groundwater Samples

To provide a preliminary indication of the risks of a waterborne outbreak in the township, 16S rRNA sequences obtained from groundwater samples in Wainfleet were screened for waterborne-pathogen containing genera as determined by a list of human pathogens. The lowest abundance of waterborne-pathogen containing genera was detected in groundwater samples collected in July. No more than 6% of the sequences were annotated to pathogen-containing genera in July (Figure 17). With the exception of 11315 and 11615, less than 7% of the samples collected in August had pathogen-containing genera (Figure 18). In contrast, all but one groundwater well sampled in the fall months contained at least 10% relative abundance of pathogen-containing genera (Figure 19). Groundwater well 11660 collected in October and November contained the highest relative abundance of these genera, comprising 41.7% of all 16S rRNA sequences.

Acinetobacter and *Yersinia* were the most common genera with waterborne pathogens in the July samples with the lowest *E. coli* counts. In contrast, *Arcobacter* was most frequently detected in groundwater well 11615 in July, which had the highest *E. coli* counts (Figure 17). In August, well 11615 also had the highest relative abundance of *Pseudomonas*, accounting for 39.5% of the 16S sequences at the genus level (Figure 18). Groundwater samples collected in September had higher relative abundances of *Arcobacter* and *Mycobacterium*, while samples collected in November had higher abundances of *Pseudomonas* and *Arcobacter* at the genus level (Figure 19).

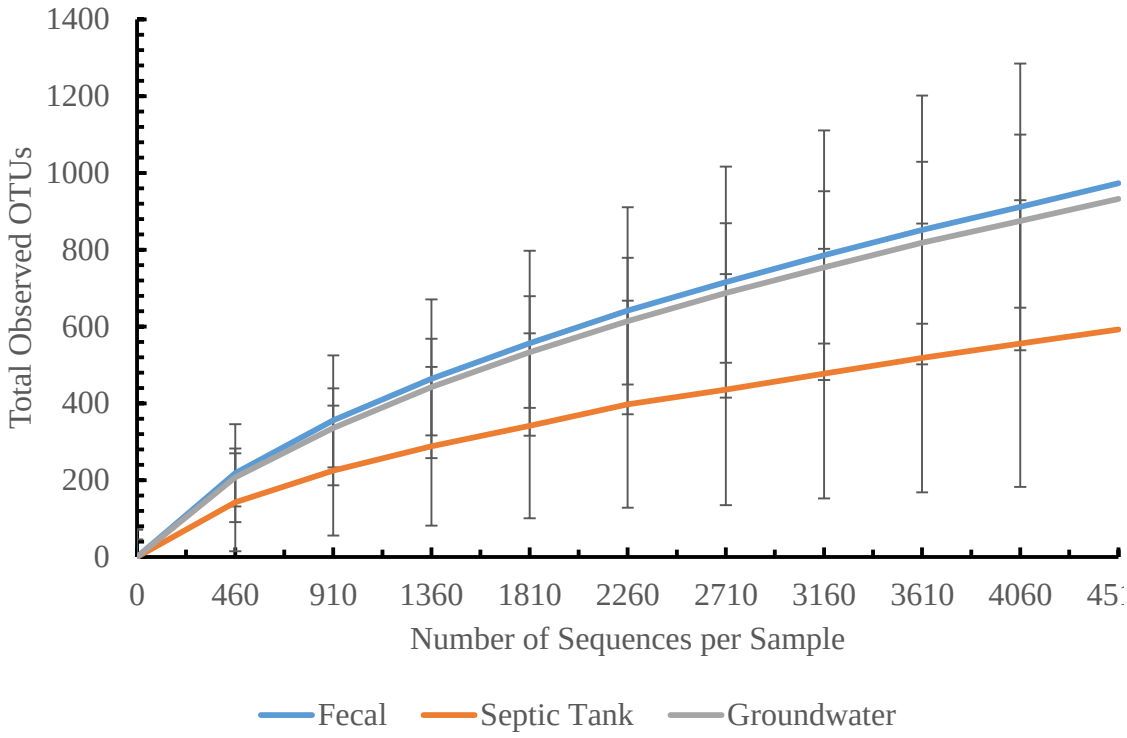

Figure 9. Rarefaction Curve Analysis of Observed OTU Diversity in Fecal and Groundwater Samples.

Rarefaction Curve analysis was conducted on all samples using the QIIME pipeline. n(Groundwater Sample) = 30; n(Fecal Sample) = 12; n(Septic Tank Sample) = 3.

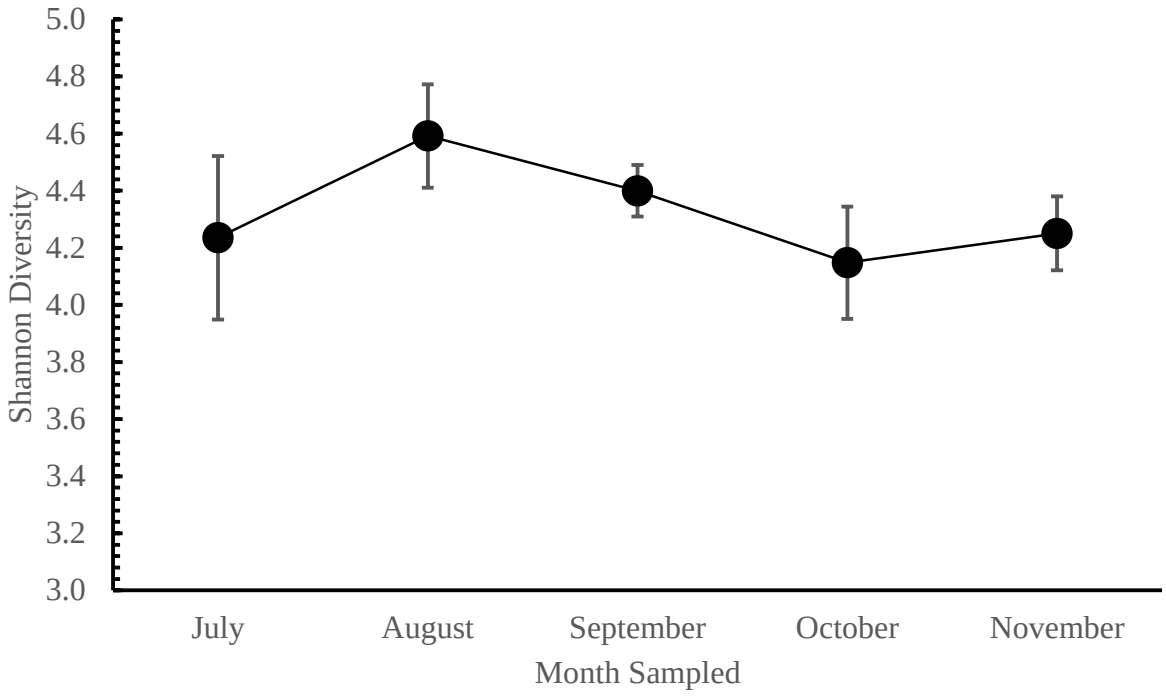

Figure 10. Shannon Diversity of Wainfleet Groundwater Wells in Monthly Intervals

Shannon Diversity is the measure of the amount of species that are detected within a groundwater sample. The Shannon Diversity of individual groundwater samples within a sampling month was calculated before the mean was obtained for that sampling month. n(July) = 6; n(August) = 6; n(September) = 4; n(October) = 3; n(November) = 3. Shannon Diversity increased slightly in August before decreasing and stabilizing in the fall months.

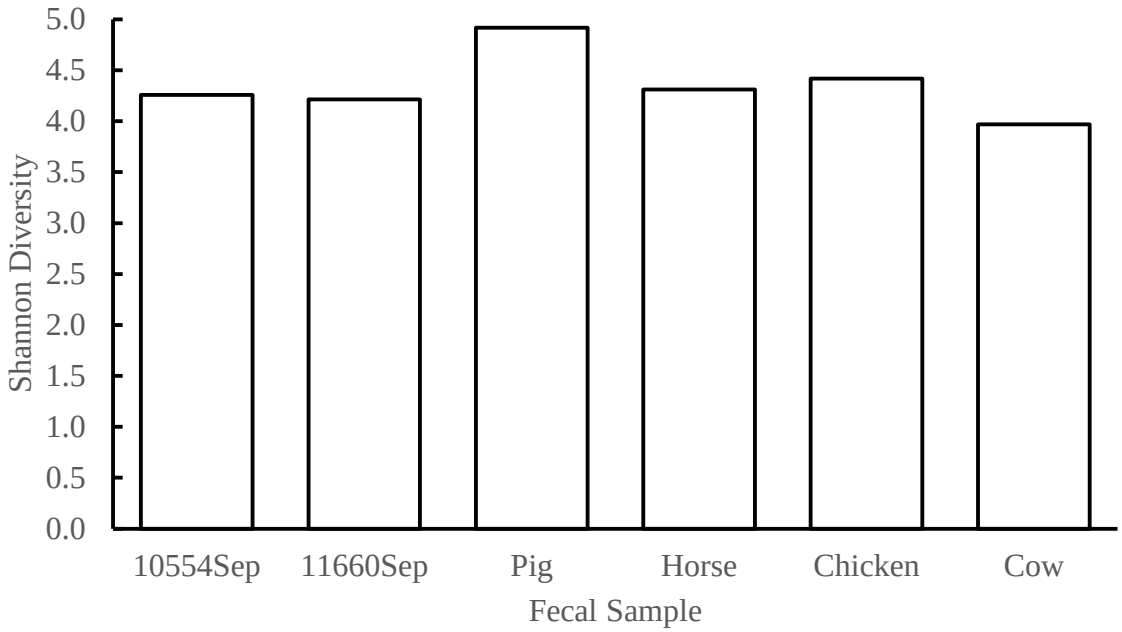

Figure 11. Mean Shannon Diversity of Wainfleet Fecal and Septic Samples

The average Shannon Diversity was calculated if more than one replicate was collected for a given fecal sample. Sep = septic tank sample. All Shannon Diversity calculations were recorded in three biological replicates. Septic tank samples were processed as single biological replicates from the house labelled with their address number, their sampling ID.

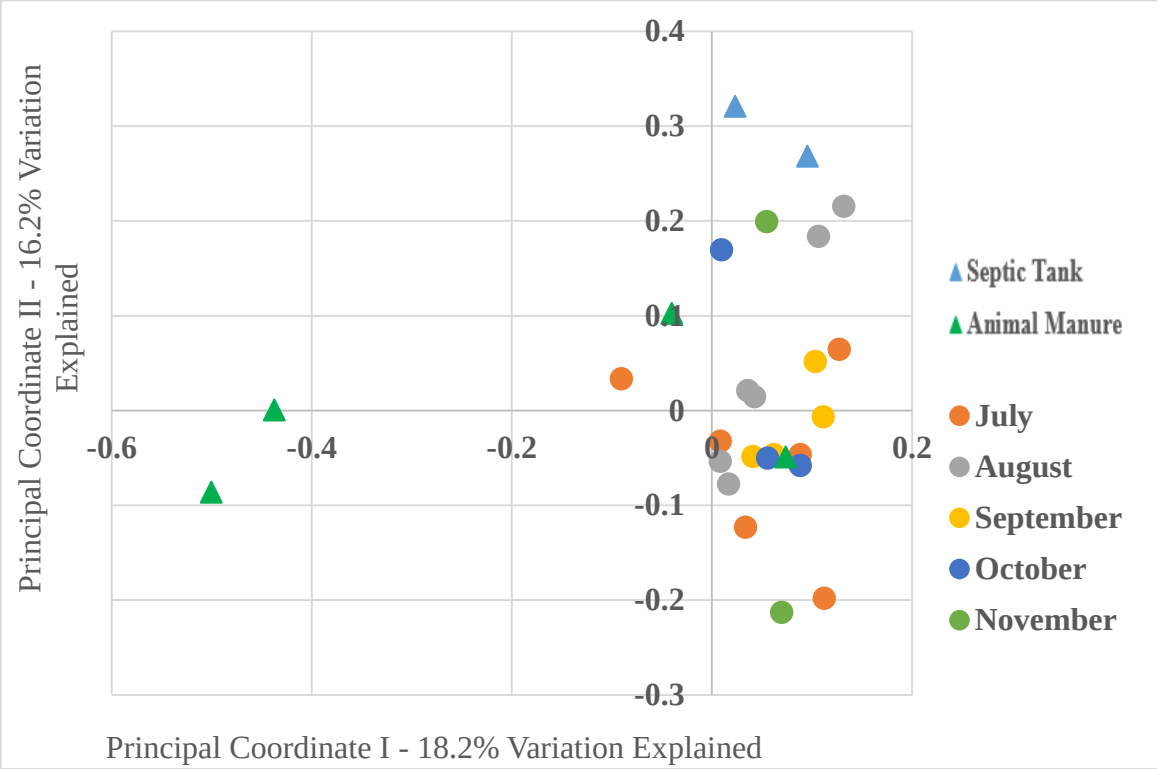

Figure 12. Principal Coordinate Analysis of Groundwater and Fecal Microbiotas based on Sampling Month

Colour-coded dots represent the sampling month for an individual groundwater well.

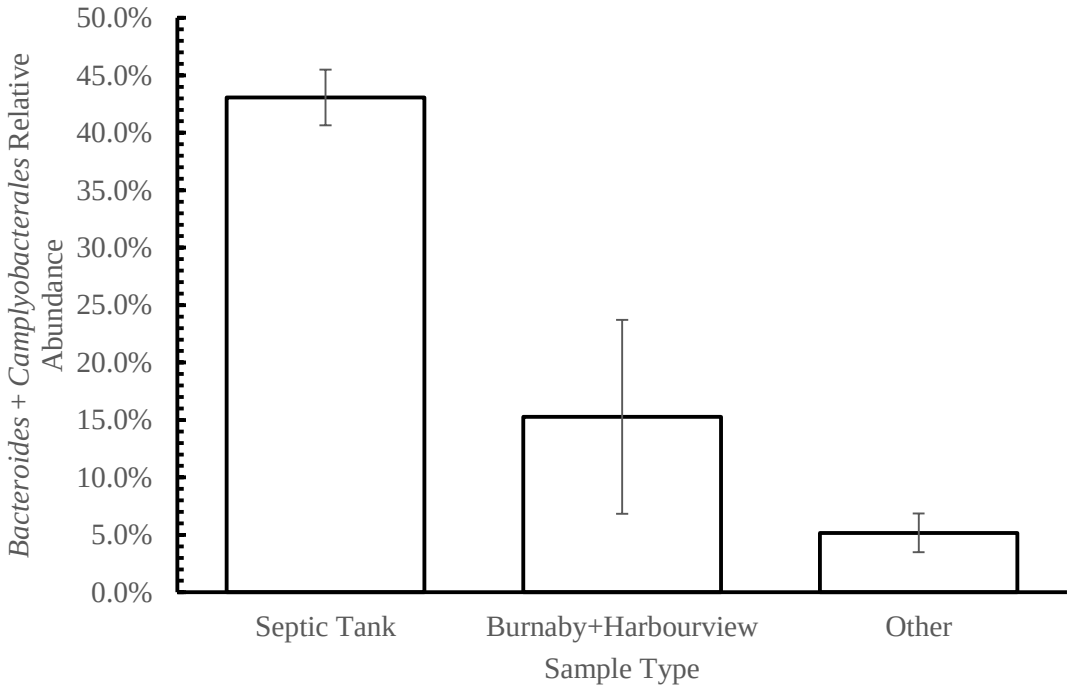

Figure 13. Relative Abundance of *Bacteroides* and *Campylobacterales* Taxa in Septic Tank and Groundwater Samples

Bar graphs were prepared to compare the relative abundances of the two most prominent microbial taxa in septic samples for septic tanks, sampled collected from 11615 and 11215, and all other groundwater wells. n(Septic Tanks) = 2; n(11615+11215) = 6; n(Other) = 16. Groundwater wells 11615 and 11215 contained the highest log-transformed *E. coli* CFU counts. The n value adds water samples collected across sampling months and excludes groundwater wells that were UV-treated

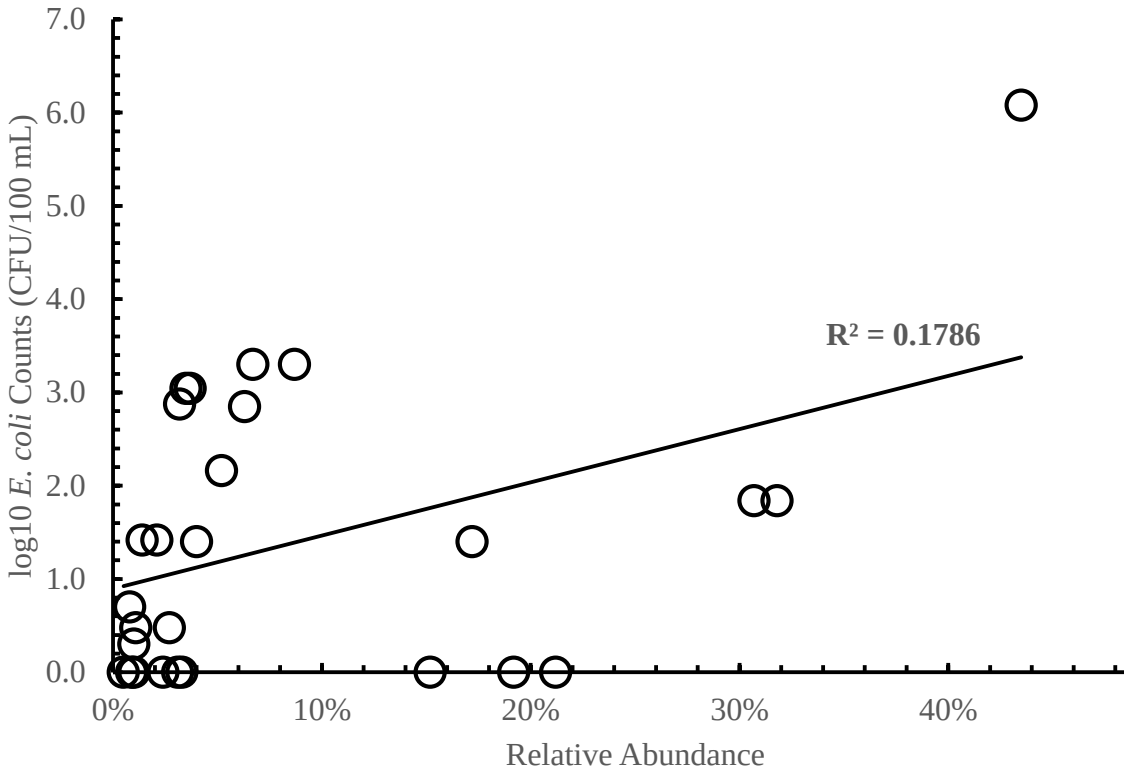

Figure 14. Correlation Analysis of $\log_{10}$ *E. coli* Counts and the Relative Abundance of *Bacteroides* and *Camplyobacterales* taxa

Bacteroides and *Campylobacterales* are the two most abundant microbial taxa in the septic tank samples. Their combined relative abundance was positively, but weakly correlated with log-transformed *E. coli* counts in private groundwater samples.

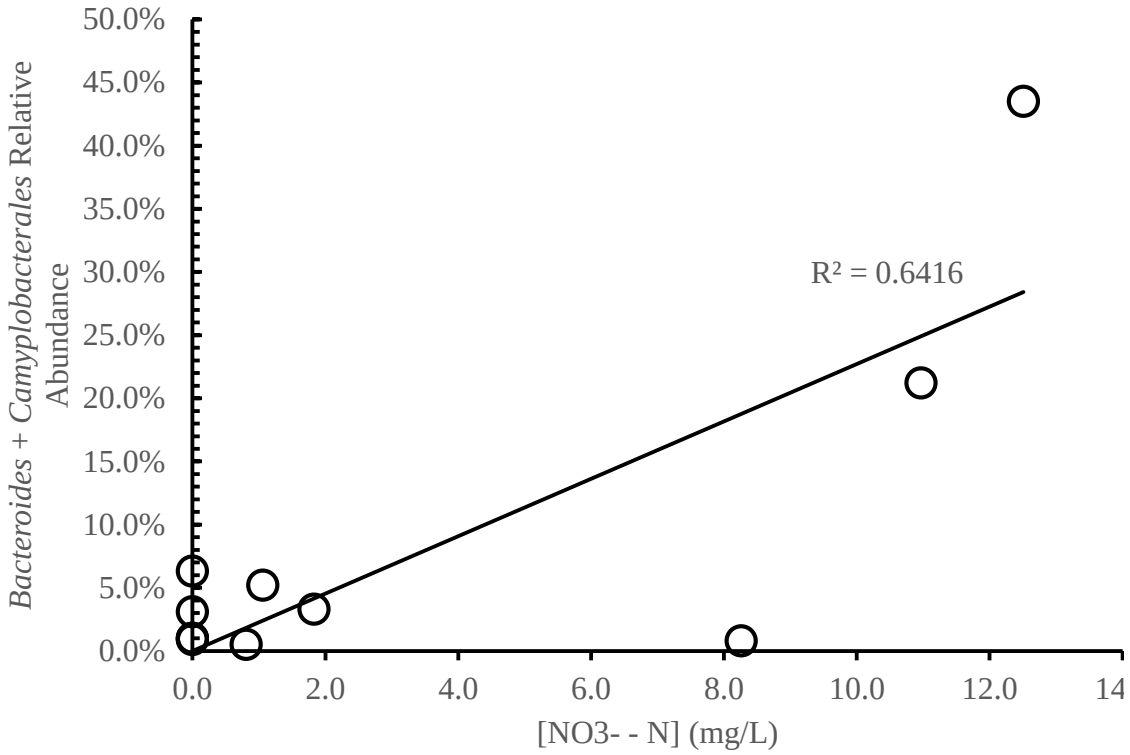

Figure 15. Correlation Analysis of Sewage Bacteria Abundances with Oxidized Nitrate Concentrations in Private Groundwater Wells

Correlation analysis was conducted between the two most abundant bacterial taxa in groundwater well samples and oxidized nitrate concentrations. Only groundwater samples whose nitrate concentrations were determined with the cadmium reduction method were considered for this study.

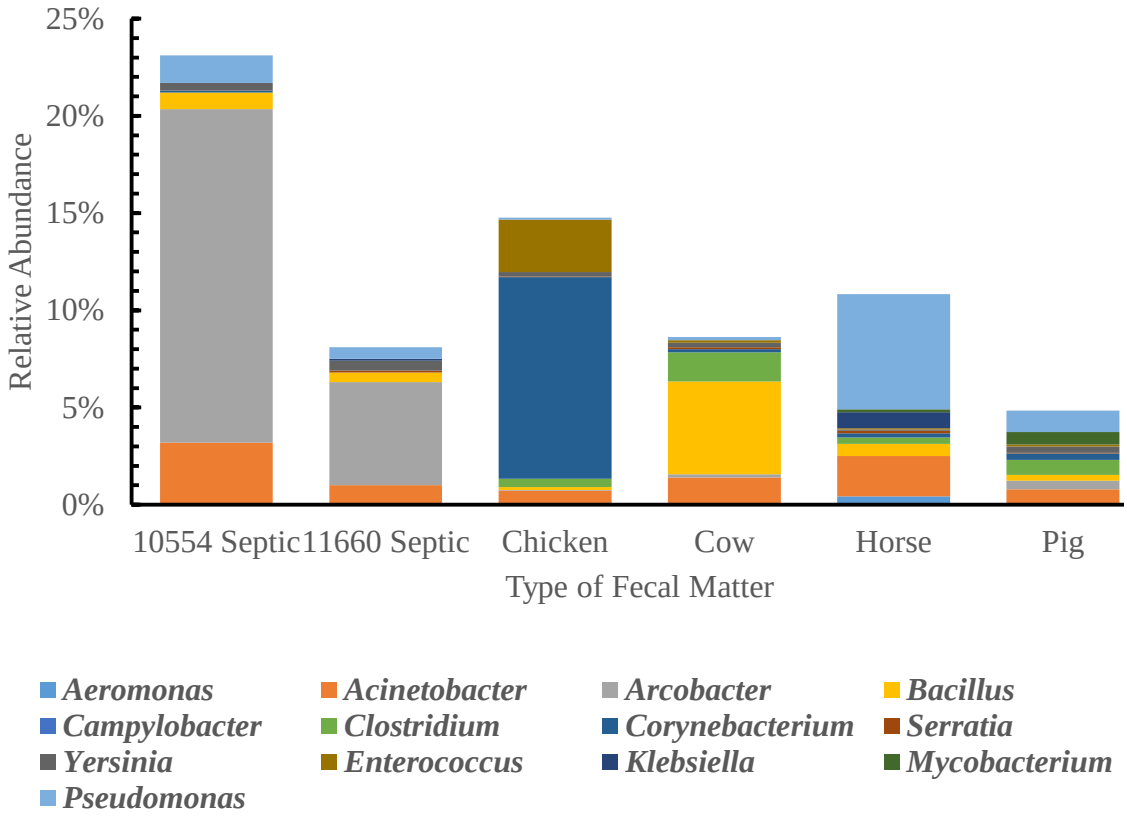

Figure 16. Relative Abundance of Waterborne Pathogens in Fecal Samples

Waterborne pathogens were collected from a database prepared by Taylor et al. (2001). The type of fecal matter that was sampled contained different genera that could belong to waterborne pathogens. Only genera that contained waterborne pathogens and appeared at higher than 0.2% abundance across all samples.

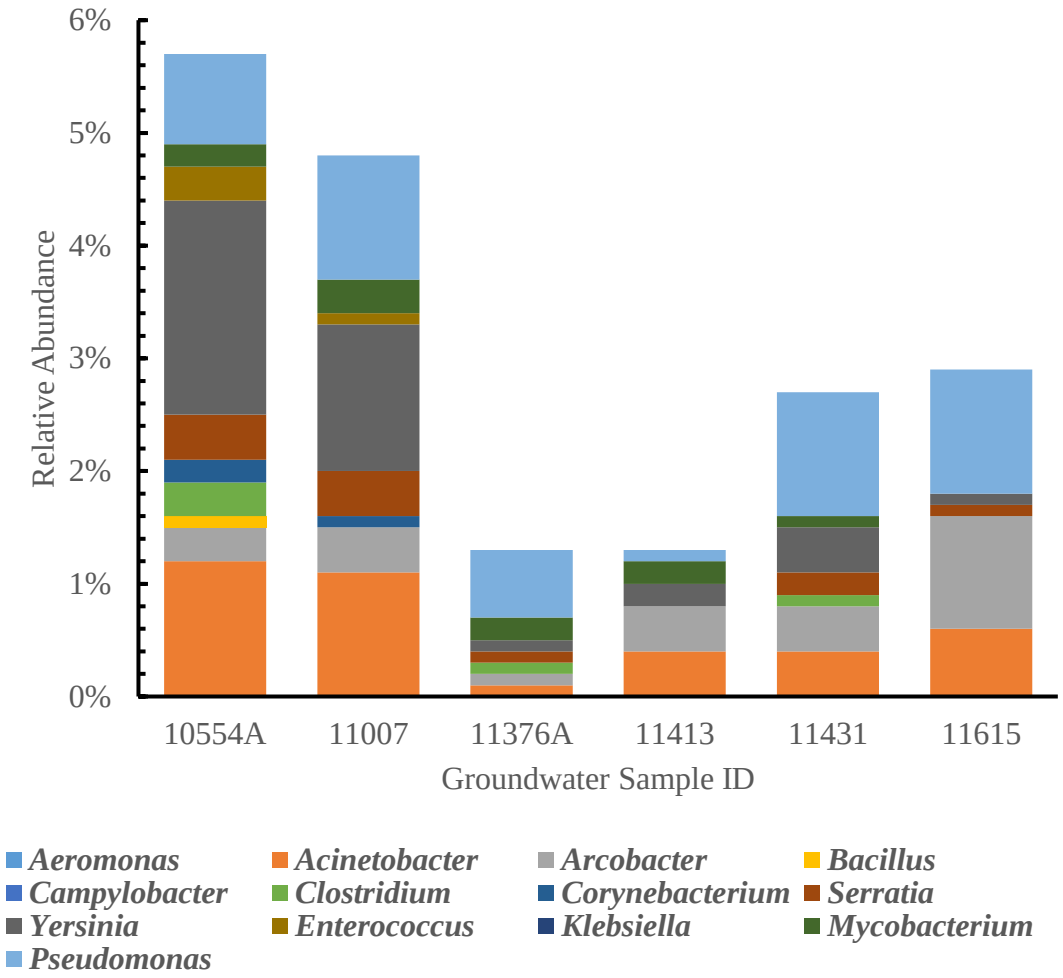

Figure 17. Relative Abundance of Waterborne Pathogen Containing Genera in July Groundwater Samples

The relative abundance of waterborne-pathogen containing taxa was determined at the genus level, the highest level of resolution for microbial groups in 16S rRNA sequencing. Waterborne pathogen genera were graphed if they contained greater than 0.1% relative abundance in at least one groundwater well in July. Samples were organized by increasing *E. coli* $\log_{10}$ counts from left to right.

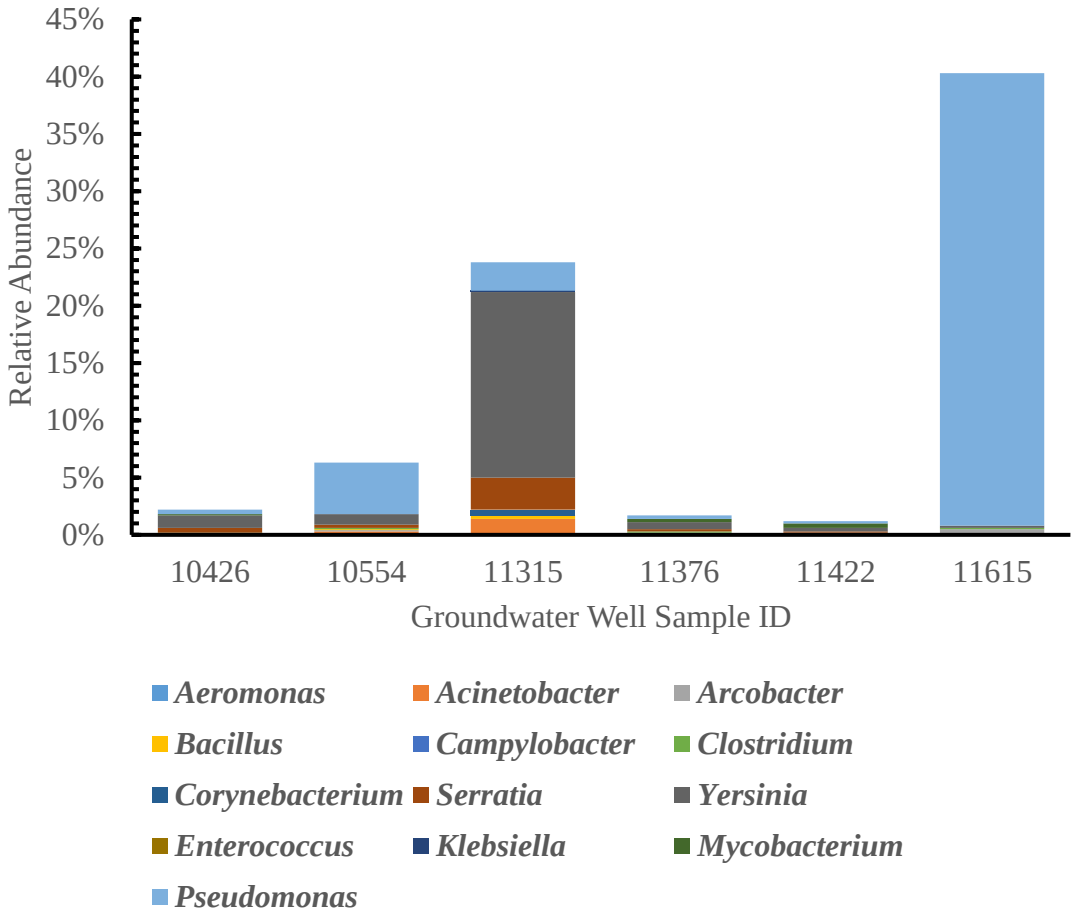

Figure 18. Abundance of Waterborne-Pathogen Containing Genera in August Groundwater Samples

The relative abundance of waterborne-pathogen containing taxa was determined at the genus level, the highest level of resolution for microbial groups in 16S rRNA sequencing. Waterborne pathogen genera were graphed if they contained greater than 0.1% relative abundance in at least one groundwater well in August. Samples were organized by increasing *E. coli* $\log_{10}$ counts from left to right.

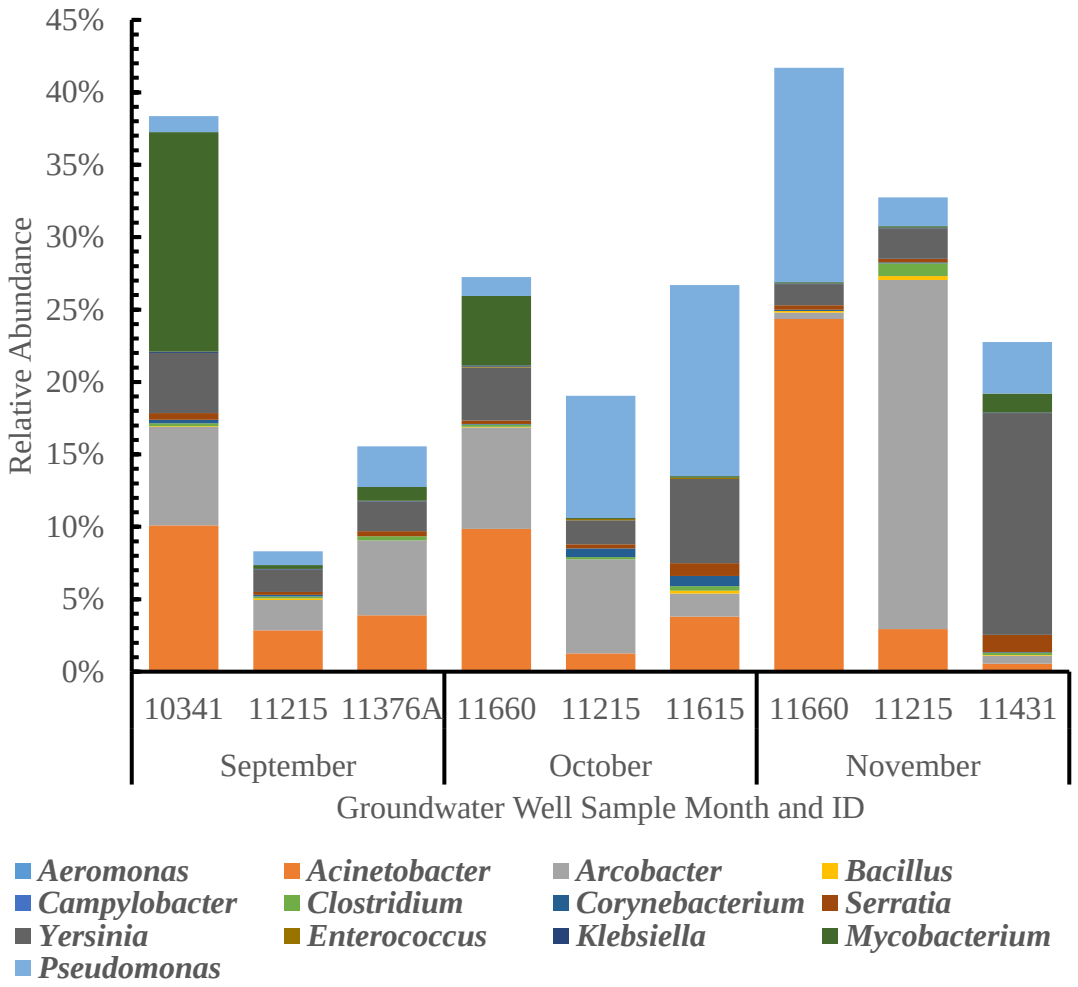

Figure 19. Relative Abundance of Waterborne Pathogen Containing Genera in Fall Groundwater Samples

The relative abundance of waterborne-pathogen containing taxa was determined at the genus level, the highest level of resolution for microbial groups in 16S rRNA sequencing. Waterborne pathogen genera were graphed if they contained greater than 0.1% relative abundance in at least one groundwater well in all fall samples. Samples were organized by increasing *E. coli* $\log_{10}$ counts from left to right, as well as the sampling month.

3.5 Discussion

Using culture-based and physiochemical assays determined that fecal contamination has been transferred into the volunteers' groundwater wells over the past year. Two of the groundwater wells, 11615 and 11215, have received an average of at least 1000 *E. coli* CFU/100 mL during the summer months. A previous study hypothesized that leaking septic tanks are transferring raw sewage into the town's aquifer, becoming the town's primary fecal contamination source. The continued presence of FIB in groundwater wells across Wainfleet raises concerns that the town's groundwater is unsafe to drink. Culture-based assays cannot pinpoint fecal contamination sources. Targeted sequencing of the 16S rRNA gene could be used to characterize microbial communities. The microbiotas of groundwater and potential fecal contamination sources can be compared to approximate the extent of fecal contamination in groundwater wells. 16S rRNA sequencing can also be used to identify fecal-specific indicators in reference fecal samples that, when detected in Wainfleet's groundwater samples, represents host-specific contamination.

3.5.1 Characterization of Manure Microbiotas

Microbial community profiles obtained by 16S rRNA sequencing can be used as references to determine the extent of fecal contamination in source water (Unno et al. 2010). The use of 16S rRNA sequencing on DNA extracts from fecal samples can also identify host-specific microbes that can be used as markers of human and animal-based contamination (Shanks et al. 2011). To identify sources of fecal contamination in groundwater wells receiving fecal inputs, rRNA 16S rRNA sequences were obtained from

STE and manure samples that were collected from four animal farms and two residential septic tanks, located in 11660 Belleview Beach Rd. and 10554 Lakeshore Road.

α-diversity metrics can be used to compare groundwater, sewage, and manure microbiotas at the community-level. These datasets can be used to determine whether a microbial community has been adequately captured. Rarefaction curves of the samples revealed that the curves plateaued for septic tank samples, but not so for animal fecal samples. Septic tank samples typically have lower total diversity than in other environments, and a study featuring WWTP effluents demonstrated a substantial decrease in total microbial diversity when river waters were mixed with wastewater effluents (Drury et al. 2013). Rarefaction curves for animal fecal samples did not plateau even when over 1000 unique OTUs was obtained. While predominant members of the fecal environments can be detected, rare taxa in these environments may not be detected, and thus there could be more fecal indicators that can be identified in fecal samples.

The two STE samples have similar microbial communities at the order level and were dominated by members of the Bacteroidetes, Firmicutes, and Proteobacteria phyla, like *Bacteroidales*, *Clostridiales*, and *Campylobacterales* (Figure S4). STE samples are stable across different geographic regions as other STE samples comprise of Bacteroidia and Epsilonproteobacteria at the class level (Tomaras et al. 2009). Both septic tanks also contain members of *Arcobacter*, a prominent bacterial group in sewage (Collado and Figueras 2011), and *Sulfurospirillum*, a group of microaerophilic, sulfur-reducing bacteria. Although *Campylobacterales* may not be an effective indicator of fecal contamination in groundwater wells, the increased abundance of these microbes in septic tanks and

groundwater wells with high *E. coli* counts raises the possibility of septic tank contamination in the most contaminated groundwater wells.

Chicken manure contains a characteristically high proportion of 16S rRNA sequences that are annotated to members of the Firmicutes phylum. Chicken gut samples high relative abundance of the Firmicutes phylum (Wei et al. 2013), which is observed in the chicken manure microbiota. Within the Firmicutes phylum, members of the *Gallicola* genus, classified within the Clostridia class, are detected in high abundance at chicken fecal samples from a Wainfleet chicken farm. The *Gallicola* genus could be explored further to determine whether they are more abundant in chicken fecal samples and absent in other animal manure samples. Collecting more chicken manure samples from other chicken farms would help to examine the distribution of 16S rRNA sequences annotated to the *Gallicola* genus in chicken manure samples across other farms.

Like chicken fecal samples, cow manure contains a higher relative abundance of members in the Firmicutes phylum compared with the Bacteroidetes phylum (Figure S5). The cow gut is previously shown to have high abundances in members of the Firmicutes phyla (Kim et al. 2011). Many of the sequences within the Firmicutes phyla belong to members of the *Turicibacter* genus, a group of bacteria found in the cow gut. 16S rRNA sequences annotated to *Turicibacter* in groundwater wells could be tested further to see whether members of *Turicibacter* can be used to identify cow manure contamination in the town's groundwater. The gut microbiotas of warm-blooded animals may comprise similar microbial community structure at the phylum level, but these communities can be delineated at lower taxonomic classifications. The animal manure microbiotas are diverse,

comprising of differentially-abundant microbial groups as determined by their 16S rRNA sequencing profiles. The unique diets that different farm animals can eat likely contributes to the distinction between the fecal microbiotas of warm-blooded animals (Ley et al. 2008b).

The pig and horse manure microbiotas on the other hand contain higher abundance of sequences annotated to the Proteobacteria phylum. These results are in disagreement with previous studies that suggested a greater relative abundance of members classified within the Firmicutes in the equine gut (Shepherd et al. 2012) and a near even ratio of Firmicutes and Bacteroidetes phyla in pig fecal samples (Looft et al. 2014). The pig and horse manure samples were exposed to extreme cold weather at the time of sampling. The exposure to the extreme cold weather for an extended period of time could have inactivated the anaerobic microbes, particularly those from the *Bacteroidales* order, which resulted in their decreased abundances in pig and horse fecal samples.

3.5.2 The Groundwater Microbiome in Wainfleet

Groundwater environments in general are poorly studied, as few studies have examined the effects of anthropogenic inputs to groundwater aquifers (Knappett et al. 2012; Smith et al. 2012). Nevertheless, anthropogenic inputs into the unconfined aquifer can result in shifts to nutrient concentrations which changes the structure and functioning of the aquifer microbiota (Hemme et al. 2010). As a first step to identify sources of fecal contamination in contaminated groundwater wells, 16S rRNA sequencing was used to survey Wainfleet's groundwater microbiota. First, the Shannon Diversity in the sampled groundwater wells did not change substantially over the course of the sampling regime

(Figure 10). Total microbial diversity in Wainfleet's groundwater is stable, much like other groundwater environments. Groundwater microbial communities are also diverse and abundant, comprising between 6-40% of the world's microorganisms (Griebler and Lueders 2009). Like animal manure samples, rarefaction curves generated for groundwater samples did not plateau (Figure 1). Only 500-mL of groundwater was collected from each groundwater well. Although groundwater microbial communities are diverse, they are sparse in groundwater, and so more volume needs to be collected and concentrated to capture the groundwater microbiome.

PCoA analysis of environmental samples can be used to compare overall microbial community structures. Samples that cluster more closely together contain more similar microbial communities than those samples that are more distant. PCoA analysis of the groundwater samples shows a single cluster of groundwater wells by sampling month, with some points lying outside this cluster. At the phylum level, all groundwater wells are majorly comprised of members within the Proteobacteria phylum (data not shown). However, at the lower taxonomic levels, groundwater microbial communities are diverse based on sampling month and groundwater well location. 41.9% of the 16S rRNA sequences in groundwater well 11315 collected in July are annotated to *Enterobacteriales*, while the same taxa comprised 0.90% of sequences in July 11422 (Figure S6). On the other hand, 11422 in July had the highest relative abundance of *Nitrospirales* taxa, comprising 14.20% of the well's 16S rRNA sequences. Although the relative abundances of microbial taxa differs between groundwater wells within sampling months, other taxa remained stable within a sampling month. In groundwater wells collected in September 2015, the relative

abundance of *Rhizobiales* 16S rRNA sequences 3.40% to 8.90%, a smaller range when compared with other microbial taxa (Figure S8). While the Wainfleet groundwater microbiotas are stable at higher taxonomic levels, changes in microbiota composition at the order level are observed. The changes in relative abundance of microbial taxa thus remains an important driver of spatial variation in source water microbiotas (Staley et al. 2013).

3.5.3 STE Markers in Groundwater Wells

The previous study examining Wainfleet's groundwater quality determined that leaking septic tanks are the primary fecal contamination source. To assess this possibility, groundwater 16S rRNA sequences were searched for microbes predominantly found in STE, namely members of the *Bacteroides* and *Campylobacterales* taxa. The groundwater wells with the highest *E. coli* counts have the highest relative abundance of *Bacteroides* and *Campylobacterales* sequences at 31.0% and 31.3% of sequences respectively (Figure 13). The *Bacteroides* taxa is classified within the *Bacteroidales* order, a group of microbes that are used as genetic markers of fecal contamination as there exists host-specific populations that can be detected with qPCR assays. The groundwater well labelled 11615 likely receives recent inputs of STE contamination, as *Bacteroides* and *Campylobacterlaes* comprised 45.0% of 16S rRNA sequences in August. The positive detection of *Campylobacterales* microbes in higher abundances in STEs and this groundwater well suggests septic tank contamination of that well. 11215 Harbourview Road may also receive septic tank contamination, as 35.6% of the 16S sequences were annotated to members of the *Bacteroides* and *Campylobacterales* taxa, although it had 63 CFU/100 mL (Figure 14).

The groundwater well 11376A collected in August is an anomaly when correlating *E. coli* counts with the relative abundances of STE microbes. While this groundwater well did not contain detectable *E. coli* or enterococci after UV treatment, it contains the highest relative abundances of *Campylobacterales* and *Bacteroidales* in August. 11376A was a groundwater well with a malfunctioning UV treatment system, and while it has inactivated conventional FIBs, STE microbes remained in the well at the time of sampling. Because of the malfunctioning UV system in this groundwater well, 11376A is not included in the linear regression analysis as an outlier point.

Even after removing the outlier 11376A point from the linear regression analysis, *E. coli* counts and STE abundances are not significantly correlated ($P > 0.05$; Figure S10). The *E. coli* may not originate from the leaking septic tanks. An average of 3.3% of STE sequences were annotated to *Enterobacteriales* at the order level among the two septic tanks (Figure S4). The groundwater aquifer may be connected to the adjacent Lake Erie through a surface-groundwater interface. The *E. coli* may be transferred into the underlying aquifer and into the groundwater wells, providing another route of FIB transmission. The lack of a discernable relationship between *E. coli* counts and the presence of STE microbes can also be the result of the *Bacteroidales* sequences as they could come from non-human sources or comprise of environmental populations. Although the detection of *Campylobacterales* could suggest the leaching of sulfur-reducing microbes from septic tanks, the *E. coli* could come from other sources like lake water.

A positive and significant correlation between nitrate concentrations and the relative abundance of STE microbes is observed for groundwater wells collected in summer

2015 ($P < 0.05$; Figure 15). The presence of nitrate ions in some of the groundwater wells could be further indication of fecal contamination caused by nearby, leaking septic tanks discharging STEs. The nitrate concentrations could also originate from lake water that seeps into the groundwater aquifer, much like the higher *E. coli* counts in some of the groundwater wells.

For this study, STE markers were detected in groundwater wells alongside *E. coli* counts and nitrate concentrations in the groundwater wells that may be receiving the highest fecal contamination loads. Many of the groundwater wells that were sampled contained *E. coli* counts near detection levels, nitrate concentrations within the provincial limits, and lowered abundances of *Campylobacterales* and *Bacteroidales* 16S rRNA sequences. Although 11615 contained elevated amounts of all three variables, no relationship exists between these variables for the other groundwater wells. The aforementioned results do not conclude that leaking septic tanks are the primary contamination source, but neither do they remove that possibility. A larger sampling size comprised of groundwater wells known to contain a $\log_{10}$ *E. coli* count between 2.0 and 6.0 are needed to determine whether the detection of *Campylobacterales* in groundwater wells constitutes septic tank contamination in groundwater aquifers.

3.5.4 Detection of Pathogen-Containing Genera in Groundwater and Fecal Samples

One of the goals of MST assays is to directly quantify waterborne pathogens to assess the risks of a disease outbreak caused by the drinking water source (Field and Samadpour 2007). Genera that contain waterborne pathogens have been detected in biosolid samples have been detected with pyrosequencing of the 16S rRNA gene (Bibby et

al. 2010). More recently, the Illumina MiSeq platform has been used to detect waterborne pathogens that survived sequential steps of the wastewater treatment process (Kumaraswamy et al. 2014). While waterborne pathogens have been examined in waste samples, they have yet to be identified in contaminated groundwater wells like those in Wainfleet.

Taxonomic groups that contain waterborne pathogens at the genus level are identified in manure, sewage, and groundwater samples using a list of waterborne pathogens as a database (Taylor et al. 2001). Both STEs contain members of *Arcobacter* and *Acinetobacter*, both of which are predominant members of sewage effluents (Collado and Figueras 2011; Vandewalle et al. 2012). Members of *Corynebacterium* are found in the highest abundances in chicken feces, while the highest relative abundance of *Bacillus* is found in cow manure (Figure 16). The high abundance of *Corynebacterium* in chicken manure has been observed in chicken broiler samples, reflecting the host's ability to degrade wood and cycle nitrogen and sulfur-related compounds (Lu et al. 2003). The presence of *Bacillus* in cow manure could represent a concern for the local cow farms as pathogens like *Bacillus cereus* could cause food poisoning when spread to source water environments (Drobniewski 1993).The *Bacillus* sequences detected in cow manure were not resolved at the species level, meaning that the *Bacillus* sequences may belong to non-pathogenic bacteria. A further assessment on the *Bacillus* populations in the cow manure may be useful to determine more robustly whether pathogenic microbes can be transferred from farms into the groundwater source. The identification of pathogen-containing genera can help to provide impetus for examining further the risks of a waterborne outbreak and

provide other genetic markers of host-specific fecal contamination in groundwater environments across Canada.

Groundwater samples that were collected from June to November are also screened for waterborne pathogens at the genus level. A higher relative abundance of *Yersinia*, which contains opportunistic human pathogens, is observed in groundwater samples with lower *E. coli* counts. *Yersinia* species can come from human and animal fecal matter, but they can also be natural residents of surface and groundwater environments (Bottone 1999). The detection of *Yersinia* cannot be used to determine fecal contamination sources. However, the elevated abundance of *Yersinia* in groundwater wells across all sampling months with 16S rRNA sequencing provides room for examining *Yersinia* populations further in groundwater environments.

In groundwater wells with higher *E. coli* counts, the highest relative abundances of *Arcobacter* and *Pseudomonas* in 11615 and 11215 has been detected in the summer and fall sampling months, along with *Acinetobacter* during the fall months in the same groundwater wells (Figure 12). All three taxa are commonly found in sewage effluents, particularly *Arcobacter* which is also detected in both STE samples collected from Wainfleet. Leaking septic tanks could be attributed to the higher relative abunances of *Arcobacter* in the groundwater wells, but they could also come from other sources like lake water which must be considered. On the other hand, the *Corynebacterium* genus is absent in all groundwater samples (Figure 17, Figure 18, Figure 19), decreasing the possibility that chicken farms are contaminating the groundwater wells that were sampled. The nearest chicken farm is located at least 5 km away from the residential housing in the township, so

the local farm could only act as a diffuse source of fecal contamination at most. The *Bacillus* genus is absent in almost every groundwater well, also eliminating cow feces as a notable source of fecal contamination in the town's aquifer.

3.5.5 Conclusions and Future Directions

This component of the study sought to provide a survey of potential fecal contamination sources and to provide another step towards identifying a fecal contamination source and *E. coli* counts in groundwater wells within Wainfleet's boil water advisory zone. One approach which can be used to potentially identify fecal contamination sources is the wholesale characterization of microbial communities in drinking water sources and manure samples by sequencing the 16S rRNA gene. Groundwater and manure samples are delineated with 16S rRNA sequencing, where Proteobacteria and Bacteroidetes are the predominant phyla for the respective sample types. The PCoA clustering analysis determined that while most groundwater samples are clustered together, some groundwater wells were clustered more closely with septic tank samples. The groundwater samples that were located closer to the septic tank samples contain a higher relative abundance of *Campylobacterales* microbes which are also found more predominantly in septic tanks and nearly absent in animal manure samples. Potential bacteria found only in animal manure have also been identified, like *Turicibacter* and *Gallicola* in cow and chicken manure. These microbes may represent fecal contamination from an animal source that was not detected in high abundances within groundwater wells.

All groundwater wells contained pathogen-containing genera, including *Yersinia* and *Arcobacter*. While the former can be found in many environments, *Arcobacter* are

microbes that are commonly found in raw and treated sewage. The detection of these microbes does not confirm the presence of sewage contamination in groundwater or directly confirm the presence of a waterborne pathogen. However, the detection of these genera could provide a venue for examining human-based fecal contamination sources in groundwater wells further.

The results put together provide more impetus towards examining leaking septic tanks as the primary fecal contamination source in groundwater wells. While the microbial community profiles obtained by 16S rRNA sequencing cannot conclude fecal contamination sources, they do provide evidence for examining human fecal contamination sources further in the town's groundwater. To robustly identify fecal contamination sources, a suite of genetic assays can be used to detect human and animal-specific fecal microbes in groundwater. The PCoA plot that was generated can also be examined further to determine which microbial groups are driving the clustering of groundwater wells with the highest *E. coli* counts with the two septic tank microbiotas. Sequencing millions of DNA fragments with shotgun sequencing can also provide a higher resolution of microbial groups, down to the species level, which can help to identify whether the pathogen-containing genera contain pathogenic microbes that are present in the groundwater samples. Drawing from this data, qPCR assays targeting genes that contribute to the risk of disease for the human host can be used to quantify the risk of infectious disease through exposure to the town's groundwater sources.

Chapter 4:

Molecular Detection of the Universal and Human-Specific *Bacteroidales* 16S Marker to Pinpoint Sources of Fecal Contamination in Selected Groundwater Wells across Wainfleet

4.1 Abstract

Boil water advisories are enacted by public health officials when fecal contamination levels exceed provincial limits. Microbial source tracking (MST) assays are used to quantify fecal contamination levels and to pinpoint contamination sources in drinking water sources. For communities relying on groundwater to provide drinking water, quantifying and pinpointing fecal contamination inputs raises awareness of the importance of maintaining drinking water sources for public health. Wainfleet is a Niagara township under the longest active boil water advisory in Canada. Culture-based, physiochemical, and 16S rRNA sequencing assays were three assays used to quantify fecal contamination levels in groundwater wells within the boil water advisory region. Individual groundwater wells received higher fecal contamination loads than others, and leaking septic tanks were determined to be primary fecal contamination source. To further support this conclusion, human *Bacteroidales* 16S rRNA copy numbers were quantified in DNA extracts from septic tank effluents (STE) and groundwater wells across Wainfleet. Human *Bacteroidales* concentrations were the highest in groundwater wells with the most FIB counts, at levels almost 50 times more than groundwater wells with less than a $\log_{10}$ *E. coli* count of 2.0. Furthermore, the concentrations of the HF183 *Bacteroidales* marker was significantly correlated with the relative abundance of *Bacteroides* ($P < 0.05$). These results, combined with culture counts and deep amplicon sequencing, confirm the contributions malfunctioning septic tanks have on groundwater contamination in Wainfleet. Together, the results should facilitate the discussion concerning waste treatment in the township which in turn will prevent fecal contamination flow into the groundwater wells.

4.2 Introduction

Rural communities across Canada, like Wainfleet, rely on groundwater aquifers as their primary drinking water source. Wainfleet is under the longest active boil water advisory in Canada because fecal coliforms and *E. coli* counts were detected above provincial limits in their residential groundwater wells. Leaking septic tanks were previously considered the primary source of fecal contamination, along with poorly constructed wells, which facilitated the transfer of fecal contaminants into the groundwater wells (verbal communication with Wainfleet staff). The previous report that examined fecal contamination levels a decade ago only had access to culture-based assays. These assays cannot identify sources of fecal contamination in source water environments.

Wholesale characterization of microbial communities can provide relative contributions of fecal contamination sources based on sequence similarities between environmental and fecal samples (Unno et al. 2010). Next-generation sequencing of 16S rRNA sequences from DNA extracts in Wainfleet's groundwater, sewage, and manure samples revealed that groundwater wells with the highest *E. coli* counts and nitrate concentrations contained increased relative abundances of bacterial taxa that contain STE-dwelling microbes like *Bacteroides* and *Campylobacterales*. The sequencing profiles raise the possibility that septic tank effluents are the primary contributors of fecal contamination loads in groundwater wells where the highest *E. coli* counts are observed.

Genetic markers for quantifying fecal contamination have been identified. These markers can be used to complement targeted-sequencing methods for identifying fecal contamination sources. The quantification of these genetic markers with qPCR could

circumvent the shortcomings of culture-based assays for pinpointing fecal contamination sources (Field and Samadpour 2007) and the weak association between FIB counts and waterborne pathogen detection (Colford et al. 2007). Quantitative-PCR (qPCR) assays of the 16S gene belonging to the human-based *Bacteroidales* microbes have been identified for quantifying fecal contamination in environmental samples. The standard qPCR marker for quantifying human fecal contamination is the HF183 marker (Bernhard and Field 2000). This marker was first developed for use with end-point PCR before qPCR assays were developed to quantify the 565-bp fragment in environmental samples (Seurinck et al. 2005). The HF183 marker has since been used to quantify human fecal contamination in tropical watersheds (Nshimyimana et al. 2014), and freshwater lakes (Staley et al. 2012). However, no publication exists for the quantification of the HF183 marker in groundwater environments.

The aim of this section is to quantify the human *Bacteroidales* marker in groundwater wells across Wainfleet with qPCR. qPCR assays using the HF183 marker can be used to measure the amount of human fecal bacteria in Wainfleet's groundwater aquifer and test that leaking septic tanks are the primary contamination source. The qPCR data, when combined with culture-based and 16S rRNA sequencing surveys, can be used to provide a comprehensive assessment of fecal contamination levels and sources in contaminated groundwater wells. The information can then be used by public health officials to facilitate groundwater purification and ensure the town's public health.

4.3 Methods

4.3.1 Sample Collection and DNA Extraction

Groundwater and fecal samples were collected from private groundwater wells, septic tanks, and animal manure as described previously for the culture counts and microbial community characterization (Chapter 2, Chapter 3). DNA was also extracted from these samples the same way as was conducted for obtaining DNA for deep amplicon sequencing of the 16S rRNA gene (Chapter 3). DNA was quantified from these extracts with the Nanodrop 2000 (Thermo Scientific). The final volume of all DNA extracts was 100 μL, and all DNA extracts were stored in -20^{o}C for future experiments.

4.3.2 Preparation of Plasmid DNA Standards for qPCR Assays

qPCR assays relying on the absolute quantitation of fecal, genetic markers require the preparation of plasmid DNA standards as references to the genomic copy number of the markers in environmental samples. To begin, primers that flanked the fragments in the 16S gene that belonged to human *Bacteroidales* populations were obtained from previous literature (Bernhard and Field 2000). PCR assays amplifying the general and host-specific markers of interest were conducted using DNA extracted from reference samples as templates. For all PCR assays, 50 μL reaction volumes were prepared, containing 5 μL of Thermopol Buffer (10x, NEB), 1 μL each of the forward and reverse primer (10 μM), 1 μL of dNTP mix (NEB, 10 mM), 1 μL of 5 U Taq Polymerase (NEB), and 2 μL of DNA template. All PCR assays were run with the following thermal cycling conditions: Initial denaturation for 5 min at 95^{o}C, 35 cycles of denaturation at 95^{o}C for 30 s, annealing at 52^{o}C, and extension at 72^{o}C for 1 min, and a final extension step at 72^{o}C for 6 min. PCR

products were loaded onto 1% agarose gels for product visualization with a UV Transilluminator. DNA was extracted from gel bands that emitted the amplicon with the PCR and Gel DNA Extraction Kit (Macherey-Negel). The PCR and gel-extraction process was repeated if the gel extracts did not contain visible DNA when 2 μL was loaded onto another gel with gel electrophoresis. This step was performed to obtain a brighter DNA amplicon for gel extraction as gel extraction yields are generally between 50-70%.

Once a concentration of at least 50 ng/μL of DNA amplicons were obtained, 2 μL of gel DNA extracts were used for the TOPO TA Cloning Reaction (Invitrogen), where DNA was ligated into the TOPO vector, following the manufacturer's instructions. *E. coli* chemical-competent cells were then transformed by the ligated TOPO Vector. Briefly, 2 μL of the ligation product was transferred into 40 μL of chemically competent *E. coli* DH5α cells and incubated on ice for 20 min. The *E. coli* cells were then heat-shocked by transferring the tubes into a 42°C water bath for 30 s and then placing the tubes back on ice. 260 μL of Recovery Medium (Lucigen) was then added to the cell preparation, transferred into a culture tube, and incubated for 1 h at 37°C. 50 μL of the culture was then streaked into an LB blue-white screening plate (LB agar, 50 μg/mL X-gal, 50 μg/mL kanamycin, 20 μg/mL IPTG). The plates were then incubated for 24 h at 37°C.

A white colony was taken from the blue-white screening plate and inoculated into 3-mL of LB broth supplemented with 50 μg/mL kanamycin. The liquid culture was incubated overnight, and DNA was extracted from 1-aliquots of the cultures with the Plasmid DNA Miniprep Kit (Norgen Biotek, Thorold, ON) and subsequently with the kit-free Alkaline Lysis miniprep protocol. DNA concentrations were then quantified with the

Nanodrop 2000 (Thermo Scientific) and the Qubit Fluorometer 2.0 (Invitrogen). DNA extracts were then diluted such that a range of 10^1-10^7 plasmid copies of the marker was present in the DNA solutions using the formula for calculating total genome copies (Equation 1).

4.3.3 qPCR Standard Curve and Genetic Marker Quantitation

qPCR assays were subsequently conducted using 2 μL of groundwater DNA extracts and plasmid standard solutions. All qPCR assays were conducted with the SsoFast EvaGreen Super Mix solution (Bio-Rad) with 10 μL reaction mix volumes. In the reaction mixes were 1 μL of DNA template, 5 μL of the SsoFast EvaGreen SuperMix (Bio-Rad), and 0.5 μL each of the forward and reverse primer (10 μM). Thermal cycling conditions were conducted as follows: Initial denaturation at 95^{o}C for 30 s and 40 cycles of denaturation at 95^{o}C for 30 s, and 60^{o}C for 10 s. Melt curve analysis was conducted immediately after the qPCR was completed, from 65-95^{o}C in increments of 0.5^{o}C for 5 s per increment. All qPCR analyses were then conducted with the CFX Analyzer (Bio-Rad), where standard and melt curves were generated and qPCR copy numbers were determined. All DNA templates were diluted 10-fold to minimize the effects of qPCR inhibitors. Concentrations obtained through these qPCR assays were then multiplied by 10 to account for the dilution factor.

$$x = \frac{6.0221 \times 10^{23} \frac{molecules}{mol}(m)}{600\frac{g}{mol}(N) \times 1 \times 10^{9} ng/g}$$

Equation 1. Calculating Genome Copy Number from Plasmids

This equation was used to dilute the plasmid DNA standards to the correct genome copy numbers to prepare the qPCR standard curve. *x* represents the gene copy number in molecules, *m* represents the total mass of the amplicon in ng, and *N* represents the length of the dsDNA plasmid in bp.

4.4 Results

4.4.1 Standard Curve Generation and Limit of Detection

qPCR standard and melt curves were prepared for quantifying human-specific 16S *Bacteroidales* markers (Figure 20). As the SsoFast Evagreen Dye was used, melt curves were also prepared. Melt curve analysis of the HF183 qPCR assay yielded a single peak (Figure 20). qPCR standard curves of the HF183 marker had an R^2 value of 0.996 and an E value of 99.0%. Melt-curve analysis of the HF183 assay showed a single peak above threshold at 84^{o}C. For both markers, the limit of quantitation was 10 genome copies/100 mL, and the limit of detection was 1 genome copy/100 mL.

4.4.2 Detection and Quantitation of Human-specific *Bacteroidales* populations in Wainfleet's Groundwater

A subset of groundwater samples were selected for quantification of the general and human-specific *Bacteroidales* marker based on the detection of septic tank markers among the 16S rRNA sequencing data. Of the 8 groundwater wells that were tested for detection of the HF183 marker,

The two septic tanks had on average 50 genome copies of the HF183 marker per 50-mL of septic tank effluent (Figure 21). HF183 marker concentrations also coincided with the high abundance of septic tank microbes, namely *Bacteroidales* and *Campylobacterales*, in the septic tank effluent (Figure 21). In groundwater wells, the HF183 marker was detected in the highest concentrations in 11376A in August, the only groundwater sample with over 100 genome copies per 500 mL of filtered groundwater (Figure 22). The concentrations of the HF183 marker in all other groundwater wells were

not substantially different, even between groundwater wells with the highest *E. coli* colonies (11615+11215) and all the other groundwater wells (Figure 21). The HF183 marker was also prevalent in wells 11615 and 11215 at different concentrations. 11615 had the highest HF183 marker concentrations in August, the same month it had the highest *E. coli* counts in the entire study (Figure 22). The HF183 marker concentration in groundwater well 11215 was also substantially higher when the sample was collected in October as opposed to September (Figure 22). However, *E. coli* counts in the same groundwater well was only marginally higher in October when compared with September. For all other groundwater wells that were sampled in July and August, the HF183 concentrations were at the limit of detection, 1 genome copy per 500 mL of filtered groundwater. Gel electrophoresis of the HF183 assay in sewage samples yielded a gel band with a smaller band size than the expected amplicon size and a faint-band representing the HF183 marker in two groundwater samples (Figure S11).

4.4.3 Correlation of *Bacteroidales* Quantification Data with Targeted Sequencing Data

An important step to unifying the different approaches available to quantifying and sourcing fecal loads in source waters is to compare the independent sets of data obtained from these experiments. For this study, quantification data obtained from qPCR assays of the general and host-specific *Bacteroidales* markers was compared with the relative abundances of septic tank microbes in selected groundwater wells across Wainfleet.

E. coli counts were the highest in groundwater well 11615 when it was collected in August (Figure 4), but the HF183 marker was present at the highest concentrations in

11376A with over 120 copies of it detected per 500 mL (Figure 24). Linear regression analysis of the two variables revealed a weak positive correlation, without statistical significance when using the Spearman's Rank Correlation Test ($P > 0.05$).

A comparison of HF183 marker concentrations with the relative abundance of septic tank markers as determined within 16S sequencing revealed more substantial trends. For instance, groundwater well 11215 had twice as high relative abundance of *Bacteroidales* and *Campylobacterales* microbes in October than September, with HF183 marker concentrations also doubled between the two time points (Figure 24). The slight decrease in HF183 marker concentrations was also concurrent with the slight decrease in *Bacteroidales* and *Campylobacterales* relative abundances in groundwater well 11615 between July and October (Figure 24). The lone exception, groundwater well 11615 in August, the relative abundances of microbial groups that were identified in septic tanks was the highest, yet the HF183 marker was not present in substantially higher concentrations for August when compared with other months (Figure 24). The relative abundance of only the *Bacteroides* 16S sequences and the concentrations of the HF183 marker were also correlated with linear regression analysis, yielding an R^2 value of 0.471 and a significant positive correlation with the Spearman's Rank Correlation test ($P < 0.05$).

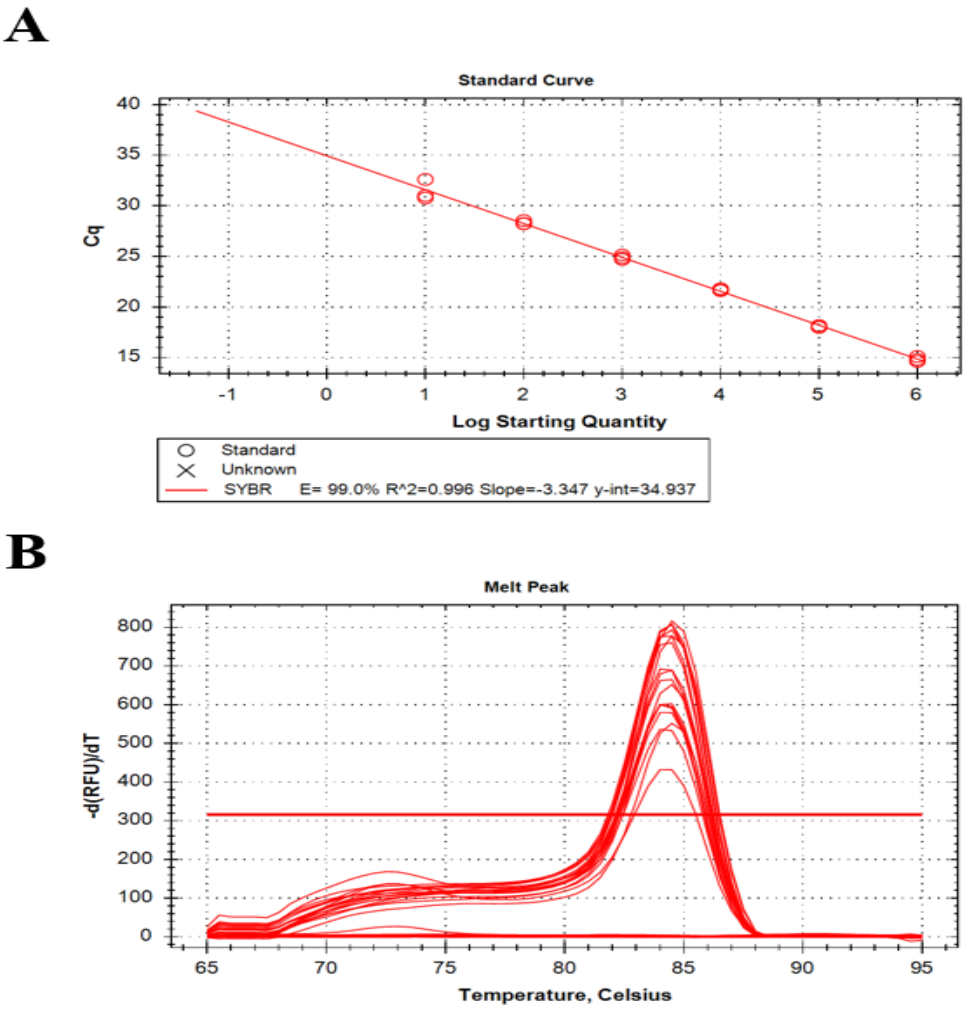

Figure 20. A) qPCR standard curve for the quantification of the HF183 *Bacteroidales* marker in environmental standards.

Crosses represent the three replicates of groundwater samples collected from sample ID 11615 in July, August, and October 2015. B) Melt curve analysis of the HF183 *Bacteroidales* qPCR assay. Temperature analysis ranged from 65ºC-95ºC in 0.5ºC intervals.

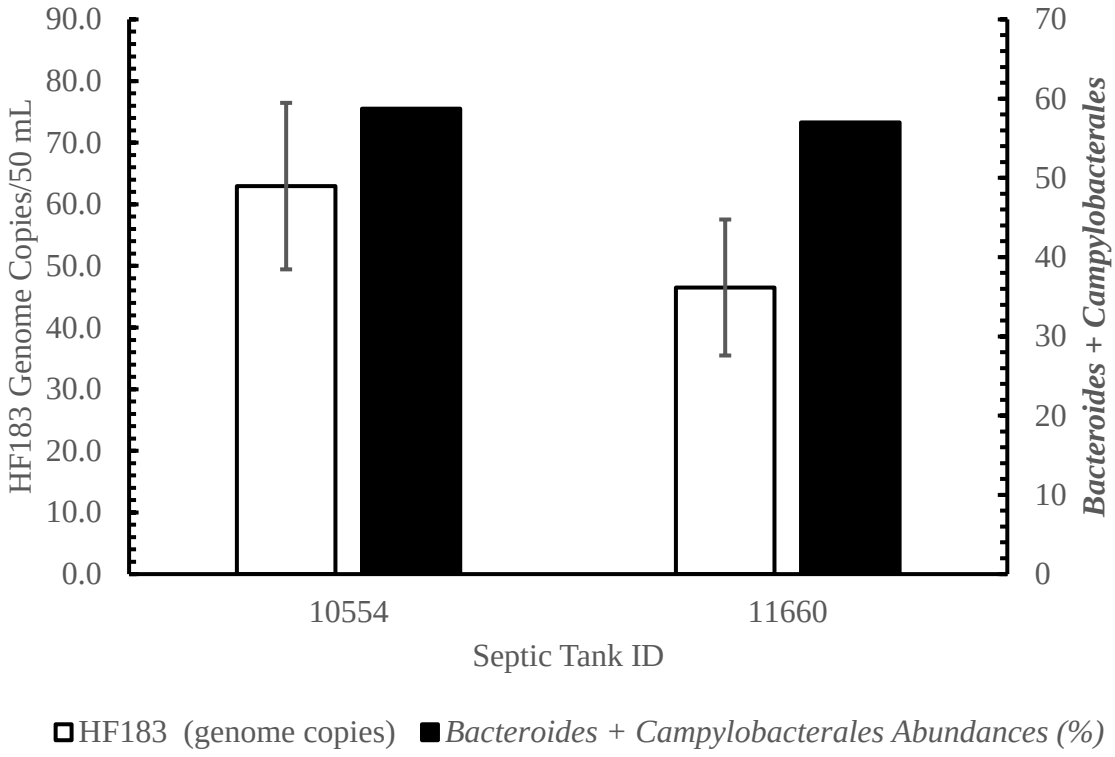

Figure 21. HF183 Quantities and Septic Tank Marker Abundances in Two Wainfleet Septic Tanks

Standard error bars were used for HF183 qPCR data for the septic tank samples. The differences in the HF183 genome copies were not statistically significant when the Mann-Whitney U test was employed ($P > 0.05$). Genome copy numbers are recorded as a concentration for every 50-mL of water sample that was filtered.

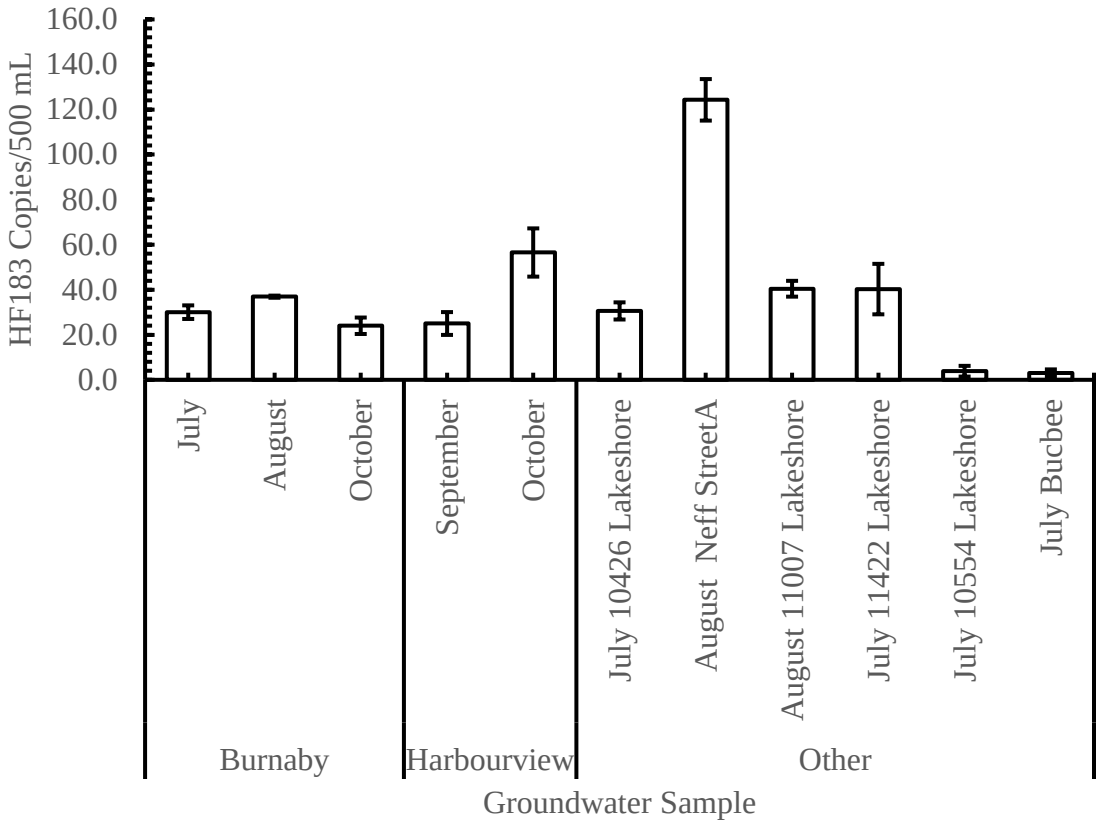

Figure 22. Quantification of the HF183 marker in a subset of Wainfleet groundwater samples using qPCR.

Bar graphs were divided into three groups: wells with sample ID 11615 and 11215, and all other groundwater wells. The other groundwater wells were obtained from a larger group of groundwater wells whose *E. coli* counts were below 10 CFU/100 mL. The "A" after a groundwater sample ID indicated that the water sample was treated with an onsite UV treatment system before sample collection. Genome copies of the HF183 marker were represented as a concentration per 500-mL of water that was filtered through a 0.45 μm pore-sized membrane filter.

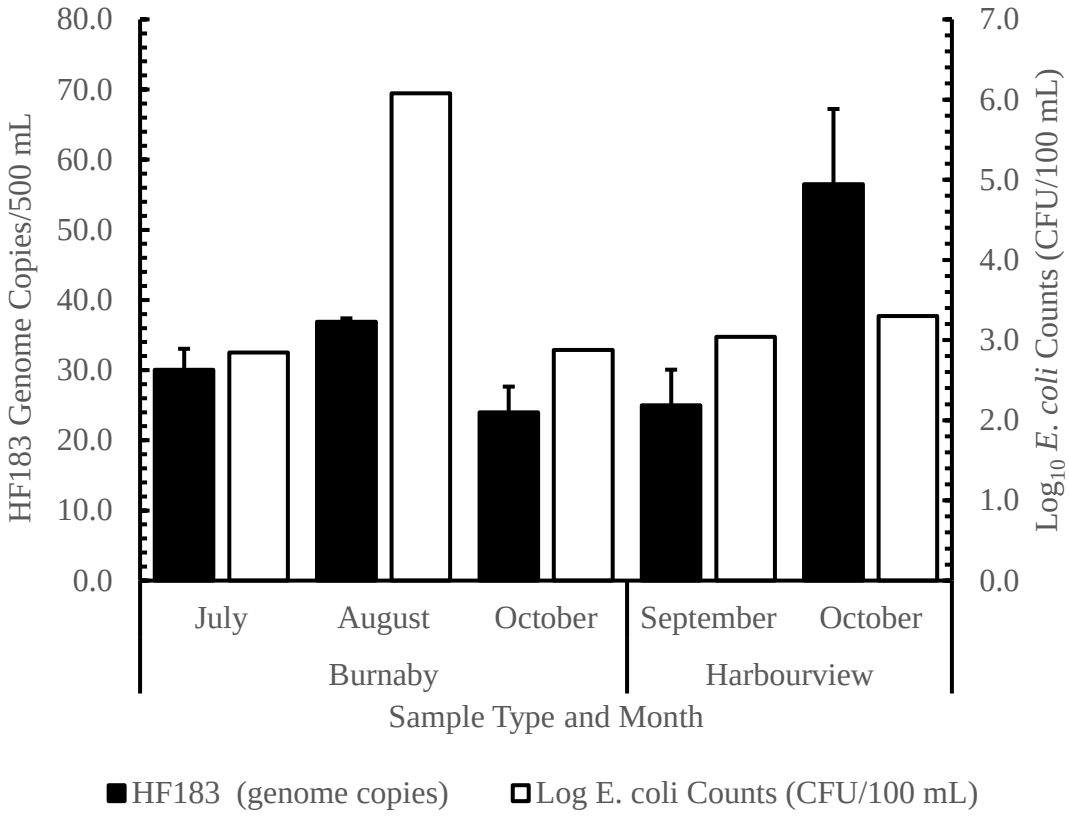

Figure 23. Comparison of HF183 Copy Numbers with *E. coli* Culture Counts in Groundwater Wells with the two Highest *E. coli* Counts

HF183 genome copy numbers were represented on the primary axis, while log-transformed *E. coli* counts are represented on the secondary y-axis. Standard error bars were used for the HF183 bar graphs, which were quantified in triplicate. Single replicates of *E. coli* counts were obtained from the groundwater wells. Genome copies of the HF183 marker were represented as a concentration per 500-mL of water that was filtered through a 0.45 μm pore-sized membrane filter.

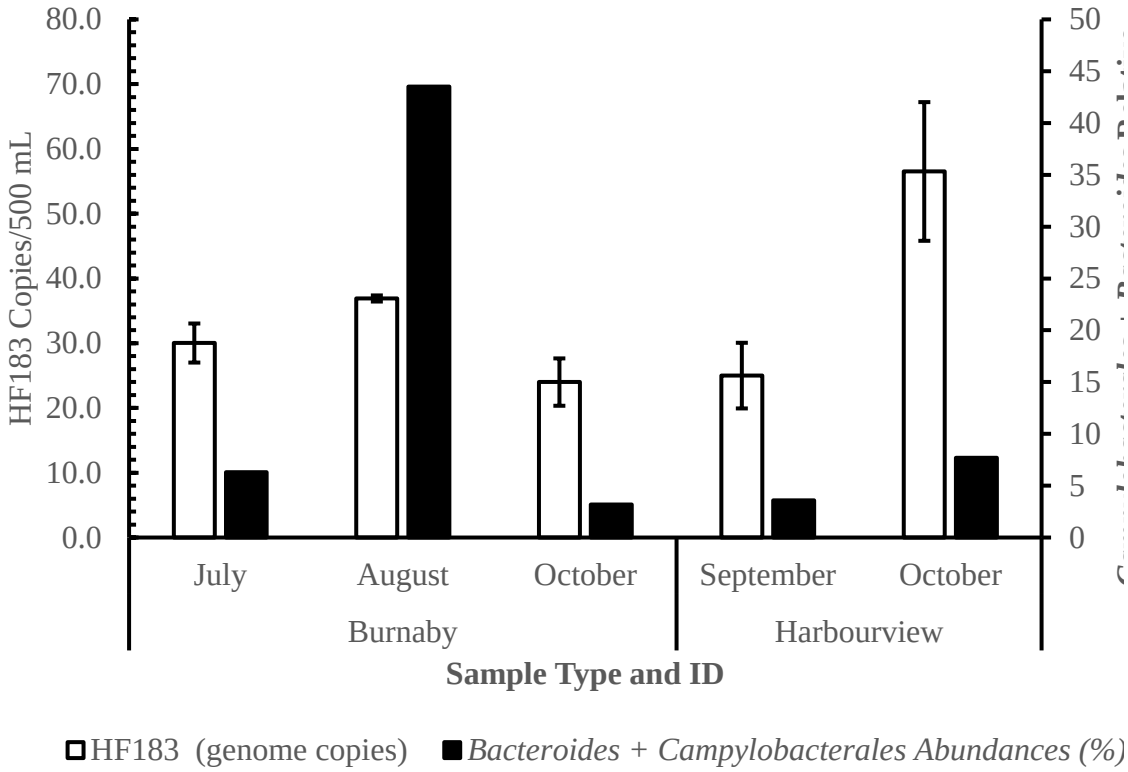

Figure 24. Bar Graphs of HF183 Copies/500 mL and the Relative Abundance of *Campylobacterales* + *Bacteroidales* Microbial Taxa in Groundwater Wells 11615 and 11215 Across Different Sampling Months

Error bars for HF183 marker quantification represent standard error bars based on triplicate measurements. Relative abundance measurements of the *Campylobacterales* and *Bacteroidales* orders were taken from sequencing data processed with QIIME. 16S rRNA copies of the HF183 marker were represented as a concentration per 500-mL of water that was filtered through a 0.45 μm pore-sized membrane filter.

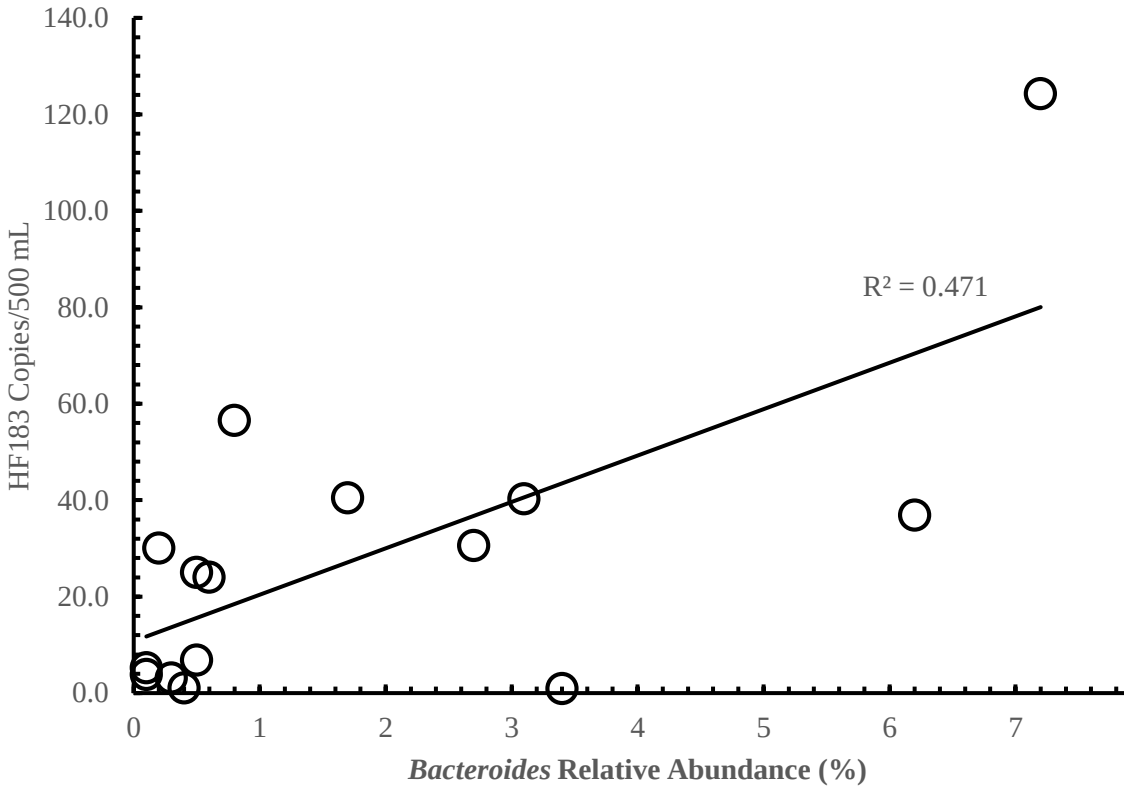

Figure 25. Scatterplot of HF183 Genome Copies vs *Bacteroides* Relative Abundance in Groundwater Samples

Sample points were obtained from Wainfleet groundwater wells that were sampled from July to November 2015. The relative abundance of *Bacteroides* sequences was obtained by targeted sequencing of the 16S rRNA gene from groundwater DNA extracts. HF183 genome copy concentrations were obtained with qPCR assays.

4.5 Discussion

In the previous components of the study, three assays were used to estimate fecal contamination levels and to start pinpointing sources of fecal contamination. Culture counts provided a preliminary survey of fecal contamination levels within a subset of the wells that were sampled. The data obtained from phosphate and nitrate concentrations provided a first indication that the elevated FIB counts could be derived from human or animal fecal contamination. The wholescale microbial community characterization of groundwater and manure-based communities within Wainfleet with 16S rRNA sequencing further pointed to leaking septic tanks as a possible contributor to groundwater contamination and the risk of waterborne infectious disease. The primary aim of this component of the source tracking study is to quantify human fecal contamination levels in residential groundwater wells using the qPCR quantification of the human *Bacteroidales* marker, HF183.

4.5.1 qPCR Standard Curves of General and Human-Specific *Bacteroidales* Markers

In qPCR assays, fluorescent dyes bind to target DNA sequences and emit fluorescence signals. When these signals reach threshold levels, the cycle number when this level is reached is used to quantify a gene marker in a sample (Ginzinger 2002). Absolute quantification of any gene marker in an environmental sample requires the preparation of a standard curve that correlates threshold cycle with gene marker concentrations. The standard curve that was generated for the HF183 marker consistently resulted in an R^2 value of 0.995, and an efficiency value of 92.2%, validating the standard curve at hand (Figure 20). Despite the large size of the amplicon (565 bp) for qPCR assays,

a robust standard curve can be produced for the quantification of the human *Bacteroidales* HF183 marker in environmental samples.

4.5.2 Human *Bacteroidales* Concentrations in Septic Tanks and Groundwater Samples

The first DNA-based marker of human fecal contamination was developed 16 years ago for the endpoint PCR detection of human *Bacteroidales* populations in environmental samples (Bernhard and Field 2000). Since then, a qPCR assay was developed which utilized dye chemistry and fluorescence to quantify these human-specific microbe (Seurinck et al. 2005) and has since been used as the standard marker of human fecal contamination.

The HF183 marker is used to detect raw human feces in environmental samples, but raw sewage can contain a microbiota distinct from freshwater and fecal samples (McLellan et al. 2010). Nevertheless, the fraction of the sewage microbiota also found in human fecal matter reflects the microbial communities typically found in the human gut (Newton et al. 2015). Importantly, this finding agrees with the detection of the HF183 marker in the effluents of two septic tanks, where an average of 50 genome copies of the HF183 marker was detected for every 50-mL of septic tank effluent (Figure 21). When compared with pristine groundwater samples where 500-mL of water was processed, septic tank samples contained on average 10 times as much of the HF183 marker (Figure 21, Figure 22). The HF183 marker thus can be used to indicate septic tank contamination should elevated levels of the marker be detected in groundwater samples.

After confirming that the HF183 marker can be detected in septic tank samples, groundwater samples collected from July to November were then screened for the human

Bacteroidales marker. HF183 marker concentrations were the highest in 11376A when collected in August, at 124 genome copies per 500-mL (Figure 22). Interestingly, this groundwater well also had the highest relative abundances of 16S rRNA sequences annotated to members of the *Campylobacterales* and *Bacteroidales* order. These results contradict the absence of *E. coli* colony counts in the groundwater sample, where the UV treatment system was activated at the time of sampling. This point represents an example where the absence of *E. coli* counts does not necessitate the absence of sewage contamination. The personnel collecting the groundwater samples has been told that the UV treatment system was malfunctioning, and that there was sewage buildup in this groundwater well for some time. The UV treatment system thus may have been able to inactivate the fecal indicators, but not enough so to remove the sewage contamination from the well.

For the other groundwater wells, HF183 marker concentrations ranged from non-detection in groundwater wells with fewer than 100 *E. coli* CFU/100 mL to 50 HF183 genome copies per 500-mL for groundwater wells with higher than 100 *E. coli* CFU/100 mL. The groundwater wells with lower *E. coli* counts also did not have detectable HF183 copy numbers. These wells may be receiving diffuse or no fecal contamination inputs during the month of sampling. Other groundwater wells, like 11215 that was collected in October, have higher HF183 concentrations and high *E. coli* counts, but a lower relative abundance of STE microbes ranging from 5% to 10% (Figure 23, Figure 24). These groundwater wells may contain other microbes that could be found in septic tanks which have not been identified by 16S rRNA sequencing.

4.5.3 Connecting qPCR Data with Culture Counts and Targeted Sequencing Datasets

To begin integrating the analyses used to quantify sewage and manure contamination in Wainfleet's groundwater wells, correlation test between HF183 marker concentrations, *E. coli* counts, and relative abundances of septic tank markers were performed. Linear regression analysis comparing *E. coli* counts with HF183 marker concentrations yielded a non-significant correlation with the Spearman Rank test ($P > 0.05$). The larger increase in *E. coli* counts compared with HF183 16S rRNA copies contributes to the non-significant correlation ($P > 0.05$) between the two variables (Figure S10). Groundwater wells with *E. coli* counts near the detection level may receive fecal contamination from diffuse sources. *E. coli* counts may not be useful for quantifying fecal contamination levels in Wainfleet's groundwater wells if *E. coli* counts are closer to detection limits. More groundwater wells should be sampled which have log-transformed *E. coli* counts ranging from 3.0 to 5.0 to determine whether human *Bacteroidales* concentrations can still be positively correlated with *E. coli* counts. Nevertheless, the HF183 marker is found in higher concentrations in groundwater wells receiving higher FIB loads. The HF183 marker is still useful for quantifying sewage contamination in groundwater.

The insignificance of the correlation was observed despite removing well 11376A from the linear regression analysis for its malfunctioning UV treatment system. The finding contradicts the results obtained from a study in the Upper Sugar Creek watershed where *E. coli* and human *Bacteroidales* concentrations were significantly correlated ($P < 0.001$)

(Drozd et al. 2013). Furthermore, a study conducted in the Gulf of Mexico determined that enterococci counts and HF183 detection had a significantly positive correlation as well ($P < 0.001$) (Gordon et al. 2013). The presence of *E. coli* in environmental samples does not necessitate the detection of waterborne pathogens or human fecal indicators (Hellein et al. 2011). However, the results herein show that culturable *E. coli* and HF183 can be used concomitantly to assess the contributions of human contaminants to source water pollution.

4.5.4 Conclusions and Future Directions

Throughout this project, diverse approaches have been implemented to survey groundwater microbial communities, quantify and source potential contaminants in groundwater wells across Wainfleet, and provide a comprehensive look into the safety of the township's primary drinking water source. Culture-based and wholescale microbial community characterization have not only shown that fecal contamination continues to seep into the town's groundwater, but also point towards leaking septic tanks as the primary source of contamination for the groundwater wells that are most impacted. The qPCR detection of has also corroborated the presence of human-based contaminants in groundwater wells with the highest *E. coli* counts and abundance of human gut microbes. Altogether, the groundwater wells most susceptible to fecal contamination receive it from leaking septic tanks. qPCR detection of the HF183 marker affirms these findings, where septic tank microbes The HF183 marker can be used to complement culture-based assays in quantifying fecal contamination loads in Wainfleet's groundwater wells.

The groundwater wells with the highest septic contamination loads either had poor groundwater well construction, as observed with 11615 (verbal information) or were

insufficiently closed off from the surrounding soils. The first courses of action would be to thus encourage efforts to construct a groundwater well dug far deeper and sufficiently enclosed to prevent fecal contaminants from seeping into the wells. The previous Wainfleet study also determined that over half of the homes that were assessed in Wainfleet did not have sufficient lot size for constructing a functional septic tank, increasing the risk of STE seeping through the thin clay soils into the underlying aquifer. Either providing a public sewage system or repairing the leaking septic tanks at least can provide a means for decreasing the amount of sewage effluent that seeps into groundwater.

Put together, this study has combined different approaches to confirming septic tanks as the primary contributor to the contamination of Wainfleet's primary drinking water source. Wainfleet's primary drinking water source is still under an active boil water advisory as of this year. With little action being done to repair the malfunctioning septic tanks and protect the groundwater wells from seeping fecal contamination, fecal contamination from human sources continues to be transferred into the wells. Culture counts and physiochemical assays point towards human or animal fecal contamination, and the use of 16S rRNA sequencing has identified leaking septic tanks as the primary and possibly lone fecal contamination source. Finally, HF183 concentrations are higher in groundwater wells with higher *E. coli* counts and more STE microbes, raising the possibility further that septic tanks are the primary fecal contamination source in Wainfleet's groundwater. Altogether, these datasets should help public health officials to prevent more fecal contamination from seeping into the town's groundwater wells, minimize the risks of a waterborne illness outbreak, and lift the boil water advisory.

Future studies should focus on the use of statistical approaches like Quantitative Microbial Risk Assessment, which was used to assess the risks of contracting human campylobacteriosis by swimming based on culture-based *E. coli* and *Campylobacter* counts (Schmidt et al. 2013). Shotgun sequencing of a subset of DNA extracts belonging to groundwater wells with the highest and lowest *E. coli* counts to determine if virulence genes and antibiotic resistance genes can also be detected. Virulence genes can be used to further assess the risks of a waterborne outbreak as the detection of pathogen-containing genera cannot precisely characterize pathogenic microbies in environmental samples. Antibiotic resistance genes can originate from human wastes and animal feeding operations where antibiotics are used to treat infectious diseases and stimulate animal growth respectively (Martinez 2008). The virulence and antibiotic-resistance genes could be used to expand upon MST assays to detect pathogenic microbes and to pinpoint sources of fecal contamination in groundwater environments.

Appendix A: Supplementary Figures and Tables

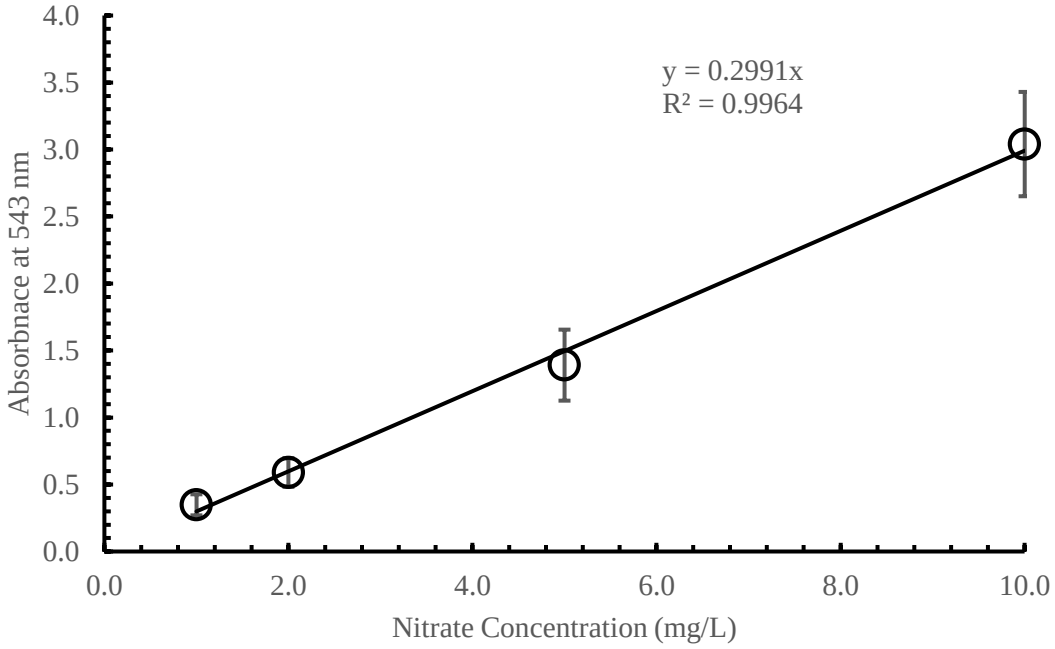

Figure S1. Cadmium Reduction Method Standard Curve

The cadmium reduction method standard curve correlates the absorbance of the water samples after treatment with the colour reagent at 543 nm with oxidized nitrate concentrations. A strong positive correlation between the two variables enabled the determination of oxidized nitrate concentrations with the cadmium reduction column. The limit of detection for the cadmium reduction method is 1 mg/L, while the maximum detectable nitrate concentration is 10 mg/L.

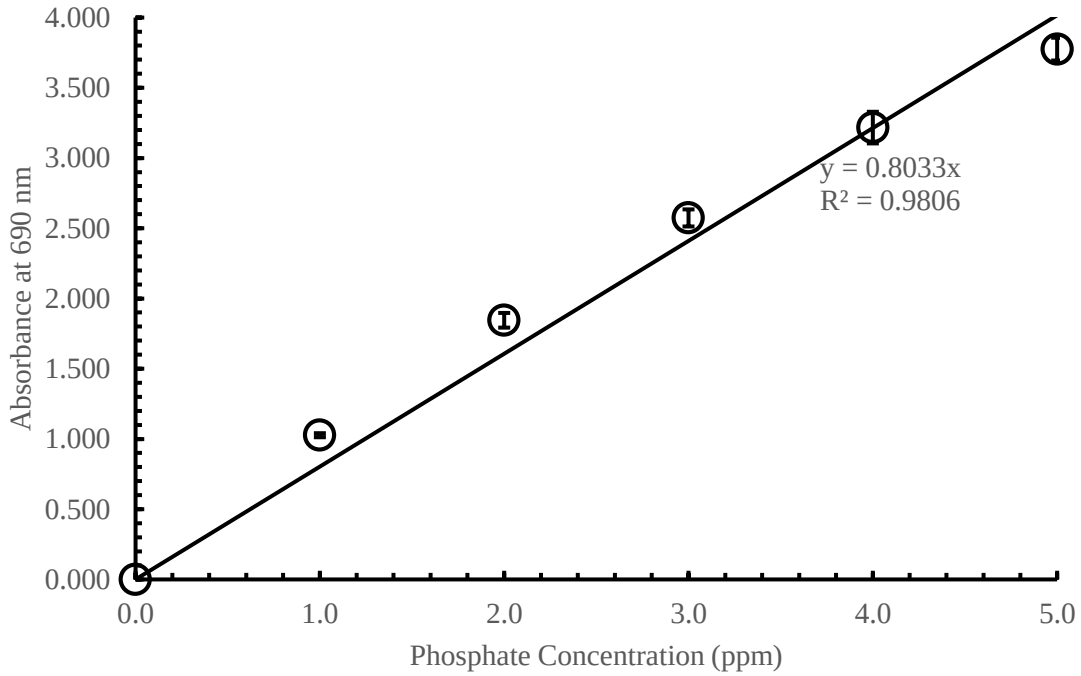

Figure S2. Stannous Chloride Method Standard Curve

The stannous chloride method correlates the absorbance of a water sample at 660 nm with the phosphate concentration. Here, a near perfect, positive correlation between the two variables was recorded. The minimum detectable concentration is 0.1 mg/L of phosphate, and the maximum concentration of phosphate that was recorded for this curve was 5.0 mg/L.

Number	1 (701)	2 (702)	3 (703)	4 (704)	5 (705)	6 (706)	7 (707)	8 (708)	9 (709)	10 (710)	11 (711)	12 (712)
A (501)	Jul11335	Jul10426	Jul11376	Jul11635	Jul10554	Jul11422	JulNeg	Aug11635	Aug11431	Aug10554A	Aug11376A	Aug11007
B (502)	Aug11435	Aug11413	AugNeg	Sept10122	Sept11235R	Sept11235R2	Sept11376AR1	Sept11376AR2	Sept10341R1	Sept10341R2	Sept11660R1	Sept11660R2
C (503)	Oct11235R1	Oct11235R2	Oct11635R1	Oct11635R2	Oct11660R1	Oct11660R2	OctNeg	Nov11235R1	Nov11235R2	Nov11431R1	Nov11431R2	ECP1
D (504)	SBChiR1	SBChiR2	SBChiR3	SBHoR1	SBHoR2	SBHoR3	SBCR1	SBCR2	SBCR3	SBPR1	SBPR2	SBPR3
E (505)	10554SepR1	10554SepR2	11660SepR1	11660SepR2	ECN1	D5Neg	D5671	D5672	D5673	D5675	D5687	D5688
F (506)	D5689	D5690	D5691	D10Neg	D10671	D10672	D10673	D10674	D10675	D10687	D10688	D10689
G (507)	D10690	D10691	D35Neg	D35671	D35672	D35673	D35674	D35675	D35687	D35688	D35689	D35690
H (508)	D35691	D35Neg	D35671	D35672	D35673	D35674	D35675	D35687	D35688	D35689	D35690	D35691

Figure S3. 96-Well Plate Layout for 16S rRNA PCR Assays

Sample layout of a 96-well plate for performing PCR assays of the 16S rRNA gene. All primers that were used have a unique adapter sequence that enables the IIllumina Sequencing platform to delineate DNA extracts from each other. The layout is structured such that there is a unique primer combination for every DNA template to be used in the study. All DNA samples starting with D represent extracts from mice feces for an unrelated study. EC represents *E. coli* controls, N being non-pathogenic and P being pathogenic.

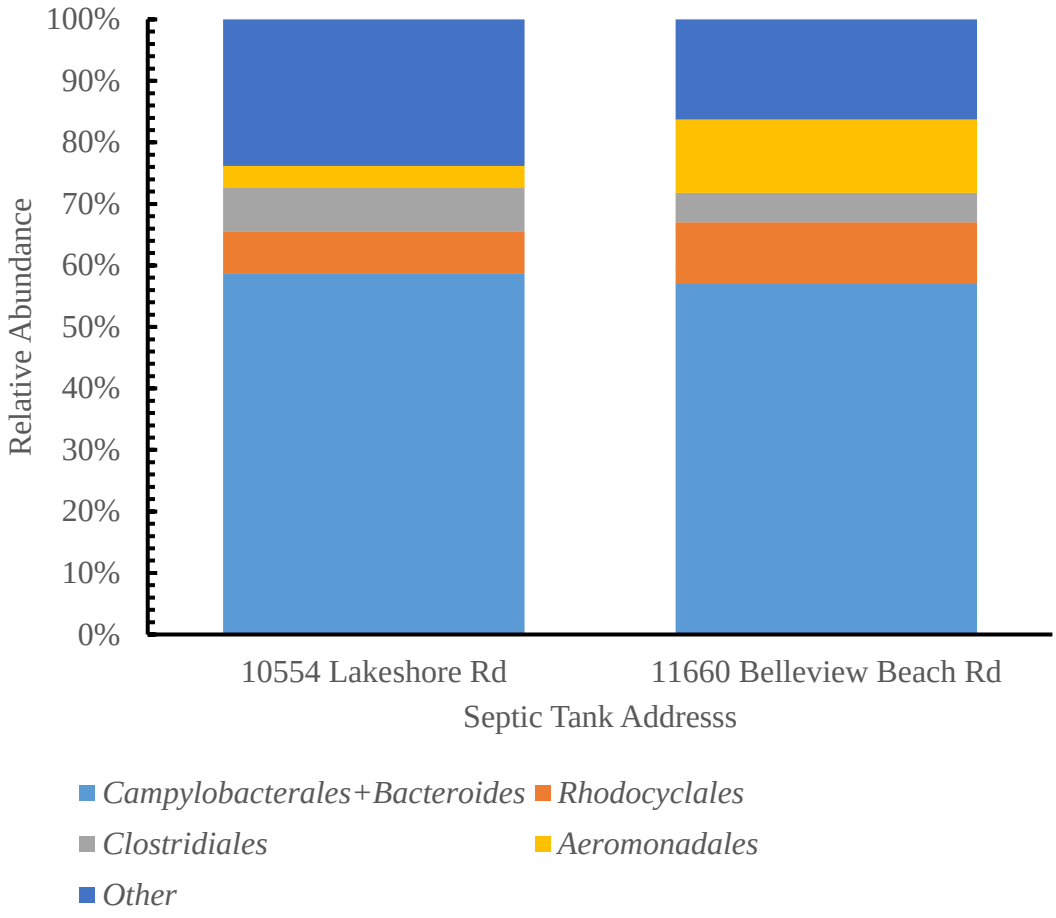

Figure S4. The Relative Abundance of Microbial Orders in Two Septic Tanks

Relative abundance stacked bar graphs of 16S rRNA sequences annotated to microbes at the order level. Only microbial taxa with at least 0.5% abundance is considered in the analysis. All microbial taxa with less than 0.5% relative abundance are classified within the "Other" microbial taxa.

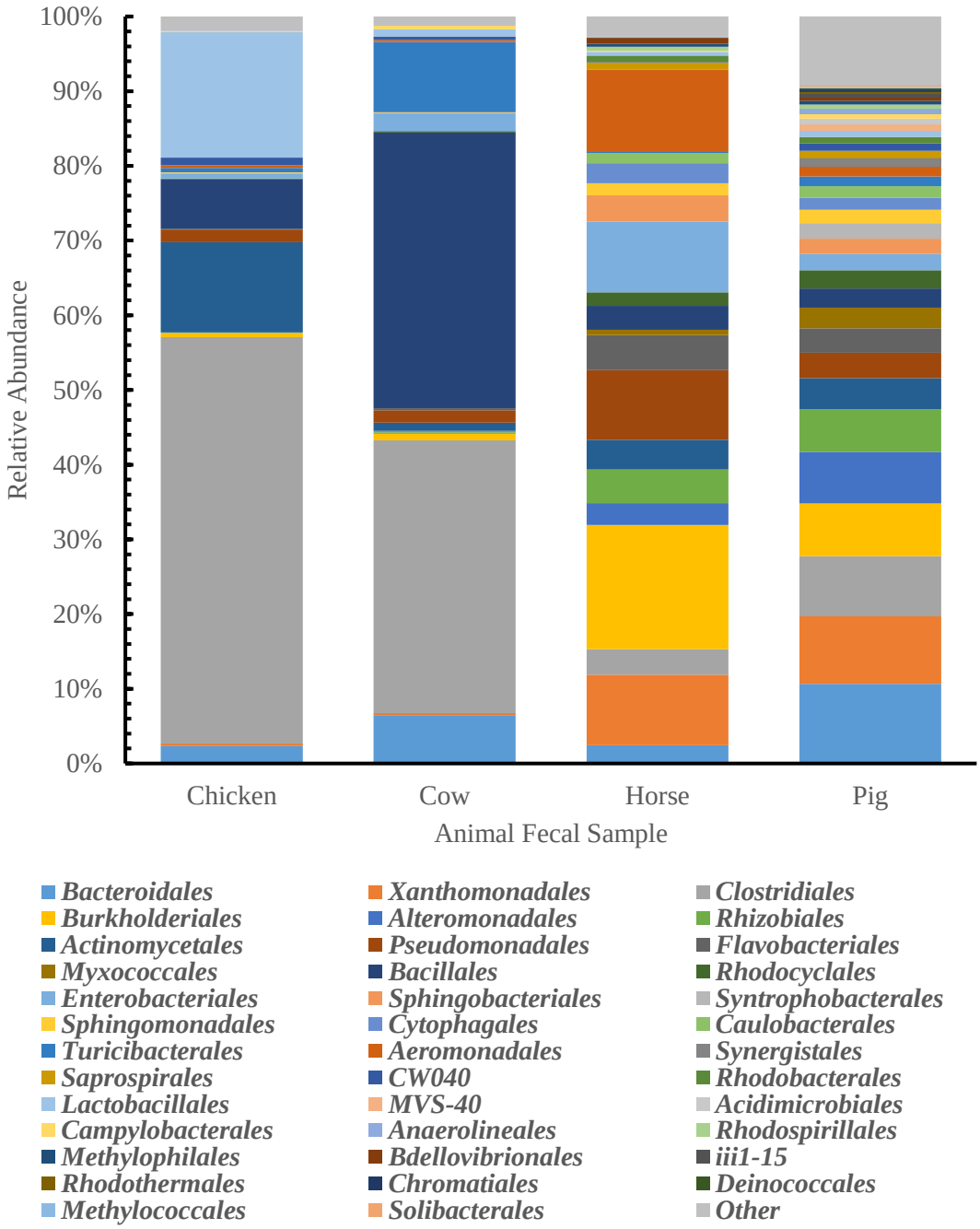

Figure S5. Relative Abundance of Microbial Taxa at the Order Level for Feces Obtained from Four Farm Animals

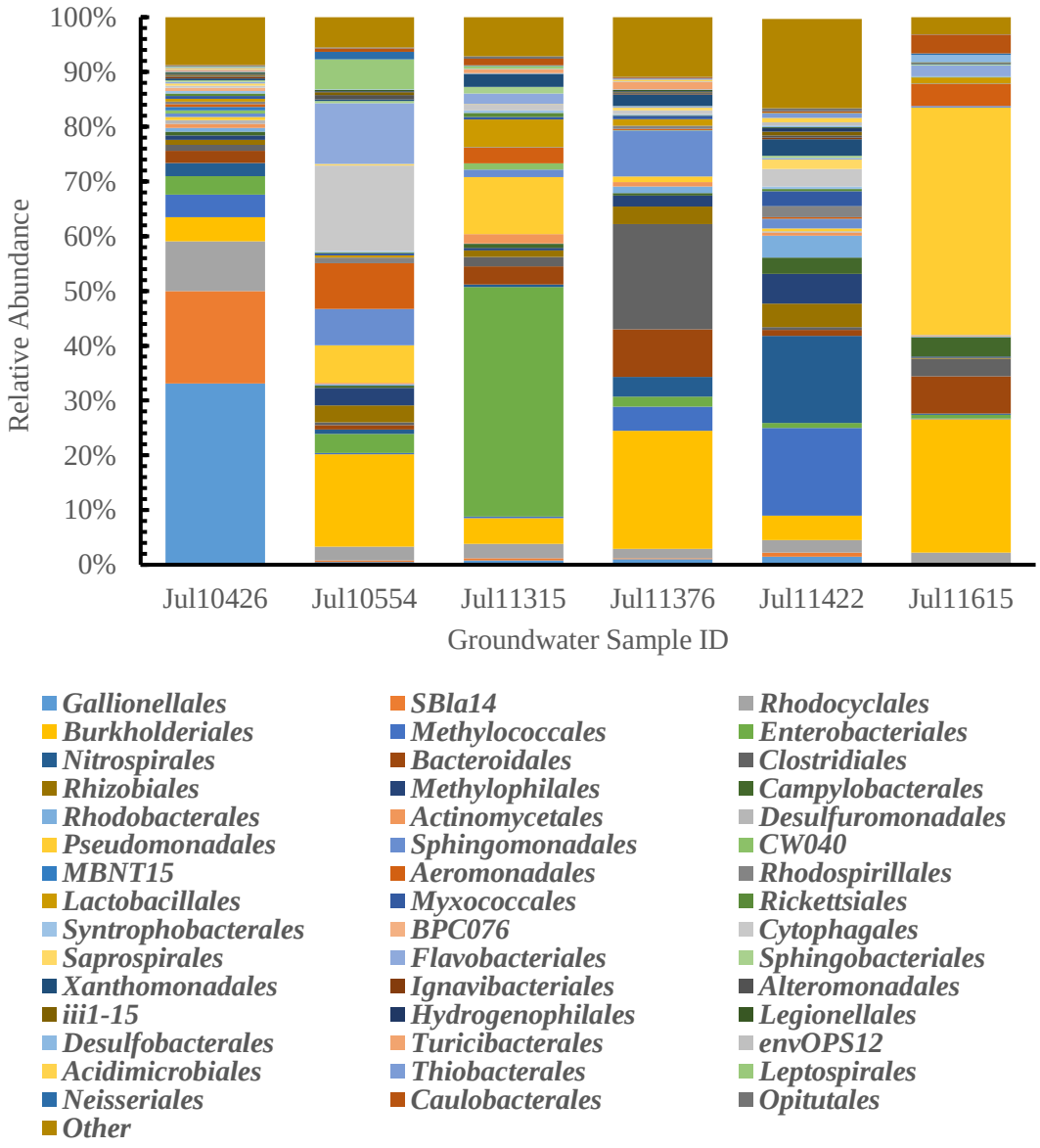

Figure S6. Relative Abundance of Microbial Orders in 6 Groundwater Samples Collected in July

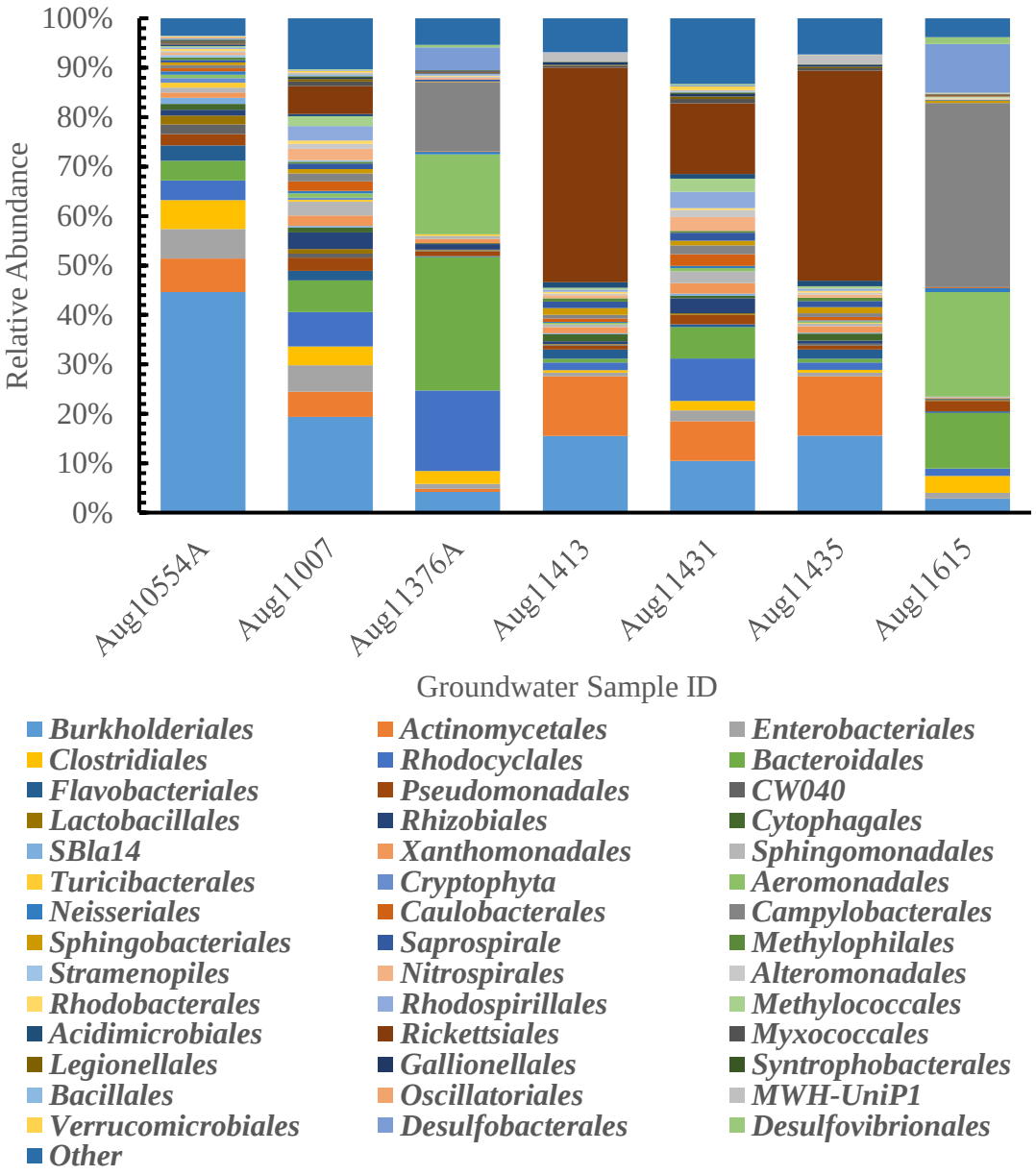

Figure S7. Relative Abundance of Microbial Orders in 6 Groundwater Samples Collected in August

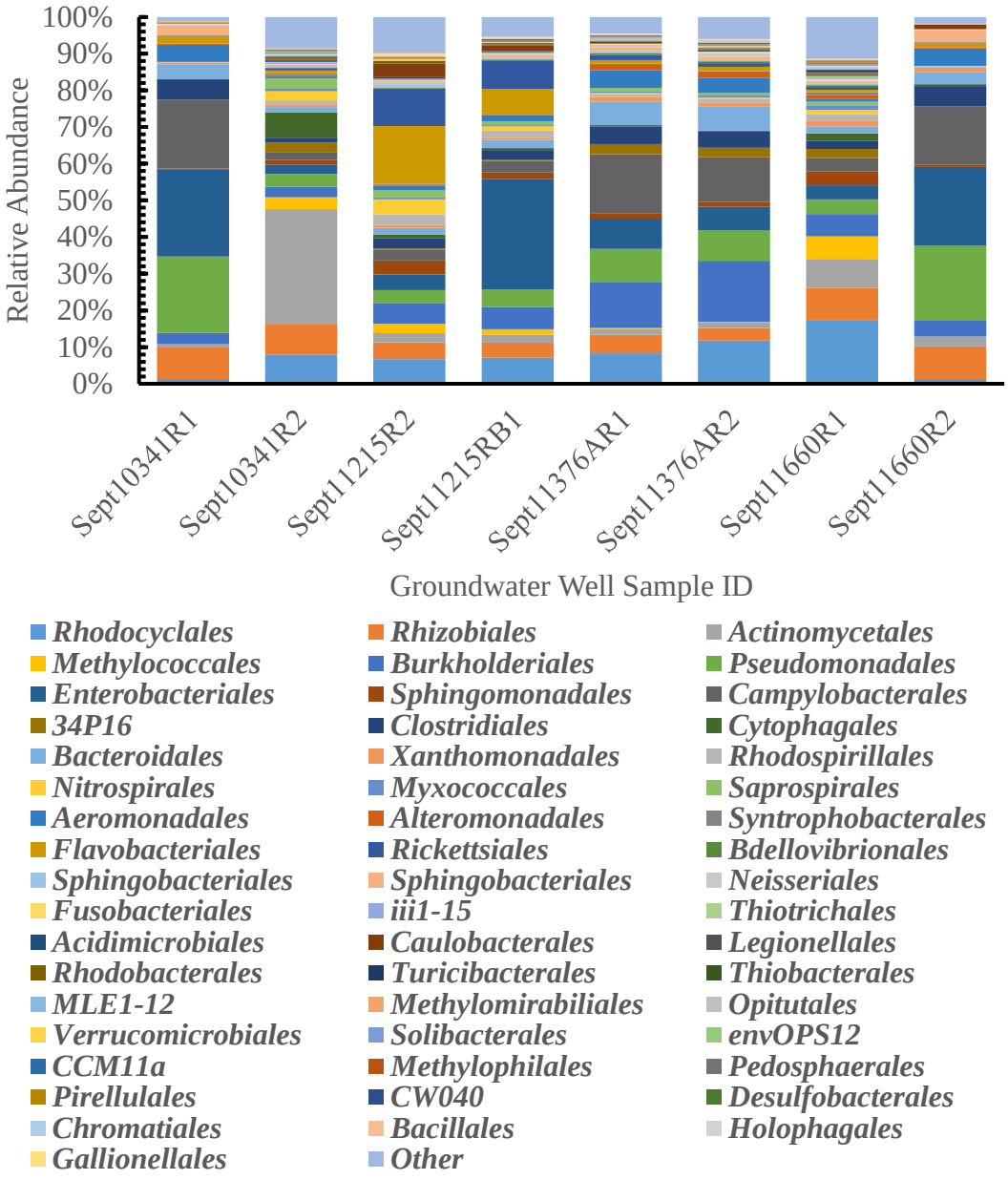

Figure S8. Relative Abundance of Microbial Taxa at the Order Level for Groundwater Samples in September.

R represents the biological replicate of the water sample, as water samples from the wells were collected twice in the September sampling event.

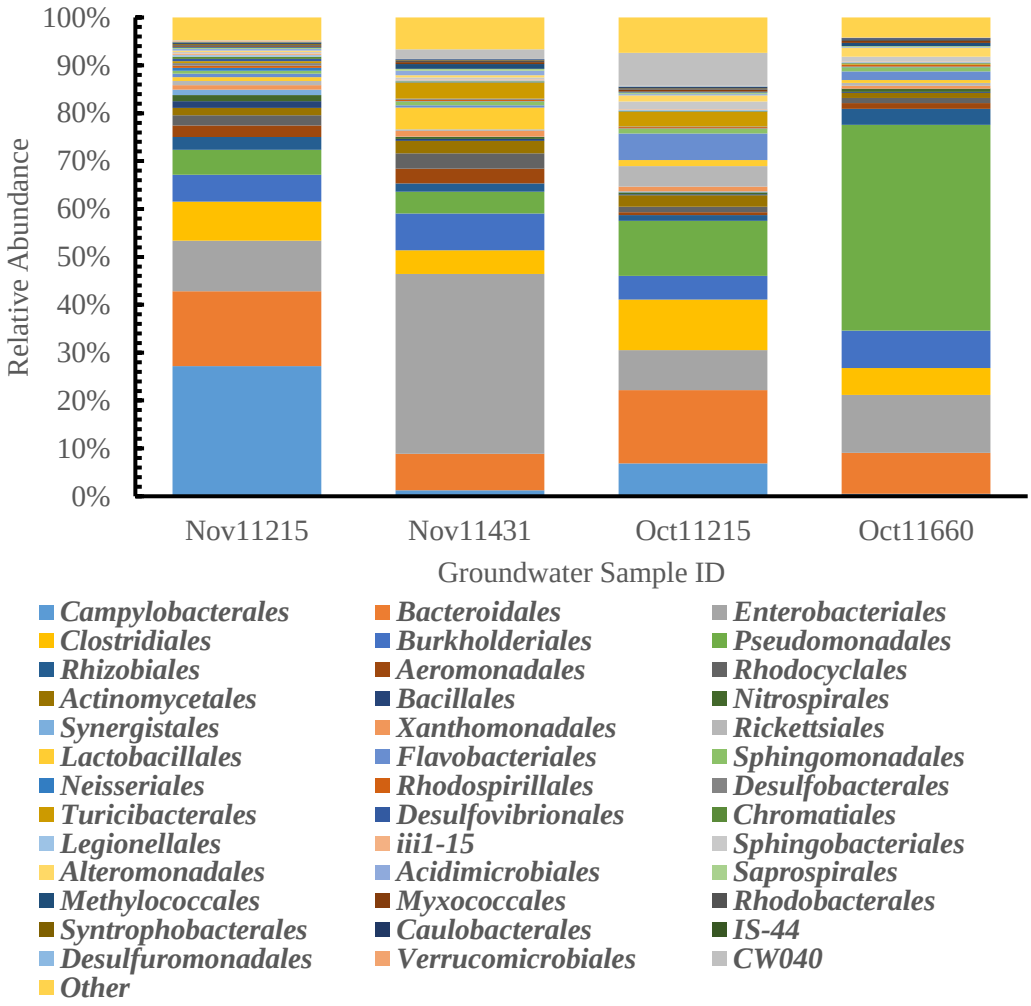

Figure S9. Relative Abundance of Microbial Taxa at the Order Level for Groundwater Samples Collected in October and November

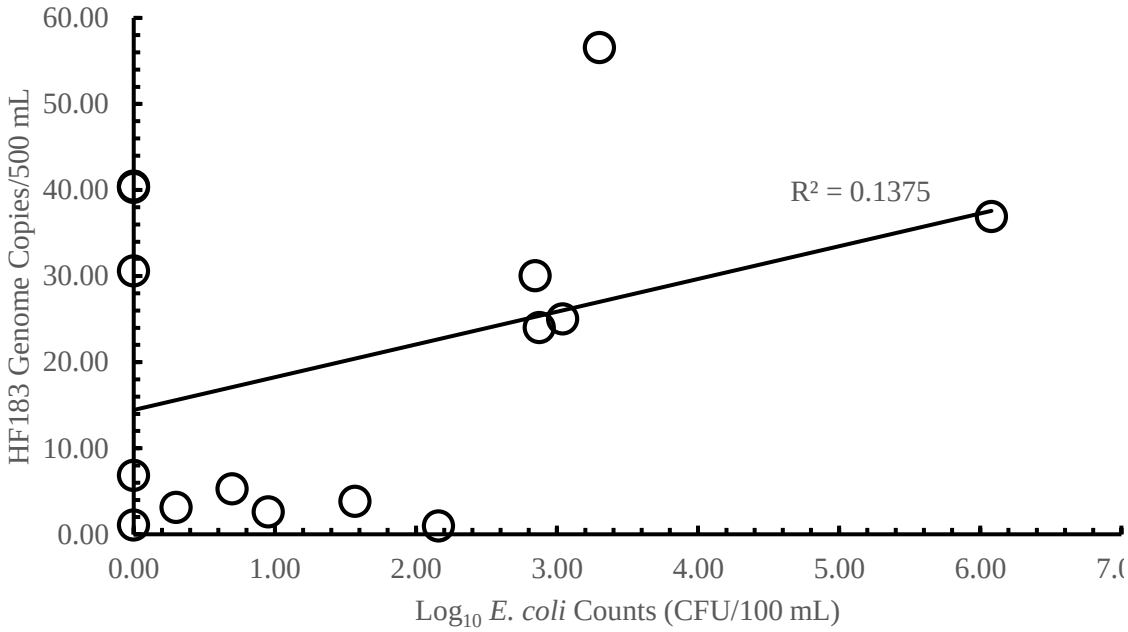

Figure S10. Linear Regression of HF183 Genome Copy Number and Log-transformed E. coli Counts

Spearman's Rank Correlation test determined that no significant correlation was detected ($P > 0.05$) between the two variables. 11376A, collected in August, was removed as an outlier point from the linear regression analysis. All samples collected were inputted for the linear regression analysis, regardless of the sampling month.

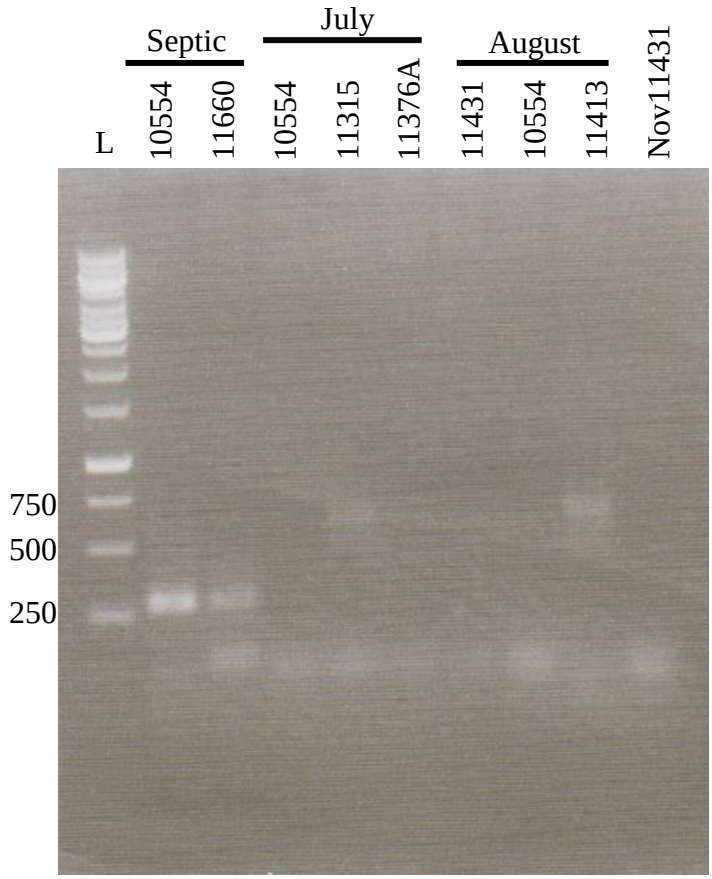

Figure S11. Gel Electrophoresis of HF183 qPCR Amplicons.

Expected amplicon size of the HF183 marker is 565 bp. A band was detected in the two septic tank wells, but with a smaller amplicon size. Faint amplicons near the bottom of the gel represent primer-dimer formation during the qPCR assay. The numbers represent the sample ID, the house number where the sample was collected. L = 1 kb ladder; Nov = November. A = groundwater well was UV-treated.

Appendix B: Standard Operating Procedures

Culture Counts of Fecal Indicator Bacteria with the Membrane Filtration Technique

The protocols were prepared to follow standard methods for quantifying fecal contamination, and are applicable to all three fecal indicator bacteria to be discussed (Rice et al. 2012).

1. Obtain water sample from the study site of interest. Transport to the lab in a cooler filled with ice packs.
2. Once samples arrive at the lab, perform 100-fold serial dilutions of the water sample by diluting 1-mL of water sample to 100 mL of 1x PBS solution (10x PBS solution recipe: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4)
3. Pass 100-mL of the water and serially diluted solutions through 0.45 μm pore-sized membrane filters with a vacuum pump.
4. Transfer the filters onto *Escherichia coli*, enterococci, and *E. coli* O157:H7 selective plates: DC agar (Oxoid), mEI agar (BD Difco), and Rainbow Agar (Biolog) respectively.
5. Incubate the selective plates for 22-24 h at 42^{o}C.
6. Count the amount of colonies that grow on the selective plates. Select the plate which yields 20-100 colonies. Multiply the value by the dilution factor to obtain total culture counts.
7. Report the results as CFU/100 mL, and log transform them if necessary.

Genomic DNA Extraction from Bacterial Isolates

This protocol was taken from Wilson (1987).

1. Prepare a 3-mL liquid culture by inoculating a single bacterial colony into 3-mL of 1x LB. Incubate the liquid culture overnight at 37ºC.
2. Transfer a 1.0-mL aliquot of the overnight culture into a 1.7-mL microcentrifuge tube. Centrifuge the liquid culture at 2,000 x *g* for 10 min. Discard the supernatant.
3. Resuspend the pellet in 567 µL of 1x TE buffer with repeated pipetting. Add 30 µL of 10% SDS and 3 µL of 20 mg/mL protease. Mix by flipping the microfuge tube up and down repeatedly. Incubate the solution for 1 h at 37ºC.
    a. As the lysis proceeds, the solution will become viscous and clear.
4. Add 100 µL of 5 M NaCl and 80 µL of CTAB/NaCl solution. Mix thoroughly by shaking the tube up and down and incubate the tube at 65ºC for 10 min.
5. Add 780 µL of 24:1 chloroform:isoamyl alcohol solution. Mix thoroughly, and spin at 10,000 x *g* for 5 min.
6. Transfer the aqueous layer of the supernatant (top layer) to a sterilized 1.7-mL microfuge tube. Add an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol solution. Mix thoroughly, and spin at 10,000 x *g* for 5 min.
7. Transfer the aqueous layer of the supernatant (top layer) to a sterilized 1.7-mL microfuge tube. Add 0.6 volume of isopropanol and flick the tube until white strings of DNA appear.
8. Pellet the precipitate by spinning the microfuge tubes at 8,000 x *g* for 1 min.
9. Remove the supernatant and add 1.0-mL 70% ethanol to the tube. Spin the tubes at 10,000 x *g* for 5 min at room temperature. Dry the ethanol by opening the tube inside a biohood.
10. Resuspend the DNA pellet in 100 µL of TE buffer. Store DNA extracts at -20ºC before downstream applications.

Miniprep Extraction of Plasmid DNA by Alkaline Lysis with SDS

The protocol was adapted from Sambrook et al. (1989).

1. Inoculate 3 mL of LB containing the appropriate antibiotic with a single colony of transformed bacteria. Incubate the culture overnight at 37°C with vigorous shaking.
2. Transfer 1.5-mL of the overnight culture into a 1.7-mL microfuge tube. Centrifuge the preparation at 14,000 x *g* for 5 min. Discard the supernatant.
   NOTE: Discard ALL of the LB medium, as traces of the medium can inhibit restriction digestion to confirm the successful ligation of the insert into the plasmid.
3. Resuspend the pellet in 100 μL of ice-old Alkaline Lysis Solution I with vigorous vortexing.
4. Add 200 μL of Alkaline Lysis Solution II to each bacterial suspension. Rapidly invert the tube five times to mix the contents. DO NOT VORTEX.
5. Add 150 μL of ice-cold Alkaline Lysis Solution III. Invert the tube several times and store the tube on ice for 3-5 min.
6. Centrifuge the bacterial lysates at 14,000 x *g* for 5 min. Transfer the supernatant to a sterile, autoclaved 1.7-mL microfuge tube.
7. Add 1 mL of 25:24:1 phenol:chloroform:isoamyl alcohol solution to the supernatant. Vortex the prep for 10 s and centrifuge at 14,000 x *g* for 10 min.
8. Transfer the upper aqueous layer to a new microfuge tube. Repeat Step 7.
9. Transfer the upper aqueous layer to a new microfuge tube. Add 1-mL chloroform to the tube and vortex for 10 s. Centrifuge the tube at 14,000 x *g* for 10 min.
10. Transfer the upper aqueous layer to a new microfuge tube. Add 1-mL of 100% ethanol and 50 μL of 3 M sodium acetate (pH 5.2) and vortex the tube to mix the contents. Place the tube on ice for 30 min.
11. Centrifuge the tubes at 14,000 x *g* for 10 min. Gently remove the supernatant.
12. Rinse the pellet with 1 mL of 70% ethanol. Centrifuge the tube at 14,000 x *g* for 10 min. Dry the pellet by placing the tube inside a biohood.
13. Resuspend the pellet in 100 μL of TE buffer. Add 1 μL of 10 mg/L RNase A to the solution. Incubate the tubes at room temperature for 30 min.
14. Add 1 volume of 25:24:1 phenol:chloroform:isoamyl alcohol solution to the supernatant. Vortex the prep for 10 s and centrifuge at 14,000 x *g* for 10 min.
15. Transfer the upper aqueous layer to a new microfuge tube. Add 1-mL of 100% ethanol and 50 μL of 3 M sodium acetate (pH 5.2) and vortex the tube to mix the contents. Place the tube on ice for 30 min.
16. Centrifuge the tubes at 14,000 x *g* for 10 min. Gently remove the supernatant.
17. Rinse the pellet with 1 mL of 70% ethanol. Centrifuge the tube at 14,000 x *g* for 10 min. Dry the pellet by placing the tube inside a biohood.
18. Resuspend the pellet in 100 μL of TE buffer. Measure DNA concentration with the Qubit Fluorometer 2.0.

PCR and Gel Extraction of a 16S gene fragment for Illumina Sequencing

This protocol is newly developed by the Schellhorn lab.

1. Obtain a sterile 96-well PCR plate. Prepare a layout for all the DNA templates that will be used to run the 16S PCR (Figure S3).
2. Each well will contain a unique combination of a forward and reverse primer with an Illumina adapter sequence tag attached to each primer. This primer pair amplifies the V3-V4 region of the 16S gene. Add 2.5 μL each of the forward and reverse primer to the correct well (Figure S3).
3. Add 2 μL of DNA template to the correct wells as laid out in step 1.
4. Prepare a master mix for the PCR assay, accounting for the following volumes within a single reaction mix:
    a. 2.5 μL 10x PCR Buffer, -$MgCl_2$ (Invitrogen)
    b. 0.5 μL 10 mM dNTP Mix Solution (NEB)
    c. 1.0 μL 10 mg/mL BSA in ddH_2O, UV Irradiated
    d. 0.25 μL Taq DNA Polymerase (Invitrogen)
    e. 0.75 μL $MgCl_2$ (Invitrogen)
5. Transfer 5 μL of the master mix to each of the wells in the 96-well plate. Add ddH_2O to a final volume of 25 μL.
6. Run the PCR assay with the following thermal cycling conditions:
    a. 95°C: 3 min
    b. 94°C: 30 s; 50°C: 30 s, 72°C: 1 min (40 cycles)
    c. 72°C: 10 min
7. Load the PCR products into individual wells of 0.9% agarose gel. Run a gel electrophoresis for 80 min at 90 V and 0.4 A.
8. Excise rows of gels containing the amplicon of interest and extract DNA from agarose gels and transfer them into a 50-mL Falcon tube.
9. Extract DNA from the agarose gels with the Nucleospin PCR and Gel Cleanup Kit (Macherey Negel). Place Falcon tubes into container filled with water microwaved to 55°C during NTI buffer step.
10. At the elution step, perform the elution step twice, passing 50 μL of Elution Buffer (Macherey-Negel) through the spin column each time.
11. Concentrate the gel DNA extracts to 25 μL with the concentrator plus vacuum centrifuge (Eppendorf).
12. Reload the DNA contents into a newly prepared 0.9% agarose gel and run gel electrophoresis
13. Perform gel DNA extraction with the gel extraction kit, following manufacturer's instructions.
14. Repeat Steps 10 and 11.

The Cadmium Reduction Method

Introduction

The aim of this document is to provide the protocol for quantifying nitrate concentrations in source water environments with the cadmium reduction method. The method is based on the reduction of nitrate ions into nitrites before a reaction with a colour reagent. The intensity of the pink colour that forms is then used to quantify the oxidized nitrate concentrations in the samples.

The Protocol

Reagent Recipes

- 2% $CuSO_4$: Dissolve 20 g of $CuSO_4 \cdot 5\ H_2O$ powder in 500-mL of ddH_2O and dilute to 1-L.
- Dilute NH_4Cl-EDTA: Dissolve 13 g NH_4Cl and 1.7 g disodium EDTA in 900-mL of ddH_2O. Dilute to 1-L. Dilute 300-mL of the stock solution to 500-mL with ddH_2O.
- Colour Reagent: Add 5 g of sulfanilamide into 50-mL 85% phosphoric acid and 400-mL ddH_2O. Add 0.5 g *N*-(1-naphthyl)-ethylenediamine dihydrochloride after dissolving the sulfanilamide.

Preparing Nitrate Standard Solutions

1. Dry KNO_3 powder for 24 h at 105^{o}C in an oven.
2. Dissolve 0.3609 g of the dried powder in ddH_2O and dilute to 500-mL. This solution constitutes 100 ppm of NO_3^- standard.
3. Dilute the appropriate volume of standard solution into a final volume of 100-mL to obtain a range of [NO_3^-] from 0-10 ppm.

Preparing the cadmium reduction column

1. Insert a small bundle of glass wool into the bottom of the cadmium reduction column. Fill the column with ddH_2O, making sure that the water does not seep through the bottom with the tap.
2. Wash 12.5 g of 20-100 mesh cadmium granules with 6 N HCl and rinse with water.
3. Swirl the granules in 50 mL of 2% $CuSO_4$ solution until the blue colour begins to fade. Decant the solution and repeat the process until a brown colloidal precipitate forms.
4. Gently wash the granules with ddH_2O until the brown precipitates are removed and until the solution begins to become less turbid.

5. Transfer the cadmium granules into the cadmium reduction column until the granules are piled three-quarters the height of the column.
6. Pass 200-mL of dilute NH_4Cl-EDTA solution through the cadmium-filled column to activate the column.
7. Ensure that the column does not dry out under any circumstances. If the column is to be stored, fill the column with NH_4Cl-EDTA

Making the Cadmium Reduction Standard Curve

1. Dilute 25-mL of each nitrate standard solution to 100-mL with dilute NH_4Cl-EDTA solution. Swirl gently to mix
2. Pass 50-mL of the 1 ppm NO_3^- solution mixed with NH_4Cl-EDTA through the cadmium reduction column to activate the granules.
3. Pour the rest of the mixed nitrate standard solutions through the cadmium reduction column. Discard the first 25-mL of the standard solution. Retain the next 50-mL of the solution in an Erlenmeyer flask.
4. Add 2-mL of colour reagent and mix the resultant solution by swirling.
5. Exactly after 10 min, record the absorbance of the solution at 543 nm with a spectrophotometer.
6. Formulate standard curve plotting absorbance at 543 nm against the nitrate concentration that was oxidized.

Quantifying Nitrate Concentrations in Water Samples

1. Dilute 25-mL of the water sample to 100-mL with dilute NH_4Cl-EDTA. Swirl gently to mix
2. Pass the diluted water sample through the cadmium reduction column. Discard the first 25-mL of the solution. Retain the next 50-mL of the solution in an Erlenmeyer flask.
3. Add 2-mL of colour reagent and mix the resultant solution by swirling.
4. Exactly after 10 min, record the absorbance of the solution at 543 nm with a spectrophotometer.
5. Convert the absorbance of the solution into the oxidized-nitrate concentrations using the standard curve.

The Stannous Chloride Method

Introduction

The aim of this document is to provide a protocol for quantifying phosphate concentrations in source water environments with the stannous chloride method. The stannous chloride method converts the phosphate ions into molybdophosphoric acid before this compound is reduced into molybdenum blue by stannous chloride. The intensity of the blue colour that is emitted by the molybdenum blue is then recorded to quantify the phosphates that were reacted.

The Protocol

The Reagents

- Ammonium molybdenum reagent: Dissolve 2.5 g of $(NH_4)_6Mo_7O_{24} \cdot 4\ H_2O$ in 17.5-mL of ddH_2O in one bottle. In another bottle, slowly add 28-mL of conc. H_2SO_4 to 40-mL of ddH_2O. Once cooled, add this mixture into the molybdate solution, and dilute to 100-mL.
- Stannous Chloride Solution: Dissolve 2.5 g of fresh $SnCl_2 \cdot 2\ H_2O$ in 100-mL glycerol. Heat in a water bath and stir with a glass rod to speed up the dissolution process.

Quantifying Phosphate Concentrations in Standards and Water Samples

1. Dissolve 0.2195 g of anhydrous KH_2PO_4 in ddH_2O and dilute the solution to 1-L. This solution represents 50 ppm of PO_4^{-3}. Dilute this stock solution to 50-mL with ddH_2O to prepare a concentration range between 0 and 10 ppm.
2. Add a single drop of phenolphthalein into the standards and water samples. Shake gently to mix.
3. Add 4.0-mL of ammonium molybdenum reagent and 500-µL of stannous chloride solution into the standards and water samples.
4. After 10 min, measure the absorbance of the blue colour from the standards at 660 nm. Generate a standard curve plotting absorbance and phosphate concentration.
5. At the same time, measure the absorbance of the water samples at 660 nm and calculate phosphate concentrations using the standard curve.

References

Anderson, K.L., Whitlock, J.E. and Harwood, V.J. (2005) Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied and Environmental Microbiology* 71, 3041-3048.

Ashbolt, N.J., Grabow, W. and Snozzi, M. (2001) Indicators of microbial water quality. *IWA Publishing*, 289-316.

Aw, T.G. and Rose, J.B. (2012) Detection of pathogens in water: from phylochips to qPCR to pyrosequencing. *Current Opinion in Biotechnology* 23, 422-430.

Bernhard, A.E. and 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. *Applied and Environmental Microbiology* 66, 4571-4574.

Bettelheim, K.A. (1998) Studies of *Escherichia coli* Cultured on RainbowTM Agar O157 with Particular Reference to Enterohaemorrhagic *Escherichia coli* (EHEC). *Microbiology and Immunology* 42, 265-269.

Bibby, K., Viau, E. and Peccia, J. (2010) Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. *Water Research* 44, 4252-4260.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Bottone, E.J. (1999) *Yersinia enterocolitica*: overview and epidemiologic correlates. *Microbes and Infection* 1, 323-333.

Bunge, J., Willis, A. and Walsh, F. (2014) Estimating the Number of Species in Microbial Diversity Studies. *Annual Review of Statistics and Its Application* 1, 427-445.

Cao, Y., Van De Werfhorst, L.C., Dubinsky, E.A., Badgley, B.D., Sadowsky, M.J., Andersen, G.L., Griffith, J.F. and Holden, P.A. (2013) Evaluation of molecular community analysis methods for discerning fecal sources and human waste. *Water Research* 47, 6862-6872.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K. and Gordon, J.I. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7, 335-336.

Chakravorty, S., Helb, D., Burday, M., Connell, N. and Alland, D. (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods* 69, 330-339.

Chiu, C.-H. and Ou, J.T. (1996) Rapid identification of *Salmonella* serovars in feces by specific detection of virulence genes, *invA* and *spvC*, by an enrichment broth culture-multiplex PCR combination assay. *Journal of Clinical Microbiology* 34, 2619-2622.

Clark, S.T., Gilbride, K.A., Mehrvar, M., Laursen, A.E., Bostan, V., Pushchak, R. and McCarthy, L.H. (2011) Evaluation of low-copy genetic targets for waterborne bacterial pathogen detection via qPCR. *Water Research* 45, 3378-3388.

Colford, J.M., Jr., Wade, T.J., Schiff, K.C., Wright, C.C., Griffith, J.F., Sandhu, S.K., Burns, S., Sobsey, M., Lovelace, G. and Weisberg, S.B. (2007) Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. *Epidemiology* 18, 27-35.

Collado, L. and Figueras, M.J. (2011) Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clinical Microbiology Reviews* 24, 174-192.

Cortas, N.K. and Wakid, N.W. (1990) Determination of inorganic nitrate in serum and urine by a kinetic cadmium-reduction method. *Clinical Chemistry* 36, 1440-1443.

Coyne, M.J. and Comstock, L.E. (2008) Niche-specific features of the intestinal *Bacteroidales*. *Journal of Bacteriology* 190, 736-742.

Crouch, S.R. and Malmstadt, H. (1967) Mechanistic investigation of molybdenum blue method for determination of phosphate. *Analytical Chemistry* 39, 1084-1089.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72, 5069-5072.

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Santo Domingo, J.W., Simpson, J.M., Walters, S.P. and Field, K.G. (2005) Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Applied and Environmental Microbiology* 71, 3184-3191.

Drobniewski, F.A. (1993) *Bacillus cereus* and related species. *Clinical Microbiology Reviews* 6, 324-338.

Drozd, M., Merrick, N.N., Sanad, Y.M., Dick, L.K., Dick, W.A. and Rajashekara, G. (2013) Evaluating the occurrence of host-specific , general fecal indicators, and bacterial pathogens in a mixed-use watershed. *Journal of Environmental Quality* 42, 713-725.

Drury, B., Rosi-Marshall, E. and Kelly, J.J. (2013) Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Applied and Environmental Microbiology* 79, 1897-1905.

Field, K.G. and Samadpour, M. (2007) Fecal source tracking, the indicator paradigm, and managing water quality. *Water Research* 41, 3517-3538.

Figueras, M.J. and Borrego, J.J. (2010) New perspectives in monitoring drinking water microbial quality. *International Journal of Environmental Research and Public Health* 7, 4179-4202.

Fisher, J.C., Eren, A.M., Green, H.C., Shanks, O.C., Morrison, H.G., Vineis, J.H., Sogin, M.L. and McLellan, S.L. (2015) Comparison of Sewage and Animal Fecal Microbiomes by Using Oligotyping Reveals Potential Human Fecal Indicators in Multiple Taxonomic Groups. *Applied and Environmental Microbiology* 81, 7023-7033.

Fong, T.T., Mansfield, L.S., Wilson, D.L., Schwab, D.J., Molloy, S.L. and Rose, J.B. (2007) Massive microbiological groundwater contamination associated with a waterborne outbreak in Lake Erie, South Bass Island, Ohio. *Environmental Health Perspectives* 115, 856-864.

Galan, J.E., Ginocchio, C. and Costeas, P. (1992) Molecular and functional characterization of the *Salmonella* invasion gene *invA*: homology of *InvA* to members of a new protein family. *Journal of Bacteriology* 174, 4338-4349.

Ginzinger, D.G. (2002) Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Experimental Hematology* 30, 503-512.

Gordon, K.V., Brownell, M., Wang, S.Y., Lepo, J.E., Mott, J., Nathaniel, R., Kilgen, M., Hellein, K.N., Kennedy, E. and Harwood, V.J. (2013) Relationship of human-associated microbial source tracking markers with Enterococci in Gulf of Mexico waters. *Water Research* 47, 996-1004.

Green, H.C., Haugland, R.A., Varma, M., Millen, H.T., Borchardt, M.A., Field, K.G., Walters, W.A., Knight, R., Sivaganesan, M., Kelty, C.A. and Shanks, O.C. (2014) Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Applied and Environmental Microbiology* 80, 3086-3094.

Griebler, C. and Lueders, T. (2009) Microbial biodiversity in groundwater ecosystems. *Freshwater Biology* 54, 649-677.

Gubala, A.J. (2006) Multiplex real-time PCR detection of *Vibrio cholerae*. *Journal of Microbiological Methods* 65, 278-293.

Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiology and Molecular Biology Reviews* 62, 1435-1491.

Gupta, R.S. (2006) Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (*Campylobacterales*). *BMC Genomics* 7, 1.

Haack, S.K., Duris, J.W., Kolpin, D.W., Fogarty, L.R., Johnson, H.E., Gibson, K.E., Focazio, M., Schwab, K.J., Hubbard, L.E. and Foreman, W.T. (2015) Genes Indicative of Zoonotic and Swine Pathogens Are Persistent in Stream Water and Sediment following a Swine Manure Spill. *Applied and Environmental Microbiology* 81, 3430-3441.

Harwood, V.J., Staley, C., Badgley, B.D., Borges, K. and Korajkic, A. (2014) Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiology Reviews* 38, 1-40.

Heathwaite, A.L., Dils, R., Liu, S., Carvalho, L., Brazier, R., Pope, L., Hughes, M., Phillips, G. and May, L. (2005) A tiered risk-based approach for predicting diffuse and point source phosphorus losses in agricultural areas. *Science of the Total Environment* 344, 225-239.

Hellein, K.N., Battie, C., Tauchman, E., Lund, D., Oyarzabal, O.A. and Lepo, J.E. (2011) Culture-based indicators of fecal contamination and molecular microbial indicators rarely correlate with *Campylobacter spp.* in recreational waters. *Journal of Water and Health* 9, 695-707.

Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S., Barry, K., Tringe, S.G., Watson, D.B., He, Z., Hazen, T.C., Tiedje, J.M., Rubin, E.M. and Zhou, J. (2010) Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J* 4, 660-672.

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'Brien, P., Deletic, A. and McCarthy, D. (2016) Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Research* 93, 242-253.

Hijnen, W., Beerendonk, E. and Medema, G.J. (2006) Inactivation credit of UV radiation for viruses, bacteria and protozoan (oo) cysts in water: a review. *Water Research* 40, 3-22.

Holman, I.P., Howden, N.J., Bellamy, P., Willby, N., Whelan, M.J. and Rivas-Casado, M. (2010) An assessment of the risk to surface water ecosystems of groundwater P in the UK and Ireland. *Science of the Total Environment* 408, 1847-1857.

Holman, I.P., Whelan, M.J., Howden, N.J.K., Bellamy, P.H., Willby, N.J., Rivas-Casado, M. and McConvey, P. (2008) Phosphorus in groundwater-an overlooked contributor to eutrophication? *Hydrological Processes* 22, 5121-5127.

Hrudey, S.E., Huck, P.M., Payment, P., Gillham, R.W. and Hrudey, E.J. (2002) Walkerton: Lessons learned in comparison with waterborne outbreaks in the developed world. *Journal of Environmental Engineering and Science* 1, 397-407.

Huws, S.A., Kim, E.J., Lee, M.R., Scott, M.B., Tweed, J.K., Pinloche, E., Wallace, R.J. and Scollan, N.D. (2011) As yet uncultured bacteria phylogenetically classified as *Prevotella*, *Lachnospiraceae* incertae sedis and unclassified *Bacteroidales, Clostridiales* and *Ruminococcaceae* may play a predominant role in ruminal biohydrogenation. *Environmental Microbiology* 13, 1500-1512.

Ibekwe, A.M., Leddy, M. and Murinda, S.E. (2013) Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. *PLoS One* 8, e79490.

Ishii, S., Nakamura, T., Ozawa, S., Kobayashi, A., Sano, D. and Okabe, S. (2014) Water quality monitoring and risk assessment by simultaneous multipathogen quantification. *Environmental Science & Technology* 48, 4744-4749.

Jarvie, H.P., Neal, C. and Withers, P.J. (2006) Sewage-effluent phosphorus: a greater risk to river eutrophication than agricultural phosphorus? *Science of the Total Environment* 360, 246-253.

Johnson, T.B., McKay, L.D., Layton, A.C., Jones, S.W., Johnson, G.C., Cashdollar, J.L., Dahling, D.R., Villegas, L.F., Fout, G.S., Williams, D.E. and Sayler, G. (2011) Viruses

and bacteria in karst and fractured rock aquifers in East Tennessee, USA. *Ground Water* 49, 98-110.

Jothikumar, N. and Griffiths, M.W. (2002) Rapid detection of *Escherichia coli* O157: H7 with multiplex real-time PCR assays. *Applied and Environmental Microbiology* 68, 3169-3171.

Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B. and Wuertz, S. (2007) 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: a Bayesian approach. *Water Research* 41, 3701-3715.

Kim, M., Morrison, M. and Yu, Z. (2011) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiology Ecology* 76, 49-63.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41, e1.

Knappett, P.S., McKay, L.D., Layton, A., Williams, D.E., Alam, M.J., Huq, M.R., Mey, J., Feighery, J.E., Culligan, P.J., Mailloux, B.J., Zhuang, J., Escamilla, V., Emch, M., Perfect, E., Sayler, G.S., Ahmed, K.M. and van Geen, A. (2012) Implications of fecal bacteria input from latrine-polluted ponds for wells in sandy aquifers. *Environmental Science & Technology* 46, 1361-1370.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R. and Kelley, S.T. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* 8, 761-763.

Kobayashi, A., Sano, D., Hatori, J., Ishii, S. and Okabe, S. (2013) Chicken- and duck-associated Bacteroides-Prevotella genetic markers for detecting fecal contamination in environmental water. *Applied Microbiology and Biotechnology* 97, 7427-7437.

Koivunen, J. and Heinonen-Tanski, H. (2005) Inactivation of enteric microorganisms with chemical disinfectants, UV irradiation and combined chemical/UV treatments. *Water Research* 39, 1519-1526.

Korajkic, A., Parfrey, L.W., McMinn, B.R., Baeza, Y.V., VanTeuren, W., Knight, R. and Shanks, O.C. (2014) Changes in bacterial and eukaryotic communities during sewage decomposition in Mississippi river water. *Water Research* 69C, 30-39.

Krolik, J., Evans, G., Belanger, P., Maier, A., Hall, G., Joyce, A., Guimont, S., Pelot, A. and Majury, A. (2014) Microbial source tracking and spatial analysis of *E. coli* contaminated private well waters in southeastern Ontario. *Journal of Water and Health* 12, 348-357.

Kumaraswamy, R., Amha, Y.M., Anwar, M.Z., Henschel, A., Rodriguez, J. and Ahmad, F. (2014) Molecular analysis for screening human bacterial pathogens in municipal wastewater treatment and reuse. *Environmental Science & Technology* 48, 11610-11619.

Lagier, J.C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G. and Maraninchi, M. (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* 18, 1185-1193.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R. and 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. *Applied and Environmental Microbiology* 72, 4214-4224.

Leblanc-Maridor, M., Beaudeau, F., Seegers, H., Denis, M. and Belloc, C. (2011) Rapid identification and quantification of Campylobacter coli and Campylobacter jejuni by real-time PCR in pure cultures and in complex samples. *BMC Microbiology* 11, 113.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D. and Knight, R. (2008a) Evolution of mammals and their gut microbes. *Science* 320, 1647-1651.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R. and Gordon, J.I. (2008b) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews: Microbiology* 6, 776-788.

Lockhart, K.M., King, A.M. and Harter, T. (2013) Identifying sources of groundwater nitrate contamination in a large alluvial groundwater basin with highly diversified intensive agricultural production. *J Contam Hydrol* 151, 140-154.

Looft, T., Allen, H.K., Cantarel, B.L., Levine, U.Y., Bayles, D.O., Alt, D.P., Henrissat, B. and Stanton, T.B. (2014) Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *ISME J* 8, 1566-1576.

Lozupone, C. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* 71, 8228-8235.

Lu, J., Sanchez, S., Hofacre, C., Maurer, J.J., Harmon, B.G. and Lee, M.D. (2003) Evaluation of Broiler Litter with Reference to the Microbial Composition as Assessed by

Using 16S rRNA and Functional Gene Markers. *Applied and Environmental Microbiology* 69, 901-908.

Lundberg, J.O., Weitzberg, E., Cole, J.A. and Benjamin, N. (2004) Nitrate, bacteria and human health. *Nature Reviews Microbiology* 2, 593-602.

Marti, R., Gannon, V.P., Jokinen, C., Lanthier, M., Lapen, D.R., Neumann, N.F., Ruecker, N.J., Scott, A., Wilkes, G., Zhang, Y. and Topp, E. (2013) Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the South Nation River basin using *Bacteroidales* microbial source tracking markers. *Water Research* 47, 2315-2324.

Martinez, J.L. (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science* 321, 365-367.

McLellan, S.L. (2004) Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. *Applied and Environmental Microbiology* 70, 4658-4665.

McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N. and Sogin, M.L. (2010) Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology* 12, 378-392.

Mieszkin, S., Furet, J.-P., Corthier, G. and Gourmelon, M. (2009) Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific *Bacteroidales* 16S rRNA genetic markers. *Applied and Environmental Microbiology* 75, 3045-3054.

Miettinen, I.T., Vartiainen, T. and Martikainen, P.J. (1997) Phosphorus and bacterial growth in drinking water. *Applied and Environmental Microbiology* 63, 3242-3245.

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L. and McLellan, S.L. (2013) A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microbial Ecology* 65, 1011-1023.

Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M. and Sogin, M.L. (2015) Sewage reflects the microbiomes of human populations. *MBio* 6, e02574.

Nshimyimana, J.P., Ekklesia, E., Shanahan, P., Chua, L.H. and Thompson, J.R. (2014) Distribution and abundance of human-specific *Bacteroides* and relation to traditional indicators in an urban tropical catchment. *Journal of Applied Microbiology* 116, 1369-1383.

O'Brien, A.D., Newland, J.W., Miller, S.F., Holmes, R.K., Smith, H.W. and Formal, S.B. (1984) Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226, 694-696.

Okabe, S., Okayama, N., Savichtcheva, O. and Ito, T. (2007) Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Applied Microbiology and Biotechnology* 74, 890-901.

Oliver, J.D. (2005) The viable but nonculturable state in bacteria. *J Microbiol* 43, 93-100.

Oliver, J.D. (2010) Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEMS Microbiology Reviews* 34, 415-425.

Parker, J.K., McIntyre, D. and Noble, R.T. (2010) Characterizing fecal contamination in stormwater runoff in coastal North Carolina, USA. *Water Research* 44, 4186-4194.

Quilliam, R.S., Williams, A.P., Avery, L.M., Malham, S.K. and Jones, D.L. (2011) Unearthing human pathogens at the agricultural–environment interface: A review of current methods for the detection of *Escherichia coli* O157 in freshwater ecosystems. *Agriculture, Ecosystems & Environment* 140, 354-360.

Ramamurthy, T., Ghosh, A., Pazhani, G.P. and Shinoda, S. (2014) Current Perspectives on Viable but Non-Culturable (VBNC) Pathogenic Bacteria. *Front Public Health* 2, 103.

Ran, Q., Badgley, B.D., Dillon, N., Dunny, G.M. and Sadowsky, M.J. (2013) Occurrence, genetic diversity, and persistence of enterococci in a Lake Superior watershed. *Applied and Environmental Microbiology* 79, 3067-3075.

Reitzer, L. (2003) Nitrogen assimilation and global regulation in *Escherichia coli*. *Annual Reviews in Microbiology* 57, 155-176.

Rice, E.W., Bridgewater, L., Association, A.P.H., Association, A.W.W. and Federation, W.E. (2012) *Standard methods for the examination of water and wastewater*. Washington, D.C.: American Public Health Association.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning*: Cold spring harbor laboratory press New York.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H. and Robinson, C.J. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75, 7537-7541.

Schmidt, P.J., Pintar, K.D., Fazil, A.M., Flemming, C.A., Lanthier, M., Laprade, N., Sunohara, M.D., Simhon, A., Thomas, J.L., Topp, E., Wilkes, G. and Lapen, D.R. (2013) Using *Campylobacter spp.* and *Escherichia coli* data and Bayesian microbial risk assessment to examine public health risks in agricultural watersheds under tile drainage management. *Water Research* 47, 3255-3272.

Seurinck, S., Defoirdt, T., Verstraete, W. and Siciliano, S.D. (2005) Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environmental Microbiology* 7, 249-259.

Shanks, O.C., Kelty, C.A., Archibeque, S., Jenkins, M., Newton, R.J., McLellan, S.L., Huse, S.M. and Sogin, M.L. (2011) Community structures of fecal bacteria in cattle from different animal feeding operations. *Applied and Environmental Microbiology* 77, 2992-3001.

Shannon, K.E., Lee, D.Y., Trevors, J.T. and Beaudette, L.A. (2007) Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Science of the Total Environment* 382, 121-129.

Shepherd, M.L., Swecker, W.S., Jr., Jensen, R.V. and Ponder, M.A. (2012) Characterization of the fecal bacteria communities of forage-fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiology Letters* 326, 62-68.

Sidhu, J.P., Skelly, E., Hodgers, L., Ahmed, W., Li, Y. and Toze, S. (2014) Prevalence of *enterococcus* species and their virulence genes in fresh water prior to and after storm events. *Environmental Science & Technology* 48, 2979-2988.

Silkie, S.S. and Nelson, K.L. (2009) Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. *Water Research* 43, 4860-4871.

Simpson, J.M., Santo Domingo, J.W. and Reasoner, D.J. (2002) Microbial Source Tracking: State of the Science. *Environmental Science & Technology* 36, 5279-5288.

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., Newton, K., Brown, M.H. and Mitchell, J.G. (2012) Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. *Environmental Microbiology Reports* 14, 240-253.

Stackebrandt, E. and Goebel, B. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* 44, 846-849.

Staley, C., Reckhow, K.H., Lukasik, J. and Harwood, V.J. (2012) Assessment of sources of human pathogens and fecal contamination in a Florida freshwater lake. *Water Research* 46, 5799-5812.

Staley, C., Unno, T., Gould, T.J., Jarvis, B., Phillips, J., Cotner, J.B. and Sadowsky, M.J. (2013) Application of Illumina next-generation sequencing to characterize the bacterial

community of the Upper Mississippi River. *Journal of Applied Microbiology* 115, 1147-1158.

Stoeckel, D.M. and Harwood, V.J. (2007) Performance, design, and analysis in microbial source tracking studies. *Applied and Environmental Microbiology* 73, 2405-2415.

Straub, T.M. and Chandler, D.P. (2003) Towards a unified system for detecting waterborne pathogens. *Journal of Microbiological Methods* 53, 185-197.

Taylor, L.H., Latham, S.M. and Mark, E. (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 356, 983-989.

Tomaras, J., Sahl, J.W., Siegrist, R.L. and Spear, J.R. (2009) Microbial diversity of septic tank effluent and a soil biomat. *Applied and Environmental Microbiology* 75, 3348-3351.

Unno, T., Jang, J., Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.-S., Chun, J. and Hur, H.-G. (2010) Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environmental Science & Technology* 44, 7777-7782.

Unno, T., Kim, J., Kim, Y., Nguyen, S.G., Guevarra, R.B., Kim, G.P., Lee, J.H. and Sadowsky, M.J. (2015) Influence of seawater intrusion on microbial communities in groundwater. *Science of the Total Environment* 532, 337-343.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* 30, 418-426.

Vandewalle, J.L., Goetz, G.W., Huse, S.M., Morrison, H.G., Sogin, M.L., Hoffmann, R.G., Yan, K. and McLellan, S.L. (2012) *Acinetobacter*, *Aeromonas* and *Trichococcus*

populations dominate the microbial community within urban sewer infrastructure. *Environmental Microbiology* 14, 2538-2552.

Vierheilig, J., Farnleitner, A.H., Kollanur, D., Bloschl, G. and Reischer, G.H. (2012) High abundance of genetic Bacteroidetes markers for total fecal pollution in pristine alpine soils suggests lack in specificity for feces. *Journal of Microbiological Methods* 88, 433-435.

Vorosmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. and Davies, P.M. (2010) Global threats to human water security and river biodiversity. *Nature* 467, 555-561.

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M. and Whittam, T.S. (2009) Cryptic lineages of the genus *Escherichia*. *Applied and Environmental Microbiology* 75, 6534-6544.

Wei, S., Morrison, M. and Yu, Z. (2013) Bacterial census of poultry intestinal microbiome. *Poultry Science* 92, 671-683.

Wilkes, G., Edge, T., Gannon, V., Jokinen, C., Lyautey, E., Medeiros, D., Neumann, N., Ruecker, N., Topp, E. and Lapen, D.R. (2009) Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Research* 43, 2209-2223.

Wilson, K. (1987) Preparation of genomic DNA from bacteria. *Current Protocols in Molecular Biology*, 2.4. 1-2.4. 5.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences* 74, 5088-5090.

Wommack, K.E., Bhavsar, J. and Ravel, J. (2008) Metagenomics: read length matters. *Applied and Environmental Microbiology* 74, 1453-1463.

Wong, S.Y., Paschos, A., Gupta, R.S. and Schellhorn, H.E. (2014) Insertion/deletion-based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments. *Environmental Science & Technology* 48, 11462-11470.