

Comprehensive Evaluation of *Bacteroidales* for Identification of Fecal Contamination
Sources in Freshwater

by

Rasanthi Udenika Wijesinghe

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 14, 2013

Key words: Microbial source tracking, water quality, qPCR
Bacteroidales, *E. coli*, DNA fingerprinting, PMA

Copyright 2013 by Rasanthi Udenika Wijesinghe

Approved by

Yucheng Feng, Chair, Professor of Soil Microbiology
Frank Owsley, Associate Professor of Animal Science
Scott Santos, Associate Professor of Biological Sciences
Ahjeong Son, Associate Professor of Environmental Engineering
C. Wesley Wood, Professor of Environmental Soil Science

Abstract

Accurate identification of fecal contamination sources in an impaired watershed is crucial for developing best management practices. Microbial source tracking (MST) is commonly used to identify fecal pollution sources and includes both library-based and non-library-based approaches. The library-based approach often involves the development of a known source library from DNA fingerprints obtained from fecal samples of known host groups, allowing the DNA fingerprints from water samples to be compared against those in the known source library. The non-library based approach involves the direct detection of host-associated markers present in the water samples. In the first study, the library-based *E. coli* rep-PCR DNA fingerprinting method and *Bacteroidales* host-associated markers were used to identify fecal contamination sources in a mixed land-use watershed. DNA extracted from 64 water samples was analyzed using end-point and quantitative PCR. A total of 1,050 *E. coli* rep-PCR DNA fingerprints obtained from water samples were then compared against the known source library, which consisted of 945 unique *E. coli* DNA fingerprints from nine host groups. All of the water samples were positive for both general *Bacteroidales* markers and *E. coli*. The rep-PCR method detected human and cattle contamination in 94% and 75% of the water samples, respectively, while end-point PCR found human and cattle markers only in 27% and 28% of the samples. qPCR, on the other hand, detected human-associated markers in 64.0% of the samples.

The second study evaluated human- and cattle-associated *Bacteroidales* genetic markers for their applicability in Alabama and identified the most suitable primer sets in qPCR assays for assessing fecal contamination in environmental samples. Four human- and seven cattle-associated genetic markers were tested and human-associated HF183 and cattle-associated CowM3 appeared to be the best human and cattle markers, respectively. DNA extracted from surface water samples was amplified with general *Bacteroidales* primers as well as human- and cattle-associated primers. The results indicated that general *Bacteroidales* genetic markers were positive for all samples, with the highest concentration being 1,180,500 gene copies/100 ml. Human-associated *Bacteroidales* markers were detected in 87% of the water samples, while only 8% of the water samples contained the cattle-associated *Bacteroidales* markers.

The major disadvantage of PCR is its inability to discriminate between DNA from live and dead cells. Propidium monoazide (PMA) is a DNA intercalating agent that can be used to detect DNA from live cells. The third study focused on the survival of *E. coli* and live *Bacteroidales* in stream water and sediment microcosms. The general *Bacteroidales* markers were detectable up to 7 and 9 days in stream water and sediment microcosms, respectively, but human markers were detected only in the first 3 days in both microcosms. During the study period, a 3-log reduction of *E. coli* was observed, with 2,500 CFU/100 ml remaining in the water microcosm at the end of 14 days. However, *E. coli* survived in the sediment for more than 75 days. Experiments with PMA revealed that about 50% of the *Bacteroidales* gene copies amplified by qPCR were actually from dead cells or extracellular DNA.

In summary, although both library-based and non-library-based MST methods can be used to detect the sources of fecal contamination in an impaired watershed, amplification of host associated *Bacteroidales* genetic markers with PMA-qPCR provides information on recent fecal pollution with a shorter turn-around time. More research should be devoted to developing viable cell-based approaches.

Acknowledgments

The author would like to thank her major Professor, Dr. Yucheng Feng, for her helpful direction and encouragement in completing this program; this work could never possibly be done without her technical support and valuable advices. She would also thank her Committee members Dr. Wes. Wood, Dr. Scott Santos, Dr. Ahjeong Son, and Dr. Frank Owsley for their technical support and advices have given throughout the study period. She is also grateful to Michele Owsley, Brenda Wood, Robin Gaverno, John Owen and lab colleagues for their generous assistance at times. Financial support of the research was provided by the Montgomery County Soil and Water Conservation District, Alabama Agricultural Experiment Station and Alabama Water Resource Research Institute. Finally, she would like to express deep appreciation to her husband, two daughters and her family for their encouragement, support and love throughout this study.

Table of Contents

Abstract	ii
Acknowledgments	v
List of Tables	viii
List of Illustrations	x
1. Literature Review	
1.1. Introduction	1
1.2. Objectives	7
1.3. Indicator Bacteria	7
1.4. Approaches to detect sources of fecal contamination	21
1.5. Polymerase chain reaction	23
1.6. Survival of bacteria in the secondary habitat	31
1.7. Current prospect and future of microbial source tracking	39
1.8. References	58
2. A comparison of library dependent and library independent methods to identify the sources of fecal contamination	
Abstract	77
2.1. Introduction	78
2.2. Methods and materials	81
2.3. Results	90
2.4. Discussion	95

2.5. Summary	100
2.6. References	103
3. Detection and quantification of <i>Bacteroidales</i> human and cattle associated genetic markers in surface water	
Abstract	126
3.1. Introduction	127
3.2. Methods and materials	129
3.3. Results	135
3.4. Discussion	140
3.5. Summary	150
3.6. References	152
4. Differentiation of <i>Bacteroidales</i> 16S rRNA genetic markers from live and dead cells and determination of their persistence in the secondary habitat	
Abstract	169
4.1. Introduction	170
4.2. Methods and materials	173
4.3. Results	180
4.4. Discussion	186
4.5. Summary	192
4.6. References	193
5. Summary and future direction	213
6. Appendices	216

List of Tables

Table 2.1 Source groups and DNA fingerprints in the known source library.....	108
Table 2.2 <i>Bacteroidales</i> primers used for fecal sources detection and quantification.....	109
Table 2.3 Selected quality control parameters used to evaluate qPCR reaction....	110
Table 2.4 Rates of correct classification (RCC) based on 3- and 9- way separation.....	111
Table 2.5 Number of <i>E. coli</i> isolates identified at the six sampling sites by month.....	112
Table 2.6 Number of <i>E. coli</i> isolates identified at the six sampling sites.....	113
Table 2.7 <i>E. coli</i> concentration found in each sampling site.....	114
Table 2.8 <i>Bacteroidales</i> concentration found in each sampling site.....	115
Table 2.9 Human associated <i>Bacteroidales</i> detected in Catoma Creek watershed during 1-year study period.....	116
Table 2.10 <i>E. coli</i> , <i>E. coli</i> fingerprints and <i>Bacteroidales</i> gene markers that tested positive in this study.....	117
Table 3.1 Primer sets tested for specificity and sensitivity	157
Table 3.2 Specificity and sensitivity of selected primers.....	158
Table 3.3 Performance information for qPCR.....	159
Table 3.4 Number of samples tested positive for general <i>Bacteroidales</i> and host associated molecular markers and <i>E. coli</i>	160
Table 3.5 Selected nutrient parameters and turbidity monitored during the study period.....	161
Table 4.1 Primers, target genes and amplicon lengths for qPCR assays.....	198

Table 4.2 Quality control parameters for AllBac and qHF183 qPCR assays.....	199
Table 4.3. Survival of <i>Bacteroidales</i> and <i>E. coli</i> in their secondary habitat..... as a percentage of initial concentration: microcosm prepared with stream water.	200
Table 4.4. Survival of <i>Bacteroidales</i> and <i>E. coli</i> in their secondary..... Habitat as a percentage of initial concentration: microcosm prepared with sediment	201
Table 4.5 Selected chemical parameters for microcosms during the study period	202

List of Figures

Fig. 1.1 Eubacterial phylogenetic tree based on 16S rRNA gene.....	74
Fig. 1.2 Proportions of <i>Bacteroides</i> species found in clinical samples.....	75
Fig. 1.3 Principle of the PMA technology.....	76
Fig. 2.1 Sampling sites in the Catoma Creek watershed.....	118
Fig. 2.2 rep-PCR fingerprints generated using the BOX A1R primer.....	119
Fig. 2.3 qPCR amplification plot and qPCR std. curve.....	120
Fig. 2.4 MANOVA analysis for 3-way and 9-way source groups.....	121
Fig. 2.5 Sources identification based on 3-and 9-way source groups.....	122
Fig. 2.6 Source distribution at different sampling sites in the Catoma Creek..... watershed	123
Fig. 2.7 Cumulative <i>E. coli</i> concentration found in each sampling site.....	124
Fig. 2.8 Cumulative general <i>Bacteroidales</i> molecular marker concentration..... found in each sampling site	125
Fig. 3.1 Sampling sites at the Parkerson Mill Creek watershed.....	162
Fig. 3.2 General <i>Bacteroidales</i> gene copy concentrations found above LLOQ at different sampling sites during the sampling period.	163
Fig. 3.3 Human associated <i>Bacteroidales</i> 16S rRNA gene copy concentrations.... found above LLOQ in different sampling sites during sampling period	164
Fig. 3.4 Cattle associated <i>Bacteroidales</i> gene copy concentrations found at..... different sampling sites during the sampling period	165
Fig. 3.5 Five-day geometric means of <i>E. coli</i> concentrations, general..... <i>Bacteroidales</i> and human associated genetic markers found during the study period.	166
Fig. 3.6 <i>E. coli</i> concentrations found at different sampling sites during..... sampling period and SSM 235 CFU/100 ml	167

Fig. 3.7 The overall correlation between general <i>Bacteroidales</i> marker..... and <i>E. coli</i> ..	168
Fig. 4.1 Verification of the effectiveness of PMA.....	203
Fig. 4.2 Persistence and survival of general <i>Bacteroidales</i> 16S rRNA gene..... copies (AllBac) in microcosms prepared with stream water	204
Fig. 4.3 Persistence and survival of human associated <i>Bacteroidales</i> 16S rRNA.... (qHF183) gene copies in microcosm prepared with stream water	205
Fig. 4.4 The effect of PMA on amplification of general <i>Bacteroidales</i> 16S rRNA AllBac markers in sediment samples	206
Fig. 4.5 Correlation between the percentages of <i>Bacteroidales</i> live cells and..... general <i>Bacteroidales</i> 16S rRNA markers	207
Fig. 4.6. Persistence and survival of general <i>Bacteroidales</i> 16S rRNA gene..... copies in microcosms prepared with 1% sediment	208
Fig. 4.7 Persistence and survival of human associated <i>Bacteroidales</i> 16S rRNA.... gene copies in microcosms prepared with 1% sediment	209
Fig. 4.8 Survival of <i>E. coli</i> in microcosms prepared with stream water..... and 1% sediment	210
Fig. 4.9 The correlation between <i>E. coli</i> and general <i>Bacteroidales</i> 16S rRNA.... (AllBac) gene markers found in (a) stream water microcosms and (b) sediment microcosms	211
Fig. 4.10 The ability of PMA to discriminate DNA from live and dead cells in..... influent and effluent of a waste water treatment plant in Auburn, AL	212

1. Introduction and Literature Review

1.1 Introduction

Water-related recreational activities such as swimming, surfing, canoeing, kayaking, and jet skiing are popular among millions of people living in the United States, and are excellent ways to engage in physical activities as part of a healthy life style. People of all ages enjoy spending time around oceans, lakes, and rivers each year. It is, however, essential to be aware of the risk of recreational water illnesses. Gastrointestinal diseases represent the main recreational water related disease, and 157,964 river/stream miles in the US are already known to be impaired or possibly impaired due to pathogens (USEPA, 2013). Fecal pollution is the major cause of recreational water pollution, and adds significantly to the nation's economic burden due to illnesses related to swimming and bathing. Further, fecal contamination is the main reason for shutting down shellfish harvesting areas, closing beaches and imposing bathing restrictions on designated beaches. Both human and animal fecal matter are responsible for fecal pollution, and can come from either point sources such as raw sewage discharges, effluent from wastewater treatment plants and industrial activities, or non-point discharge sources such as agriculture, forestry, wildlife or urban runoff. In order to implement effective remediation practices, it is important to accurately identify the sources of fecal contamination

Microbial source tracking (MST) is a useful technique for detecting the presence of enteric organisms in their secondary habitat and differentiating the sources of these

organisms. Currently, MST is the most popular way of determining the sources of fecal contamination in surface water. MST methods can be divided into two main approaches: library dependent and library independent methods. The nucleic acid-based, library dependent approach typically involves developing a DNA fingerprint library using DNA from *E. coli* or *Enterococcus* obtained from known host groups such as humans, cattle, dogs, horses, chickens, deer, other domestic animals and wildlife. *E. coli* or *Enterococcus* DNA fingerprints obtained from surface water samples are then compared against this known source library to determine the sources of the fecal contamination in water. The size of the library is a significant factor in this process because larger libraries will provide more reliable results than smaller libraries. In addition, spatial variability and temporal variability also affect the accuracy of source identification. This is a technically demanding and time consuming approach, so most MST research is moving towards non-library based approaches such as the identification of host specific genes such as the 16S rRNA gene.

The selection of the proper organism for MST is vital and should be undertaken with great care. Several factors must be considered before selecting bacteria for monitoring fecal contamination in a watershed or a MST study, including the survival characteristic of the target organism, its regrowth in the secondary habitat, abundance in fecal matter, pathogenicity, ease of cultivation under regular laboratory conditions and correlation with pathogenic organisms in the water. The USEPA recommends using *E. coli* and *enterococci* as indicators of fecal contamination for fresh water and *enterococci* for marine water (USEPA, 2012). *E. coli* has been used to monitor fecal pollution of fresh water for more than 30 years and is considered a good indicator because it has a high

growth rate, most of its strains are non-pathogenic, it is easy to grow under normal laboratory conditions, and there is a positive correlation between gastrointestinal diseases suffered by swimmers and *E. coli* concentrations in those waters (Haile et al., 1999; Prüss, 1998; Wade et al., 2003). However, *E. coli* not only survives for longer in the secondary habitat, it can also regrow (Flint, 1987; Ishii et al., 2006; Byappanahalli et al., 2006). This makes *E. coli* a poor indicator for identifying recent fecal pollution because of its ability to survive in both stream water and in sediments. Turbulent water currents such as those commonly experienced during storms can resuspend the sediment and add settled cells into the overlaying water (Eichmiller et al., 2013). Hence, high *E. coli* numbers, especially during storm periods, do not necessarily furnish up-to-date information on recent water pollution.

Bacteroidales has been suggested as a potential alternative indicator organism. Members of the order *Bacteroidales* are Gram negative, rod shaped, bile-resistant, and non-spore forming obligate anaerobes living in the intestinal tract of warm-blooded animals (Wexler, 2007). They are abundant in feces, with about 25% to 30% of the human gut population represented by the phylum Bacteroidetes (Yang et al., 2009a). Most *Bacteroidales* strains are harmless, and only a few are opportunistic pathogens (Salyers, 1984). As members of *Bacteroidales* are obligate anaerobes, a shorter life span in the secondary habitat is expected. This is a key feature for microbial source tracking because water samples testing positive for these organisms will indicate recent water pollution. However, a study with *B. thetaiotaomicron* revealed that its genome contained gene sequences of Complex I (NADH-quinone oxidoreductase) and Complex II (succinate dehydrogenase) aerobic respiratory pathways. These aerobic respiratory

pathways suggest the capability of these organisms to survive longer in an oxygenated environment than other obligate anaerobes (Xu et al., 2003). Thus, studying the survival of *Bacteroidales* in the secondary habitat is fundamental if its use is to become widespread in the MST field.

Nucleic acid based approaches, i.e., the amplification of specific genes, have been used to determine the prevalence of *Bacteroidales* in the secondary habitat (Dick et al., 2005a; Bower et al., 2005; Shanks et al., 2008, 2009; Haugland et al., 2005; Layton et al., 2006; Kreader 1998; Bernhard and Field, 2000a,b). Several previous studies have found that general *Bacteroidales* 16S rRNA gene markers can persist in stream water for eight to 24 days, although their persistence depends mainly on the temperature and presence of predators in the water (Seurinck et al., 2005; Okabe 2007; Bell et al., 2009).

End-point PCR and quantitative PCR (qPCR) are the two main types of DNA amplification techniques used in the MST field. qPCR involves both qualitative and quantitative analyses of gene fragments, while end-point PCR provides only qualitative information. In addition, qPCR is a rapid, sensitive and efficient process where results can be obtained within a few hours of sample collection. However, the development of an accurate standard curve is vital for this technique because the calculations of gene copy numbers depend on developing a standard curve based on known concentrations of genomic or plasmid DNA. Thus, qPCR is more technically demanding and expensive than end-point PCR. The selection of either end-point PCR or qPCR for a source tracking study mainly depends on the precise research goals and the availability of funds. In addition to PCR techniques mentioned above, nested PCR and multiplex PCR can be used to improve the detection efficiency of the target. Nested PCR offers an effective

way to amplify molecular markers, which are not abundant in the environmental samples. First, the target is amplified with order specific primers. The PCR products are then amplified again with host specific primers, which are nested in the first primer set. This is a particularly effective method of detecting a smaller amount of targets present in water. Multiplex PCR involves simultaneous amplification of two or more targets in the same PCR reaction. This technique can be used to reduce cost of PCR runs and gel electrophoresis (Guan et al., 2013).

The main drawback of these DNA amplification techniques is their inability to distinguish between the DNA associated with live cells and dead cells. Most reported MST studies have not differentiated *Bacteroidales* DNA from live or dead cells. Since DNA from dead cells can persist for a long time in the extracellular environment, this ability to differentiate between DNA from live and dead cells is essential. Even after cell death has occurred, DNA can persist for as long as two weeks in stream water or more than 16 weeks under sterile conditions at 4°C (Josephson et al., 1993). Exacerbating the problem, DNA from both live and dead cells can be amplified during PCR, which leads to overestimation of specific bacterial populations in ecosystems. As with bacteria, DNA has a tendency to settle with sediments, persists for long periods of time, and enters the stream water during a storm event due to the resuspension of sediments. Therefore, PCR may tend to overestimate *Bacteroidales* populations in water.

Techniques used to discriminate between live and dead cells include culturable cell counts, fluorescent dye assays combined with flow cytometry, fluorescence microscopy and quantitative assays using a fluorescence micro plate reader, (Nebe-von-Caron et al., 2000; Kramer et al., 2009). Most of these techniques suffer from limitations

that prevent them being used simply to separate live cell DNA from dead cell DNA. The quantification of mRNA is a precise way to determine the number of live cells in environmental water (Walters and Field, 2009), but mRNA has a short half-life and is unstable in the environment, making it a technically demanding and expensive approach (Josephson et al., 1993). Propidium monoazide (PMA) and ethidium monoazide (EMA) are two dyes that can be used to detect live cells from dead cells. Both are DNA intercalating agents: the azide group in each intercalates with DNA by producing strong covalent bonds in the presence of bright visible light and subsequently inhibits PCR amplification. The use of EMA/PMA to detect live pathogens in clinical samples is commonly performed when conducting qPCR analyses (Kobayashi et al., 2009; Kramer, 2009). However, this technique has not yet been widely utilized for environmental samples, perhaps due to a problem with insufficient light penetration through the particulate matter and high levels of suspended solids present in surface water samples, which may interfere with the photo-induced cross linking of PMA/EMA to DNA (Varma et al., 2009; Wagner et al., 2008). Bae and Wuertz (2009) successfully applied this method to quantify the live and dead *Bacteroidales* present in the effluent and influent of a sewage treatment plant. PMA has also been used effectively to differentiate live *Enterococcus* and *Bacteroidales* from dead cells in wastewater samples collected from a public wastewater treatment plant (Varma et al., 2009). The overall goal of the work reported here was thus to evaluate the use of *Bacteroidales* to monitor fecal contamination and identify the sources of fecal contamination in surface waters.

1.2 Objectives

1. Compare the use of library dependent and library independent methods for identifying the sources of fecal contamination;
2. Detect and quantify human and cattle associated *Bacteroidales* genetic markers in surface water; and
3. Differentiate *Bacteroidales* 16S rRNA genetic markers from live and dead cells and determine their persistence in the secondary habitat.

1.3 Indicator Bacteria

A. Indicator bacteria for identification of fecal contamination in surface water

Waterborne pathogenic organisms are responsible for gastrointestinal diseases for humans, and sometimes these diseases can be fatal. About 4 billion cases of diarrhoea occur each year around the world, and that leads to death of about 2.2 million people (WHO, 2013). Enumeration of each pathogen to identify their presence in a water body is unrealistic; therefore, the use of indicator organisms to detect those pathogens in water is a reasonable approach to address this problem. Indicator bacteria are a group of organisms that live in the intestinal track of warm-blooded animals and enter the secondary habitat with feces. Their presence in surface water indicates possible contamination with feces. Selection of proper indicator bacteria is crucial, and there are several aspects to be considered. These indicator organisms should: 1) be present in proportion to fecal contamination, 2) show the same performance in all water types, 3) be safe to work with in the laboratory, 4) survive in the environment as long as pathogens live, 5) not replicate in the water or sediments, and 6) not show false positive or false negatives results. In the United States, total coliform (TC), fecal coliform (FC), *E. coli*,

and *Enterococcus* are commonly used indicator organisms. Recently, more attention has been focused on anaerobic fecal indicator bacteria such as *Bacteroidales* (WHO, 2001).

Use of indicator organisms to identify the fecal pollution in water has a long history. In 1885, a biologist named Theodor Escherich first found a group of microorganisms which lived in both water and sewage. That group was named *Bacillus coli* (renamed as *Escherichia coli* in 1919) (Brenner et al., 2005). By 1891, the concept that water pollution due to fecal matter is dangerous to human health was initiated (WHO, 2001). In 1914, for the first time, Public Health Service Drinking Water Standards were adapted in the United States. *E. coli* was used as the indicator organism, and they were tested with the Multiple-Tube Fermentation Test method, which is now referred to as the Most Probable Number (MPN) procedure. This was an easy method to perform, but it required 48 hours to obtain the results. Another drawback was the necessity for a series of identification tests to confirm *E. coli*. Several years later, the total coliform (TC) group was used as an indicator instead of *E. coli*. This group was used as indicator for more than 40 years until apprehended TC was deemed not reliable organism of the high level of false positive results. In the 1960s, the fecal coliform (FC) group was introduced as an indicator. FC was as an indicator until the 1980s; however, it also indicated a high level of false positive results. For the second time, *E. coli* was selected as fecal indicator bacteria for fresh surface water about 90 years after its first selection. Simultaneously, *Enterococcus* species were selected to monitor fecal contamination in water (USEPA, 1986). Now, enzyme and biochemical based tests such as modified m-TEC media and Colilert-18 methods are available for easier and faster identification of *E. coli* within 18-24 hours (WHO, 2001). Until the 1990s biologists assumed that *E. coli*

could not re-grow in secondary habitats. Ishii et al. (2006) reported that *E. coli* had the ability to grow in the non-sterile, non-amended soils. This was the first report of *E. coli* re-growth in secondary habitats. Later, this finding was confirmed by other researchers (Brennan et al., 2010; Byappanahalli et al., 2006). Thus, lack of fecal specificity raises the question about the credibility of *E. coli* as an indicator. Nevertheless, the USEPA still recommends the use of *E. coli* as the indicator organism to identify the fecal contamination in recreational water (USEPA, 2012).

B. Major types of indicator bacteria and their role in water quality assessment

B.1 Total coliforms (TC): The total coliform group consists of 19 genera (including *Escherichia*, *Klebsiella*, *Enterobacter* and *Citrobacter*) and 80 species. They are aerobes or facultative anaerobes, Gram negative, rod shaped, non-spore forming bacteria living mainly in the intestinal track of warm blooded animals. They can also reside in soil and surface water and on leafy plants. Some species have non-fecal origins and can re-grow under favorable conditions. TC bacteria can ferment lactose and produce gas and acid at 35°C in 24-48 hours. The most probable number method (MPN) and membrane filtration method are currently used standard methods to quantify TC in water samples (USEPA 2002a; WHO, 2001). TC can grow on enrichment media (mEndo) at 35°C and show gold-green sheen within 24 hours (Oram, 2013). Currently, commercial kits are available to quickly and efficiently detect TC, such as the Colilert-18 method (IDEXX Laboratories, Inc.). In this method, *E. coli* and total coliforms can be identified by using nutrient indicators known as 4-methyl-umbelliferyl β -D glucuronide (MUG) and O-nitrophenyl β -D galactopyranoside (ONPG). Coliforms use their β -galactosidase enzyme to metabolize ONPG and change it from colourless to yellow (www.IDEXX.com).

B.2 Fecal coliforms (FC). Fecal coliforms are a sub group of total coliforms; they include the genera of *Escherichia*, *Klebsiella* and *Citrobactor*. FC are Gram negative, non-spore forming rod shaped bacteria which produce both acid and gas by fermenting lactose at 44.5°C within 48 hours (Reynold, 2003). Since FC have the ability to live at high temperatures, they are known as thermotolerant coliforms. Some of the members in this group have environmental origins, and some have the ability to re-grow in secondary habitats. Their presence in water indicates possible occurrence of pathogenic bacteria in that water such as *Shigella*, *Cholera*, *Salmonella* and viruses including hepatitis A and Norwalk group viruses in that water (Reynolds et al., 2003, Simpson et al, 2002). Even though USEPA recommends *E. coli* as the standard indicator bacteria to identify fecal contamination in recreational waters, some states are still using FC as the indicator bacteria in recreational water. The most probable number method (MPN) and membrane filtration method are the currently used standard methods to enumerate FC in water samples (USEPA, 2002c; WHO, 2001). FC has the ability to ferment lactose and produce gas at 44.5°C within 24 hours. Thus, gas production in EC medium indicates the presence of fecal coliforms in water samples. In the membrane filtration, m-FC is used as the medium, which contains aniline blue as an indicator. FC incubated at 44.5°C for 22-24 hours form colonies in a range of blue color on this medium. The blue color shows FC's ability to ferment lactose to acid and their presence in water (Geldreich et al., 1965).

B.3 *Escherichia coli*: *E. coli* is a Gram negative, rod shaped, facultative anaerobic bacterium living in the intestinal track of warm blooded animals. About 0.1% of the total bacteria in an adult gut microbiome is represented by *E. coli*. It belongs to the family *Enterobacteriaceae*, which contains human pathogens such as *Salmonella*, *Shigella* and

Yersinia. E. coli can exist in nature as commensal strains as well as pathogenic strains, e.g., Serotype O157:H7. Based on genomic information, *E. coli* can be divided into four main phylogenetic groups: A, B₁, B₂ and D (Carlos et al., 2010). Virulence genes were more ubiquitous in groups B₂ and D. Group B₁ isolates are primarily from nonhuman sources and Group B₂ isolates primarily from human sources. Group B₂ isolates are genetically distant and show adaptation to the host. Isolates from B₁ are more genetically similar and are not well adapted to any particular host. Therefore, environmental survival is significantly higher in B₁ isolates than B₂ isolates (White et al., 2011). This study further revealed that *E. coli* host specificity is not defined by genome wide sequence changes, but the presence or absence of specific genes and associated promoter elements are mainly responsible for this.

E. coli has a biphasic lifestyle as host independent and host associated phases. Therefore, *E. coli* should have different survival strategies and may have evolved gradually. Stomach of humans and other animals has pH as low as 2.5. Microorganisms enter the stomach with ingested food, and this low pH creates a severe stress to them. Most microorganisms cannot tolerate this low pH and die out. Only surviving members enter the intestinal tract that provides a comfortable environment for them to survive, persist and grow. Once they enter the open environment, they again have to face harsh conditions such as low nutrient level, UV radiation, extreme temperature, etc. Growth and survival of *E. coli* in the secondary habitat is mainly restricted by availability of nutrients and energy sources. Most of the *E. coli* may have conserved key evolutionary adaptations in their genomes (Van Elsas et al., 2011).

E. coli has a rapid growth rate and is easy to culture. USEPA conducted a series of studies to identify the relationship between the presence of indicator organisms in the water and the incidence of intestinal illness of swimmers in those waters (USEPA, 1986). In fresh water, there is a high correlation between *E. coli* and illnesses. Therefore, USEPA recommended *E. coli* as an indicator organism for fresh surface water (USEPA, 1986). In 1986, USEPA established the recreational water quality criteria (RWQC) where the 5-day geometric mean for full body contact activities should be less than 126 CFU/100 mL for 30 days, and the single sample maximum (SSM) should be less than 235 CFU/100 mL. In 2012, these RWQC was revised, but the geometric mean value was still valid. The SSM criterion was removed and a new criterion was introduced, which is known as statistical threshold value (STV): 410 CFU/100 mL. According to this new criterion, a maximum of 10% of the samples can violate STV within a 30-day period (USEPA, 2012).

Currently available methods to enumerate *E. coli* in water can be divided into two groups: the membrane filtration and the most probable number (MPN) methods. Modified m-TEC method is a filtration technique introduced by USEPA in 2002b. This is a fast (24-hour) and easy (one-step) method to enumerate *E. coli* in fresh water. The modified m-TEC medium contains a chromogen, 5-bromo-6-chloro-3 indoly- β -D-glucuronide. *E. coli* produces an enzyme known as β -D-glucuronidase, which can catabolize glucuronic acid and develop a red or magenta colored compound within 24 hours at 44.5°C. This is a characteristic test to identify *E. coli* (USEPA, 2002b). The following chemical tests need to be used to further verify those colonies as *E. coli*: Simmon citrate agar (Difco), 1% tryptone (Difco), EC broth (Difco) and methyl red-

Voges-Proskauer (Difco) reagents. *E. coli* does not utilize citrate as a substrate. They grow at 44.5°C, produce gas and florescence under UV light in the EC-MUG, produce indole with tryptone and produce acid end product when growing in methyl red-Voges-Proskauer broth (USEPA, 2002b). Traditionally, multiple tubes were used for the most probable number procedure. However, commercial kits known as Colilert-18 method (IDEXX Laboratories, Inc.) are now available. In this method, *E. coli* can be identified by using a nutrient indicator known as 4 methyl–umbelliferyl β -D glucuronide (MUG). *E. coli* uses β -glucuronidase to metabolize MUG and produces a yellow compound which fluoresces under UV light (IDEXX.com).

B.4 *Enterococcus*: *Enterococcus* species are non-motile, Gram-positive bacteria living mainly in the intestinal track of mammals, and also in soil, food, water, plants, birds and insects. This group was earlier categorized under the genus *Streptococcus* but later recognized as an independent genus. They can grow at pH 9.6, temperatures between 10°C and 45°C and in 6.5% NaCl. Hence, they survive longer in the marine environment and water treatment. Both membrane filtration and MPN methods are used to enumerate *Enterococcus*. The modified *Enterococcus* procedure was introduced by USEPA, which is fast (24-hour) and easy (one-step) to enumerate *Enterococcus* in fresh and marine waters (USEPA, 2002c). Colonies with a blue halo on the membrane on modified mEI Agar are recognized as *Enterococcus*. Traditionally, the most probable number method was performed using multiple tubes. However, now a commercial kit is available and that is known as Enterolert-E method (IDEXX Laboratories, Inc.). Enterolert-E uses nutrient-indicator (Defined Substrate Technology–DST) to detect

Enterococcus. When *Enterococcus* metabolizes this nutrient-indicator, the resulting product fluoresces under UV light (IDEXX.com).

C. Potential indicator organisms

C.1 Human enteric viruses. This group includes the Norwalk virus, rotavirus, hepatitis A virus, enterovirus, Adenovirus and others. These ubiquitous viruses have a longer than normal survival period (2 to 130 days) in secondary habitats, and are resistant to waste water treatment processes such as chlorination and UV radiation (Jiang et al., 2000). Other indicator viruses are bacteriophages, such as F-specific RNA coliphages, which infect *E. coli*. Coliphage Serotypes II and III are associated with humans and serotypes I and IV are generally associated with animals (Scott et al., 2002; USEPA, 2005). A study with human specific Adenovirus and coliphage in coastal water found a strong correlation ($r=0.99$) between these two groups. However, this study failed to show a good correlation between Adenovirus and fecal indicators (Jiang et al., 2000). F-RNA bacteriophage was used as a fecal indicator by Gourmelon et al.,(2007), where 21% of individual human fecal samples, 60% of pig slurries and all sewage samples contained F RNA bacteriophages. Also, 82% of the water samples contained these bacteriophages. Use of viruses as indicators is costly and time consuming; therefore, this method is not popular among researchers.

C.2 *Bifidobacterium*. *Bifidobacterium* species are obligate anaerobes, and they mainly living in in human intestines. Since *Bifidobacterium* species are rarely found in other animals, they are considered good indicators to detect human fecal pollution in surface water. The main disadvantage is that there is still no clear evidence of how long *Bifidobacterium* can survival in secondary environments (Scott et al., 2002).

C.3 *Bacteroidales*. Members of the Order *Bacteroidales* are obligate anaerobic, Gram-negative, non-motile, rod shaped, non-spore forming bacteria living in the intestinal tracts of mammals. These bacteria are abundant in feces, representing more than 25-30% of the gut population (Salyers, 1984). *Bacteroidales* are a good indicator of recent fecal contamination of water because they are obligate anaerobes and cannot survive long in secondary habitats. Unlike other indicator organisms, which are facultative anaerobic bacteria, *Bacteroidales* cannot be easily grown in the laboratory. Therefore, non-culturable methods such as identification of specific genes can be used to determine their presence in water.

C.4 The Genus *Bacteroides*

C.4.1. Classification. *Bacteroides* species belong to Domain Bacteria, Phylum Bacteroidetes, Class Bacteroidia, Order *Bacteroidales*, Family *Bacteroidaceae* and Genus *Bacteroides* (NCBI.gov). Figure 1.1 illustrates the relationship of Phylum Bacteroidetes with other phyla such as Chloroflexi, Gemmatimonadetes and Firmicutes. This tree was developed based on 16S rRNA gene sequences and according to this tree, Phylum Bacteroidetes diverged in the early stages of the evolutionary process; therefore, they are not closely related to other Gram-negative bacteria such as *Escherichia coli* (Woese, 1987; Pace, 2009).

C.4.2. Differentiation of the genus *Bacteroides* from other closely related genera.

Earlier, *Prevotella* and *Bacteroides* species were categorized under the same genus, but they are phenotypically and genotypically heterogeneous. Hence, they have been taxonomically separated into two genera *Bacteroides* and *Prevotella*. The basic test used to distinguish these two groups is 20% bile test. *Bacteroides* can grow in the broth

containing 20% bile or bile esculin agar whereas *Prevotella* is sensitive to 20% bile or bile esculin agar. Some phenotypic analyses, such as the antibiotic resistance, indole, and sugar fermentation tests are used to differentiate *Bacteroides* from their closely related genera (Krieg et al., 2011).

C.4.3. Nutrition and metabolism. *Bacteroides* species utilize simple sugars as their energy sources, but the human colon lacks simple sugars due to absorbance by the small intestine. In the absence of sugar, they utilize polysaccharides as their energy source. All species of *Bacteroides* have the ability to degrade cellulose to produce simple sugars. Therefore, *Bacteroides* are responsible for the most of polysaccharide digestion in the colon. *Bacteroides* ferment glucose via the Embden-Meyerhof pathway (Salzer, 1984). Whole genome sequencing studies have found other alternative metabolic pathways *Bacteroides*. Six main enzymes are responsible for breaking down of polysaccharides: α -glucosidase; β -glucosidase, α -galactosidase, β -galactosidase, β -N-acetyl glucosamididase, and α -L-fucosidase. Some studies have suggested that *Bacteroides* are not able to use protein as a sole source of energy (Salzers, 1984).

C. 4.4. Natural habitat and clinical significance. *Bacteroides* are abundant in the human gut; the phylum Bacteroidete represent about 25-30% of the human colon bacterial population. *B. vulgatus*, *B. distasonis* and *B. thetaiotaomicron* are commonly present in the human microbiome; *B. fragilis*, *B. ovatus*, *B. eggerthii*, and *B. uniformis* are found at lower concentrations (Salzers, 1984). Fig. 2 shows the proportions of *Bacteroides* species found in clinical samples. According to this study, *B. fragilis* is the dominant species found in clinical samples, followed by *B. thetaiotaomicron* and *B. distasonis* (Wexler, 2007). These organisms play an important role in the human gut by

breaking down polysaccharides into simple sugars that the host can absorb. Some of the species are opportunistic pathogens, such as *B. fragilis*, *B. distasonis*, *B. thetaiotaomicron*, and *B. ovatus*. They cause blood stream infections and abscesses of the brain, lung, or abdominal cavity (Xu et al., 2003; Salyers, 1984). They have a high growth rate compared with most other organisms in the colon. The generation time when grown in the presence of monosaccharide is 35 hours; however, their growth rate may be slower in the colon due to competition with other organisms (Salyers, 1984).

C.4.5. Antibiotic and drug resistance. *Bacteroides* produce antibacterial compounds Bacteriocins, which inhibit their competitors. For example, *B. fragilis* produces Bacteriocins which can inhibit the RNA polymerase of another strain of *B. fragilis*. This mechanism may help reduce the competition between strains sharing the same ecological space. According to clinical studies, intestinal species secrete higher levels of this protein than non-intestinal isolates. *Bacteroides* are resistant to bile in the intestine and play a key role in recycling and absorbing bile to the host (Wexler, 2007). They show very good resistance to antibiotics compared with other anaerobes. TetQ (resistance to tetracycline) and ErmF (resistance to erythromycin) are some of the most common antibiotic resistant genes in *Bacteroides* (Xu et al., 2003). Several studies found that genes resistant to tetracycline, clindamycin and erythromycin are carried on plasmids, and can transfer from one strain to another (Tally et al., 1979). *B. fragilis* and *B. thetaiotaomicron* could transfer antibiotic (ampicillin, amoxicillin, cephalothin, tetracycline, minocycline, and chloramphenicol) resistant genes to *E. coli* strain K12 (Mancini and Behme 1977). They also contain 60 proteins that are related to drug efflux

systems; thus, their active transport system plays a major role in antibiotic resistance (Xu et al., 2003).

C.4.6. Cell wall and cell membrane. *Bacteroides* are Gram-negative bacteria. The Gram-negative cell wall consists of a thin, peptidoglycan layer, about 5-10 nm thick, adjacent to the cytoplasmic membrane. In addition to the peptidoglycan layer, the Gram-negative cell wall contains an outer membrane, a lipid-protein bilayer, which possesses proteins, phospholipids, and lipopolysaccharides. This outer membrane separates the external environment from the periplasm. The outer membrane is porous and allows exchanges of amino acids, sugars, iron etc. There are several outer membrane proteins involved in the specific uptake of metabolites (maltose, vitamin B12, nucleosides) and iron from the medium. Thus, the outer membranes of the Gram-negative bacteria provide a selective barrier to external molecules and also prevent the loss of metabolite-binding proteins and hydrolytic enzymes (nucleases, alkaline phosphatase) from the periplasm. The outer membrane contains vesicles, some related to virulence which excretes toxins, and some are used to attach to host cells. The cell wall contains various types of complex lipopolysaccharides (LPS), which are often used interchangeably with endotoxins. Endotoxins play an important role in the pathogenesis of many Gram-negative bacterial infections. They use pili and fimbriae for adhesion to the host cells (Beveridge, 1999; Wexler, 2007). Unlike other bacteria, the *Bacteroides* cell wall contains sphingolipids. These lipids are common in mammal cells but not in bacterial cells. A clinical study found that *Bacteroides* sphingolipids may contribute to inhibition of neutrophil function in mammal cells (Kato et al., 2002).

C.4.7. Mobile elements. Mobile elements found in *Bacteroides* are plasmids, transposons, and conjugative transposons. These mobile elements are important in exchanging genetic elements such as antibiotic resistant genes. Most of the strains of *Bacteroides* contain one or two plasmids varying in size from 3-7 Kb, and most of these plasmids contain antibiotic resistance genes. During cell division, plasmids can either integrate into the chromosome or replicate independently. Transposons are conjugative and mobilizable but cannot replicate independently; during cell division they integrate into the chromosomal DNA and copy along with the genome. Conjugative transposons are ubiquitous among *Bacteroides* and are largely responsible for resistance to tetracycline and erythromycin (Wexler, 2007).

C.4.8. *Bacteroides* genome. *Bacteroides* cells contain one circular chromosomal genome; some strains possess one or two plasmids. Chromosomal genome size is around 5 Mb while plasmid size varies from 3-7 kb (Salyers, 1984). Xu et al. (2003) sequenced the *B. thetaiotaomicron* genome and found that the chromosomal genome was 6.26 Mb while that of plasmid genome was 33 kb. *B. thetaiotaomicron* contained 4779 protein coding genes on the chromosomal genome and 38 on the plasmid. In addition, the chromosomal genome contained sequences for 71 tRNA and 5 rRNA operons. Gram-negative bacteria contained four types of protein secretion systems namely Type I, Type II, Type III and Type IV. *B. thetaiotaomicron* possesses Type I, II, and IV systems and lacks the Type III system, which is a characteristic of virulent bacteria (Xu et al., 2003).

Bacteroides have a greater ability to survive in aerated environments than other obligate anaerobes. The *B. thetaiotaomicron* genome encodes for three complete glycolytic pathways: Embden-Meyerho-Parnas, Entner-Doudoroff, and Oxidative

Pentose Phosphate. Also the genome contains sequences of Complex I (NADH-quinone oxidoreductase) and Complex II (succinate dehydrogenase) of the aerobic respiratory pathway. Genes for these aerobic respiratory pathways allows *Bacteroides* to survive in the oxygenated environment for a few days (Xu et al., 2003).

When compared against the GeneBank database, *B. thetaiotaomicron* has the greatest similarity to the genome of *Chlorobium tepidum*, free living photosynthetic green-sulfur bacteria. The next closest similarity is with *Clostridium perfringens*, a Gram-positive normal gut bacterium. This finding contradicts with 16S rRNA results which suggest *Bacteroides* and *Clostridium* are not related (Xu et al., 2003).

The *Bacteroides fragilis* genome was sequenced in 2005 and its chromosomal genome was found to be 5.2 Mb, coding for 4274 genes. The GC content is 43%, and protein coding density is 88.1%. *Bacteroides fragilis* has 19 rRNA operons and 73 tRNA operons. The size of the plasmid was 36.5 Kb with 32% GC content (Cerdeño-Tárraga et al., 2005).

Results of whole genome sequencing are readily used in the medical field. However, the information is not widely used in the other fields such as environmental microbiology. The 16S rRNA is one of the most conserved genes with a few variable regions in the genome. *Bacteroides* 16S rRNA sequence is about 1534 bp in length, and it has variable regions between 32 bp to 708 bp. Primers have been developed to target these upstream and downstream primer regions; amplification of these regions give the information of total *Bacteroidales* present in water samples. Some host-specific sequences are within these variable regions. Therefore, amplification of the regions between host specific primer sites gives information about the sources of fecal

contamination in a particular water-body. Currently, primers are available to identify sources such as humans, cattle, swine, dogs and horses. There is significant interest in using *Bacteroidales* as fecal indicator bacteria to identify sources of fecal pollution, because of their lack of ability to grow in the secondary habitat, and their dominant presence in feces. However, until completing whole genome sequencing, we assumed that *Bacteroidales* contained about five 16S rRNA operons per genome (Bernhard and Field, 2000a). Sequencing of whole genomes revealed that number of 16S rRNA genes present in an organism may vary from species to species. Therefore, interpretation of *Bacteroidales* data, especially comparing these gene copy numbers with *E. coli* CFU in water samples, should be done with caution.

1.4 Approaches to detect sources of fecal contamination

Fecal source tracking approaches can be divided into two main categories: chemical and biological methods.

A. Chemical methods. Recognition of human-associated chemicals in water is a straight forward approach to detect human signatures in a particular water body. The most frequently detected compounds in surface water are caprostanol (fecal steroids), cholesterol (plant and animal steroids), insect repellants, caffeine, antibacterial disinfectants, fire retardants and nonionic detergent metabolites. All these chemical compounds indicate the influence of human activities in the water (Koplin et al., 2002).

B. Biological methods. Biological methods or microbial source tracking techniques (MST) involve detecting enteric organisms present in their secondary habitat and differentiating the sources of these organisms. Various genotypic and phenotypic methods have been used for source tracking purposes. MST approaches can be divided

into two categories: library based and non-library based techniques. Each category is further divided into two groups: genotypic and phenotypic methods (Sargeant et al, 2011).

B.1 Library based phenotypic methods. A library can be developed with the phenotypic characteristics of indicator bacteria; this is particularly useful for differentiation between human and non-human sources. Common methods include antibiotic resistant analysis (ARA) (Harwood et al., 2000; Wiggins et al., 2003) and carbon source utilization pattern (CUP) analysis (Souza et al., 1999; Hagedorn et al., 2003).

B.2 Library based genotypic methods. A library is developed using DNA fingerprints of a target organism isolated from the known host groups. *E. coli* and enterococci are often the target organism used in these methods. The DNA fingerprints developed from the environmental samples are compared with the known source library to identify the sources of fecal contamination, i.e. repetitive sequence-based PCR (rep-PCR) (Dombek et al., 2000; Wijesinghe et al. 2009; Murugan et al., 2013), ribotyping (USEPA, 2005; Parveen et al., 1999; Carson et al., 2001,2005), pulse field gel electrophoresis (PFGE) (USEPA 2005; Stiles, 2003; Parveen et al., 2001; McLellan et al., 2001), random amplified polymorphic DNA (RAPD) analysis, and amplified fragment length polymorphism (AFLP) analysis.

B.3 Non-library based phenotypic methods. Host specific phenotypic characteristics of microorganisms are used for identification processes, i.e., fecal coliform/streptococci (FC/FS) ratio (Geldreich, 1969). Other examples include host

specific indicator organisms, such as viruses, F⁺ coliphage serotyping and enterotoxin biomarkers.

B.4 Non-library based genotypic methods. Host-specific molecular markers such as 16S rRNA genes, mitochondrial genes and some other functional genes are used in the non-library based genotypic methods. These molecular markers have host specificity, which permits the detection of fecal contamination sources in water. This approach is mainly used to detect obligate anaerobes, such as *Bacteroides-Prevotella* group and the genus *Bifidobacterium* in water samples. These bacteria have a host specificity and low survival rate, and thus, serve as indicators of recent fecal pollution (USEPA, 2005; Scott et al., 2002).

There is no perfect method to detect the sources of fecal contamination in a watershed. Each method has advantages as well as disadvantages. Selection of the appropriate method to detect the sources of fecal contamination should be based on research goals, time frame and the funds availability. All these genotypic methods involve amplification of a fragment of the genome by polymerase chain reaction.

1.5 Polymerase chain reactions (PCR)

A. Principles of PCR

The amount of DNA present in a cell is very small, which is about 2.5 fg of DNA/cell (Button and Robertson, 2001). This concentration is too small to be used successfully to detect the target gene. Polymerase chain reaction was an innovative discovery by Kary Mullis in 1983 and allows the amplification of a single copy of DNA to millions of copies after about a 30-cycle reaction (Mullis, 1990; Saiki et al., 1985). PCR reactions can be divided into three phases: exponential, linear and plateau.

In the exponential phase, all the reagents are fresh and available; if the reaction efficiency is 100%, PCR product is doubling and accumulating in each reaction cycle. The reaction is very specific and precise. This is the phase where real-time PCR takes its measurements. As reactions progress some reagents are being consumed and the reaction slows in the linear phase. The PCR products do not double in each reaction, and reaction components are consumed. In the plateau phase, the reaction has stopped and no more products are being made. The PCR products will begin to degrade if left long enough. End-point PCR measurements are taken at this point, and it is known as end point detection (Applied Biosystem)

Taking measurements during the exponential phase is more accurate. Three replicates beginning with the same DNA quantities, result in different amplified quantities at the plateau phase in contrast to the exponential phase when three replicates have the same DNA quantities (Applied Biosystem).

B. Different types of PCR

End-point PCR, touchdown PCR, nested PCR, multiplex PCR, hot start PCR and real time PCR or quantitative PCR (qPCR) are the most common types of PCR techniques used to amplify DNA.

B.1 End-point PCR. End-point PCR is the beginning of this technique, which only furnishes qualitative information. This is an inexpensive and simple approach to amplify DNA. qPCR is an improved version of end-point PCR that facilitates quantitative and qualitative information. It that has more advantages than end-point PCR, i.e., high precision, high sensitivity, wide dynamic range of detection, no post-PCR processing (such as gel electrophoresis and handling of ethidium bromide), increase in fluorescent

signal directly proportional to the number of gene copies increase, and expressed results as numbers (Life technologies).

B.2 Touchdown PCR. Touchdown PCR is used to amplify a specific sequence from a complex genome, which helps to correct mis-priming of the oligonucleotide primers to the target. In this technique, the annealing temperature of the reaction is decreased 1°C in every second cycle from 65°C until touchdown at 55°C. If there is any difference in melting temperature (T_m) between correct and incorrect annealing temperatures, it will be corrected during this 2-fold per cycle (Don et al., 1991; Roux, 1995).

B.3 Nested PCR. Nested PCR is used to eliminate some unwanted products while at the same time considerably increases the sensitivity. This nested PCR is used to improve the detection efficiency when the initial DNA concentration is very low and cannot be detected by the end-point PCR. This method involves two sets of primers. The first set of primers, i.e. species-specific primer, amplifies the particular gene fragment of the genome. Then, these amplified products are subjected to an additional round of amplification with the host-specific primers, which are nested within the first set of primers (Roux, 1995).

B.4 Multiplex PCR. Multiplex PCR is another DNA amplification technique, which facilitates the amplification of multiple targets in a single procedure. There are two types of multiplex assays: single template PCR and multiple template PCR. In single template multiplex PCR, several specific regions within the template are amplified with several sets of primers. In contrast, specific regions in multiple templates are amplified with several sets of primers in multiple template PCR. There are several factors to be

considered when selecting or designing primers for multiplex PCR: 1) primer length should be short, i.e., 18-22 bp in length; 2) primers should have a similar melting temperature (T_m) with a temperature variation of 3°-5°C being acceptable; 3) primers should have specificity to amplify only the target sequence; and 4) should avoid primer dimer formation, which leads to unspecific amplification. This method is ideal for saving money, time and templates and commonly used in the medical field for pathogen identification, high throughput SNP genotype analysis, mutation analysis, gene deletion analysis and template quantification (Premier Biosoft). This technique can be used in the MST field with end-point PCR as well as real-time PCR. Different host-specific primers (i.e. human-specific and cattle-specific primers) can simultaneously be used to amplify host specific sequences present in water samples.

B.5 Hot start PCR. Hot start PCR is used to reduce primer dimers and non-specific amplification. These non-specific products can be the result of PCR that occurred during the short incubation period. The main aim of this hot start is to withhold at least one of the essential components in the PCR mixture until the temperature in the first cycle rise above the melting temperature (T_m) of the reactants. Taq polymerase is the most commonly used component withheld until reaction temperature is above T_m (Roux et al., 1995).

B.6 QuantitativePCR. qPCR is novel approach of DNA amplification, which gives the qualitative as well as quantitative information. qPCR quantification of gene copy numbers are based on the quantitative relationship between the amount of starting target sample and amount of PCR product at any given cycle number. It detects the accumulation of amplicons during the reaction, which is measured at the exponential

phase of the reaction. Detection of PCR products can be performed with TaqMan assay or SYBR Green assay. SYBR Green is a fluorescent dye that binds the minor grooves of double stranded DNA. Once this dye binds to the DNA, it emits fluorescence. The more double stranded amplicons produced, the more fluorescent signals will be emitted (Applied Biosystem).

Currently, real-time quantitative PCR is used for both relative and absolute quantification. The relative quantification approach measures the changes in DNA target concentrations relative to another gene, often referred to as an internal control or a reference target. This is a commonly used method in gene expression studies to determine the treatment effects such as how many fold gene copy numbers increased due to the treatment. This approach is not common in the MST field (Sivaganesan et al., 2010; Schmittgen and Livak, 2008). Efficiency cCrrection Method and Comparative C_T are two types of relative gene quantification methods. The Efficiency Correction Method calculates the relative expression ratio from the real time PCR efficiencies and the C_T . The Comparative C_T method assumes that efficiency of the PCR is close to 1 and the PCR efficiency of the target gene is similar to the internal control gene (Schmittgen and Livak, 2008).

Absolute quantification is used when a precise quantity of amplicons or gene copy numbers are required. This approach is popular in the bacterial source tracking field. Real-time PCR detects the accumulation of amplicons during the reaction, and these accumulated products are measured at the exponential phase of the PCR (Applied Biosystem,1). These quantification results are based on standard curves developed with known gene copy numbers of the target with genomic DNA or plasmid DNA. Once a

calibration curve is developed, the concentration of the target marker in an unknown sample can be estimated because there is a quantitative relationship between the amount of starting DNA/RNA and the amount of DNA product at the quantification cycle (C_q) number. Accuracy of the initial DNA concentration is highly dependent on the accuracy of the calibration curve. Therefore, validation of a calibration curve is crucial. Factors such as nucleotide base composition, template DNA concentration, dilution preparation and DNA stability during storage can introduce variations to a calibration curve (Sivaganesan et al., 2010).

Either 'single' standard curve or 'master' standard curve can be used for qPCR quantification. Single standard curve is generated with each PCR instrument, and the target gene copy numbers are calculated directly using this curve. Since a standard curve is running with every reaction, available reaction wells for samples are limited. Therefore, single standard curve is expensive and time consuming. This approach is suitable for a few samples, or few experimental runs (Sivaganesan et al., 2010). Master calibration is another approach; multiple instrument runs are used to generate a single master calibration curve and the rest of the samples run without a standard curve. The sample quantification is performed based on the master curve. The main concern here is that different instrument runs yield almost the same slope, but significantly different intercepts. 'Pooled' and 'mixed' methods are the improved versions of 'single' and 'master' calibration approaches. The 'pooled' calibration curve is developed with all standard curves of the same instrument runs for the same set of primers. This approach helps minimize the run-to-run variability due to day-to-day reagent mixing, pipetting, thermal cycling, fluorescence, dilution preparation, and storage of calibration curve

standards, etc. The ‘mixed’ strategy is the combination of ‘master’ and ‘pooled’ approaches, utilizing the common slope and the different intercepts, which vary from run-to-run. A Bayesian statistical modeling technique combined with a Monte Carlo Markov Chain simulation method was used by Sivaganesan et al. (2010) to generate ‘single’, ‘pooled’, ‘master’ and mixed calibration curves. The results of this study suggested that the selection of the calibration curve depends on the number of instrument runs needed to complete the study and the level of precision required meet the research goals. The ‘single’ approach is suitable for a small scale studies (one instrument run), the ‘pooled’ approach is suitable for the samples which require 4-6 instrument runs, and the ‘mixed’ manner is applicable for large scale studies including seven or more instrumental runs (Sivaganesan et al., 2010).

qPCR is a simple and precise method for determining the amount of target sequence or gene present in a sample. However, it is important to critically evaluate the quality of PCR data, especially among different laboratories. The following are some important factors that must be considered for real-time PCR evaluation.

a. **Efficiency of PCR.** PCR efficiency should be established by means of a standard curve because it provides simple, rapid, and reproducible indication of the assay efficiency, the analytical sensitivity, and robustness of the assay. Amplification efficiency is determined from the log linear portion of the standard curve; PCR efficiency = $10^{-1/\text{slope}} - 1$. The same standard curve with low and high amplification efficiency yields two different slopes and therefore, two different C_q values. PCR efficiency depends on the assay, master mix performance and sample quality. Achieving

an efficiency of 100% for every reaction is not possible all the time; therefore, efficiency between 90%-110% is acceptable (Applied Biosystem; Bustin et al., 2009).

b. **Dynamic range.** The efficiency of the standard curve varies with the dynamic range of the standard curve; therefore, well distributed dynamic range is essential. An assay with 5-log standard dilutions and 100% PCR efficiency yields low variation of standard curve efficiency from 94% to 106%. The standard curve with a 100% PCR efficiency and narrow dynamic range, i.e., testing the dilution series with a single log provide a wide range of efficiency, 70% to 170% (Applied Biosystem).

c. **R² value.** The coefficient of determination (R²) provides the information about the relationship between independent and dependent variables; it predicts how good one value is at predicting another. In real-time PCR, R² should be above 0.99 (Applied Biosystem).

d. **Accuracy.** How close experimental results are to the actual value (Bustin et al., 2009).

e. **Analytical sensitivity.** The minimum number of copies in a sample that can be measured accurately by an assay. Frequently, sensitivity is expressed as the limit of detection (LOD) the concentration that can be detected with a reasonable certainty with a given analytical procedure. The most sensitive theoretical LOD that can be achieved is 3 copies for real-time quantitative PCR (Bustin et al., 2009).

f. **Lower limit of quantification (LLOQ).** The minimum number of gene copies that can be quantified with an acceptable level of precision and accuracy.

g. **Repeatability.** The precision and the robustness of an assay or the measure of closeness of replicates. The results show the short-term precision or intra-assay variance.

This is measured by standard deviation (SD) or coefficient of variance (CV). If PCR is 100% efficient, standard deviation of replicates would be <0.167 (Bustin et al., 2009).

h. **Reproducibility.** Reproducibility shows the long-term precision or inter-assay variance between runs or between different laboratories. Reproducibility is typically expressed as standard deviation (SD) or coefficient of variation (CV) of copy numbers or concentrations (Bustin et al., 2009).

i. **Analytical specificity or detection of non-specific amplification.** The qPCR assay should detect the appropriate target sequence rather than other non-specific targets in samples. Sometimes, SYBR Green assays tend to amplify amplicon-independent sequence regions known as primer dimers, and those products can be detected with a melting or dissociation curve conducted after each reaction. A peak of the melting curve at lower temperatures corresponds to the primer dimers, and one at high temperatures represents the specific product. These primer dimers are prevalent in no template control (NTC) and low concentration wells (Mehra and Hu, 2005; Bustin et al., 2009).

1.6 Survival of bacteria in the secondary habitat

A. Differentiation of live and dead bacteria

Bacterial viability can be divided into four groups: viable (active and readily culturable), dormant (inactive but ultimately culturable), active but non-culturable and dead (inactive and non-culturable) (Kell et al., 1998). The plate counting method is one way of quantifying viable cells, but most environmental microorganisms are nonculturable under normal laboratory conditions. In situ hybridization combined with flow cytometry (FCM) has been used to detect live and dead cells. Fluorescent dye assays can be used to discriminate live cells from dead cells in combination with FCM,

microscopy and micro plate method (Invitrogen.com). The most commonly used technique is live/dead cell staining, discrimination with microscope (Nebe-von-Caron et al., 1998). Most of these methods allow counting and quantifying live and dead cells, but do not facilitate separation of species or strains (Kramer, 2009). These techniques have various limitations and cannot be used to separate DNA from live and dead cells. Quantification of mRNA is a precise way to determine live cells present in environmental water (Walters and Field, 2009); but mRNA has a short half-life, and it is unstable in the environment. Therefore, this is a technically demanding and expensive approach (Josephson et al., 1993).

B. Role of PMA and EMA in distinguishing live from dead cells

Propidium monoazide (PMA) and ethidium monoazide (EMA) are two types of dyes that can be used to detect live cells excluding dead cells. PMA/EMA is a DNA intercalating agent; in the presence of bright visible light, the azide group intercalates with DNA and produces strong covalent bonds (Fig.1.3). These strong covalent bonds prevent double stranded DNA separation during denaturation, and this inhibits PCR amplifications. EMA/PMA has the ability to penetrate only cell walls and cell membranes of dead cells, not those of live cells. The unbound EMA/PMA reacts with the water molecules in solution and inactivates its reactivity. Consequently, PMA/EMA treatment facilitates selective amplification of live cell DNA in the bacterial population (Nogva et al., 2003; Rudi et al., 2004; Nocker and Camper, 2006; Nocker et al., 2006; Nocker et al., 2007; Vesper et al., 2007; Rawsthorne et al., 2009; Kramer et al., 2009). Organisms already tested with EMA/PMA-PCR are *E. coli* O157:H7, *Listeria monocytogenes*, *Micrococcus luteus*, *Mycobacterium arium*, *Pseudomonas syringae*,

Salmonella typhimurium, *Serratia marcescens* (Nocker et al., 2006b) and *Clostridium perfringens* (Wagner et al., 2008). Another study by Kramer et al. (2009) used probiotic bacteria such as *Lactobacillus acidophilus* and *bifidobacterium animalis* to compare plate counting, real-time PCR with PMA treatment and fluorescent staining combined with FCM. The results showed PMA treatment in combination with real-time PCR is a reliable and effective approach to discriminate live probiotic cells from dead ones.

Studies with *E. coli*, *Campylobacter jejuni* and *Listeria monocytogenes* found that EMA caused a significant loss of live cells. Intact cells of bacterial species can take up more EMA than others, resulting in damaged live cells (Nocker et al., 2006; Flekna et al., 2007). As a result, EMA caused a reduction in viable cells and led to underestimation of the particular group of microorganisms. To overcome this problem PMA was introduced later, which has a better selectivity and can enter only the dead cells (Nocker et al., 2007). However, Chang et al. (2010) found no obvious difference between EMA/PMA penetrations into the viable cells, and that the required PMA concentration was about 4-times higher than that of EMA to achieve the same effect. This study suggested that the use of EMA is more cost effective compared with PMA for investigating *Legionella* in water samples. PMA treatment is very useful and common in clinical studies to reduce false-positive PCR results. Amplification of 16S rRNA and *tuf* genes in *Staphylococcus aureus* and *Staphylococcus epidermidis* treated with antibiotics, vancomycin or gentamicin was effectively inhibited by PMA (Kobayashi et al., 2009; Kramer et al., 2009). However, this technique has not yet been widely extended to environmental samples, perhaps due to insufficient light penetration through particulate matter and high levels of suspended solids present in the surface water, which may interfere with the

photo-induced cross linking of PMA/EMA to DNA (Varma et al., 2009, Wagner et al., 2008). Bae and Wuertz (2009, 2012) effectively applied this method to quantify live and dead *Bacteroidales* present in effluent and influent of a sewage treatment plant. Additionally, PMA was successfully used to differentiate live *Enterococcus* and *Bacteroidales* from dead ones in wastewater (Varma et al., 2009).

C. Factors affecting survival of *E. coli* and *Bacteroidales* in the secondary habitat

Most of the enteric bacteria such as *Bacteroidales* and *E. coli* have a biphasic life pattern as host-associated and host-independent phases. Their survival in both phases are different. Survival of *E. coli* and *Bacteroidales* in the host independent or secondary habitat depends on many factors, such as availability of nutrients and energy sources, moisture, temperature, salinity, predators, pH, heavy metals, biological oxygen demand (BOD), chemical oxygen demand (COD) and availability of anaerobic pockets etc. (Van Elsas et al., 2011). There are some studies about enteric bacterial survival; among these studies, some have similar results and others show different survival patterns. A few studies found that fecal bacteria can persist, survive and regrow in the secondary habitat (Lee et al., 2006; John et al., 2008), and other studies found that they can only survive and persist in the secondary habitat. Anderson et al. (2005) did not find evidence of FC and enterococcus growth in fresh water microcosms. FC had a longer survival period in the fresh water than enterococci. This study also found that some FC strains have abilities to persist longer in the secondary environment than counterpart strains under the same conditions. Okabe and Shimazu (2007) found that *Bacteroidales-Prevotella* molecular markers can persistence at low temperatures and high salinity, but they did not observe any increase in these markers. Fecal bacteria survival outside of the host organism

depends on the mesocosm conditions; their persistence was highest in contaminated soils, followed by wastewater and dog feces (John et al., 2005).

E. coli can survive in different environments to different extents. *E. coli* O157:H7, a human pathogen that persists at pH 2.5, can survive in manure-amended soils for more than 100 days (Franz et al., 2008; Jiang et al., 2002) and 109 days in the water (Avery et al., 2008; McGee et al., 2002). Additionally, it can survive in cattle hides and carcasses, but their survival in the hide can be as short as nine days (Arthur et al., 2011). *E. coli* decay rates were the same in the microcosms amended with fresh human and cow feces (Walters and Field, 2009). Below are the some important factors that affect the survival of fecal bacteria.

C.1 Temperature. Temperature is an important factor that influences the survival of fecal indicators and has a positive effect on the decay rates. *Bacteroidales* molecular markers persist for a longer time at 10°C than at higher temperatures (Schulz and Childers, 2011; Bell et al., 2009; Okabe and Shimazu 2007). Human specific *Bacteroidales* 16S rRNA molecular markers persisted for more than 24 days at 4°C and 12°C in microcosms (Seurinck et al., 2005). Kreader (1998) detected a strong effect of temperature on the survival of *B. distasonis*; the molecular markers were detected for at least 2 weeks at 4°C, 4 to 5 days at 14°C, 1 to 2 days at 24°C and only 1 day at 30°C. At low temperatures, activity of predators and degradation of molecular markers are comparatively low; PCR targets persisted for at least a week in the filtered or sterilized water at 24°C but only 1-2 days in non-filtered river waters (Kreader, 1998; Okabe and Shimazu, 2007). Craig et al. (2004) found that low temperature enhances the survival of *E. coli*, where *E. coli* survived for more than 28 days at 10°C in sediments. Temperature

fluctuations adversely affect *E. coli* survival; their survival in manure under fluctuating temperature was less than the constant temperature (Semenov et al., 2007). Since activity of predators is low at low temperature, enteric bacteria can survive longer.

C.2 Dark versus light conditions. *E. coli* in microcosms amended with fresh human and cow feces survived longer in the dark, but that trend was not statistically significant (Walters and Field, 2009). Other studies showed that sunlight adversely affect the survival of *E. coli* and fecal coliforms in the environmental samples (Sinton et al., 2002). Sunlight intensity and UV radiation are the main factors that affect the death of bacteria in a stream. However, Walters and Field (2009) conducted their study at the end of fall; there was not enough sunlight intensity or UV radiation to kill bacteria. If they had conducted the study during the summer, they may have gotten different results. Persistence of some molecular markers depends on the light and dark conditions of the microcosm. Cattle specific CF128F markers persisted for a longer period than CF193F under the light condition. End-point PCR detected CF128F markers until day 14 under dark and light conditions and that of CF193F markers detected up to Day 3 under light and up to 6 days under dark conditions. Human specific markers, HF183F and HF134F, showed no difference in persistence in the dark and light environments. Both human markers could not be detected after Day 7 (Walters and Field, 2009).

C.3 Salinity. Schuiz and Childes (2011) found that, when the microcosm is under dark condition, decay rates of *Bacteroidales* are inversely related to salinity. But Okabe and Shimazu (2007) found that high salinity enhances the survival of *Bacteroidales*; high salinity may kill or inhibit the predators in the microcosm. Okabe and Shimazu (2007) used unfiltered sea water samples to prepare microcosms which had salinities of 0, 10, 20

and 30 ppt and compared *Bacteroidales* and fecal indicators persistence in these microcosms. *B. fragilis* marker persistence was constant for 11 days in non-filtered sea water, while that in non-filtered river water was 5 days (Okabe and Shimuzu, 2007). In addition, no decay of 16S rRNA genetic markers was found in both filtered river and seawater. Based on these results, they have hypothesized that salinity has an indirect effect on the persistence of *Bacteroidales* genetic markers; high salt concentrations may control the activities of predators or other indigenous microorganisms in water samples, which indirectly contributed to salinity controls on predation. Predators and indigenous microorganisms play a vital role in the decay of *Bacteroidales* molecular markers (Okabe and Shimuzu, 2007). Unlike *Bacteroidales* molecular markers, decay rates of fecal coliforms in both water column and the sediment of saltwater was greater than that of fresh water. Saltwater has a negative effect on fecal coliform persistence (Anderson et al., 2005; Craig et al., 2004). However, the study of Okabe and Shimuzu (2007) showed that salinity significantly affected the decay rate of total coliforms, but there was no effect on prevalence of fecal coliforms.

C.4 Survival in sediment verses water column. Fecal indicator bacteria (FIB) can survive for a longer period in the sediment than in the water column. Most of the nutrients such as Ca^{2+} , Mg^{2+} , K^+ , and NH_4^+ are adsorbed to the sediments, and FIB can utilize these nutrients for their survival and regrowth. A study by Davies et al. (1995) found FIB concentration in the sediment about 100 to 1,000 times higher than in the overlaying water. In addition, FIB adsorbed by sediment particles may be protected from many factors such as predators, heavy metal toxicity, salinity and UV radiation. Fecal coliforms in sediments can be about 2,200 times greater than that in the water column

(Crabill et al., 1999). However, at 10°C, both the water column and the sediment contained same proportion of FIB. These results showed that the sediment and water column are equally good reservoirs of fecal pollution at low temperatures (Pote et al., 2009). At higher temperatures, i.e, 20°C, fecal decay rates in sediments were lower than in the water column (Pote et al., 2009; Craig et al., 2004). At an Australian beach, *E. coli* was detected for only 7-days in the water column, but it was detected even after 28 days in the sediment (Craig et al., 2004). Lee et al. (2006) detected *E. coli* growth in beach sediments and overlaying water; *E. coli* concentration in water increased by 3-order magnitude in the sediment and water microcosm, while its concentration reduced by 2-3-order magnitude in the microcosm contained only water. Growth in sediment is greatly affected by the presence of predators. When predators are present, *E. coli* can persist, but their growth would not occur. In addition, organic matter content of the sediment also enhances the growth of these organisms (Lee et al., 2006).

C.5 Particle size of sediments and organic matter content. Availability of carbon sources such as organic matter is probably the main critical factor affecting the survival of any bacteria in the open environment. Sediment or soil particle size also affects their survival. Smaller particles have a larger surface to volume ratio. Thus, small particles can adsorb more nutrients such as Ca^{2+} , Mg^{2+} , and NH_4^+ , which are essential for bacterial growth and survival. Some studies found that large particle size and low organic matter content of the sediment have negative effects on *E. coli* survival and persistence in the secondary habitat (Craig et al., 2003; Lee et al., 2006; Pote et al., 2009). However, aggregate size of the fecal matter does not affect biological decay rate of *Bacteroidales* 16S rRNA gene markers (Bell et al., 2009).

1.7 Current prospects and the future of microbial source tracking

A. The development of the MST field

Bacterial host specificity is the main assumption behind MST. Bacterial variability among hosts may be due to various factors such as host diet, climate, geographic location, etc. In addition, host age and health especially affect within population variation (Shanks et al., 2008). The variation between ruminant and non-ruminant may be caused by the unique anatomy and physiology of the ruminant digestive system. This difference may provide different evolutionary pathways for *Bacteroidales* from the inhabitants of non-ruminant hosts (Dick et al., 2005b).

The concept of using anaerobic bacteria such as *Bacteroides-Prevotella* and *Bifidobacterium* as alternative indicators was suggested more than a decade ago. As these organisms are obligate anaerobes, their cultivation under ambient laboratory conditions is difficult. Therefore, molecular based approaches such as detection of specific genes are the most common and easy way to detect these organisms in water samples. Kreader (1995) conducted a study to determine host specificity of *Bacteroides* species. Three PCR primer sets specific to *Bacteroidales* 16S rRNA gene were designed to detect the host specificity of *B. distasonis*, *B. thetaiotaomicron* and *B. vulgatus*. Results of this study show that *Bacteroides* species are not specific only to a particular host group; these three species were more dominant in the human microbiome than in other animals such as cattle, swine, goats and sheep. *B. vulgatus* was less common in human microbiome than two other species. *B. vulgatus* was prevalent in house pets but two other species were not as common in house pets. About 63% of both human and house pet samples contained *B. vulgatus*. *B. thetaiotaomicron* was present only in horse samples. The rest of the animal

groups (i.e., cattle, swine, goats and sheep) contained all three species but in low numbers than in humans (Kreider, 1995).

Bernhard and Field (2000a) developed the 16S rRNA gene based method to detect *Bacteroides-Prevotella* and *Bifidobacterium* in surface water. A community profile of 16S rRNA gene developed with *Bacteroides-Prevotella* and *Bifidobacterium* with length heterogeneity PCR (LH-PCR) and terminal restriction fragment length polymorphism analysis (T-RFLP) revealed that the host specificity of *Bacteroidales* lineage. This community profile showed unique host specific patterns, and also that these patterns were reproducible. The molecular markers of *Bifidobacterium* were not as detectable and prevalent in all host groups as the *Bacteroides-Prevotella* group (Bernhard and Field, 2000a).

The 16S rRNA gene is the most commonly used molecular marker to detect the order *Bacteroidales* in environmental samples. The variable regions, 32F upstream and 708F downstream of 16S rRNA, contain specific primer regions for the Order *Bacteroidales*. This variable region contained human-specific primer regions 183F/708R and 134F/654R, as well as ruminant-specific primer regions such as 128F/708R and 193F/708R (Bernhard & Field, 2000a,b). Initially, DNA amplification was performed with end-point PCR, which provides only qualitative information. Later, real-time or quantitative PCR was introduced and that provided the qualitative as well quantitative information of sources and degree of pollution. The amplicon length of the PCR product is one of the main factors that determine the amplification efficiency of real-time PCR. If the amplicon length is large, those fragments do not show high replication efficiency. The amplicon length of the HF183 product is 525 bp, and that may not replicate

efficiently during the real-time PCR. Therefore, selecting a new set of primers or a new downstream primer is essential to yield a smaller amplicon length. Seurinck et al. (2005) designed new downstream primer (265R) for human-specific 183F upstream primer. This new primer produces a smaller amplicon (82 bp), which is especially suitable for real-time PCR.

Carson et al. (2005) developed a new primer set with the amplicon length of 542 bp, which was specific to *Bacteroidales thetaiotaenicon*. About 92% of human, 100% sewage and 16% dog feces were positive for this marker. Since a considerable percentage of dog feces were positive, this marker was not popular among other researchers.

The AllBac assay (296F/412R) was developed to detect total *Bacteroidales* present in water. Amplicon length of the target region is 116 bp; therefore, this set of primers is more suitable for real time PCR than Bac 32F/708R. The 16S rRNA sequence contained the primer regions for BoBac and HuBac, which was designed to detect ruminant and human fecal contamination. BoBac has 100% sensitivity and specificity, while HuBac showed 100% sensitivity, but had 32% false positive rate with cross reactivity in swine feces (Layton et al., 2006).

Kildare et al. (2007) developed another 16S rRNA gene-specific assay to detect total *Bacteroidales* (BacUni-UCD), human-associated (BacHum-UCD), cow-associated (BacCow-UCD) and dog-associated (BacCan-UCD) genetic markers in surface water. All samples from humans, cats, dogs, seagull, cows and wastewater effluents were positive for BacUni-UDC assay. The human assay has the ability to discriminate cow fecal samples from human samples with no cross reactivity in cat, horse or seagull

samples. However, this BacHum-UCD could not amplify all human fecal samples; only 66.6% of the human stool samples were amplified. In addition, 12.5% of dog fecal samples were also amplified by this primer set. BacCow-UCD has the ability to fully discriminate cattle feces from human stool. This assay amplified 38% of horse feces, but it did not amplify fecal samples from other host groups. The BacCan-UCD assay was developed to detect dog feces in water samples. This assay discriminated dog feces from cow and seagull feces. However, it did not show good specificity; 22% of human, 14% of cat and 28% of wastewater samples were amplified by this BacCan-UCD assay. Also, it did not detect all dog fecal samples only 5/8 dog fecal samples were positive for this assay (Kildare et al. (2007).

Mieszkin et al. (2010) developed primers to detect fecal pollution caused by pigs, and the primer set has high specificity (100%) and sensitivity (98%). Dick et al. (2005) designed primers for pigs and horses. These primers were also highly specific with a detection limit of 100 copies/reaction.

In addition to the 16S rRNA gene, other functional genes such as energy metabolism and electron transport (CowM2) and degradation of surface polysaccharide and lipopolysaccharide (CowM3) are also host-specific. Shanks et al. (2008) used these genes to develop host specific markers (i.e., CowM2 and CowM3) to identify cattle pollution in stream water. CowM2 and CowM3 are the only molecular markers that have ability to discriminate ruminants from cattle. The specificity of these primers was tested with 16 animal species; only cattle fecal matter was amplified. These molecular markers have good sensitivity, ranging between 98-100% in qPCR.

Weidhaas et al. (2011) used the 16S rRNA gene of *Brevibacterium* sp. LA35 to track water and soil pollution caused by poultry litter. The LA35 gene marker was correlated with fecal indicators such as fecal coliforms, *E. coli* and *Enterococcus*, and this gene marker was about 1 to 2 logs greater than fecal indicator concentrations. One of the major disadvantages in MST is the inability to detect the diluted target in water; therefore, the presence of high concentration of a particular gene is advantageous for detecting the target in environment sample. The LA35 gene marker has a positive correlation with inorganic elements such as Cu, P and Zn, which are associated with poultry litter (Weidhaas et al., 2011).

Organisms such as human enteric virus, fecal streptococci, Bifidobacteria and Sulphite-reducing clostridia, have been evaluated for use as fecal indicators in surface water. However, they have not been successful for various reasons such as low sensitivity, high cost, low reliability, or unsafe to use under normal laboratory conditions. Thus, these organisms are not considered to be successful indicators to identify the fecal contamination in water. Bovine entero-virus (BEV), human enteroviruses (HEV) and human adenoviruses (HAdV) were the other group of organisms used to detect fecal contamination in the stream water (Fong et al., 2005). These enteroviruses have less sensitivity; only 78% of cattle samples were positive for BEV. BEV-like sequences were also detected in sheep, goats and horses (Jimenez-Clavero et al., 2005).

B. Evaluation of available genetic markers

A good host-specific marker should be present in all samples of the target group and should be absent in all non-target groups. Gawler et al. (2007) introduced two parameters to evaluate the markers: sensitivity and specificity. Sensitivity refers to the ability to detect the target (positive) samples [true positives/(true positives + false negatives)]; specificity describes the ability to not detect non-target (negative) samples [true negatives/(true negatives + false positives)]. Among the human specific molecular markers, HF183F/708R markers have the highest sensitivity, specificity and robustness. Ahmed et al (2007) evaluated the sensitivity and specificity of HF183 and HF134 markers with 207 fecal samples, which included 13 host groups and 52 human samples from septic tanks and sewage treatment plants. This study was conducted in Southeast Queensland, Australia and found that specificity and sensitivity of HF183 were 100%, and that of HF134 were 95.5% and 97.3%, respectively (Ahmed et al., 2007). Another study showed 100% sewage samples were positive for general *Bacteroidales* as well as human specific (183F/708R) markers (Bower et al., 2005).

The new assay developed for qPCR (183F/265R) based on the HF183 primer region, which is designated as qHF183, also revealed the higher assay specificity and sensitivity. qPCR detected human signatures in a water sample up to 1 ng wet feces per liter of freshwater (Seurinck et al., 2005). Detection limits of Bac32, HF134 and HF183 were 1×10^{-14} , 1×10^{-12} and 1×10^{-11} g of fecal DNA, respectively (Lamendella et al., 2007). Van de Werfhorst (2011) compared host specificity and sensitivity of HF183 and BacHum-UCD. BacHum-UCD had 100% sensitivity for sewage, human fecal and septage samples, but it also amplified 83% of cat, 75% of dog, 33% of gull, and 40% of

raccoon or rat fecal samples. HF183 amplified 63% of human samples, 67% of septage samples and 100% of sewage samples. Additionally, it amplified 8% of cat samples while all other non-human samples were negative for this assay.

A comparison of general *Bacteroidales*, human-specific and cow-specific assays was conducted with fecal samples (human, cow, horse, dog, cat, seagulls and waste water treatment plant (WWTP)) collected from different locations in California (Kildare et al., 2007). Total *Bacteroidales* assays, BacUni-UCD and AllBac, amplified 100% of all the samples tested, while Bac32 amplified 100% of horses, cats, seagulls and WWTP samples; 94.4% human samples and 87.5% of cow and dog samples. All three human-associated assays, HF183, HuBac and BacHum-UCD, showed 100% sensitivity to WWTP samples but less sensitivity to human stool samples, 61.6%, 88.9% and 66.7%, respectively. A total 25% of dog samples and 14% of cat samples were positive for the HF183 assay, 14% of dog samples were positive for the BacHum-UCD assay, and all fecal samples except seagulls showed cross reaction with HuBac markers (Kildare et al., 2007). Cattle-specific assays, BacCow-UCD and BoBac, showed 100% sensitivity, while BacCow-UCD amplified 37.5% of horse samples and BoBac amplified 11.2% of human samples (Kildare et al., 2007).

Bovine-specific molecular markers were evaluated for their performances such as sensitivity, geographical robustness and specificity (Shanks et al., 2010). Performance of Bac2, Bac3, CF128, CF193 markers were analysed with end-point PCR while CowM2, CowM3 and BoBac markers were tested with real-time PCR. Among BoBac, CowM2 and CowM3, BoBac is the most abundant molecular marker followed by CowM3 and CowM2. CF128, CF193, and BoBac have the assay specificity of 76%, 99.9% and

47.4%, respectively. This indicates the less specific BoBac markers are more abundant. Bac2, Bac3, CowM2 and CowM3 showed high specificity, greater than 98.9%, which suggests that using genes other than 16S rRNA may improve the assay specificity. CF128 and BoBac were the most sensitive and geographically robust assays; however, they demonstrated the lowest specificity. Bac2, Bac3, CowM2 and CowM3 are highly specific but less sensitive. This study suggested that bovine assay performance is not consistent over the geographic region; therefore, prior characterization of assay performance for each watershed of interest is vital (Shanks et al., 2010).

In another study, 75 cattle fecal samples were used to evaluate CF128 and CF193 markers (Bower et al., 2005). CF128 showed greater sensitivity than CF193. All of the cattle fecal samples were positive for CF128, while only 85% of samples had signals with CF193 (Bower et al., 2005). Lamendella et al. (2007) found similar results as Bower et al. (2005). The CF128 assay exhibited higher sensitivity than CF193; 96% of the cattle fecal samples were positive for the CF128 marker and 90% for the CF193 marker (Lee et al., 2008). More than 90% of cattle samples were positive for the general *Bacteroidales* marker, Bac32. The detection limits of cattle markers, Bac32 and CF128, were 4.17×10^{-7} and 4.17×10^{-6} g of feces, respectively (Lee et al., 2008).

C. The correlation between *Bacteroidales* genetic markers and fecal indicator bacteria

USEPA has still not recognized the use of *Bacteroidales* as an indicator organism. This is mainly due to the lack of research that defines the correlation between the presence of *Bacteroidales* and occurrence of pathogens in the surface water. Therefore, assessment of the correlations between *Bacteroidales*, *E. coli* and other pathogens is important. Some studies showed a good correlation between *Bacteroidales* and *E. coli* or

fecal coliforms. Savichtcheya et al. (2007) reported the correlation coefficient between total coliform and all *Bacteroidales* in stream water was 0.86 while that between fecal coliform and human specific *Bacteroidales* was 0.74. Further, *Bacteroidales* 16S rRNA genetic markers had a positive correlation with *E. coli* O157 ($r > 0.52$) and *Salmonella* ($r > 0.56$) (Savichtcheya et al., 2007). Another study in France reported that the correlation coefficient between *E. coli* and pig specific *Bacteroidales* markers was 0.77 (Mieszkin et al., 2010). Gourmelon et al. (2007) found a significant correlation between *E. coli* and *Bacteroidales* human-specific molecular markers, but they could not find a significant correlation between *E. coli* and ruminant markers. Positive correlations between *Enterococcus* and Bac32 and CF128 molecular markers were reported by Lee et al. (2008).

A study of Santoro et al. (2007) reported a different conclusion; no correlation between *Bacteroidales* and fecal indicator bacteria (FIB) in individual samples. On the other hand, long term data analysis instead of individual sample data analysis showed that high FIB abundance was correlated with human specific markers.

D. Application of *Bacteroidales* to detect the sources of fecal contamination in surface water

A number of studies have successfully used *Bacteroidales* molecular markers to detect fecal contamination in surface water. Both ruminant and human markers were used to detect human and ruminant fecal pollution in Plum Creek, Nebraska (Lamendella et al., 2007). The CF128 marker was detected in 96% of fecal and 54% of water samples, whereas CF193 was found in 90% and 28% of fecal and water samples, respectively. Ruminant fecal pollution is more prevalent than human fecal pollution in this creek. Only 3-5% of water samples showed human markers, while sediments contained higher

percentage of human markers (19-24%). A total of 81% of water samples were positive for Bac32F, the detection limit of Bac32F, HF134F and HF183F were 1×10^{-12} , 1×10^{-9} and 1×10^{-10} g of sewage DNA, respectively (Lamendella et al., 2007).

Sources of fecal contamination in the Great Lake coastal beaches were evaluated with general *Bacteroidales*, human-specific and cattle-specific markers. General *Bacteroidales* markers were prevalent in all water samples where about 1/3 of the samples contained *E. coli* concentrations less than 235 CFU/100 ml. Human markers were found at three of the seven beaches, and *E. coli* concentrations were as high as 5,800 CFU/100 ml. None of the nine beaches were positive for cattle-specific markers (Bower et al., 2005).

Human fecal pollution on the Orange Beach, CA was detected with the HF183 assay (Santoro and Boehm, 2007). According to this study, temperature, turbidity, salinity or oxygen level did not affect the human marker's presence in water. However, high and low tide levels affected total and fecal coliform concentrations in the water. During the low tides, total and fecal coliform concentrations were higher than during high tides (Santoro and Boehm, 2007).

A study in Brittany, France used the qHF183 primer to detect human fecal contamination downstream of Daulas town (Mieszkin et al., 2009). qHF183 molecular markers were quantified in 90% of water samples, the concentrations ranged from 3.6 to 4.6 \log_{10} copies per 100 ml. Ruminant-specific BacR molecular markers were detected in all water samples collected around the cattle farms. The concentrations ranged from 4.6 to 6.0 \log_{10} copies per 100 ml. Two pig markers, Pig-1-Bac and Pig-2-Bac, were detected

in 25% and 62.5% of water samples collected around the pig farms, respectively (Mieszkin et al., 2009).

Lee et al. (2008) conducted a study to compare two different watersheds with different management practices. Multiple host specific PCR assays were used to detect the ruminant origin bacterial population in these watersheds. CF128 was detected in 65% of the water samples, while non-16S rRNA gene markers, Bac1 and Bac5, were found in 32% and 37% of the water samples, respectively. CF128F and Bac32F showed the seasonality of their occurrence. Bac32F occurred more frequently in summer and spring than in winter, while CF128 occurred more often in winter than fall.

Bacteroidales assays were also used to detect fecal contamination in Kenya, the first time on African continent (Jenkins et al., 2009). Total *Bacteroidales* present in feces of humans, cows and donkeys, wastewater and River Njoro were monitored with two universal *Bacteroidales* primers: BacUni-UCD and general *Bacteroidales* (TNA) molecular markers. BacUni-UCD had 100% sensitivity for all samples while TNA assay sensitivity varied from 0% wastewater to 100% human feces. Human specific BacHum-UCD and HF183 did not amplify cow and donkey fecal samples (100% specificity), but both assays have a poor sensitivity: only 25% and 58% human fecal samples were amplified with BacHum-UCD and HF183 assays, respectively. The BacCow-UCD assay showed 100% specificity and 93% sensitivity. According to this study, cattle are the main source of fecal contamination in River Njoro, where 78% of samples contained cattle markers. Only 11% of water samples contained both BacHum and HF183 markers (Jenkins et al., 2009).

E. Constraints to be considered in the application of host-specific markers

E.1 PCR inhibition. Wastewater and fecal samples contain various organic and inorganic substances, which may not be removed during DNA extraction. Some of these chemicals can interfere with PCR and inhibit qPCR amplifications. Therefore, PCR inhibition should be assessed when detecting host-specific markers. There are several ways to detect PCR inhibition; one way is to spike a known amount of plasmid DNA to the sample before PCR. Inhibition can be calculated using following equation. Percent recovery = [(measured DNA copies in water sample spiked with 2.5 uL (10^5 copies) plasmid DNA) - (measured copies in unspiked water sample)*100] / 2.5×10^5 (Layton et al., 2006).

E.2 Bias and artifacts in PCR. There is bias inherently associated with PCR amplification. PCR bias may be caused by differences in primer binding energy, formation of heteroduplex, or formation of chimeras. In studies of community structure analysis, the elimination of PCR bias and artefact is essential to obtain reliable data from PCR-based analysis (Kanagawa, 2003). The other major reasons for skewing template to product ratio are PCR selection and PCR drift. PCR selection may be caused by preferential denaturation due to overall low GC content and higher binding efficiency at GC rich fragments. PCR drift is mainly caused by stochastic variation in the early cycle of reactions (Polz and Cavanaugh, 1998).

E.3 Presence of *Bacteroidales*. The Order *Bacteroidales* is not the dominant group in every animal group. About 25-30% of the human microbiome consisted of the phylum Bacteroidete; their persistence in the human feces is abundant (Yang et al., 2009a). Thus, *Bacteroidales* is a good candidate to detect human fecal contamination in water.

However, *Bacteroidales* are not prevalent in certain host groups such as birds. A study was conducted in Ontario and Ohio to assess the presence of *Bacteroides* in Canadian geese, and the results indicated that only 10% of the Canadian geese gut microbiome contains the *Bacteroidales* 16S rRNA gene markers. However, 16S rRNA gene sequences of Clostridia and Bacilli species were prominent in these fecal samples, which were 33.7% and 38.1%, respectively (Lu et al., 2009). *Bacteroidales* are not prevalent in some other birds such as chicken and gulls. A metagenomic study found that chicken gut populations contained Clostridia (21%), Bacteroidetes (15%) and Bacilli (17%) (Lu et al., 2007). Further, analysis of 16S rRNA sequences from gull feces found that gull biota contained Bacilli (37%), Clostridia (17%), Gammaproteobacteria (11%) and Bacteroidetes (1%) (Lu et al., 2008). Therefore, researchers should rethink the application of the Order *Bacteroidales* as the sole organism for MST. It is desirable to become familiar with possible contamination sources in the watershed. If birds are a possible source in a watershed, selection of two MST organisms such as *Bacteroidales* and *Clostridium* or *Bacillus* may resolve this problem. Since the avian population is an important source group, especially in a beach area, selection of proper and representative microbial groups is important.

F. Advanced approaches and future of MST

Knowing the microbial community structure in human and animal guts is vital because it helps predict host specificity and survival characteristic in the secondary habitat. Active/live microbes interact with their immediate abiotic and biotic environments and play an active role; therefore, it is important to know survival of organisms in their secondary habitat (Klein, 2007). The bigger the population, the more

efficient the population evaluation is. Bacterial variation is not caused by sex, or meiotic recombination and self-propagating 'selfish DNA,' or elements like introns. The absence of non-coding sequences (introns) in bacteria causes the efficient passing of mutation in the genome from generation to generation. Natural selection plays an important role of producing an ideal organism for a particular environment. In addition, the age of the population, the size of the population and the constantly changing environment also affect variations in the bacterial community (Klein, 2007).

DNA sequencing is one of the most innovative ways to study microorganisms present in the natural assemblage and their evolutionary relationships. DNA sequencing would provide information about the genomic variability of members in the same species. These sequences clearly show that microorganisms are diverse at the genetic level and no two genomes are identical. Analysis of bulk DNA provides more information about the microbial community, this process is known as metagenomic analysis. Handelsman et al. (1998) defined Metagenomics as “the genomic analysis of the collective microbial assemblage found in an environmental sample.” Chen and Pachter (2005) defined Metagenomics as "the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species.”

There are several approaches to develop metagenomic libraries, the initial methods were based on the cloning of environmental DNA followed by sequencing. Currently, many studies focus on high-throughput sequencing methods such as 454 pyrosequencing, Illumina and SOLiD. High-throughput DNA sequencing is very important, especially for community structure analysis, because it facilitates the

exploration of phylogenetic and metabolic diversity of microbes in the environment. The 16S rRNA gene is the most frequently used target in high-throughput sequencing to get the phylogenetic information about bacterial groups. These different DNA sequencing techniques have been used to evaluate human gut microbial community structure and some of these studies are summarized below.

F.1 Use of next generation sequencing for gut microbiome analysis

Sanger (1975) introduced DNA sequencing, which is known as dideoxy chain termination technology or the Sanger method. The Sanger method is used in various fields for DNA sequencing. Arumugam et al., (2011) developed a metagenomic library based on Sanger sequencing data from feces collected from a total of 22 European individuals, 13 Japanese and two Americans. In addition, pyrosequencing data from fecal matter of two more Americans were added to this library – totaling 39 individuals. This study and some previous studies also found that Firmicutes and Bacteroidetes were the dominant phyla in the human gut microbiota, and *Bacteroides* are the most variable and abundant genus across the sample (Arumugam et al., 2011; Kurokawa et al., 2007; Tap et al., 2009). Sequences of the metagenomic library revealed that human guts contained three robust enterotypes clusters: *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2), and *Ruminococcus* (enterotype 3). Arumugam et al.(2011) did not find the connection between enterotype and nation, ethnic or continent specificity, or association with sex, weight or health of individuals. These results suggest that existence of a well-adapted, limited number of host-specific bacteria might respond differently to diet and drug intake (Arumugam et al., 2011). These outcomes are

compatible with other human gut population studies using pyrosequencing (Turnbaugh et al., 2009; Qin et al., 2010).

The relationship between long term dietary habits and gut microbial enterotypes was evaluated using fecal samples collected from 98 individuals from Europe, United States and Africa by the pyro sequencing of 16S rRNA gene (Wu et al., 2011). Long term diet habits demonstrated a strong correlation with the enterotype composition in microbiome. *Bacteroides* enterotypes were strongly associated with animal proteins and saturated fat, while *Prevotella* enterotypes were associated with carbohydrates and simple sugar. These results suggested that *Bacteroides* may prevalent in the people of western countries where meat consumption is high; *Prevotella* may prevalent in African countries which have carbohydrate based diet patterns (Wu et al., 2011).

The metagenomic study with pyrosequencing found that the human microbiome is consistent with more than 9,000,000 unique genes in this bacterial community (Yang et al., 2009a) Microbes in the human gut play an important role in food digestion, nutrient and absorption. The change in the human gut bacterial community is associated with several diseases as diabetes, hypertension and obesity (Yang et al., 2009a). Bacteria from eight of 13 phyla are present in the human gut; Phylum Bacteroidetes (30%) and Phylum Firmicutes (65%) are the major phyla in the human microbiome. In addition, 1.6% of Proteobacteria, 1.5% of Verrucomicrobia and 1.4% of Actinobacteria were found in the human gut. (Yang et al., 2009a).

Shanks et al. (2011) used pyrosequencing to analyze microbial communities in cattle feces. This technique was used to analyse 633,877 sequences from hyper variable region of 16S rRNA genes obtained from 32 individual cattle feces samples. The

community structure varied greatly depending on the feeding operations. Overall, 55% of the cattle gut population is consisted of Phylum Firmicutes. Phylum Bacteroidetes was the next most abundant group (25%); Tenericutes (2.5%) and Proteobacteria (2.5%) were present to a lesser extent. The Phylum Bacteroidetes was abundant in cattle fed with unprocessed grain while the phylum Firmicutes was dominant in cattle fed with processed grain and forage. In addition, management practices more strongly affect the population variability than their geographic variability (Shanks et al., 2011).

Application of next generation sequencing for MST.

Use of next generation sequencing techniques in MST field is not yet common. Cross amplification of the molecular markers is one of the common concerns in the MST studies. Therefore, use of two different bacterial groups with host specificity may provide improved assessment of the sources of fecal contamination. The phylum Firmicutes is most abundant in the human gut where 65% of bacteria belong to this group (Yang et al., 2009b). Family *Lachnospiraceae* belongs to the phylum Firmicutes, and this group also has host-specific molecular markers. The next generation sequencing technology was used to develop new genetic markers to detect the members of Family *Lachnospiraceae*. Sequence analysis of 37 sewage samples confirmed that *Lachnospiraceae* is ubiquitous in human sewage (Newton et al., 2011). This new genetic marker is named as Lachno2, which is the second most abundant phylotype in the *Lachnospiraceae*. A qPCR Lachno2 assay was developed and used along with *Bacteroidales* HF183 primers. Human fecal pollution in Milwaukee's harbor water was identified through microbial pyrotag sequence analysis of the V6 region of the 16S rRNA gene. However, this Lachno2 assay is at the developing stage. Results of this

study showed Lachno2 does not amplify cattle target. Importantly, the tight correlation between *Bacteroidales* and Lachno2 assay suggests the ability of using both molecular markers to provide strong evidence of human fecal contamination in surface water (Newton et al., 2011).

Unlike PCR based approaches, the metagenomic library with pyro sequencing provides a more complete view of community structure. PCR based libraries and metagenomic libraries usually do not give same results. This may be due to PCR selection, amplifying certain genome fragments over other genome fragments. The Metagenomic libraries were successfully used to analyse the bacterial community structure in the Delaware River; three methods, i.e., 16S rRNA gene library, fluorescence insitu hybridization (FISH) and PCR clonal library methods, were compared. The results obtained from these three methods were different. The Cytophages like group was dominant in the metagenomic library (54%); beta proteobacteria was dominant in the PCR clonal library (50%); while the FISH library had two dominant groups as Beta-proteobacteria (25%) and Actinobacteria (26%). The library coverage increases with the genetic distance for both libraries. At the high genetic distance, both libraries have the same coverage (100%). Coverage of the metagenomic library was higher than coverage of the PCR clonal library at all the levels (Cottrell et al., 2005).

Next generation sequencing of *Bacteroidales* 16S rRNA gene is the current focus of MST studies. It has lower sequencing cost, flexibility, accuracy, easy automation, less time consumption compared with other techniques, and also produces millions of DNA sequences (Ronaghi, 2001; http://genomics.org/index.php/DNA_sequencing). Data analysis is one of the major constrains in the application of next generation sequencing

technique. Unno et al. (2011) developed an online MST tool, PyroMiST (<http://env1.gist.ac.kr/~aeml/MST.html>). Total bacteria or *Bacteroidales* pyrosequencing data are compared against this database to detect fecal contamination in waterways. PyroMist takes less time for data analysis and is user friendly compared with other data analysis methods.

1.8 References

1. **Ahmed, W., J. Stewart, T. Gardner, D. Powell, P. Brooks, D. Sullivan, and N. Tindale.** 2007. Sourcing faecal pollution: A combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments. *Water Res.* **41**:3771-3779.
2. **American Society for Testing and Materials.** 1985. Standard test method for particle-size analysis of soils D 422-63 (1972). Annual Book of ASTM Standards 04.08:117-127. American Society for Testing Materials, Philadelphia.
3. **Anderson, M. L., J. E. Whitlock, and V. J. Harwood.** 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Appl. Environ. Microbiol.* **71**:3041-3048.
4. Applied Biosystem.1.Real time PCR vs. traditional PCR.
http://www6.appliedbiosystems.com/support/tutorials/pdf/rtpcr_vs_tradpcr.pdf
5. Applied Biosystem.2. Real time PCR-understanding Ct.
http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generalddocuments/cms_053906.pdf
6. **Arthur, T. M., X. W. Nou, N. Kalchayanand, J. M. Bosilevac, T. Wheeler, and M. Koohmaraie.** Survival of *Escherichia coli* O157:H7 on cattle hides. *Appl. Environ. Microbiol.* **77**:3002-3008.
7. **Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Dore, J. Weissenbach, S. D. Ehrlich, and P. Bork.** Enterotypes of the human gut microbiome. *Nature* **473**:174-180.
8. **Avery, L. M., A. P. Williams, K. Killham, and D. L. Jones.** 2008. Survival of *Escherichia coli* O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Sci. Total Environ.* **389**:378-385.
9. **Bae, S., and S. Wuertz.** 2009. Discrimination of viable and dead fecal *Bacteroidales* bacteria by quantitative PCR with propidium monoazide. *Appl. Environ. Microbiol.* **75**:2940-2944.

10. **Bae, S., and S. Wuertz.** 2012. Survival of host-associated *Bacteroidales* cells and their relationship with *Enterococcus* spp., *Campylobacter jejuni*, *Salmonella enterica* Serovar *Typhimurium*, and Adenovirus in freshwater microcosms as measured by propidium monoazide-quantitative PCR. *Appl. Environ. Microbiol.* **78**:922-932.
11. **Bell, A., A. C. Layton, L. McKay, D. Williams, R. Gentry, and G. S. Sayler.** 2009. Factors influencing the persistence of fecal *Bacteroides* in stream water. *J. Environ. Qual.* **38**:1224-1232.
12. **Bernhard, A. E., and K. G. Field.** 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**:1587-1594.
13. **Bernhard, A. E., and K. G. Field.** 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **66**:4571-4574.
14. **Beveridge, T. J.** 1999. Structures of Gram-negative cell walls and their derived membrane vesicles. *J. Bacteriol.* **181**:4725-4733.
15. **Bower, P. A., C. O. Scopel, E. T. Jensen, M. M. Depas, and S. L. McLellan.** 2005. Detection of genetic markers of fecal indicator bacteria in Lake Michigan and determination of their relationship to *Escherichia coli* densities using standard microbiological methods. *Appl. Environ. Microbiol.* **71**:8305-8313.
16. **Brennan, F. P., F. Abram, F. A. Chinalia, K. G. Richards, and V. O'Flaherty.** 2010. Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil. *Appl. Environ. Microbiol.* **76**:2175-2180.
17. **Brenner D, J., N.R. Krieg, J.R. Staley, G. Garrity.** 2005. *Bergey's Manual® of Systematic Bacteriology. The Proteobacteria, Part B: The Gammaproteobacteria. Volume 2. ISBN: 978-0-387-24144-9 .*
18. **Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer.** 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**:611-622.
19. **Button, D.K., B. Robertson** 2001. Determination of DNA content of aquatic bacteria by flow cytometry. *Appl. Environ. Microbiol.* **67**:1636-45.
20. **Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky, and S. Ishii.** 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ. Microbiol.* **8**:504-513.

21. **Caldwell, J. M., M. E. Raley, and J. F. Levine.** 2007. Mitochondrial multiplex qPCR as a source tracking method in fecal-contaminated effluents. *Environ. Sci. Technol.* **41**:3277-3283.
22. **Carlos, C., M. Pires, N. Stoppe, E. Hachich, M. Sato, T. Gomes, L. Amaral, and L. Ottoboni.** 2010. *Escherichia coli* phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. *BMC Microbiology* **10**:161.
23. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Appl. Environ. Microbiol.* **67**:1503-1507.
24. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl. Environ. Microbiol.* **69**:1836-1839.
25. **Carson, C. A., J. M. Christiansen, H. Yampara-Iquise, V. W. Benson, C. Baffaut, J. V. Davis, R. R. Broz, W. B. Kurtz, W. M. Rogers, and W. H. Fales.** 2005. Specificity of a *Bacteroides thetaiotaomicron* marker for human Feces. *Appl. Environ. Microbiol.* **71**:4945-4949.
26. **Cerdeno-Tarraga, A. M., S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. G. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell, and J. Parkhill.** 2005. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**:1463-1465.
27. **Chang, B., T. Taguri, K. Sugiyama, J. Amemura-Maekawa, F. Kura, and H. Watanabe.** 2010. Comparison of ethidium monoazide and propidium monoazide for the selective detection of viable *Legionella* cells. *Jpn. J. Infect. Dis.* **63**:119-123.
28. **Chen, K., and L. Pachter.** 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS. Comput. Biol.* **1**:106-112.
29. **Cottrell, M. T., L. A. Waidner, L. Y. Yu, and D. L. Kirchman.** 2005. Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environ. Microbiol.* **7**:1883-1895.
30. **Crabill, C., R. Donald, J. Snelling, R. Foust, and G. Southam.** 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. *Water Res.* **33**:2163-2171.

31. **Craig, D. L., H. J. Fallowfield, and N. J. Cromar.** 2004. Use of macrocosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with in situ measurements. *J. Appl. Microbiol.* **96**:922-930.
32. **Davies, C. M., J. A. H. Long, M. Donald, and N. J. Ashbolt.** 1995. Survival of fecal microorganisms in marine and fresh water sediments. *Appl. Environ. Microbiol.* **61**:1888-1896.
33. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005a. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3179-3183. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005a. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3179-3183.
34. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005b. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3184-3191.
35. **Dick, L. K., E. A. Stelzer, E. E. Bertke, D. L. Fong, and D. M. Stoeckel.** 2010. Relative decay of *Bacteroidales* microbial source tracking markers and Cultivated *Escherichia coli* in Freshwater Microcosms. *Appl. Environ. Microbiol.* **76**:3255-3262.
36. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl. Environ. Microbiol.* **66**:2572-2577.
37. **Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick.** 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**:4008-4008.
38. **Eichmiller, J. J., R. E. Hicks, and M. J. Sadowsky.** 2013. Distribution of genetic markers of fecal pollution on a freshwater sandy shoreline in proximity to wastewater effluent. *Environ. Sci. Technol.* **47**:3395-3402.
39. **Flekna, G., P. Stefanic, M. Wagner, F. J. M. Smulders, S. S. Mozina, and I. Hein.** 2007. Insufficient differentiation of live and dead *Campylobacter jejuni* and *Listeria monocytogenes* cells by ethidium monoazide (EMA) compromises EMA/real-time PCR. *Res. Microbiol* **158**:405-412.

40. **Flint, K. P.** 1987. The long-term survival of *Escherichia coli* in river water. *J. Appl. Bacteriol.* **63**:261-270.
41. **Fong, T. T., and E. K. Lipp.** 2005. Enteric viruses of humans and animals in aquatic environments: Health risks, detection, and potential water quality assessment tools. *Microbiol. Molecul. Biol. Rev.* **69**:357-371.
42. **Franz, E., A. V. Semenov, A. J. Termorshuizen, O. J. de Vos, J. G. Bokhorst, and A. H. C. van Bruggen.** 2008. Manure-amended soil characteristics affecting the survival of *E. coli* O157:H7 in 36 Dutch soils. *Environ. Microb.* **10**:313-327
43. **Fu, L. L., J. B. Shuai, Y. B. Wang, H. J. Ma, and J. R. Li.** Temporal genetic variability and host sources of *Escherichia coli* associated with fecal pollution from domesticated animals in the shellfish culture environment of Xiangshan Bay, East China Sea. *Environ. Pollut.* **159**:2808-2814.
44. **Gawler, A. H., J. E. Beecher, J. Brandao, N. M. Carroll, L. Falcao, M. Gourmelon, B. Masterson, B. Nunes, J. Porter, A. Rince, R. Rodrigues, M. Thorp, J. M. Walters, and W. G. Meijer.** 2007. Validation of host-specific *Bacteroidales* 16S rRNA genes as markers to determine the origin of faecal pollution in Atlantic Rim countries of the European Union. *Water Res.* **41**:3780-3784.
45. **Geldreich, E. E., H. F. Clark, C. B. Huff, and L. C. Best.** 1965. Fecal coliform organism medium for the membrane filter technique. *J. Am. Water Works Assoc.* **57**:208-214.
46. **Geldreic E, and B. A. Kenner.** 1969. Concepts of fecal Streptococci in stream pollution. *J. Water Pollut. Control. Fed.* **41**:R336-341.
47. **Gourmelon, M., M. P. Caprais, C. Le Mennec, S. Mieszkin, C. Ponthoreau, and M. Gendronneau.** 2007. Application of library-independent microbial source tracking methods for identifying the sources of faecal contamination in coastal areas. *Environ. Sci. Technol.* **61**:1401-1409.
48. **Guan, Z. P., Y. Jiang, F. Gao, L. Zhang, G. H. Zhou, and Z. J. Guan.** 2013. Rapid and simultaneous analysis of five foodborne pathogenic bacteria using multiplex PCR. *Eur. Food Res. Technol.* **237**:627-637.
49. **Hagerdorn, C., K. Porter, and A. H. Chapman.** 2003. Bacterial source tracking to identify sources of fecal pollution in the Potomac and Anacostia Rivers and Rock Creek, Washington, D.C. Final project report. EPA/600-R-05-064.
50. **Haile, R. W., J. S. Witte, M. Gold, R. Cressey, C. McGee, R. C. Millikan, A. Glasser, N. Harawa, C. Ervin, P. Harmon, J. Harper, J. Dermand, J. Alamillo, K. Barrett, M. Nides, and G. Y. Wang.** 1999. The health effects of

- swimming in ocean water contaminated by storm drain runoff. *Epidemiology*. **10**:355-363.
51. **Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman.** 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem. Biol.* **5**:R245-R249.
 52. **Harwood, V. J., J. Whitlock and V. Withington.** 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. *Appl. Environ. Microbiol.* **66**:3698-3704.
 53. **Haugland, R. A., S. C. Sieftring, L. J. Wymer, K. P. Brenner, and A. P. Dufour.** 2005. Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Res.* **39**:559-568.
 54. **Haugland, R. A., M. Varma, M. Sivaganesan, C. Kelty, L. Peed, and O. C. Shanks.** 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected *Bacteroidales* species and human fecal waste by qPCR. *Syst. Appl. Microbiol.* **33**:348-357.
 55. **Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl. Environ. Microbiol.* **72**:612-621.
 56. **Jenkins, M. W., S. Tiwari, M. Lorente, C. M. Gichaba, and S. Wuertz.** 2009. Identifying human and livestock sources of fecal contamination in Kenya with host-specific *Bacteroidales* assays. *Water Res.* **43**:4956-4966.
 57. **Jiang, H. X., D. H. Lu, Z. L. Chen, X. M. Wang, J. R. Chen, Y. H. Liu, X. P. Liao, J. H. Liu, and Z. L. Zeng.** 2000. High prevalence and widespread distribution of multi-resistant *Escherichia coli* isolates in pigs and poultry in China. *Vet. J.* **187**:99-103.
 58. **Jimenez-Clavero, M. A., E. Escribano-Romero, C. Mansilla, N. Gomez, L. Cordoba, N. Roblas, F. Ponz, V. Ley, and J. C. Saiz.** 2005. Survey of bovine enterovirus in biological and environmental samples by a highly sensitive real-time reverse transcription-PCR. *Appl. Environ. Microb.* **71**:3536-3543.
 59. **Josephson, K. L., C. P. Gerba, and I. L. Pepper.** 1993. Polymerase chain reaction detection of nonviable bacterial pathogens. *Appl. Environ. Microbiol.* **59**:3513-3515.
 60. **Kanagawa, T.** 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* **96**:317-323.

61. **Kato, M., K. Tanaka, Y. Muto, K. Watanabe, and K. Ueno.** 2002. Effects of *Bacteroides* phosphosphingolipids on murine neutrophils. *Anaerobe* **8**:23-28.
62. **Kell, D. B., A. S. Kaprelyants, D. H. Weichart, C. R. Harwood, and M. R. Barer.** 1998. Viability and activity in readily culturable bacteria: a review and discussion of the practical issues. *Anton. Leeuw. Int. J. G* **73**:169-187.
63. **Kildare, B. J., C. M. Leutenegger, B. S. McSwain, D. G. Bambic, V. B. Rajal, and S. Wuertz.** 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: A Bayesian approach. *Water Res.* **41**:3701-3715.
64. **Klein, D. A.** 2007. Seeking microbial communities in nature: a postgenomic perspective. *Microbe* **2**:591-595.
65. **Kobayashi, H., M. Oethinger, M. J. Tuohy, G. S. Hall, and T. W. Bauer.** 2009. Improving clinical significance of PCR: SSE of propidium monoazide to distinguish viable from dead *Staphylococcus aureus* and *Staphylococcus epidermidis*. *J. Orthop. Res.* **27**:1243-1247.
66. **Koplin D.W., E. T. F., M.T. Meyer, E.M. Thurman, S.D Zaugg, L.B. Barber and H.T Buxton.** 2002. Pharmaceuticals, hormones, and other organic wastewater concentrations in the U.S. streams, 1999-2000: A national reconnaissance. *Environ. Sci. Technol.* **36**:1202-1211.
67. **Kramer, M., N. Obermajer, B. B. Matijasic, I. Rogelj, and V. Kmetec.** 2009. Quantification of live and dead probiotic bacteria in lyophilised product by qPCR and by flow cytometry. *Appl. Microbiol. Biotechnol.* **84**:1137-1147.
68. **Kreader, C. A.** 1998. Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Appl. Environ. Microbiol.* **64**:4103-4105
69. **Krieg, N. R. L., W., W.B Whitman, B.P. Hedlund, B.J. Paster, J.T. Staley, N. Ward, D. Brown, A. Parte,** 2011. *Bergey's Manual of Systematic Bacteriology*. **4**. ISBN 0-387-95042-6.
70. **Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V. K. Sharma, T. P. Srivastava, T. D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D. S. Ehrlich, K. Itoh, T. Takagi, Y. Sakaki, T. Hayashi, and M. Hattori.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**:169-181.
71. **Lamendella, R., J. W. S. Domingo, D. B. Oerther, J. R. Vogel, and D. M. Stoeckel.** 2007. Assessment of fecal pollution sources in a small northern-plains

- watershed using PCR and phylogenetic analyses of *Bacteroidetes* 16S rRNA gene. FEMS Microbiol. Ecol. **59**:651-660.
72. **Lamendella, R., N. VerBerkmoes, and J. K. Jansson.** 2012. 'Omics' of the mammalian gut - new insights into function. Curr. Opin. Biotechnol. **23**:491-500.
 73. **Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Saylor.** 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based qPCR assays for estimation of total, human, and bovine fecal pollution in water. Appl. Environ. Microbiol. **72**:4214-4224.
 74. **Lee, C. M., T. Y. Lin, C. C. Lin, G. A. Kohbodi, A. Bhatti, R. Lee, and J. A. Jay.** 2006. Persistence of fecal indicator bacteria in Santa Monica Bay beach sediments. Water Res. **40**:2593-2602.
 75. **Lee, Y. J., M. Molina, J. W. S. Domingo, J. D. Willis, M. Cyterski, D. M. Endale, and O. C. Shanks.** 2008. Temporal assessment of the impact of exposure to cow feces in two watersheds by multiple host-specific PCR assays. Appl. Environ. Microbiol. **74**:6839-6847.
 76. Life technologies. Real time PCR handbook.
http://find.lifetechnologies.com/Global/FileLib/qPCR/RealTimePCR_Handbook_Update_FLR.pdf
 77. **Lu, J. R., J. S. Domingo, and O. C. Shanks.** 2007. Identification of chicken-specific fecal microbial sequences using a metagenomic approach. Water Res. **41**:3561-3574.
 78. **Lu, J., J. W. Santo Domingo, R. Lamendella, T. Edge, and S. Hill.** 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl. and Environ. Microbiol. **74**:3969-3976.
 79. **Lu, J., J. W. Santo Domingo, S. Hill, and T. A. Edge.** 2009. Microbial diversity and host-specific sequences of Canada Goose feces. Appl. Environ. Microbiol. **75**:5919-5926.
 80. **Mancini, C., and R. J. Behme.** 1977. Transfer of multiple antibiotic resistance from Bacteroidales-fragilis to *Escherichia coli*. J. Infect. Dis. **136**:597-600.
 81. **McGee, P., D. J. Bolton, J. J. Sheridan, B. Earley, G. Kelly, and N. Leonard.** 2002. Survival of *Escherichia coli* O157:H7 in farm water: its role as a vector in the transmission of the organism within herds. J. Appl. Microbiol. **93**:706-713.
 82. **McLellan, S. L., A. D. Daniels, and A. K. Salmore.** 2001. Clonal populations of thermotolerant Enterobacteriaceae in recreational water and their potential

- interference with fecal *Escherichia coli* counts. Appl. Environ. Microbiol. **67**:4934-4938.
83. **Mehra, S., and W. S. Hu.** 2005. A kinetic model of quantitative real-time polymerase chain reaction. Biotechnol. Bioeng. **91**:848-860.
 84. **Mieszkin, S., J. F. Yala, R. Joubrel, and M. Gourmelon.** 2010. Phylogenetic analysis of *Bacteroidales* 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by qPCR. J. Appl. Microbiol. **108**:974-984.
 85. **Mullis, K.** 1990. The unusual origin of the polymerase chain reaction. Sci. Am. 56-65.
 86. **Murugan, K., P. Prabhakaran, S. Al-Sohaibani, and K. Sekar.** 2013. Identification of source of faecal pollution of Tirumanimuttar River, Tamilnadu, India using microbial source tracking. Environ. Monit. Assess. **184**:6001-6012.
 87. **Nebe-von-Caron, G., P. J. Stephens, C. J. Hewitt, J. R. Powell, and R. A. Badley.** 2000. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. J. Microbiol. Methods **42**:97-114.
 88. **Newton, R. J., J. L. VandeWalle, M. A. Borchardt, M. H. Gorelick, and S. L. McLellan.** 2011. *Lachnospiraceae* and *Bacteroidales* Alternative Fecal Indicators Reveal Chronic Human Sewage Contamination in an Urban Harbor. Appl. Environ. Microbiol. **77**:6972-6981.
 89. **Nocker, A., and A. K. Camper.** 2006a. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. Appl. Environ. Microbiol. **72**:1997-2004.
 90. **Nocker, A., C. Y. Cheung, and A. K. Camper.** 2006b. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J. Microbiol. Methods **67**:310-320.
 91. **Nocker, A., K. E. Sossa, and A. K. Camper.** 2007. Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. J. Microbiol. Methods **70**:252-260.
 92. **Nocker, A., P. Sossa-Fernandez, M. D. Burr, and A. K. Camper.** 2007. Use of Propidium Monoazide for Live/Dead Distinction in Microbial Ecology. Appl. Environ. Microbiol. **73**:5111-5117.

93. **Nogva, H. K., S. M. Dromtorp, H. Nissen, and K. Rudi.** 2003. Ethidium monoazide for DNA-based differentiation of viable and dead bacteria by 5'-nuclease PCR. *Biotechniques* **34**:804-808.
94. **Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* **74**:890-901.
95. **Okabe, S., and Y. Shimazu.** 2007. Persistence of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers in environmental waters: effects of temperature and salinity. *Appl. Microbiol. Biotech.* **76**:935-944.
96. **Oram, B.** 2013. Total Coliform Bacteria Testing. <http://www.water-research.net/coliform.htm>
97. **Pace, N. R.** 2009. Mapping the Tree of Life: Progress and Prospects. *Appl. Microbiol. Biotechnol.* **73**:565-576.
98. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**:3724-3730.
99. **Parveen, S., N. C. Hodge, R. E. Stall, S. R. Farrah, and M. L. Tamplin.** 2001. Phenotypic and genotypic characterization of human and nonhuman *Escherichia coli*. *Water Res.* **35**:379-386.
100. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin.** 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for d Human and nNonhuman sources of fecal pollution. *Appl. Environ. Microbiol.* **65**:3142-3147.
101. **Pote, J., L. Haller, R. Kottelat, V. Sastre, P. Arpagaus, and W. Wildi.** 2009. Persistence and growth of faecal culturable bacterial indicators in water column and sediments of Vidy Bay, Lake Geneva, Switzerland. *J. Environ. Science-China* **21**:62-69.
102. **Premier Biosoft.** http://www.premierbiosoft.com/tech_notes/multiplex-pcr.html).
103. **Pruss, A.** 1998. Review of epidemiological studies on health effects from exposure to recreational water. *Int. J. Epidemiol.* **27**:1-9.
104. **Qin, J. J., R. Q. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. H. Li, J. M. Xu, S. C. Li, D. F. Li, J. J. Cao, B. Wang, H. Q. Liang, H. S. Zheng, Y. L. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. M. Zhu, C. Yu, S. T. Li, M. Jian, Y. Zhou, Y. R. Li, X. Q. Zhang,**

- S. G. Li, N. Qin, H. M. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, and S. D. Ehrlich. 2010.** A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**:59-U70.
105. **Rawsthorne, H., C. N. Dock, and L. A. Jaykus.** 2009. PCR-based method using propidium monoazide to distinguish viable from nonviable *Bacillus subtilis* spores. *Appl. Environ. Microbiol.* **75**:2936-2939.
106. **Reynolds, K. A.** 2003. Coliform bacteria: A failed indicator of water quality. <http://www.wcponline.com/column.cfm?T=T&ID=2349>
107. **Ronaghi, M.** 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**:3-11.
108. **Roux, K. H.** 1995. Optimization and troubleshooting in PCR PCR Meth. Appl. **4**:S185-S194.**Roux, K. H.** 1995.
109. **Rudi, K., H. K. Hoidal, T. Katla, B. K. Johansen, J. Nordal, and K. S. Jakobsen.** 2004. Direct qPCR quantification of *Campylobacter jejuni* in chicken fecal and cecal samples by integrate cell concentration and DNA purification. *Appl. Environ. Microbiol.* **70**:790-797.
110. **Saiki RK, S. S., Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N.** 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**:1350-4.
111. **Salyers, A. A.** 1984. *Bacteroides* of the human lower intestine track. *Annu. Rev. Microbiol.* **38**:293-313.
112. **Sanger, F., and A. R. Coulson.** 1975. Rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**:441-46.
113. **Santoro, A. E., and A. B. Boehm.** 2007. Frequent occurrence of the human-specific *Bacteroides* fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators. *Environ. Microbiol.* **9**:2038-2049.
114. **Sargeant, D, W. R. K., and Scott Collyard, W. S. D. o. Ecology, and W.-. Olympia.** 2011. Review and Critique of Current Microbial Source Tracking (MST) Techniques. www.ecy.wa.gov/biblio/1103038.html
115. **Savichtcheva, O., Okayama N., Okabe S.** 2007. Relationships between *Bacteroides* 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Res.* **41**:3615-3628.

116. **Schmittgen, T. D., and K. J. Livak.** 2008. Analyzing qPCR data by the comparative C-T method. *Nature Protocols* **3**:1101-1108.
117. **Schulz, C. J., and G. W. Childers.** 2011. Fecal Bacteroidales diversity and decay in response to variations in temperature and salinity. *Appl. and Environ. Microbiol.* **77**:2563-257.
118. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002. Microbial source tracking: current methodology and future directions. *Appl. Environ. Microbiol.* **68**:5796-5803.
119. **Semenov, A. V., A. H. C. van Bruggen, L. van Overbeek, A. J. Termorshuizen, and A. M. Semenov.** 2007. Influence of temperature fluctuations on *Escherichia coli* O157:H7 and *Salmonella enterica* serovar Typhimurium in cow manure. *FEMS Microbiol. Ecol.* **60**:419-428.
120. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. D. Siciliano.** 2005. Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with qPCR for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* **7**:249-259.
121. **Shanks, O. C., E. Atikovic, A. D. Blackwood, J. R. Lu, R. T. Noble, J. S. Domingo, S. Seifring, M. Sivaganesan, and R. A. Haugland.** 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **74**:745-752.
122. **Shanks, O. C., C. A. Kelty, M. Sivaganesan, M. Varma, and R. A. Haugland.** 2009. Quantitative PCR for genetic markers of human fecal pollution. *Appl. Environ. Microbiol.* **75**:5507-5513.
123. **Shanks, O. C., K. White, C. A. Kelty, S. Hayes, M. Sivaganesan, M. Jenkins, M. Varma, and R. A. Haugland.** 2010. Performance assessment PCR-based assays targeting *Bacteroidales* genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **76**:1359-1366.
124. **Shanks, O. C., C. A. Kelty, S. Archibeque, M. Jenkins, R. J. Newton, S. L. McLellan, S. M. Huse, and M. L. Sogin.** 2011. Community structures of fecal bacteria in cattle from different animal feeding operations. *Appl. Environ. Microbiol.* **77**:2992-3001.
125. **Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner.** 2002. Microbial source tracking: State of the science. *Environ. Sci. Technol.* **36**:5279-5288.
126. **Sinton, L. W., C. H. Hall, P. A. Lynch, and R. J. Davies-Colley.** 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste

- stabilization pond effluent in fresh and saline waters. *Appl. Environ. Microbiol.* **68**:1122-1131.
127. **Sivaganesan, M., R. A. Haugland, E. C. Chern, and O. C. Shanks.** 2010. Improved strategies and optimization of calibration models for qPCR absolute quantification. *Water Res.* **44**:4726-4735.
 128. **Stiles, N.** 2003 Fall. Bacterial source tracking uncovers culprits of contamination of watersheds. *Small Flows Quarterly* **4.4**:24-29.
 129. **Souza, C., M. Rocha, A. Valera, and L.E. Eguiarte.** 1999. Genetic structure of natural population of *Escherichia coli* in wild hosts on different continents. *Appl. Environ. Microbiol.* **65**:3373-3385.
 130. **Tap, J., S. Mondot, F. Levenez, E. Pelletier, C. Caron, J. P. Furet, E. Ugarte, R. Munoz-Tamayo, D. L. E. Paslier, R. Nalin, J. Dore, and M. Leclerc.** 2009. Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**:2574-2584.
 131. **Tally, F. P., A. B. Onderdonk, D. R. Snyderman, S. L. Gorbach, and M. H. Malamy.** 1979. In vivo transfer of clindamycin resistance in *Bacteroides*. *Clin. Res.* **27**:A481-A481.
 132. **TechNotes 9.** 2013. Monitoring for microbial pathogens and indicators-National Nonpoint Source Monitoring Program (http://www.bae.ncsu.edu/programs/extension/wqg/319monitoring/TechNotes/technotes9_monitoring_microbial_pathogens.pdf).
 133. **Turnbaugh, P. J., V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight, and J. I. Gordon.** 2009. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Sci. Transl. Med.* **1**.DOI: 6ra1410.1126/scitranslmed.3000322.
 134. **Unno, T., D. Y. W. Di, J. Jang, Y. S. Suh, M. J. Sadowsky, and H. G. Hur.** Integrated Online System for a Pyrosequencing-Based Microbial Source Tracking Method that Targets Bacteroidetes 16S rDNA. *Environ. Sci. & Technol.* **46**:93-98.
 135. **USEPA.** 2013. Watershed Assessment, Tracking & Environmental Results. http://iaspub.epa.gov/waters10/attains_nation_cy.control#total_assessed_waters
 136. **USEPA.** 2012. Recreational Water Quality Criteria. **820-F-12-061.** <http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/factsheet2012.pdf>.

137. **USEPA.** 2005. Microbial source tracking guide document. **EPA/600-R-05-064.**
138. **USEPA.** 2002a. Method 1604: Total coliforms and *Escherichia coli* in water by membrane filtration using a simultaneous detection technique (MI Medium). **EPA 821-R-02-024.**
139. **USEPA.** 2002b. Method 1603: *Escherichia coli* (*E. coli*) in Water by Membrane Filtration Using Modified membrane-Thermotolerant *Escherichia coli* Agar (Modified mTEC). **EPA 821-R-02-023.**
140. **USEPA.** 2002c. Method 1680: Fecal Coliforms in Biosolids by Multiple-Tube Fermentation Procedures. **EPA-821-R-02-026.**
141. **USEPA.** 1986. Ambient Water Quality Criteria EPA 440/5-84-002.
142. **Van De Werfhorst, L. C., B. Sercu, and P. A. Holden.** 2011. Comparison of the Host Specificities of Two *Bacteroidales* Quantitative PCR Assays Used for Tracking Human Fecal Contamination. *Appl Environ Microbiol* **77**:6258-6260.
143. **Van Elsas, J. D., A. V. Semenov, R. Costa, and J. T. Trevors.** 2011. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* **5**:173-183.
144. **Varma, M., R. Field, M. Stinson, B. Rukovets, L. Wymer, and R. Haugland.** 2009. Quantitative qPCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Res* **43**:4790-4801.
145. **Vesper, S. J., M. E. Rogers, A. N. Neely, and R. A. Haugland.** 2007. Opportunistic *Aspergillus* pathogens measured in home and hospital tap water by quantitative PCR (QPCR). *J Water Health* **5**:427-431.
146. **Vogel, J. R., D. M. Stoeckel, R. Lamendella, R. B. Zelt, J. W. S. Domingo, S. R. Walker, and D. B. Oerther.** 2007. Identifying fecal sources in a selected catchment reach using multiple source-tracking tools. *J Environ. Qual.* **36**:718-729.
147. **Wade, T. J., N. Pai, J. N. S. Eisenberg, and J. M. Colford.** 2003. Do US Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environ Health Perspect* **111**:1102-1109.
148. **Wagner, A. O., C. Malin, B. A. Knapp, and P. Illmer.** 2008. Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. *Appl. Environ. Microbiol.* **74**:2537-2539.

149. **Walters, S. P., and K. G. Field.** 2009. Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. *Environ. Microbiol.* **11**:1410-1421.
150. **Weidhaas, J. L., T. W. Macbeth, R. L. Olsen, and V. J. Harwood.** 2011. Correlation of Quantitative PCR for a Poultry-Specific *Brevibacterium* Marker Gene with Bacterial and Chemical Indicators of Water Pollution in a Watershed Impacted by Land Application of Poultry Litter. *Appl. Environ. Microbiol.* **77**:2094-2102.
151. **Wexler, H. M.** 2007. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**:593-621.
152. **White, A. P., K. A. Sibley, C. D. Sibley, J. D. Wasmuth, R. Schaefer, M. G. Surette, T. A. Edge, and N. F. Neumann.** 2011. Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. *Appl. Environ. Microbiol.* **77**:7620-7632.
153. **WHO.** 2001. Water quality: Guidelines, standards and health. http://www.who.int/water_sanitation_health/dwq/whoiwa/en/
154. **WHO.**2013. http://www.who.int/water_sanitation_health/diseases/diarrhoea/en/
155. **Wiggins, B. A., P. C. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner.** 2003. Antibiotic resistant analysis for representativeness testing of multiwatershed libraries. *Appl. Environ. Microbiol.* **69**:3399-3405.
156. **Wijesinghe, R. U., Y. Feng, C. W. Wood.** 2009. Population dynamics and genetic variability of *Escherichia coli* in a mixed land-use watershed. *J. Water Health* **07**:484-496.
157. **Woese, C. R.** 1987. Bacterial Evolution. *Microbiol. Rev.* **51**:221-271.
158. **Wu, G. D., J. Chen, C. Hoffmann, K. Bittinger, Y. Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Z. Li, F. D. Bushman, and J. D. Lewis.** Linking long-term dietary patterns with gut microbial enterotypes. *Science* **333**:105-108.
159. **Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. I. Gordon.** 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**:2074-2076.

160. **Yang, X., L. Xie, Y. X. Li, and C. C. Wei.** 2009a. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *Plos One* **4**:e6074.doi:10.1371/journal.pone.0006074.
161. **Yang, L., X. Lu, C. W. Nossa, F. Francois, R. M. Peek, and Z. Pei.** 2009b. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. *Gastroenterology* **137**:588-597.

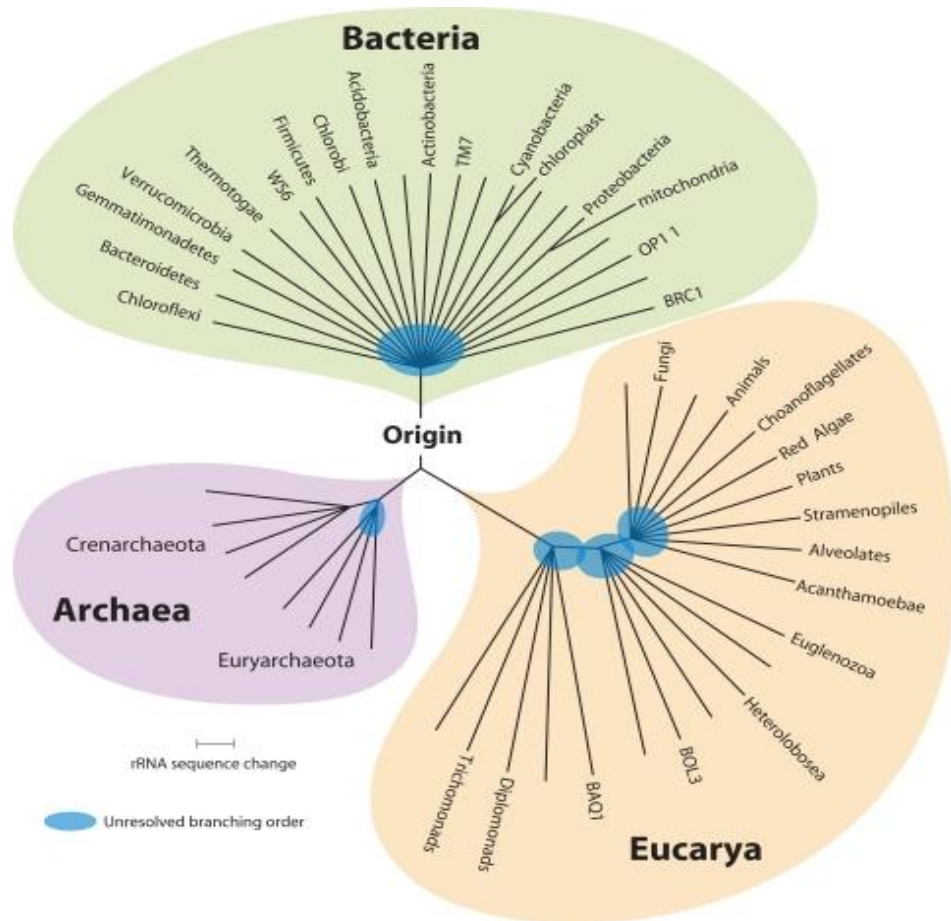


Fig. 1.1 Eubacterial phylogenetic tree based on 16S rRNA gene (Pace 2009)

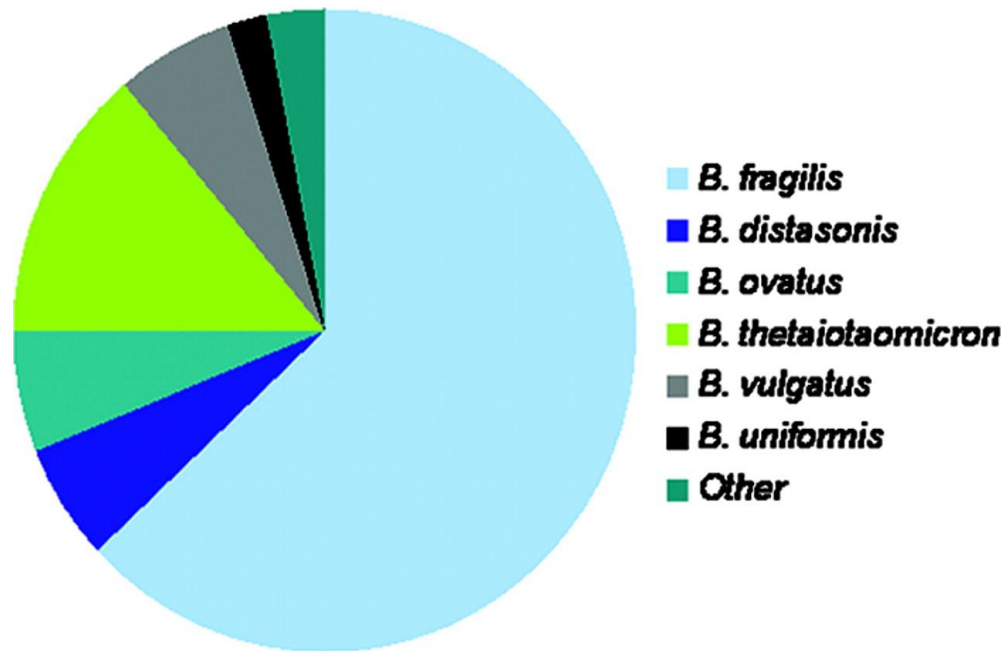


Fig. 1.2. Proportions of *Bacteroides* species found in clinical samples (Wexler, 2007)

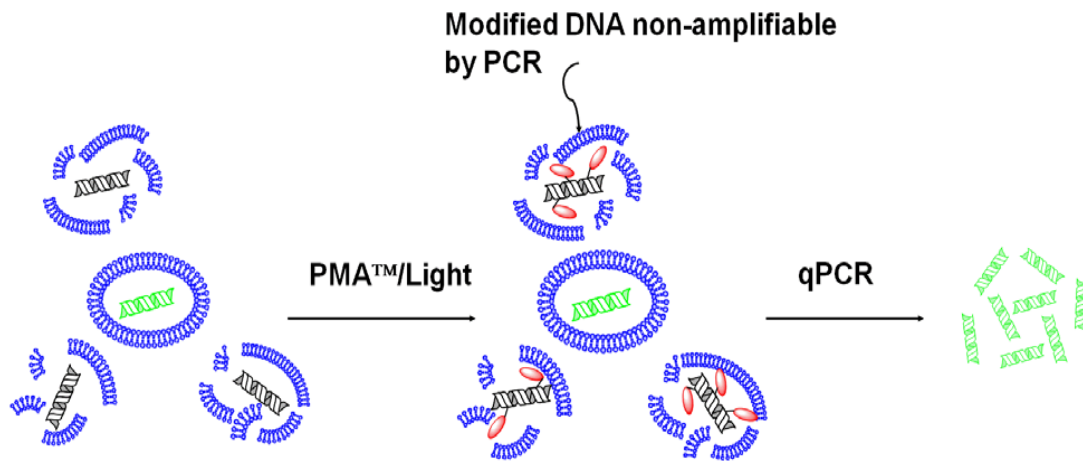


Fig. 1.3. Principle of the PMA technology (Biotium product flyer http://www.biotium.com/product/applications/search/price_and_info.asp?itm=40013).

2. Comparison of library dependent and library independent methods for identifying sources of fecal contamination

Abstract

Library-based and non-library-based approaches have been used in microbial source tracking (MST) studies to identify fecal pollution sources. In this study, one method was selected from each category and their results compared to identify the fecal pollution sources in a mixed land-use watershed. Water samples were collected over a one-year period from six locations in the Catoma Creek watershed, Alabama. DNA extracted from the water samples was amplified with general *Bacteroidales*, human- and cattle-associated primers using end-point PCR and quantitative PCR (qPCR). End-point PCR found human and cattle markers in 26.5% and 28.1% of the samples, respectively, while qPCR detected human markers in 64.0% of the samples. A total of 1,050 *E. coli* rep-PCR DNA fingerprints obtained from water samples were compared against the known source library, which consisted of 945 unique *E. coli* DNA fingerprints from nine animal host groups. The library-based method detected human and cattle fecal pollution in 93.8% and 73.4% of the samples, respectively. Based on these results, both methods can successfully be used to detect the sources of fecal contamination, but the use of multiple methods such as the combination of library-based and non-library-based methods used here may produce more robust and reliable results.

2. 1. Introduction

Fecal pollution of surface water adversely affects both recreational and economic activities. In 2011, the United States experienced 26,921 beach closings and advisories at coastal, inland and Great Lakes beaches and bays due to elevated fecal indicator bacteria counts (Natural Resources Defence Council, 2012); 43% of beaches had one or more advisories (USEPA, 2012). Furthermore, fecal polluted water unfavourably affects the shellfish industry because shellfish concentrate bacteria and virus in their tissues, rendering them unsafe for consumption. Fecal contamination of water resources adds financial burdens to the national economy for treating people with swimming related diseases, dealing with outbreaks due to fecal contaminated food, in this case shellfish, and economic losses due to the closure of beaches.

Currently, molecular based techniques are popular for the detection of fecal contamination sources. These molecular based techniques can be divided into two main categories: library based approaches and non-library based approaches. The main advantage of a library-based approach is that it facilitates the detection of any warm-blooded animal group that could be responsible for fecal pollution in a watershed, although the higher rate of false positives can be a major drawback for this technique, which is also time consuming and technically demanding (Harwood et al., 2003). Furthermore, the size of a library greatly affects the accuracy and percentage of source identification, with larger libraries providing better source identification and more accurate results than smaller libraries. Temporal variability and spatial variability also

need to be addressed with this approach (Jenkins et al., 2003; Hartel et al., 2002; Scott et al., 2002; Gordon, 2001). Once a DNA fingerprinting library is developed for a particular watershed, the library can be used to detect the sources of fecal contamination in that watershed. It is questionable, however, whether the same library can be used to detect fecal contamination sources in another watershed some distance away. Library-based microbial source tracking methods were particularly popular in the late 1990s and early 2000s (Carson et al., 2003; Dombek et al., 2000; Harwood et al., 2000; McLellan et al., 2003; Wijesinghe et al., 2009).

A number of non-library based methods were introduced in the early 2000s and have subsequently been used by many research groups (Bernhard and Field, 2000a,b; Shanks et al., 2007; Layton et al., 2006; Okabe et al., 2007; Dick et al., 2005a; Kildare et al., 2007; Savichtcheva 2007). Nucleic acid based non-library approaches involve the direct detection of host-associated genetic markers present in the water samples. Members of the order *Bacteroidales* are the most promising target organisms for this non-library based approach because of their shorter survival period in the secondary environment and abundance in fecal matter. Although about 25–30% of the microbial population in human intestines belong to the phylum Bacteroidetes (Yang et al., 2009), members of *Bacteroidales* is not abundant in most of the avian microbiome. *Clostridium* and *Bacillus* species are more abundant in chickens and gull than *Bacteroidales* (Lu et al., 2007, 2008; Lamendella et al., 2007), and goose fecal matter has been found to consist of 38% bacilli, 33.7% clostridia, and 10.1% *Bacteroidales* (Lu et al., 2009).

Detection of host-associated genetic markers is faster, more efficient, cheaper and less time consuming than library-based methods. This approach allows the detection of

sources of fecal contamination within 24 hours of sample collection compared with days to weeks for the library-based approach. However, the main disadvantage is that host-associated markers are not available for all the animal groups of interest, and some assays show false negative results due to PCR inhibition. In addition, some currently available markers show spatial variability (Shanks et al., 2006; Lamendella et al., 2012) and cross reactivity with other host groups (Shanks et al., 2009, 2010; Fremaux et al., 2009).

This study was conducted in Alabama, where 10,907 of the state's 77,272 stream/river miles are currently monitored for water quality. In 2010, 878 river/stream miles were on Alabama's 303(d) list of impaired and threatened waters (stream/river segments and lakes) due to elevated fecal bacteria, while 26.7% of monitored river/stream miles did not support their designated usage (ADEM, 2010). According to a beach water quality report published by the Natural Resources Defence Council (2012), Alabama beaches were ranked 13th out of the 30 coastal states for water quality, with 6% of beach water samples failing to achieve national recommended beach water quality standards. Storm water runoff, sewage overflow, agricultural runoff and other sources such as wildlife, septic tank malfunction and boating waste are the main contributors to this contamination (Natural Resources Defence Council, 2012). The accurate detection of sources of fecal contamination is essential, especially in a watershed with mixed land uses, because without it fecal contamination cannot be controlled effectively. To determine the appropriate method for microbial source tracking, a method comparison study was therefore conducted for the Catoma Creek watershed in Montgomery, Alabama, where humans, livestock, pets and wildlife represent the possible sources of fecal contamination. The objectives of this study were to:

- 1) evaluate a previously developed *E. coli* known source repetitive sequence-based polymerase chain reaction (rep-PCR) DNA fingerprint library, and identify the sources of fecal contamination in Catoma Creek water using this known source library;
- 2) detect general *Bacteroidales*, human- and cattle-associated molecular markers in this watershed using end-point PCR and quantitative PCR (qPCR);
- 3) compare the performance of library based and non-library based methods for identifying sources of fecal contamination in this watershed.

2.2 Materials and Methods

Study area. The Catoma Creek watershed (Fig. 2.1) is located in Montgomery County, Alabama, most of which lies in the Blackland Prairie physiographic region. Catoma Creek, a tributary of the Alabama River, has a linear length of 68 km and drains an area of 932 km², covering more than 50% of land in the county. The present use classification of Catoma Creek is “Fish and Wildlife.” A 37-km segment of the creek from the Alabama River to Ramer Creek is listed on the Alabama 303(d) List due to nutrient enrichment and elevated fecal coliform concentrations (ADEM, 2002). The watershed mainly consists of forest (54.5%), agriculture (36.2%) and urban (9.3%) land uses. There are several potential sources of fecal contamination in the watershed, including septic tank malfunction in the rural areas, leakage from sewer lines, urban runoff, runoff from pasture and agriculture fields, and wildlife.

Collection of fecal samples and isolation of *E. coli* from fecal samples. Fresh fecal samples were collected from humans, horses, dogs, cattle, deer, wild turkeys, waterfowl, beavers and chickens within Montgomery County using BBL culture swabs (BD

Biosciences, Sparks, MD). Samples were kept on ice until transported to the laboratory and processed within one day. These samples were first streaked on 100 x 15 mm MacConkey agar plates and incubated overnight at 37°C. Dark pink and morphologically different single colonies (3-6 per plate) were selected and streaked again on MacConkey agar. Further purification and verification of *E. coli* were performed as described in Appendices (S. 2.1). In addition to fecal samples donated by human volunteers, *E. coli* isolates from anonymous individuals were also obtained from a Montgomery hospital. Table 2.1 shows the number of fecal samples obtained from each source group, number of *E. coli* DNA fingerprints developed and number of unique DNA fingerprints in each source group used for library development.

Sewage samples collection for host specificity testing. A total of four raw sewage samples were collected from sewage treatment plants in Auburn, Alabama. Samples were collected in sterile 1-L high-density polyethylene (HDPE) bottles, kept on ice, transported to the laboratory and processed on the same day. Sewage samples were concentrated by centrifugation. Briefly, two 45 mL samples from each bottle were centrifuged at 2750 g for 15 minutes at 4°C. The supernatant was decanted and pellets in each tube were resuspended in 4.5 mL of phosphate buffered saline (PBS). The replicates were combined and samples were stored at -20°C until DNA extraction.

Cattle fecal samples collection for host specificity testing. Cattle fecal samples were collected from the Wilson Beef Teaching Center at Auburn University and other privately owned cattle farms near Auburn, AL. Samples were collected in sterile plastic bags, kept on ice and transported to the laboratory. A subsample (10 g) of fecal matter was mixed with 10 mL of sterile water in a 50 mL sterile tube and vortexed for 10 minutes at

maximum speed to homogenize the fecal matter. These samples were stored at -20°C until DNA extraction.

Collection of water samples and *E. coli* enumeration. Water samples were collected monthly from six locations in the watershed for the period March 2007 to February 2008 (Fig. 2.1). One location was on the main stem of Catoma Creek, designated as CW; two locations (RHM and RSP) were on the largest tributary, Ramer Creek; and the remaining locations were on three other tributaries of Catoma Creek: Little Catoma (LT), Baskins Mill Creek (BM), and White Slough (WS). Duplicate water samples were collected using 1-L sterile high density polyethylene bottles, kept on ice, transported to the laboratory, and processed within six hours of collection. Water temperature was measured on site; pH, EC and turbidity were measured in the laboratory. Additional 100 mL samples were collected from each site quarterly and stored at -20°C for nutrient analyses. Total phosphorus and metal ion concentrations (Ca, K, Mg, Al, Zn, Cu, Mn, and Fe) were determined using ICAP (SPECTRO CIROS, Germany), and total organic C and total N were measured with a TOC-V Combustion Analyser (Shimadzu, Columbia, MD). Rainfall data were obtained from the National Oceanic Atmospheric Administration (NOAA)'s website (<http://www.ncdc.noaa.gov>).

E. coli concentrations in the water samples were analysed by membrane filtration, followed by cultivation on modified mTEC agar (USEPA, 2002). Single colonies magenta/red in color were selected, and culture purification performed using MacConkey agar (Difco, Detroit, MI) and CHROMagar (Chromagar Microbiology, Paris, France). Further, *E. coli* identification and preservation are described in Appendices (S.2.1).

Extraction of DNA from fecal and water samples. Water samples were filtered under vacuum on the day of sample collection. From March to July 2007, the 100 mL water samples were filtered through 0.22 µm membrane filters, and from August 2007 to February 2008 the 200 mL water samples were filtered through 0.45 µm membrane filters. The filter membranes were folded and placed in sterile Whirl Pak bags, labelled, and stored at -80°C until DNA extraction. Before DNA extraction, membrane filters were cut into small pieces with a sterile pair of scissors. DNA was extracted using the Ultra DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA) following the manufacturer's instructions. An aliquot (300 µL) of concentrated sewage or cattle fecal slurry was used for the DNA extractions. The DNA concentration was measured using a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE).

***E. coli* rep-PCR DNA fingerprinting.** Repetitive sequence-based PCR (rep-PCR) DNA fingerprints of the *E. coli* isolates were obtained using the BOX A1R primer (5'-CTA CGG CAA GGC GAC GCT GAC G-3') (Rademaker and de Bruijn, 1997) and *E. coli* whole cells as the templates for PCR. PCR reactions and gel electrophoresis were performed according to Wijesinghe et al. (2009) (S.2.2).

Each gel contained a no template control, *E. coli* strain K-12 (ATCC 10798) as the positive control and 1 kb plus ladders (Fig. 2.2). The positive control was used to determine the repeatability and overall performance of the PCR. *E. coli* K-12 rep-PCR DNA fingerprints were analysed with cluster analysis, the Cosine similarity coefficient was used to develop the clusters, and the dendrogram was developed utilizing the UPGMA method. If a positive control showed less than 90% similarity with the rest of the positive controls, that gel was considered to be a lower performing gel and discarded.

If bands were observed in the negative controls or were un-reproducible for the positives controls, the PCR runs were repeated.

Amplification of *Bacteroidales* gene markers using end-point PCR. General

Bacteroidales and human- and cattle-specific *Bacteroidales* genetic markers were amplified with Bac32, HF183 and CowM3 primers (Table 2.2) as described by Bernhard and Field (2000a,b) and Shanks et al. (2008). The 25 μ l PCR reaction mixture contained 1 μ L of undiluted DNA (ranging from 2.2 ng/ μ L to 50.3 ng/ μ L) as the template, 1X PCR buffer without MgCl₂ (Promega, Madison, WI), 1.5 mM MgCl₂ (Promega), 0.2 mM of each dNTP (Promega), 0.16 μ g/ μ l of bovine serum albumin (BSA) (Sigma, St. Louis, MO), 0.5 μ M of each primer (IDT, Coralville, IA), 4-units Taq DNA polymerase (Promega) and 15.9 μ l of DNase and RNase free water (Promega). PCR was performed using a Biometra T-Gradient thermocycler (Whatman, Göttingen, Germany).

Touchdown PCR was used for the Bac32 general *Bacteroides* marker and HF183 human-associated marker. Thermocycling conditions were as follows: the initial denaturation of 3.5 minutes at 94°C was followed by 45 S of denaturation at 94°C, and 30 S of extension at 72°C, then one cycle at an annealing temperature of 65°C for 45 S, one cycle at an annealing temperature of 64°C for 45 S, the next cycle at an annealing temperature of 63°C for 45 S, then seven cycles starting with an annealing temperature at 62°C for 45 S, decreasing by 0.5°C per cycle. Finally, 29 cycles at an annealing temperature of 55°C for 45 S were followed by a final extension at 72°C for five minutes (McQuaig et al., 2012).

Themocycling conditions for the CowM3 marker were as follows: an initial denaturation at 94°C for two minutes, then 30 cycles of denaturation at 94°C for one

minute, annealing at 60°C for one minute, extension at 72°C for one minute and a final extension at 72°C for 6 min. A 1 µL aliquot of amplified PCR products was re-amplified using the same PCR conditions.

To detect the amplified products using the Bac32 and HF183 molecular markers, an 8 µL aliquot of PCR product was mixed with 2 µL of 6X loading dye (Promega). A 10 µL aliquot of each mixture was resolved using 1.5% agarose gel (Continental Lab Products, San Diego, CA) in 0.5X TBE buffer. PCR products amplified with the cattle associated marker were resolved using 2.0% low melting agarose gel (OmniPur, Lawrence, KS). The gels were electrophoresed at room temperature for two hours at 105 V and stained with ethidium bromide having a final concentration of 0.5 µg/mL (Fisher Biotech) in 0.5X TBE buffer for one hour. The gel image was captured using a Gel Logic 200 imaging system (Eastman Kodak Co., Rochester, NY). Water samples were considered positive for a particular molecular marker if PCR product bands were aligned with those of the positive control. All end-point PCR runs contained a no template control (NTC) and those showing amplification in negative controls were repeated.

Preparation of qPCR standards. DNA extracted from sewage samples was amplified with general *Bacteroidales* primers 32F/708R and human-associated 183F/708R primers and resolved on 1.5% agarose gel (Bernhard and Field, 2000a,b). Gel images were used to confirm that there was only a single band present for fragment sizes of 686 bp and 525 bp, respectively, for 32F/708R and 183F/708R. The remaining PCR amplified products were cleaned with the DNA Clean & Concentrator Kit (Zymo Research, Orange, CA) and cloned into the TOPO 2.1 cloning vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Plasmid DNA was extracted and cleaned with the QIAprep

Spin Miniprep Kit (Qiagen Sciences, Germantown, MD). The plasmid DNA was amplified with 32F/708R and 183F/708R primers to verify the presence of recombinant DNA prior to sequencing using an ABI 3100 DNA Genetic Analyser with M13 primers. Each sequence was checked to confirm the presence of primer sites for general *Bacteroides* primers (32F/708R) and AllBac primers (296R/412F) as well as human-associate primers HF183 (183F/708R) and qHF183 (183F/265R) (S. 2.3 and S. 2.4). Plasmid DNA concentration was measured with a NanoDrop ND-1000 UV spectrophotometer. Plasmid DNA gene copy numbers were calculated as shown in Chapter 3.

qPCR assay. The 15 μL qPCR reaction mixture contained 5 μL of 10x diluted template DNA (2.2 ng/ μL to 50.3 ng/ μL); 1.25 μL of PCR grade water (Promega); 0.5 μL of BSA (2%, Sigma); 7.5 μL of 2X Power SYBR Green master mix (Applied Biosystems, Carlsbad, California), each 0.3 μL of 10 μM forward and reverse primers, 0.15 μL (1unit/ μL) of AmpErase Uracil N-Glycosylase (UNG) (Life Technologies). The general *Bacteroidales* 16S rRNA genetic markers were amplified with AllBac primers, and the human-associated *Bacteroidales* 16S rRNA genetic markers were amplified with qHF183 primers (Fig. 2.3). The amplification was performed in Applied Biosystems' StepOne real-time PCR instrument as follows: initial denaturation at 95°C for 10 min, then 40 cycles of denaturation at 95°C for 15 S, primer annealing at 60°C for 30 S and extension at 72°C for 30 S. Subsequently, the melt curve analysis was performed using a temperature gradient of 0.5°C per minute from 95°C to 60°C.

Data analysis

***E. coli* rep-PCR DNA fingerprints.** BioNumerics version 5.0 software (Applied Maths, Kortrijk, Belgium) was used for gel normalization, band identification, library development, and source identification. Each gel was normalized using a 1 kb Plus DNA ladder in the range of 200 to 4000 bp as an external reference standard, which allowed the comparison of multiple gels. DNA fingerprinting patterns were compared using a densitometric curve-based method with the Cosine coefficient, and dendrograms were developed using the unweighted pair group method with arithmetic averages (UPGMA). A similarity score of 90% was used as the cut off for the same strain types. Duplicate strain types were eliminated from the library.

Library reliability was evaluated based on three primary criteria: reproducibility, accuracy and robustness. First, the reproducibility of the DNA fingerprints was evaluated using 117 DNA fingerprints from *E. coli* K-12 on 117 gels. These isolates were analysed with cluster analysis; the Cosine similarity coefficient was used to develop the clusters and the dendrogram was developed with the UPGMA method. Jackknife analysis was used to determine the stability of the defined groups. This method is based on removing one entry from the library and seeking to identify it against the different groups. A high percentage of entries assigned correctly to the original group indicates good internal stability of the library. Jackknife analysis with maximum similarity was used to calculate the rate of correct classification (RCC) of each host group and the average rate of correct classification (ARCC) of the library. Multivariate analysis of variance (MANOVA) was also used to evaluate the significance of user-delineated groups in the library. BioNumeric software was utilized to perform Wilk's Lambda likelihood ratio test (L

parameter) to determine whether the groups were indeed drawn from the same population. Here, a low 'L' value would indicate that entries in different groups had been drawn from different populations, or the presence of the groups was statistically significant.

DNA fingerprint patterns obtained from the water isolates of the Catoma Creek watershed were compared against those in the known source library in order to identify the sources of fecal contamination. Unknown water isolates were classified using a 3-way split library (humans, wildlife and domestic animals) and a 9-way split library (humans, cattle, dogs, horses, chickens, deer, beavers, waterfowls and wild turkeys). ID Bootstrap analysis with a sample size of 100 and 1000 iterations was used to classify isolates based on maximum similarity and to calculate the probability that each isolate had been correctly classified. The BioNumerics manual defines the quality quotient for the source identification based on the internal heterogeneity of the library units, which is an indication of the level of confidence of the source identification: a quality quotient < 1.5 indicates that source identification is probable, while 1.5 to 2.0 indicates it is possible, and > 2.0 indicates that it is improbable. In this study, a quality quotient of 1.5 was taken to be the cut off value for source identification.

***Bacteroidales* genetic markers.** Several quality control parameters were used to evaluate the performance of qPCR (Table 2.3). All DNA samples were run in duplicate with at least five point standard curves, 10 to 10^5 , in duplicate and no template controls (NTC) (Fig. 2.3). The lowest concentration of a particular genetic marker in the linear range of the qPCR standard curve was considered the lower limit of quantification (LLOQ) of each qPCR assay (Seurinck et al., 2005). The limit of detection (LOD) was calculated

according to the method described by Dick et al (2010): the LOD was defined as the fifth percentile among the observed threshold cycles (C_q) across all NTC and DNA extraction blank (lab blank) reactions. Gene copy numbers above the LLOQ were reported as quantifiable numbers, gene copies below the LLOQ but above the LOD were reported as detectable but not quantifiable, and gene copy numbers below the LOD were reported as non-detectable.

2.3 Results

Evaluation of the *E. coli* DNA fingerprint library. Individual fecal samples (736 in total) collected from nine host groups were used to obtain 1,341 *E. coli* isolates and their rep-PCR DNA fingerprints. Since the presence of multiple identical DNA fingerprints from a single fecal sample in a known source library is known to result in biases or artificial classification of unknowns, only one identical fingerprint from each fecal sample was used in the library and the remainder eliminated. DNA fingerprints obtained from the same fecal samples having a similarity of more than 90% were considered identical. As a result, about 30% of the fingerprints were removed, and the final library contained 945 unique DNA fingerprints.

Reproducibility of the DNA fingerprints between PCR runs was evaluated using *E. coli* K-12 (ATCC 10798), which was included in each gel as a positive control. The similarity of 117 *E. coli* K-12 DNA fingerprints was calculated, and the interquartile range, the difference between the first quartile and third quartile of a set of data, for K-12 from all the gels was between 83.85 and 94.25%.

Accuracy means the likelihood that a particular fingerprint will be grouped with the correct host group. This can be tested in two ways, using either challenge isolates or

random isolates. Jackknife analysis facilitates the random assignment of isolates to a host group. The rates of correct classification (RCC) for the three broad groups – humans, domestic animals and wildlife – were 87.4%, 81.0% and 79.9%, respectively, with an average rate of correct classification (ARCC) of 82.8% (Table 2.4a). The 9-way split library showed a lower ARCC of 71.2%, with the human group having the highest RCC of 87.4% and dogs showing the lowest RCC of 47.2% (Table 2.4b). Figures 2.4a and 2.4b show the MANOVA analysis of the 3-way and 9-way source groups, respectively. The ‘L’ value for the 3-way library was 0.002, while the 9-way library had a larger ‘L’ value of 0.01. Both the 3-way and 9-way source group separations were significant ($P=0.001$). These very low ‘L’ values indicate the internal stability of both libraries, with the 3-way library being even more stable than the 9-way library.

Identification of fecal contamination sources using the known source library. A total of 1,050 *E. coli* DNA fingerprints obtained from water samples collected during the year-long study were used for fecal contamination source identification. The 3-way library identified 909 (86.4 %) of the *E. coli* DNA fingerprints from the water samples, leaving only 13.6% of the *E. coli* DNA fingerprints unidentified. The largest number of isolates, 468 (44.6%), belonged to domestic animals, while almost similar numbers of isolates were categorized as wildlife and human, at 229 (21.8%) and 221 (21.0%), respectively (Fig. 5a). The 9-way known source library classified fewer isolates than the 3-way library: 850 (80.9%) isolates were classified into the 9-source groups, while 19.1% remained unidentified. Here, the isolates belonging to humans (20.3%) were the largest single group, followed by cattle (13.4%), dogs (12.0%), chickens (9.5%), waterfowl (9.0%), horses (8.2%), deer (4.0%), beavers (3.0%) and wild turkeys (1.4%) (Fig. 5b).

The human/domestic animal/wildlife source distributions for each site are presented in Figure 2.6. The proportion of domestic animal sources was the highest group for each site, with the samples collected at Site RSP having the highest percentage of unidentified isolates compared with other sites, while those from site BM had the lowest percentage of unidentified isolates.

The highest percentage of *E. coli* isolates were identified in June 2007, 94.3%, while much lower percentages of source identification were found in September 2007 (67.5%) and February 2008 (67.3%) (Table 2.5). The highest number of human associated fingerprints was identified in March 2007, where 34 samples were of human origin, while the lowest was in August 2007, where only five DNA fingerprints were categorized as human (Table 2.5). The total number of *E. coli* isolates used for fingerprint development varied from 191 at both Site RHM and Site CW to 213 at Site BM. The flow at the RSP site was intermittent and water was available in only four sampling months; hence only 68 *E. coli* isolates were obtained from this site for source identification, of which 86.8% were identified. Source identification for the remaining sites varied from 77.2% for Site LT to 84.3% for Site RHM (Table 2.6).

***Bacteroidales* genetic markers.** Suitable human and cattle specific *Bacteroidales* genetic markers for Alabama were selected based on a screening procedure (data are shown in Chapter 3). Bac32, HF183 and CowM3 primers were selected for the end-point PCR, while AllBac and qHF183 primers were selected for the qPCR amplification. LLOQ for both the AllBac and qHF183 assays were 10 gene copies/5 μ L of template DNA. The R^2 values for all the standard curves were above 0.993 and the PCR amplification efficiency ranged from 98.2% to 106% (Table 2.3).

Comparison of library based and non-library based methods. Tables 2.7 and 2.8 show the *E. coli* concentrations and general *Bacteroidales* gene copies identified by sampling month and site. *E. coli* and general *Bacteroidales* markers were found in all water samples. The lowest *E. coli* concentration occurred in November 2007 at Site LT (5 CFU/100 ml), while the highest *E. coli* numbers were recorded at Site RSP in January 2008 (2,350 CFU/100 ml) (Table 2.7). The lowest gene copy number of general *Bacteroidales* markers was found at Site CW in October 2007 (1,447 gene copies/100 mL), and the highest number was reported at Site RHM in December 2007 (185,191 gene copies/100 ml) (Table 2.8). *E. coli* concentrations were relatively low during the summer months and high during the winter period (Fig. 2.7) but student t test results showed that this *E. coli* distribution variation was not statistically significant (P=0.54). AllBac markers were abundant throughout the year, although their concentrations were higher during the winter months than in the spring months (Table 2.8). Sites WS and CW had lower AllBac marker concentrations than the other sites (Fig. 2.8), while *E. coli* concentrations were lowest at Sites LT and CW.

All the qHF183 marker gene copies were below the limit of quantification. Table 2.9 shows the qHF183 markers detected at each sampling site for each month. No human markers were detected in the months of November and December 2007, and cooler months had lower detection frequencies than warmer months. From April to October 2007, all the sites were positive for qHF183 markers. It appears that there was no relationship between the qHF183 marker occurrence and rainfall in this watershed.

The library based method identified more human and cattle signatures in the water samples than the non-library based methods. According to the *E. coli* DNA fingerprint

method, 93.8% of the samples were positive for human fecal sources. End-point PCR did not amplify as many human-associated markers as qPCR, identifying only 26.5% of samples as containing human markers compared to the figure suggested by qPCR of 64%. The library based method revealed that 73.4% of the samples were of cattle fecal origin, while non-library based end-point PCR detected cattle markers in only 28% of the samples (Table 2.10).

Correlation between *E. coli* and general *Bacteroidales* markers. In general, there was no good correlation between *E. coli* and general *Bacteroidales* concentration ($r=0.16$). These correlations varied from month to month, ranging from -0.32 in May and June 2007 to 0.98 in December 2007 (Table S. 2.1). A low-level correlation did, however, appear when separating out the results for the warmer months and the cooler months (May, June, August and September 2007 samples showed a negative correlation between these two groups) except for July 2007, which had a correlation coefficient of 0.93 . Furthermore, there was no good site specific correlation between *E. coli* and general *Bacteroidales*; the lowest correlation coefficient was found at Site LT ($r=0.09$) while the highest correlation coefficient ($r= 0.53$) was found at Site CW (Table S. 2.2). The Catoma Creek watershed covers a relatively large area of 932 km^2 and the sampling sites were distributed throughout the watershed. Rainfall data were not available for individual sampling sites (Table S2.3), and since rainfall was measured at only one location, these rainfall data may not adequately represent all the sampling sites. In addition, other watershed characteristics, such as land use patterns, soil types and topography, should also be considered because those factors may influence the effective runoff.

2.4 Discussion

This study compared the performance of library based and non-library based methods to detect the sources of fecal contamination in a mixed land use watershed. The library comparison study was conducted for a one-year period from March 2007 to February 2008. *E. coli* rep-PCR DNA fingerprint library previously developed for the Catoma Creek watershed was used to detect the source of fecal contamination in this watershed using the library based method. During the study period, the positive and negative aspects of both methods became very clear, with the main drawback of the library based approach being the gel-to-gel variability, the need to first develop and then continuously update a large known source library. The size of the library is a crucial factor for effective source identification, as it must contain an adequate number of diverse DNA fingerprints from all possible fecal contamination sources in the watershed. However, there is no standard size for such a library because this depends on the method of choice and the size of the contributing population (Stoeckel and Harwood, 2007). Jackknife analysis demonstrated the internal stability of our library, with the ARCC of the 3-way library being 82.8% and that of the 9-way library 71.2%. Our initial library contained only 414 unique DNA fingerprints from eight source groups, with an ARCC of 87.5%, but as more DNA fingerprints were added to the library, finally reaching 945 unique isolates from 9-source groups, its ARCC dropped to 71.2%. Some small libraries with clonal isolates have shown higher ARCC values because clonal isolates artificially increase the ARCC in their libraries (Dombek et al., 2000). Additionally, when a library has a relatively small number of isolates and source groups, higher ARCCs are often observed (Carson et al., 2003; McLellan et al., 2003). This was demonstrated in a rep-

PCR DNA fingerprint study conducted by Johnson et al. (2004), who reported that their clonal library contained 2,466 fingerprints from 12 source groups, with an ARCC of 82.2%; once they removed the clonal isolates the ARCC dropped to 60.5%. A study by Harwood et al. (2003) found that the libraries need to be reasonably large (more than 300 isolates) to avoid artefacts due to source-independent grouping. Our library appeared to be sufficiently large and genetically diverse for effective source identification in the Catoma Creek watershed.

The results of this study revealed that the ability to identify sources increases with increasing library size. The initial decloned library contained only 414 DNA fingerprints from eight host groups, and 502 water DNA fingerprints were compared against this initial library, leaving 41.1% of the isolates unidentified. Later, more DNA fingerprints were added to each host group and the library finally contained 945 genetically diverse *E. coli* DNA fingerprints. A total of 1,050 water DNA fingerprints developed from 2007 March to 2008 February were compared against the newly decloned library, after which only 19.1% of the water isolates remained unidentified. These results clearly show that the size of the library greatly affects the rate of source identification. Furthermore, the diversity of the library is also critical; our initial library contained 167 (40.3%) DNA fingerprints from humans, 136 (32.8%) from domestic animals, and 111 (26.8%) from wildlife. In the final version of the library, *E. coli* DNA fingerprints from humans and domestic animals increased to 358 (37.8%) each, while wildlife isolates increased to 229 (24.2%). The percentage identified as human thus increased from 18.4% in the initial 8-way library to 20.3% in the improved 9-way library. Sources identified as domestic animals improved from 18% to 43.1%, while the wildlife percentage dropped from 23%

to 17.5%. The same number of DNA fingerprints from human and domestic animal *E. coli* were present in the library, while fewer wildlife fingerprints were in the library. If more DNA fingerprints representing wildlife were to become available, this is likely to improve the percentage of wildlife identification.

The non-library based methods tested involved the direct identification of host-associated gene markers. This process was faster, less laborious and less expensive than the library based method. Gene amplification with end-point PCR was not as sensitive as qPCR when detecting targets. In this study, 64% of the water samples contained qHF183 markers above the LOD, but end-point PCR amplification of HF183 with touchdown thermocycling conditions successfully amplified only 26.5% of the samples. This may be due to the low sensitivity of end-point PCR. Different thermocycling conditions were used on end-point PCR in an attempt to improve the detection efficiencies; touchdown PCR is known to enhance the detection of general *Bacteroidales* and human-associated markers, while re-amplification of PCR products of CowM3 improves the detection of cattle markers in the samples. Especially where qPCR is not available, touchdown PCR and the re-amplification of PCR products are generally good options for improving the sensitivity of detection.

Higher percentages of *E. coli* from human and cattle sources were found using the library-based approach than percentages of human- and cattle-associated markers detected using the non-library based approach. One of the major drawbacks of the DNA fingerprinting methods is the gel-to-gel variability and high rate of false positive results (Harwood et al., 2003). On the other hand, false negatives are clearly a problem for host-associated gene detection methods. PCR inhibition is a major problem, as this prevents

amplification of the target and is difficult to overcome completely. Sample dilution is one of the common methods used to reduce PCR inhibition, but the dilution of already low DNA concentrations can result in gene copy numbers in the water samples falling below the LOD or LLOQ. This may explain why most of the samples in this study dropped below the LLOQ for human-associated markers.

Due to the nature of probability, it is unreasonable to expect the amplification of each and every sample, especially at the lower end of a standard curve. If the DNA concentration is low, it is therefore preferable to run more DNA replicate samples for a non-library based method. Collecting more water samples from which to extract DNA from each sampling site should yield more positive results than the methods used in the present study.

Moreover, the DNA samples extracted from the filter papers were used for end-point PCR within a few weeks of extraction, with the remaining DNA being stored at a temperature of -80°C and not used for qPCR amplification for almost 4 years. It is possible that fresh DNA would give better amplification than DNA stored under these conditions, thus producing better detection efficiencies than those reported here.

Furthermore, neither the DNA extraction efficiency nor the recovery percentages were evaluated for this study. Since 100% DNA recovery is not possible currently, only a small portion of the DNA present in the water samples was amplified. In this study, the water was filtered by passing it through a filter membrane, which was then cut into small pieces, placed in a the tube containing beads and bead solution and vortexed at maximum speed for 10 minutes. This may be not the best way to remove all the detached cells and DNA from the filter membrane and the use instead of a beat beater may yield better

results. The MO BIO ultra-clean DNA isolation kit was used for DNA extraction, but an earlier study reported the average DNA recovery efficiency of an ultra-clean DNA isolation kit from sediment to be as low as 14.9% (Mumy and Findlay, 2004). In contrast, Halliday and coworkers (2010) found the DNA extraction efficiency of *Enterococcus* in sand samples to reach 55.5%–66.2% using the MO BIO PowerSoil DNA isolation kit. Future studies should therefore determine the DNA recovery achieved by different techniques and select the method likely to produce the highest DNA yield. For example, salmon DNA can be used as a control to optimize DNA recovery (USEPA, 2010).

E. coli can persist longer in their secondary habitat than *Bacteroidales*, >75-days verses 9-days (Chapter 4), so it is not possible to guarantee that results obtained from an *E. coli* library based approach accurately represent current fecal contamination conditions or whether some of the contamination actually occurred several weeks earlier. In a subsequent study (Chapter 4), *Bacteroidales* human-associated HF183 markers were found to persist for as little as 3 days in stream water and sediment.

This study focused primarily on detecting human- and cattle-associated markers in water samples, but other molecular markers are available to detect fecal contamination in watersheds, namely dogs (Dick et al., 2005a; Kildare et al., 2007), horses (Dick et al., 2005b), geese (Fremaux et al., 2010; Lu et al., 2009), chickens (Lu et al., 2009), pigs (Dick et al., 2005b; Okabe et al., 2007; Mieszkin et al., 2010); sheep (Lu et al., 2007) and elk/deer (Dick et al., 2005a). However, most of these markers exhibit poor specificity and/or sensitivity (Boehm et al., 2013). In addition, *Bacteroidales* are not distributed evenly in all organisms, being dominant in the human microbome but poorly distributed in birds such as gulls or geese (Lu et al., 2009, 2008, 2007; Yang et al., 2009). This

problem can be overcome to some extent by adapting two or more MST methods for the same watershed (Vogel et al., 2007). For example, two MST approaches, *Bacteroidales* host-specific 16S rRNA gene markers and F-specific bacteriophage genotyping, have been used successfully to identify human and non-human sources in surface water (Gourmelon et al., 2010) and studies conducted in Lake Ontario in Canada and Queensland in Australia have successfully used both library based and non-library based methods for microbial source tracking along beaches with multiple sources (Ahmed et al., 2007 Edge et al., 2010).

2.5 Summary

Overall, the non-library based approach was found to be more effective than the library based method in identifying sources of fecal contamination as being either human or non-humans. Especially in designated beach areas or shellfish harvesting areas, where results need to be obtained quickly because decisions to close beaches or halt shellfish harvesting due to poor water quality must be made as quickly as possible. Both end-point and qPCR methods are fast and accurate, with results obtained within 24 hours of sample collection. Although qPCR provides information about sources of fecal pollution as well as quantitative concentration estimates, it is technically demanding and expensive. Local water resource managers may not be able to afford qPCR. Therefore, end-point PCR methods such as touchdown PCR and re-amplification of PCR products offer useful alternatives, especially if the target DNA concentration is low. Many molecular markers are now available with which to detect fecal contamination from other possible source groups in a watershed, but problems can arise due to the spatial variability, poor sensitivity and specificity of these markers. Therefore, it is imperative to select and adapt

suitable molecular markers for a particular watershed. It is also important to note that non-library based methods may underestimate the concentrations present, especially if problems arise due to PCR inhibition and poor DNA recovery.

When source identification predominantly involves a large area with mixed land use patterns, the use of the library based DNA fingerprinting method comes into its own. Once a suitable library has been constructed, this method is capable of investigating all the fecal contamination source groups of interest in a particular watershed. This approach can provide information on long term patterns of fecal contamination sources in the watershed, which is particularly significant for those charged with making many major management decisions, for example developing total maximum daily loads (TMDL) for a particular watershed, improving sewage treatment plants or building new plants, repairing sewage carrying lines or developing a riparian buffer system on livestock farms to prevent animal faeces entering the surface water. The results of this study demonstrated that the size and diversity of a DNA fingerprint library greatly affects the accuracy of source identification. The study also found, however, that the rep-PCR DNA fingerprint method is time consuming and results cannot be obtained immediately.

There is clearly no single perfect method for detecting the sources of fecal contamination in a watershed in all cases. Based on the study findings, both approaches can successfully be used to detect the sources of fecal contamination if the correct method is selected to suit the specific situation. If it is possible to conduct the source tracking study using multiple methods, i.e., by combining library based methods and non-library based methods, the results obtained will be both robust and reliable. In this study, both library based and non-library based methods identified human waste as the main

source of fecal contamination in the Catoma Creek watershed. Unfortunately, source tracking is expensive and the availability of sufficient funding to maintain a current library imposes a major constraint. Thus, the selection of the most appropriate method will depend on the type of watershed, the primary research goals, the availability of funds and the time frame.

2.6 References

1. **ADEM.** 2002. Draft TMDL development for Catoma Creek, AL.03150201-080-. www.adem.state.al.us/PublicNotice/Nov/Catoma%20CreekDO%20TMDL%20
2. **ADEM.** 2010. Parkerson Mill Creek Watershed Management Plan. http://www.auburn.edu/administration/facilities/organization/campus-planning-space-management/campus-planning/landscape-master-plan/documents/PMCPLAN_LR.pdf
3. **Ahmed, W., J. Stewart, T. Gardner, D. Powell, P. Brooks, D. Sullivan, and N. Tindale.** 2007. Sourcing faecal pollution: A combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments. *Water Res.* **41**:3771-3779.
4. **Bernhard, A. E., and K. G. Field.** 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**:1587-1594.
5. **Bernhard, A. E., and K. G. Field.** 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **66**:4571-4574.
6. **Boehm, A. B., L. C. Van De Werfhorst, J. F. Griffith, P. A. Holden, J. A. Jay, O. C. Shanks, D. Wang, and S. B. Weisberg.** 2013. Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res.* (In press)
7. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl. Environ. Microbiol.* **69**:1836-1839.
8. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005a. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3179-3183.
9. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005b. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3184-3191.
10. **Dick, L. K., E. A. Stelzer, E. E. Bertke, D. L. Fong, and D. M. Stoeckel.** 2010. Relative decay of *Bacteroidales* microbial source tracking markers and cultivated *Escherichia coli* in freshwater microcosms. *Appl. Environ. Microbiol.* **76**:3255-3262.

11. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl. Environ. Microbiol.* **66**:2572-2577.
12. **Edge T. A, S. Hill, P. Seto and J. Marsalek.** 2010. Library-dependent and library-independent microbial source tracking to identify spatial variation in faecal contamination sources along a Lake Ontario beach (Ontario, Canada). *Water Sci. Tech.* **62**: 719-727.
13. **Fremaux, B., J. Gritzfeld, T. Boa, and C. K. Yost.** 2009. Evaluation of host-specific *Bacteroidales* 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Res.* **43**:4838-4849.
14. **Fremaux, B., T. Boa, and C. K. Yost.** 2010. Quantitative real-time PCR assays for sensitive detection of Canada goose-specific fecal pollution in water sources. *Appl. Environ. Microbiol.* **76**:4886-4889.
15. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiol.SGM.* **147**:1079-1085.
16. **Gourmelon, M., M. P. Caprais, C. Le Mennec, S. Mieszkin, C. Ponthoreau, and M. Gendronneau.** 2010. Application of library-independent microbial source tracking methods for identifying the sources of faecal contamination in coastal areas. *Environ. Sci. Technol.* **61**:1401-1409.
17. **Halliday E, J. F. Griffith, and R. J. Gast.** 2010. Use of an exogenous plasmid standard and quantitative PCR to monitor spatial and temporal distribution of *Enterococcus* spp. in beach sands. *Limnol, Oceanogr. Methods* **8**:146-154.
18. **Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. *J. Environ. Qual.* **31**:1273-1278.
19. **Harwood, V. J., J. Whitlock and V. Withington.** 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. *Appl. Environ. Microbiol.* **66**:3698-3704.
20. **Harwood VJ, B. Wiggins, C. Hagedorn, R. D. Ellender, J. Gooch, J. Kern, M. Samadpour, A. C. Chapman, B. J. Robinson and B. C. Thompson.** 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *J Water Health.* **1(4)**:153-166.

21. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. *J. Environ. Qual.* **32**:305-309.
22. **Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. *Appl. Environ. Microbiol.* **70**:4478-4485.
23. **Kildare, B. J., C. M. Leutenegger, B. S. McSwain, D. G. Bambic, V. B. Rajal, and S. Wuertz.** 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: A Bayesian approach. *Water Res.* **41**:3701-3715.
24. **Lamendella, R., J. W. S. Domingo, D. B. Oerther, J. R. Vogel, and D. M. Stoeckel.** 2007. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene. *FEMS Microbiol. Ecol.* **59**:651-660.
25. **Lamendella, R., N. VerBerkmoes, and J. K. Jansson.** 2012. 'Omics' of the mammalian gut - new insights into function. *Curr. Opin. Biotechnol.* **23**:491-500.
26. **Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Sayler.** 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* **72**:4214-4224.
27. **Lu, J. R., J. S. Domingo, and O. C. Shanks.** 2007. Identification of chicken-specific fecal microbial sequences using a metagenomic approach. *Water Res.* **41**:3561-3574.
28. **Lu, J., J. W. Santo Domingo, R. Lamendella, T. Edge, and S. Hill.** 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. *Appl. Environ. Microbiol.* **74**:3969-3976.
29. **Lu, J., J. W. Santo Domingo, S. Hill, and T. A. Edge.** 2009. Microbial diversity and host-specific sequences of Canada Goose feces. *Appl. Environ. Microbiol.* **75**:5919-5926.
30. **McLellan, S. L., A. D. Daniels, and A. K. Salmore.** 2003. Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. *Appl. Environ. Microbiol.* **69**:2587-2594.

31. **McQuaig S., J. Griffith and V. J. Harwood.** 2012. Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. *Appl. Environ. Microbiol.* **78(18):**6423-6432.
32. **Mieszkin, S., J. F. Yala, R. Joubrel, and M. Gourmelon.** 2010. Phylogenetic analysis of *Bacteroidales* 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J. Appl. Microbiol.* **108:**974-984.
33. **Mumy, K. L., and R. H. Findlay.** 2004. Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR. *J. Microbiol. Methods* **57:**259-268.
34. **Natural Resources Defence Council, 2012.**
<http://www.nrdc.org/water/oceans/ttw/faq.asp>
35. **Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* **74:**890-901.
36. **Rademaker, J. L. W. and F. J. de Bruijn.** 1997. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis, p.151-171. In G. Caetano-Anolles and P. M. Gresshoff (ed.), *DNA markers: protocols, applications, and overviews*. J. Wiley and Sons, New York, N.Y.
37. **Savichtcheva, O., N. Okayama and S. Okabe.** 2007. Relationships between *Bacteroides* 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Res.* **41:**3615-3628.
38. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002. Microbial source tracking: current methodology and future directions. *Appl. Environ. Microbiol.* **68:**5796-5803.
39. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. D. Siciliano.** 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. *Environ. Microbiol.* **7:**249-259.
40. **Shanks, O. C., C. Nietch, M. Simonich, M. Younger, D. Reynolds, and K. G. Field.** 2006. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Appl. Environ. Microbiol.* **72:**5537-5546.

41. **Shanks, O. C., J. W. Santo Domingo, J. R. Lu, C. A. Kelty, and J. E. Graham.** 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Appl. Environ. Microbiol.* **73**:2416-2422.
42. **Shanks, O. C., E. Atikovic, A. D. Blackwood, J. R. Lu, R. T. Noble, J. S. Domingo, S. Seifring, M. Sivaganesan, and R. A. Haugland.** 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **74**:745-752.
43. **Shanks, O. C., C. A. Kelty, M. Sivaganesan, M. Varma, and R. A. Haugland.** 2009. Quantitative PCR for genetic markers of human fecal pollution. *Appl. Environ. Microbiol.* **75**:5507-5513.
44. **Shanks, O. C., K. White, C. A. Kelty, S. Hayes, M. Sivaganesan, M. Jenkins, M. Varma, and R. A. Haugland.** 2010. Performance assessment PCR-based assays targeting *Bacteroidales* genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **76**:1359-1366.
45. **Stoeckel, D. M., and V. J. Harwood.** 2007. Performance, design, and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* **73**:2405-2415.
46. **USEPA.** 2002. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermo tolerant *Escherichia coli* agar (modified mTEC). **EPA 821-R-02-023.**
47. **USEPA.** 2010. Method B: *Bacteroidales* in water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) assay. EPA-822-R-10-003. <http://water.epa.gov/scitech/methods/cwa/bioindicators/upload/methodb2010.pdf>
48. **USEPA.** 2012. Recreational water quality criteria. **820-F-12-061.** <http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/factsheet2012.pdf>
49. **Vogel, J. R., D. M. Stoeckel, R. Lamendella, R. B. Zelt, J. W. S. Domingo, S. R. Walker, and D. B. Oerther.** 2007. Identifying fecal sources in a selected catchment reach using multiple source-tracking tools. *J. Environ. Qual.* **36**:718-729.
50. **Wijesinghe, R. U., Y. Feng, C. W. Wood, D. M. Stoeckel and J. N. Shaw.** 2009. Population dynamics and genetic variability of *Escherichia coli* in a mixed land-use watershed. *J. Water Health* **7(3)**:484-496.
51. **Yang, X., L. Xie, Y. X. Li, and C. C. Wei.** 2009. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *Plos One* **4**. e6074. doi:10.1371/journal.pone.0006074.

Table 2.1. Source groups and DNA fingerprints in the known source library

Source group	# of fecal samples	# of total fingerprints	# of unique fingerprints 9-way library	# of unique fingerprints 3-way library
Human	369	378	358	Human -358
Cattle	56	139	82	
Dog	53	144	89	
Chicken	50	144	91	
Horse	49	138	96	Domestic-358
Deer	54	117	80	
Beaver	18	56	21	
Waterfowl	61	144	94	
Wild turkey	26	81	34	Wildlife-229
Total	736	1,341	945	945

Table 2.2. *Bacteroidales* primers used for fecal source detection and quantification

Assay	Primer Sequences (5'-3')	Target	Size	Reference
Bac32	32F: ACGCTAGCTACAGGCTT 708R: CAATCGGAGTTCTTCGTG	16S rRNA gene	676 bp	Bernhard and Field, 2000a
AllBac	296F: GAGAGGAAGGTCCCCAC 412R: CGCTACTTGGCTGGTTCAG	16S rRNA gene	106 bp	Layton et al., 2006
HF183	183F: ATCATGAGTTCACATGTCCG 708R: CAATCGGAGTTCTTCGTG	16S rRNA gene	525 bp	Bernhard and Field, 2000b
qHF183	183F: ATCATGAGTTCACATGTCCG 265R: TACCCGCCTACTATCTAATG	16S rRNA gene	82 bp	Seurinck et al., 2005
CowM3	M3 F: CCTCTAATGGAAAATGGATGGTATCT M3 R: CCATACTTCGCCTGCTAATACCTT	9-O-acetyesterase secretory protein gene	122 bp	Shanks et. al, 2008

Table 2.3. Selected quality control parameters used to evaluate qPCR reactions

Quality control parameter	AllBac	HF183
qPCR efficiency	99.2% - 106.0%	98.15%-102.0%
R ² for the standard curve	0.998 – 0.999	0.993-1.0
Lower limit of quantification	10 gene copies	10 gene copies
Limit of detection	4 gene copies/5 µl	1 gene copies/5 µl
Composite standard curve	Y= -3.28X +34.22 (R ² =1.0, E =101.7%)	Y= -3.29X + 32.65 (R ² =1.0, E = 101.5%)
Checked for PCR inhibition	Yes	Yes

Table 2.4a. Rates of correct classification (RCC) based on 3-way separation

	Domestic	Human	Wildlife
Domestic	81.0	6.70	13.1
Human	10.6	87.4	6.99
Wildlife	8.38	5.87	79.9

ARCC = 82.8%

Table 2.4b. Rate of correct classification based on 9-way separation

	Horse	Human	WF	Cattle	Deer	WT	Chicken	Beaver	Dog
Horse	78.1	0.84	4.26	3.66	5.00	0.00	3.30	4.76	7.87
Human	5.21	87.4	3.19	6.10	7.50	8.82	7.69	19.1	23.6
Waterfowl	6.25	2.79	72.3	12.2	7.50	8.82	4.40	0.00	4.49
Cattle	1.04	1.40	6.38	67.1	2.50	0.00	1.10	4.76	2.25
Deer	0.00	1.12	6.38	1.22	70.0	14.7	0.00	0.00	0.00
Wild turkey	0.00	0.28	1.06	0.00	0.00	67.6	2.20	0.00	1.12
Chicken	5.21	2.23	4.26	6.10	3.75	0.00	76.9	0.00	12.4
Beaver	0.00	1.68	0.00	1.22	0.00	0.00	0.00	71.4	1.12
Dog	4.17	2.23	2.13	2.44	3.75	0.00	4.40	0.00	47.2

ARCC = 71.2%

Table 2.5. Numbers of *E. coli* isolates identified by month

Months	Mar.	Apr.	May	2007					Oct.	Nov.	Dec.	2008	
				June	July	Aug.	Sept.	Jan.				Feb.	
Human	34	16	23	13	23	5	13	8	11	23	17	27	
Cattle	13	15	7	15	9	8	4	10	7	13	21	19	
Dogs	16	12	7	21	2	2	8	16	14	4	13	11	
Chickens	6	14	8	12	9	5	8	8	5	8	10	7	
Horses	6	5	8	16	11	7	7	4	5	3	9	5	
Deer	4	6	1	1	2	4	5	1	4	4	5	5	
Waterfowls	4	10	7	4	14	14	6	9	3	6	8	10	
W. turkeys	1	2	0	0	0	1	5	2	0	0	4	0	
Beavers	3	0	0	1	0	3	0	4	19	1	1	0	
Isolates identified	87	80	61	83	70	49	56	62	68	62	88	84	
Total isolates	106	95	70	87	89	70	83	80	78	89	105	98	
Percentage identified	82.1	84.2	87.1	94.3	79.5	70.0	67.5	77.5	87.2	69.7	83.0	67.3	

Table 2.6. Numbers of *E. coli* isolates identified at the six sampling sites

Sources	RSP	RHM	BM	LT	WS	CW	Total
Human	12	35	54	35	38	39	213
Cattle	11	20	33	26	26	25	141
Dogs	10	24	16	18	28	30	126
Chickens	13	20	14	21	17	15	100
Horses	5	13	19	19	15	16	86
Deer	3	9	6	4	8	10	42
Waterfowls	3	20	24	17	13	17	95
Wild turkeys	2	2	2	7	1	1	15
Beavers	0	18	2	2	6	4	32
Isolated identified	59	161	170	149	152	157	850
Total isolates	68	191	213	193	194	191	1050
% identified	86.8	84.3	79.8	77.2	78.4	82.2	81.0

Table 2.7. *E. coli* concentrations (CFU/100 ml) found at each sampling site during the 1-year study period.

Sampling months	Sampling sites					
	RSP	RHM	BM	LT	WS	CW
Mar-07	130	40	45	67	43	23
Apr-07	655	660	490	675	72	1,254
May-07	N/A	34	893	17	41	46
Jun-07	N/A	203	560	55	52	101
Jul-07	N/A	109	175	40	21	42
Aug-07	N/A	6	1,825	20	63	9
Sep-07	N/A	12	334	14	49	68
Oct-07	N/A	23	54	26	315	29
Nov-07	N/A	192	390	5	33	16
Dec-07	N/A	730	96	26	27	26
Jan-08	2,350	802	738	205	266	735
Feb-08	199	208	213	71	38	133

N/A: No water available in the stream

Table 2.8. *Bacteroidales* concentrations (gene copies/100 ml) found at each sampling site during the 1-year study period.

Sampling month	Sampling sites					
	RSP	RHM	BM	LT	WS	CW
Mar-07	11,607	11,721	3,128	14,255	6,290	1,643
Apr-07	1,560	20,600	5,107	10,136	2,115	5,948
May-07	N/A	3,468	2,278	26,073	4,028	2,487
Jun-07	N/A	10,189	21,185	87,124	7,167	38,126
Jul-07	N/A	42,119	116,062	36,804	11,763	8,628
Aug-07	N/A	9,963	9,874	27,471	3,789	<i>615,588</i>
Sep-07	N/A	86,319	19,051	5,910	19,849	6,195
Oct-07	N/A	29,183	14,007	21,012	19,827	1,447
Nov-07	N/A	23,159	21,150	4,797	19,932	4,237
Dec-07	N/A	185,191	46,172	59,707	29,999	32,647
Jan-08	49,482	14,577	35,170	14,174	26,571	31,246
Feb-08	45,298	36,430	36,561	18,544	26,259	24,272

N/A: No water available in the stream. Number in italics indicates an outlier that was removed from the data analysis.

Table 2.9. Human-associated *Bacteroidales* detected in the Catoma Creek watershed during the 1-year study period.

	RSP	RHM	BM	LT	WS	CW	Total detected	% detected
Mar-07	LOD	LOD	UND	UND	LOD	LOD	4	66.7
Apr-07	LOD	LOD	LOD	LOD	LOD	LOD	6	100.0
May-07		LOD	LOD	LOD	LOD	LOD	5	100.0
Jun-07		LOD	LOD	UND	LOD	LOD	4	80.0
Jul-07		LOD	LOD	LOD	LOD	LOD	5	100.0
Aug-07		LOD	LOD	UND	LOD	LOD	4	80.0
Sep-07		LOD	LOD	LOD	LOD	LOD	5	100.0
Oct-07		LOD	LOD	LOD	LOD	LOD	5	100.0
Nov-07		UND	UND	UND	UND	UND	0	0.0
Dec-07		UND	UND	UND	UND	UND	0	0.0
Jan-08	UND	UND	UND	UND	LOD	UND	1	16.7
Feb-08	UND	UND	LOD	UND	LOD	UND	2	33.3
Total detected	2	8	8	5	10	8	41	
% detected	50	66.7	66.7	41.7	83.3	66.7	64.1	

LOD: above the LOD but below the LLOQ.

UND: undetected.

Table 2.10. *E. coli*, *E. coli* DNA fingerprints and *Bacteroidales* gene markers that tested positive in this study.

Site	<i>E. coli</i>	<i>E. coli</i> DNA fingerprints		<i>Bacteroidales</i> gene markers				
				End-point PCR amplification			qPCR amplification	
		Human	Cattle	Bac32	HF183	CowM3	AllBac	qHF183
RSP	4/4	4/4	3/4	4/4	1/4	1/4	4/4	2/4
RHM	12/12	11/12	10/12	12/12	4/12	3/12	12/12	8/12
BM	12/12	12/12	7/12	12/12	2/12	4/12	12/12	8/12
LT	12/12	11/12	9/12	12/12	4/12	4/12	12/12	5/12
WS	12/12	10/12	9/12	12/12	4/12	3/12	12/12	10/12
CW	12/12	12/12	9/12	12/12	2/12	3/12	12/12	8/12
Total	64/64	60/64	47/64	64/64	17/64	18/64	64/64	41/64
%	100%	93.8%	73.4%	100%	26.5%	28.0%	100%	64.0%

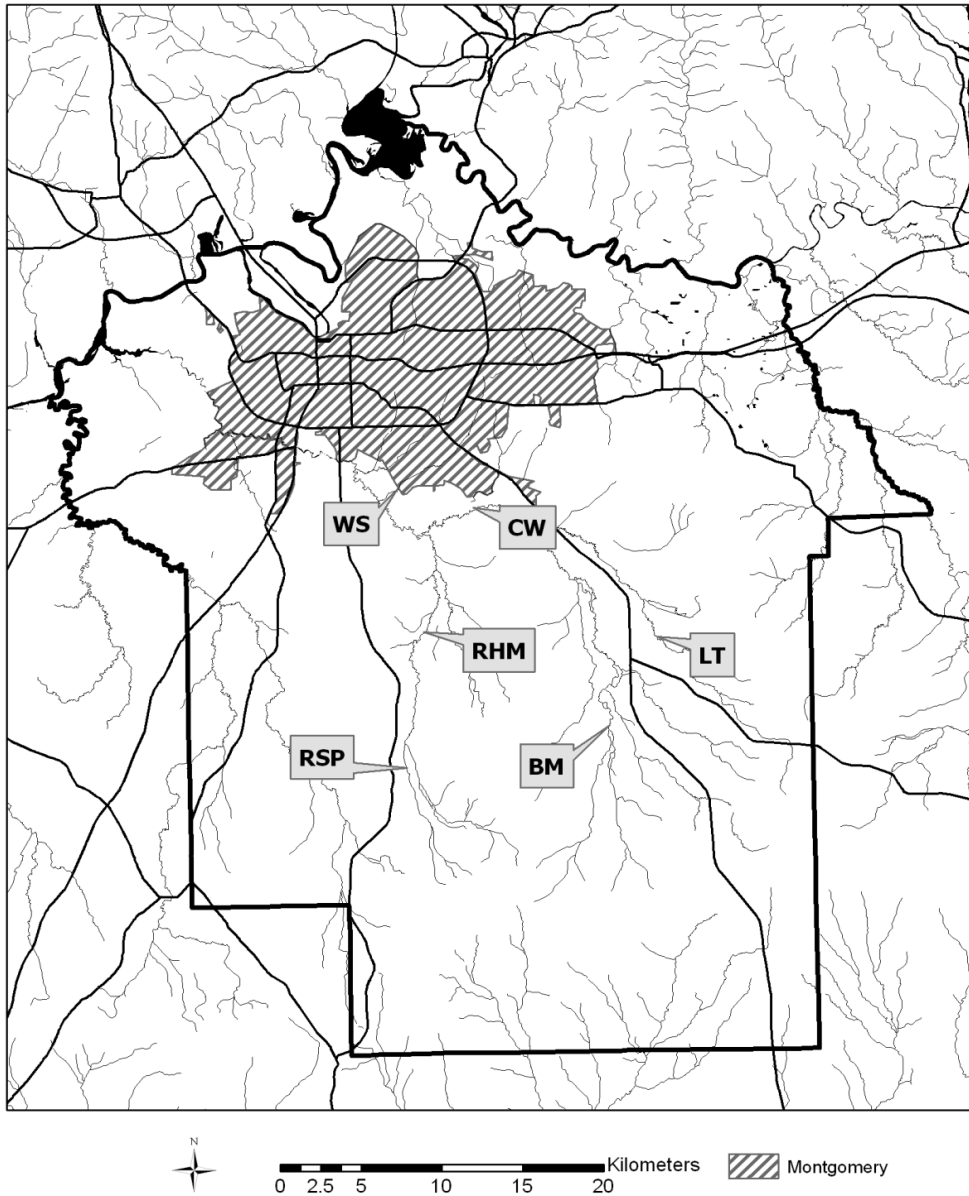


Fig. 2.1. Sampling sites in the Catoma Creek watershed: Catoma Creek at Woodley Road (CW), White Slough (WS), Little Catoma Creek (LT), Ramer Creek at Sprague Junction Road (RSP), Hobbie Mill Road (RHM), and Baskins Mill (BM).

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

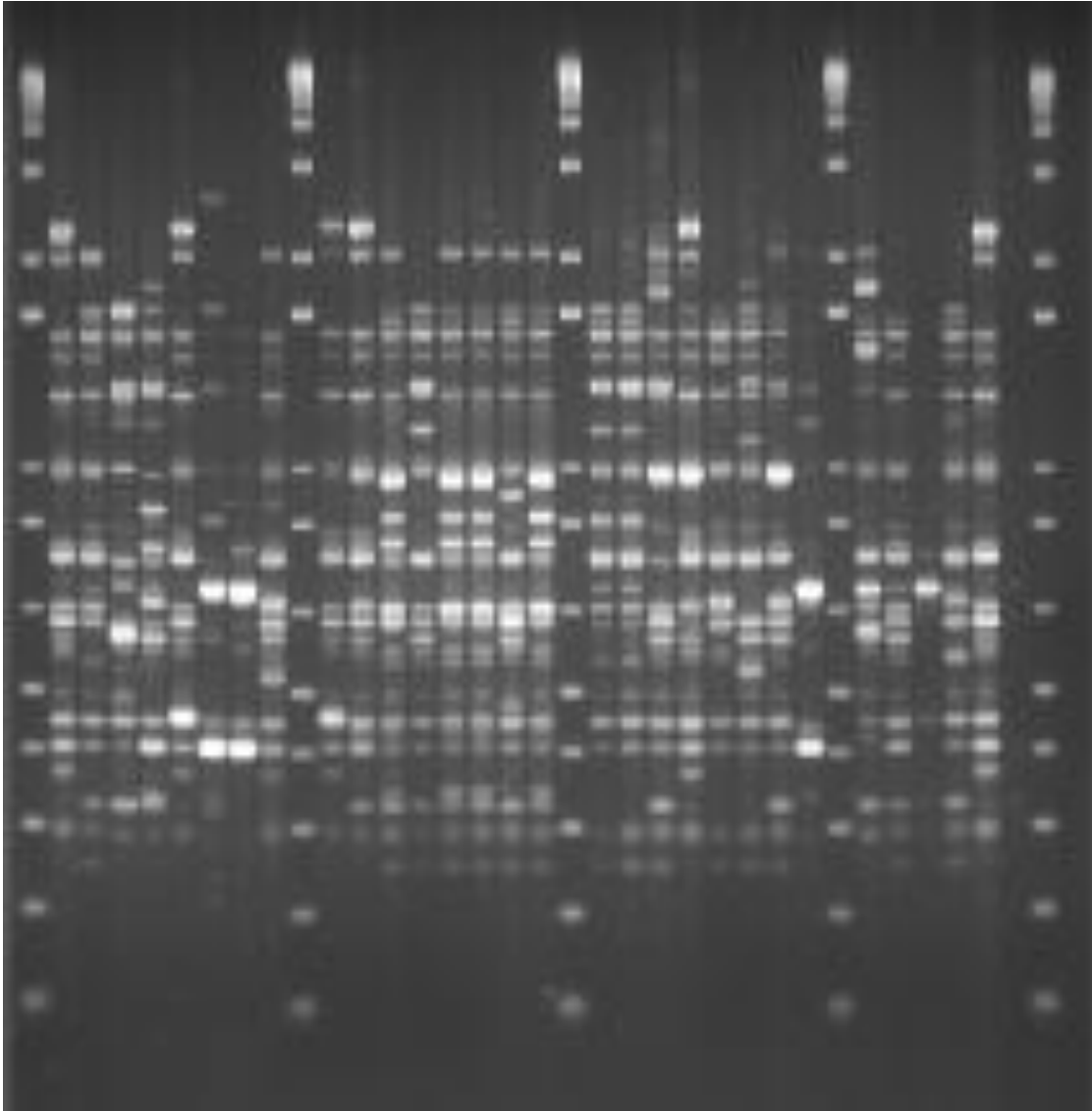


Fig. 2.2. *E. coli* rep-PCR fingerprints generated using the BOX A1R primer. Lanes 1, 10, 19, 28, and 35 contain 1-kb Plus DNA ladder. Lanes 2 and 33 contain *E. coli* K-12, and lane 34 is the negative control. Lanes 3-5 contain dog isolates, lanes 6-9, 11 and 12 chicken, lanes 13-18 horse, lanes 20-25 cattle, and lanes 26, 27, and 29 -32 wild turkey.

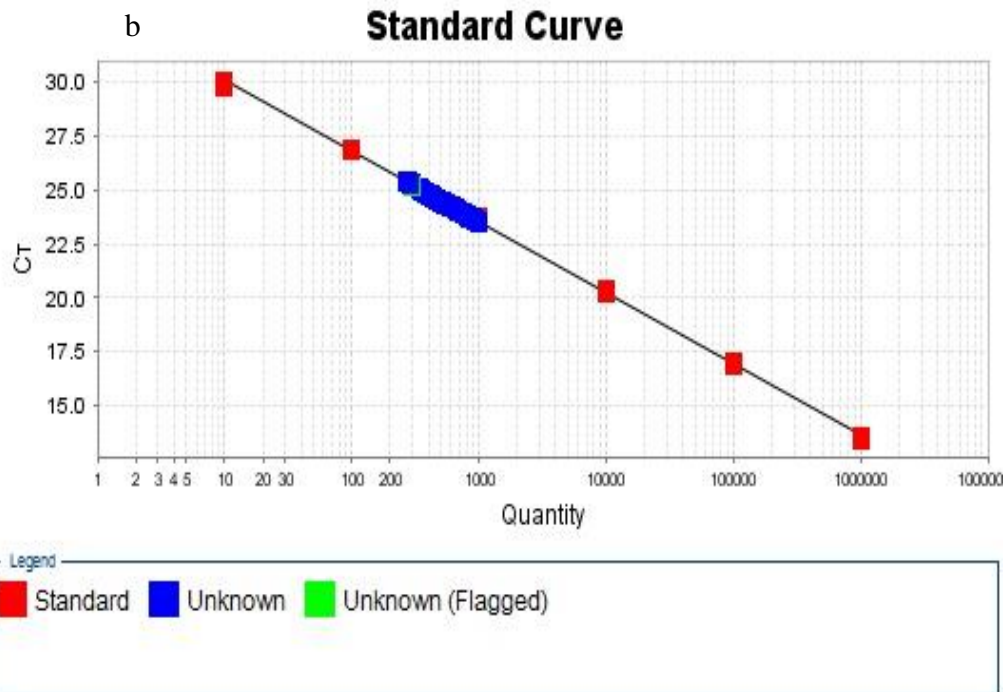
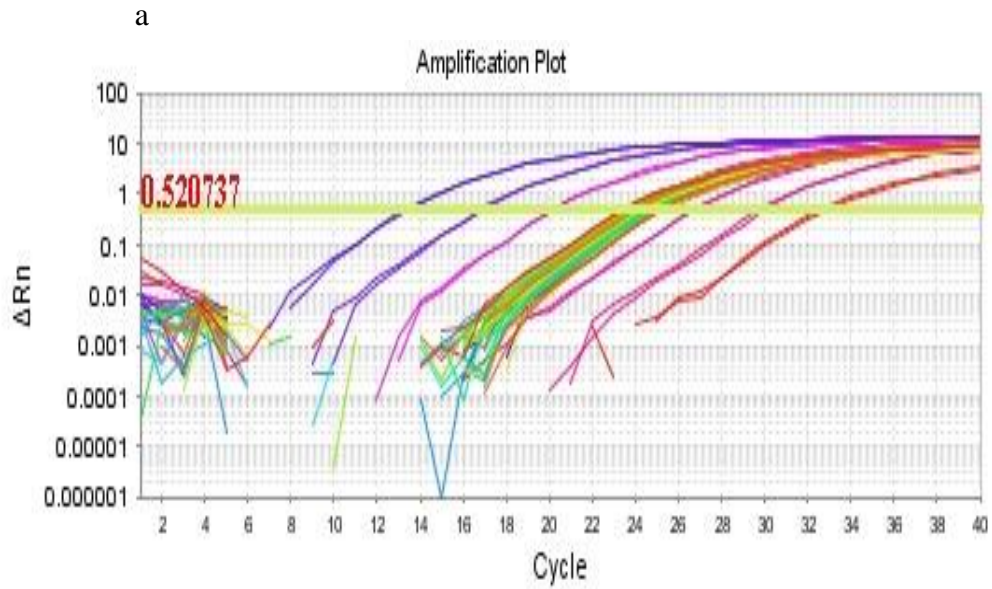


Fig. 2.3a. qPCR amplification plot: DNA from Sites RHM, BM, LT, WS and CW in June 2007 amplification with AllBac primers; Fig 2.3b. qPCR std. curve: DNA from Sites RHM, BM, LT, WS and CW in June 2007 amplification with AllBac primers.

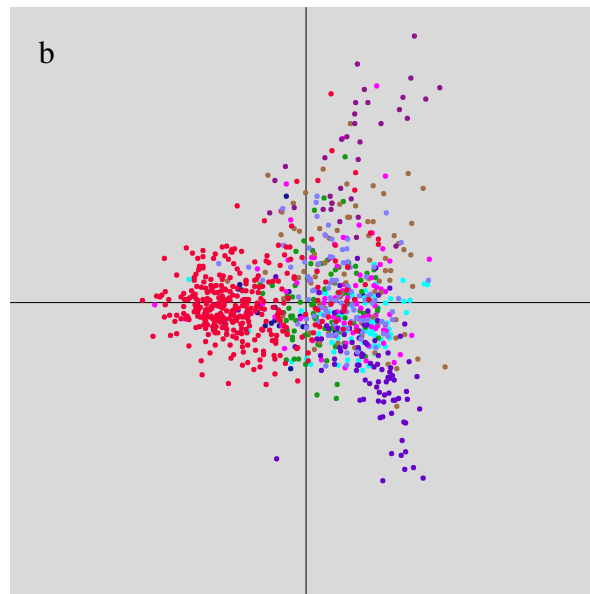
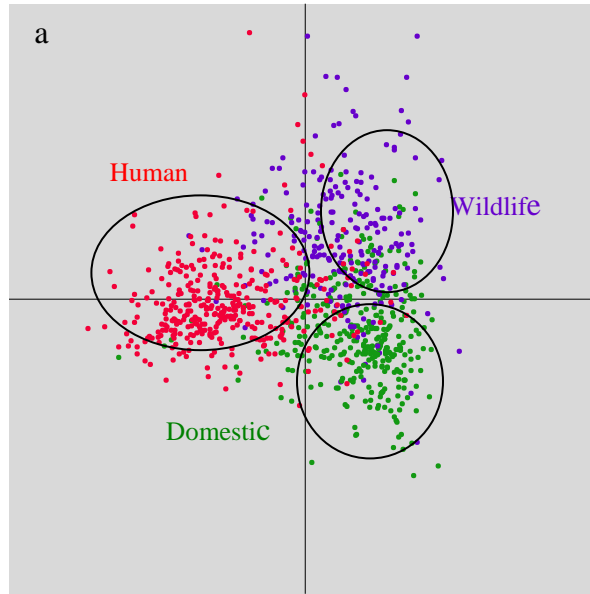


Fig. 2.4. MANOVA analysis for 3-way (a) and 9-way (b) source groups.

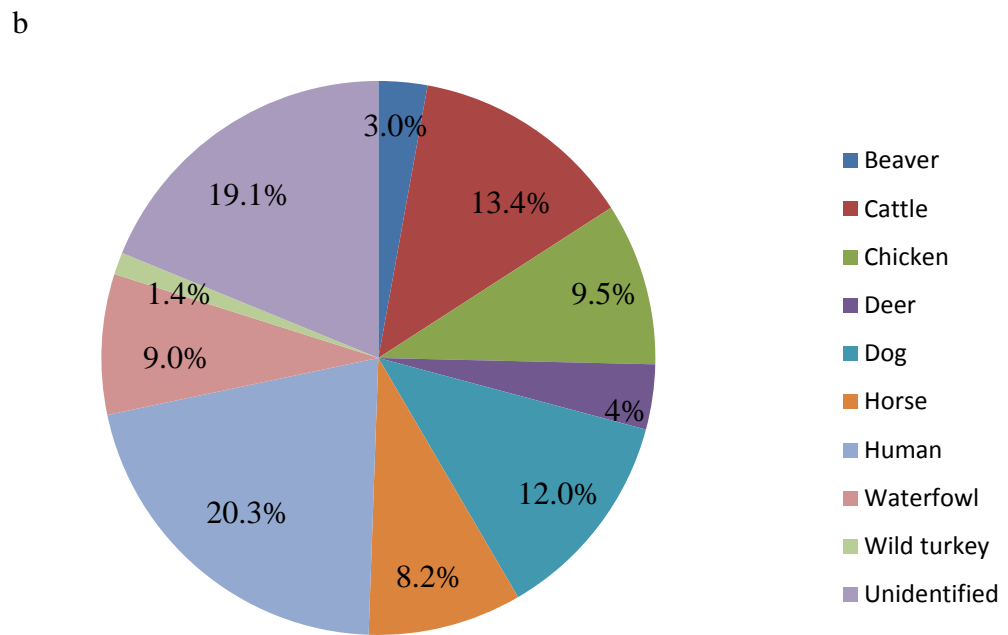
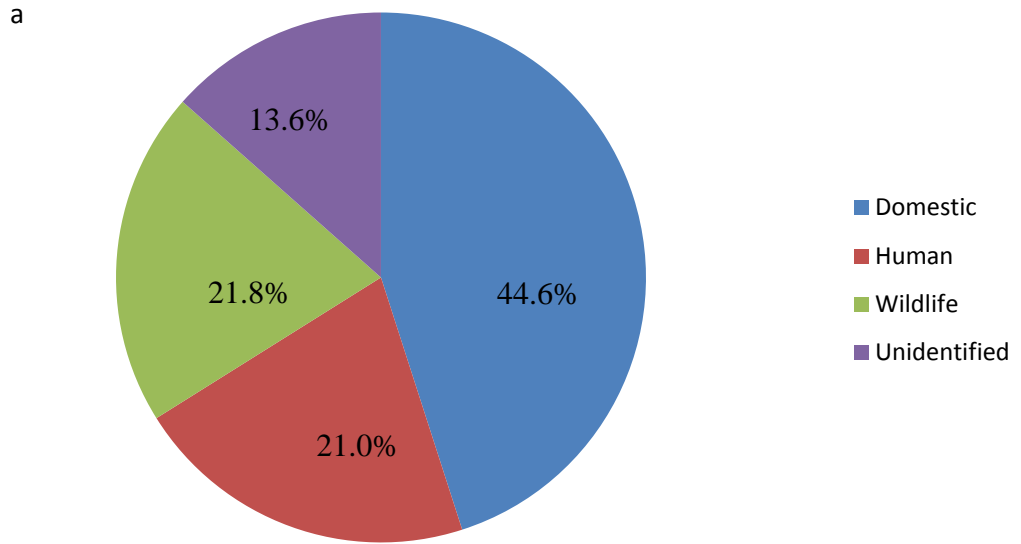


Fig. 2.5. Sources identification based on 3-way (a) and 9-way (b) source groups.

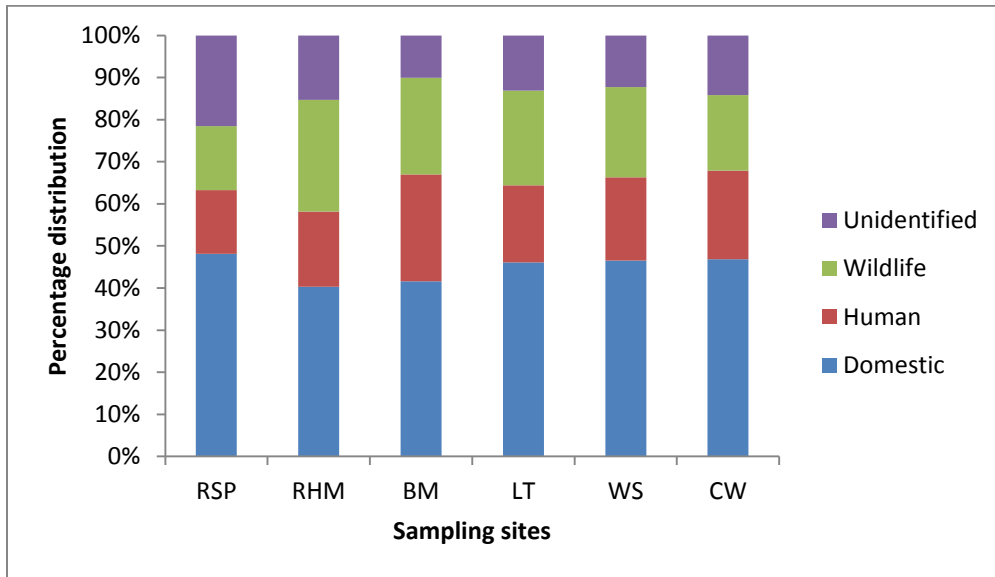


Fig. 2.6. Source distribution at different sampling sites in the Catoma Creek watershed.

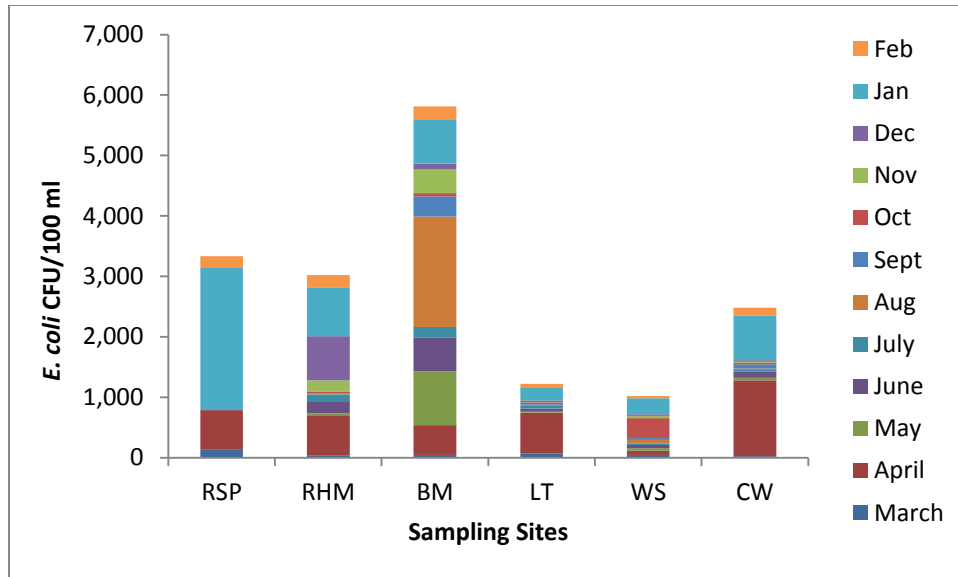


Fig. 2.7. Cumulative *E. coli* concentration found at each sampling site.

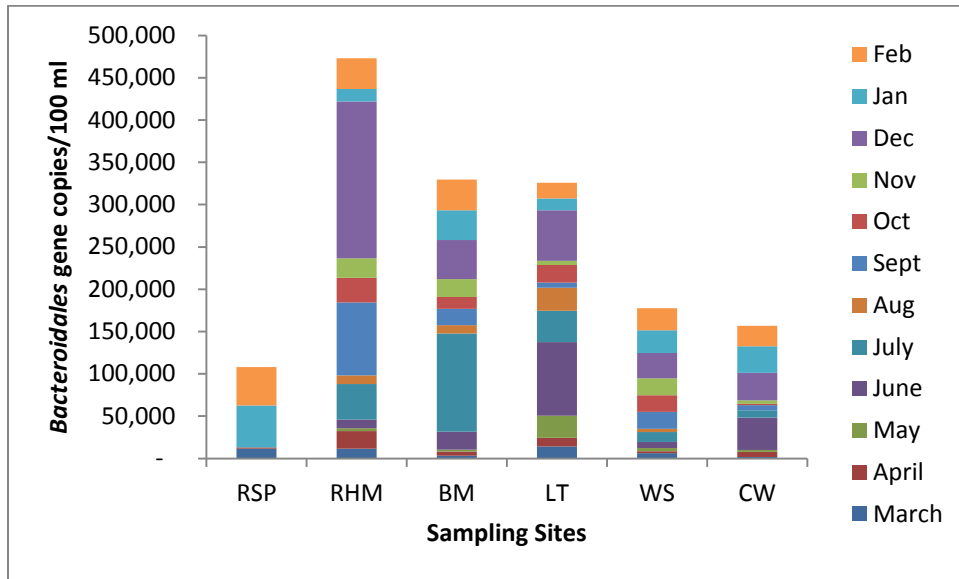


Fig. 2.8. Cumulative general *Bacteroidales* molecular marker concentration found at each sampling site. (Note: AllBac concentrations were checked for outliers using GraphPad software (<http://www.graphpad.com/quickcalcs/Grubbs1.cfm>) and one outlier for Site CW in August was removed from the data analysis at the significance level of 0.05).

3. Detection and Quantification of *Bacteroidales* Human and Cattle Associated Genetic Markers in Surface Water

Abstract

Selection of proper molecular markers for the detection of fecal contamination is imperative for conducting microbial source tracking in an impaired watershed. This study evaluated the utility of human- and cattle-associated *Bacteroidales* genetic markers as markers in quantitative PCR (qPCR) assays to assess fecal contamination in environmental samples in Alabama. Four human- and seven cattle-associated genetic markers were tested and HF183, targeting the 16S rRNA gene of *Bacteroidales*, and CowM3, targeting a gene of a surface protein, appeared to be the best human and cattle markers, respectively. Water samples from an urban stream were used to test the performance of the HF183 and CowM3 assays and general *Bacteroidales* marker and *E. coli* concentrations were also determined. Human-associated *Bacteroidales* genetic markers were detected in 87% of the water samples, while 8% of the samples contained cattle-associated markers. General *Bacteroidales* markers and *E. coli* were present in all samples and there was a positive correlation between these two parameters. Based on this study, HF183 and CowM3 show promise as suitable markers for fecal source identification in Alabama watersheds.

3.1. Introduction

Contamination of surface water by human and animal fecal matter increases the risk of exposure to pathogenic bacteria, viruses and protozoa. Accurate identification of these sources makes it possible to reduce contamination in a cost effective manner. Recently, the detection of host associated gene markers such as 16S rRNA genes and other functional genes has been the focus of a number of microbial source tracking studies (Bernhard and Field, 2000a,b; Shanks et al., 2006b, 2008). The order *Bacteroidales* is the main target of this approach because its members are strict anaerobes that live in the intestinal track of warm-blooded animals. These organisms are abundant in fecal matter, where the phylum Bacteroidetes represents 25–30% of the gut population (Salyers, 1984; Yang et al., 2009). Compared with *E. coli*, the current indicator bacteria, the Order *Bacteroidales* has a shorter survival period in the secondary habitat (Dick et al, 2010; Walters and Field, 2009). A recent study by our group also revealed that general *Bacteroidales* markers persisted in stream water microcosms for less than 10 days, whereas *E. coli* survived for more than 75 days in the same water (unpublished data). Therefore, the detection of *Bacteroidales* gene markers in water samples indicates recent fecal contamination of a particular water body. However, microcosm and mesocosm conditions cannot accurately reproduce real stream conditions. To the best of our knowledge, to date there has been no study conducted in real-stream conditions to determine the survival of *Bacteroidales* in their secondary habitat.

Several general markers targeting the hypervariable V2 region of the 16S rRNA

gene of *Bacteroidales* are available and allow for the discrimination of the order *Bacteroidales* from other bacteria (Bernhard and Field, 2000a; Layton et al., 2006; Kildare et al., 2007). Host associated markers have also been developed to detect pollution caused by humans (Bernhard and Field, 2000b, Seurinck et al., 2005; Layton et al., 2006; Kildare et al., 2007; Shanks et al., 2009; Okabe et al., 2007; Shanks et al., 2007), ruminants (Bernhard and Field 2000b; Layton et al., 2006), bovine (Shanks et al., 2006b, 2008), dogs (Kildare et al., 2007; Dick et al. 2005a), horses (Dick et al., 2005b), geese (Fremaux et al, 2010; Lu et al., 2009), chickens (Lu et al., 2009), pigs (Okabe et al., 2007; Dick et al., 2005b; Mieszkin et al., 2010), sheep (Lu et al., 2007), and elk/deer (Dick et al., 2005a).

Though the current trend is to use qPCR to detect the sources of fecal contamination in watersheds, this approach has both advantages and disadvantages compared with end-point PCR. The main advantage of qPCR is its ability to obtain quantitative as well as qualitative data, while end-point PCR provides qualitative data only; qPCR is also faster and more efficient than end-point PCR. Several gene markers that cannot be detected using end-point PCR, especially when present at low concentrations, can be detected using qPCR. However, qPCR is more technically demanding and expensive than end-point PCR, so for situations where researchers cannot afford to use qPCR, other PCR techniques such as touchdown PCR and re-amplification of PCR products may be utilized to improve assay sensitivity.

Although several markers have been developed to detect fecal contamination sources, most of these markers show geographical variability (Shanks et al., 2006a; Lamendella et al., 2007). It is therefore necessary to evaluate molecular markers in order

to determine the best for specific geographic locations. Although many *Bacteroidales* source tracking markers have been developed for various parts of the USA, as yet there have been no reported studies of the applicability of these markers in Alabama. The objectives of this study were therefore to:

1. Evaluate previously published human- and cattle-associated *Bacteroidales* molecular markers and select suitable ones for detecting the sources of fecal contamination in Alabama watersheds,
2. Detect and quantify general *Bacteroidales* and human- and cattle-associated genetic markers in surface water using qPCR and end-point PCR, and
3. Establish the correlation between the occurrence of general *Bacteroidales* markers and *E. coli* concentration.

3.2 Materials and Methods

A) Selection of suitable primers for Alabama watersheds

Sewage sample collection. A total of 20 raw sewage samples were collected from 18 different sewage treatment plants in Alabama located within a 50 miles radius of the City of Auburn. Samples were collected in sterile 1 L high-density polyethylene (HDPE) bottles, kept on ice, transported to the laboratory and processed on the same day. Sewage samples were then concentrated by centrifugation. Briefly, two 45-mL samples from each bottle were centrifuged at 2750 g for 15 minutes at 4°C. The supernatant was decanted and pellets in each tube were resuspended in 4.5 mL of phosphate buffered saline (PBS). The replicates were combined and samples stored at -20°C until DNA extraction.

Cattle fecal sample collection. Cattle fecal samples were collected from the Wilson Beef Teaching Center of Auburn University and other privately owned cattle farms near Auburn, AL. Samples were collected into sterile plastic bags and transported on ice to the laboratory. A subsample (10 g) of fecal matter was mixed with 10 ml of sterile water in a 50 ml sterile tube and vortexed for 10 minutes at the maximum speed to homogenize the fecal matter. These samples were stored at -20°C until DNA extraction.

DNA extraction. An aliquot (300 µl) of concentrated sewage and cattle fecal slurry was used for DNA extraction using the PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA) according to the manufacturer's instructions. The DNA concentration was measured using a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Evaluation of markers. Marker evaluation was conducted based on two criteria: specificity and sensitivity. Sensitivity is the percentage of target samples that are positive [sensitivity = true positives/(true positives + false negatives)], while specificity measures the ability of an assay to discriminate the target animal host from other animal sources [specificity = true negatives/(false positives + true negatives) (Gawler et al., 2007)]. Seven cattle- or ruminant-associated *Bacteroidales* primers and four human-associated *Bacteroidales* primers were evaluated (Table 1). Except for HuBac, BoBac and Hum 336, all the primers were tested for at least three sets of cattle and sewage samples collected over the period July 2007 to August 2008. All the samples that showed negative reactions for the sensitivity test were again amplified using either the same PCR conditions or with modifications such as a different annealing temperature, or touchdown PCR, as noted.

B) Application of selected primers to detect the sources of fecal contamination in an Alabama watershed

Water sample collection. This study was conducted in the Parkerson Mill Creek watershed, Auburn, AL. The creek is located within the lower Tallapoosa River basin and is designated for fish and wildlife. A 6.85-mile segment of this creek has been on the Alabama 303(d) List of Impaired Waters since 2008 due to elevated fecal bacteria (ADEM, 2011). Duplicate water samples were collected for five consecutive days in the second weeks of November and December 2009 and January 2010, a total of 15 sampling days. These samples were collected in 1 L sterile high-density polyethylene bottles from four locations at Parkerson Mill Creek (Fig. 3.1), transported to the laboratory on ice, and processed within six hours of collection. A total of 500 mL of water was vacuum filtered through a 0.45 µm membrane filter and stored at -20°C until DNA extraction using the MO BIO PowerSoil DNA Isolation kit. *E. coli* concentrations in the water samples were analyzed using the IDEXX Colilert-18/Quanti Tray method (IDEXX, Westbrook, MA). Stream water temperature was measured on site, and pH, electrical conductivity (EC) and turbidity were measured daily in the laboratory. On the 5th day of each sampling period, an additional 100 mL of water was collected and stored at -20°C for nutrient analysis. Total phosphorus and the metal ion concentrations (Ca, Mg, K, Al, Zn, Cu, Mn, and Fe) were determined using an inductively coupled plasma atomic emission spectrometer (SPECTRO CIROS, Germany), and total organic C and total N were measured with a TOC-V Combustion Analyzer (Shimadzu, Columbia, MD).

Amplification of *Bacteroidales* gene markers using end-point PCR

General, human and cattle *Bacteroidales* genetic markers were amplified with Bac32, HF183 and CowM3 primers (Table 3.1), as described by previous researchers (Bernhard and Field, 2000a,b; Shanks et al., 2008). The 25 µl PCR reaction mixture contained 1 µl of undiluted DNA (ranging from 0.54 ng/µl to 19.5 ng/µl) as the template, 1X PCR buffer without MgCl₂ (Promega, Madison, WI), 1.5 mM of MgCl₂ (Promega), 0.2 mM each dNTP (Promega), 0.16 µg/µl of bovine serum albumin (BSA) (Sigma, St Louis, MO), 0.5 µM of each primer (IDT, Coralville, IA), 4-units Taq DNA polymerase (Promega) and DNase/RNase free water (Promega). PCR was performed using a Biometra T-Gradient thermocycler (Whatman, Göttingen, Germany).

End-point PCR was used for detection of Bac32 general *Bacteroidales* and HF183 human-associated and CowM3 cattle-associated markers in Parkerson Mill Creek water samples. Thermocycling conditions for Bac32 and CowM3 were as follows: the initial denaturation at 94°C for two minutes, then 30 cycles of denaturation at 94°C for one minute, annealing at 60°C for one minute, extension at 72°C for one minute and final extension at 72°C for six minutes. Same thermocyclic conditions were used for HF183 except annealing temperature was 63°C for one minute.

To detect the amplified products using the Bac32 and HF183 molecular markers, 8 µl aliquot of PCR product was mixed with 2 µl of 6X loading dye (Promega). A 10 µl aliquot of each mixture was resolved using 1.5% agarose gel (Continental Lab Products, San Diego, CA) in 0.5X TBE buffer. PCR products amplified with the cattle specific marker were resolved using 2.0% low melting agarose gel (OmniPur, Lawrence, KS). The gels were electrophoresed at room temperature for two hours at 105 V and stained

with ethidium bromide having a final concentration of 0.5 µg/mL (Fisher Biotech) in 0.5X TBE buffer for one hour. The gel image was captured using a Gel Logic 200 imaging system (Eastman Kodak Co., Rochester, NY). Water samples were considered positive for a particular molecular marker if PCR product bands were aligned with those of the positive control. All end-point PCR runs contained a no template control (NTC) and those showing amplification in negative controls were repeated.

Preparation of qPCR standards. DNA extracted from sewage samples was amplified with general *Bacteroidales* primers 32F/708R and human-associated 183F/708R primers and resolved on 1.5% agarose gel (Bernhard and Field, 2000a,b). Gel images were used to confirm that there was only a single band present for fragment sizes of 686 bp and 525 bp, corresponding respectively to 32F/708R and 183F/708R. The remaining PCR amplified products were cleaned with a DNA Clean & Concentrator Kit (Zymo Research, Orange, CA) and cloned into the TOPO 2.1 cloning vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Plasmid DNA was extracted and cleaned with a QIAprep Spin Miniprep Kit (Qiagen Sciences, Germantown, MD). The plasmid DNA was amplified with 32F/708R and 183F/708R primers to verify the presence of recombinant DNA prior to sequencing using an ABI 3100 DNA Genetic Analyser with M13 primers. Each sequence was checked to confirm the presence of primer sites for general *Bacteroidales* primers (32F/708R) and AllBac primers (296R/412F) as well as human-associate primers HF183 (183F/708R) and qHF183 (183F/265R). Plasmid DNA concentration was measured with a NanoDrop ND-1000 UV spectrophotometer, and the gene copy numbers were determined using the formula given below.

$$\text{Gene copy No.} = \frac{6.02 \times 10^{23} * \text{plasmid DNA concentration (ng/}\mu\text{l)} * \text{template DNA (}\mu\text{l)}}{(\text{Molecular wt. of plasmid DNA} + \text{molecular wt. of gene insert})(\text{g/mole})}$$

The molecular weight of the plasmid DNA was obtained from the TOPO 2.1 cloning vector product manual. The molecular weight of the inserted gene fragment of *Bacteroidales* was determined using the calculator provided on Northwestern University's website (www.basic.northwestern.edu/biotools/oligocalc.html).

qPCR assay. The 15 μl qPCR reaction mixture contained 5 μl of template DNA, 1.4 μl of PCR grade water (Promega), 0.5 μl of 2% BSA, 7.5 μl of 2X Power SYBR Green master mixture (Applied Biosystems, Carlsbad, California), and 0.3 μl each of the 10 μM forward and reverse primers. The general *Bacteroidales* 16S rRNA genetic marker was amplified with AllBac 296F and 412R primers, the human-associated *Bacteroidales* 16S rRNA genetic marker with 183F and 265R primers, and the cattle-associated *Bacteroidales* marker with CowM3 primers. The amplification was performed using an Applied Biosystems StepOne real time PCR instrument with the following thermo cycle conditions: initial denaturation at 95°C for 10 min, then 40 cycles of denaturation at 95°C for 15 S, primer annealing at 60°C for 30 S and an extension at 72°C for 30 S. Subsequently, the melt curve was analyzed with a temperature gradient of 0.5°C per minute from 95°C to 60°C. All the samples were run in duplicate with at least five point standard curves in duplicate and no template controls (NTC). The lowest concentration of a particular genetic marker in the linear range of qPCR standard curve was considered to be the lower limit of quantification (LLOQ) of each qPCR assay (Seurinck et al., 2005). The limit of detection (LOD) was defined as the fifth percentile among observed quantification cycles (C_q) across all NTC and DNA extraction blank (lab blanks)

reactions (Dick et al., 2010). Gene copy numbers above the LLOQ were reported as quantifiable, gene copies below the LLOQ and above the LOD were reported as detectable but not quantifiable, and gene copy numbers below the LOD were reported as non-detectable.

3.3 Results

A. Marker selection

Table 3.1 shows the *Bacteroidales* human- and cattle-associated markers used for evaluation. If any assay showed less than 100% sensitivity, those assays were rerun with the same fecal samples to verify the results and the sensitivity was recalculated if necessary. Specificity was determined using DNA extracted from cattle feces or sewage. Among all the human assays tested, only HF183 showed 100% sensitivity and 100% specificity against cattle fecal samples. The other three human-associated markers, Hum 163, Hum 336 and HuBac, all demonstrated poor specificity/sensitivity (Table 3.2). These markers were therefore not evaluated further.

Bac1, Bac2, Bac3, BoBac, CF128 and CF 193 are commonly used to detect cattle fecal pollution, but our results showed that CF193 failed to amplify DNA from any of the cattle samples. CF128 and BoBac had 100% sensitivity but both primers had low specificity, at 15% and 50%, respectively (Table 3.2). CowM3 primers showed 100% sensitivity and 100% specificity against sewage samples. Based on these results, HF183 was selected to detect human fecal pollution in subsequent experiments and the CowM3 marker was selected to detect cattle fecal contamination.

B. Detection of fecal contamination sources in Parkerson Mill Creek

qPCR assays were performed to quantify general, human-associated and cattle-associated *Bacteroidales* markers. The quality of qPCR data was evaluated mainly based on the published guidelines for real-time PCR experiments (Table 3.3). The coefficients of determination (R^2) for all assays were above 0.99, and assay efficiencies were between 94% and 106%. The lower limit of quantification (LLOQ) for the AllBac assay was 2000 copies/100 ml of water and those for both qHF183 and CowM3 were 20 gene copies/100 mL of water. The limits of detection (LOD) for AllBac, qHF183 and CowM3 assays were 336, 13 and 10 gene copies per 100 mL of water, respectively.

A total of 60 water samples were collected from the four sites, each in duplicate, over the three-month period and used for identification of fecal contamination sources at Parkerson Mill Creek. Both end-point PCR and qPCR detected general *Bacteroidales* markers in all samples (Table 3.4). The qPCR results showed that 91% of samples contained general *Bacteroidales* markers above the LLOQ, while 100% of the samples were above the LOD (Table 3.4). Overall, samples collected in November and January showed the highest concentrations of general *Bacteroidales* gene markers, while the December samples had the lowest concentrations in most of the samples (Fig. 3.2). The highest level for the general *Bacteroidales* gene markers measured in this study was 1,180,500 gene copies per 100 mL of water at Site H on the 5th sampling day in November. Site Q had higher general *Bacteroidales* markers than any of the other sites.

End-point PCR detected lower percentages of human markers in water samples than were found with qPCR. Overall, end-point PCR showed only 14.2% of samples as containing human associated HF183 markers. None of the December samples were

positive for HF183 and only 20% and 22.5% of the November and January samples, respectively, contained human markers (Table 3.4). Based on qPCR amplification, 87.5% of the water samples were found to contain the human-associated marker at levels above the LOD and 79% were above the LLOQ (Fig. 3.3). In January, all the water samples contained levels of the human marker above the LOD and 95% of the samples were above the LLOQ. The lowest percentage of qHF183 markers was detected in the water samples collected in December, where only 72.5% and 60% of the samples were above the LOD and LLOQ, respectively. Ninety percent of November's samples were above the LOD and 82.5% were above the LLOQ. The highest number of qHF183 gene markers were detected at Site Q on the 4th sampling day in January: 35,118 gene copies/100 ml. Site H had the lowest percentage of human marker detection, with only 73% of the samples containing human markers. A total of 96.7% of the samples collected from Site Q contained human markers above the LOD, while 87% of the samples were above the LLOQ (Fig. 3.3).

Compared with human markers, cattle markers were detected at lower frequencies by both end-point and quantitative PCR. Only 5% of the samples contained CowM3 markers according to end-point PCR amplification, rising to 8% with qPCR, although only 6% of the samples were above the LLOQ. End-point PCR was able to detect CowM3 markers only in November, while qPCR detected cattle markers during both the November (12.5%) and December (10%) sampling periods. Neither end-point PCR nor qPCR detected any cattle markers in January (Fig. 3.4).

All the water samples contained *E. coli* (Table 3.4), with Site Q having the highest *E. coli* concentration (5,346 CFU/100 ml) in November. The geometric means for

the *E. coli*, general *Bacteroidales* and human *Bacteroidales* for all four sites are shown in Fig. 3.5. A total of 92% of the water samples were above the 5-day geometric mean criterion of 126 CFU/100 ml that must not be exceeded in bodies of water to be used for swimming and full body contact activities (USEPA, 1986, 2012). The samples collected from Site B in January 2010 were the only samples whose geometric means fell below 126 CFU/100 ml. Parkerson Mill Creek is not generally used for this type of activity, but is instead designated for fish and wildlife, so these sampling sites were evaluated based on the EPA's single sample maximum (SSM) criterion for fish and wildlife, 576 CFU/100 mL (USEPA, 1986). A total of 32% of the samples collected in this watershed were above the SSM criterion for fish and wildlife, and 60%, 30% and 5% of the samples collected in November, December and January, respectively, violated this criterion. Site D had the lowest number of violations (20%) and Site B the highest (40%) for fish and wildlife (Fig. 3.6).

Water samples were collected in the late fall/early winter months when frequent rainfall events occur in the study area (Appendix 3.1). In November 2009 there was no significant rainfall, which can contribute to runoff, for 10 days prior to sampling, but heavy rain occurred during the 2nd and 3rd days of sampling. December had three rainy days during the sampling period; 25 mm (1") rainfall events occurred on two consecutive days 7 days before collecting samples, and 18 mm (0.7") rainfall occurred the day before sample collection began. January was relatively dry, with no rain occurring during the sampling period or for the 10 days leading up to it. Rainfall significantly affected the distribution of host associated gene marker and *E. coli* concentrations in this watershed. The highest *E. coli* concentrations were found in November, followed by December and

then the lowest concentrations in January. The Student t-test revealed these *E. coli* distributions in different sampling months to be significant ($P < 0.001$). AllBac concentrations were high during the November sampling period, but the concentration distribution over the three months was not statistically significant ($P > 0.05$). Cattle-associated markers were found in November and December, but were not detected in January. In contrast, all the samples collected in January contained human markers above the LOD, while 95% of the samples were above the LLOQ. December had the lowest percentage of samples containing human-associated markers above the LOD. The Student t-test found that the human associated marker distributions in the December and January samples were statistically significant ($P < 0.05$).

Fig. 3.7 shows an overall correlation coefficient (r) of 0.68 ($p < 0.0001$) between *E. coli* and the AllBac marker. For the individual sites, Site B showed the lowest correlation ($r = 0.60$) while Site D showed the highest correlation ($r = 0.82$) between *E. coli* and general *Bacteroidales* markers.

The physiochemical parameters of the water samples collected during the study period are summarized in Table 3.5. The water pH varied from 6.67 to 8.02 and the average water temperatures ranged between 2.0°C and 18.5°C. The lowest turbidity (1.50 NTU) was found at Site B and the highest turbidity (102 NTU) was at Site H. Since turbidity varied over such a wide range, turbidity was divided into two groups—low turbidity (the lowest turbidity reported at each site during the 5-day sampling period in each month) and high turbidity (the highest turbidity reported at each site during the 5-day sampling period in each month)—to examine the correlation between turbidity and *E. coli* and general *Bacteroidales* gene markers. The concentrations of Ca, Mg, K, organic C

and total N did not change significantly based on either rainfall or sampling site, although total P concentrations were high during the November sampling period and low during the January sampling period. The C/N ratio was the highest for Site H, at 45.1, in January while the lowest ratio, also in January, was at Site B, at 10.0 (Table 3.5).

Table S 3.1 shows the Pearson's product moment correlation between the geometric mean of AllBac and *E. coli* with each month's physiochemical parameters. The AllBac markers showed no significant correlation with any of the physiochemical parameters, but *E. coli* had a significant correlation between total P ($r=0.64$, $p=0.02$) and high turbidity level ($r=0.81$, $p=0.001$).

3.4 Discussion

This study evaluated four human- and seven cattle-associated markers based on marker specificity and sensitivity criteria. The evaluation results for the HF183 marker were comparable with most of those reported previously (Seurinck et al., 2005; Ahmed et al., 2007; Fremaux et al., 2010; Dick et al., 2010; Mieszkin et al., 2010; Peed et al., 2011), namely 100% sensitivity and 100% specificity against cattle fecal matter. As the cattle fecal *Bacteroidales* 16S rRNA gene showed 100% specificity for HF183 primers, it was not deemed necessary to test it further with other source groups. Several studies have been conducted in other parts of the United States as well as in other countries to evaluate this primer. Not only have most of these studies agreed on the 100% specificity of this marker, their results have further revealed that HF183 and qHF183 are not geographically constrained (Seurinck et al., 2005; Ahmed et al., 2007). Thus, it is reasonable to assume that HF183 will not amplify *Bacteroidales* 16S rRNA originating from other source groups. HF183 was originally identified in Oregon and its sensitivity

as a marker tested against three sewage and 13 human fecal samples, successfully amplifying all three sewage samples and 11 out of the 13 human fecal samples. Primer specificity was tested against cattle, cats, deer, dogs, ducks, elks, goats, llamas, pigs, seagulls and sheep, and HF183 did not amplify any of these non-target groups (Bernhard and Field, 2000b).

The amplicon length of HF183 is 525 bp, which is too long for qPCR amplification because it may result in low efficiency. Therefore, a new reverse primer was designed in Belgium that decreases the amplicon length to make it suitable for qPCR amplification. The new amplicon length is 82 bp and it was named qHF183. This primer was evaluated for sensitivity and shown to amplify 6 out of 7 human fecal and all four sewage samples tested. When evaluated for specificity, no amplification of dog, horse, cow or pig feces occurred (Seurinck et al., 2005).

Another study in Queensland, Australia, found that HF183 was able to amplify all 52 sewage and septic tank samples tested. A specificity test revealed that HF183 did not amplify *Bacteroidales* 16S rRNA genes originating from ducks, kangaroos, cattle, horses, dogs, chickens, pigs, pelicans, goats, deer, wild birds and sheep (Ahmed et al., 2007). A Canadian study also revealed 100% specificity of HF183, finding no cross amplification with DNA from cows, pigs, chickens, geese, moose, deer, bison and goats, with 95% sensitivity (Fremaux et al., 2010).

A recent study conducted in 27 laboratories in the United States and Europe evaluated the performance of 41 microbial source tracking methods, based primarily on the specificity and sensitivity of each assay. Seven of the laboratories in this inter-laboratory study evaluated HF markers and reported that HF183, qHF183 with SYBR

Green and Taqman assays have sensitivities of between 92% and 100%, although one lab did report an HF183 sensitivity of 50%. The results from one lab showed the qHF183 SYBR Green assay to have low specificity, 28%, but three other laboratories showed specificities ranging from 80% to 100%. The qHF183 SYBR Green assay also showed some cross reactivity with deer. The main objective of this study was to examine inter-laboratory variability on different assay performance, however, and it did not address the effect of spatial variability on assay performance (Boehm et al., 2013).

A study in California compared BacHum-UCD and HF183 assays for specificity and sensitivity. HF183 was able to amplify all sewage, 63% human fecal and 67% septage samples. When the specificity of these assays was evaluated, HF183 did not amplify any dog, gull or raccoon samples, although 8% of the cat samples were amplified by HF183 primer. BacHum-UCD showed 100% sensitivity against sewage, human fecal and septage samples, but poor specificity against cat, dog, gull and raccoon fecal samples (Van De Werfhorst et al., 2011). Several other studies have also demonstrated the high specificity and sensitivity of HF183 primers, which have been successfully used to detect human fecal contamination in surface water (Dick et al., 2010; Mieszkin et al., 2010; Peed et al., 2011; Fremaux et al., 2009; Lamendella et al., 2007; Gourmelon et al., 2010; Sauer et al., 2011). Based on these results, qHF183 was selected as the molecular marker in the current study with which to detect fecal contamination in Parkerson Mill Creek.

The selection of a good molecular marker to detect cattle fecal pollution in this watershed was more complex, as unlike HF183, cattle molecular markers have shown considerable spatial variability (Shanks et al., 2006a; Lamendella et al., 2007). The specificity and sensitivity of seven commonly used ruminant- or cattle-associated primers

were therefore evaluated. CF193 and CF128, which were developed in Oregon with reportedly high specificity and sensitivity (Bernhard and Field, 2000a,b) were less effective in Alabama. In our study, the CF193 primer showed very poor sensitivity, failing to amplify any of the ten cattle samples tested and although CF128 was able to amplify all 41 cattle samples with 100% sensitivity, it showed poor specificity, amplifying only 17 out of 20 sewage DNA samples. The developers of these markers used only a relatively few samples to draw their specificity and sensitivity conclusions, and their samples did not represent large geographical regions.

An earlier intensive study was conducted to determine the prevalence of several cattle associated markers in different geographic regions in the United States (Shanks et al., 2010) revealing strong disagreements between previously reported specificity and sensitivity values for CF128, CF193, Bac2 and Bac3 assays, which they ascribed to geographical variability. These results suggest that before adopting any molecular marker for source identification, it is vital to perform extensive testing to confirm its suitability for a particular watershed. Although the current study was conducted in Auburn, Alabama, the primers tested were not developed in Alabama or even the southeastern USA, so the first part of the study examined the performance of a number of potentially suitable primers for detecting human and cattle fecal pollution in Alabama watersheds.

One marker comparison study over a wider geographical area (West Virginia, Wyoming, Ohio, Virginia, Delaware, Georgia, Florida) found the overall specificities of CF193 and CF128 to be 99.9% and 76.0% and the sensitivities 68% and 85%, respectively (Shanks et al., 2010). In this study, samples collected from Georgia showed high prevalence (sensitivity) of CF193 and CF 128 (97%) in cattle herds. Their sensitivity

results were comparable with ours for CF128, where CF128 showed 100% sensitivity. But Shanks et al. (2010) did not test specificity against human fecal or sewage, where we found CF128 has poor specificity against sewage. None of these studies conducted specificity test on human fecal or sewage in Alabama or the southeastern USA. While a study in Canada found that CF128 was able to differentiate 92% of ruminant samples from non-ruminant samples, cross reactions occurred in 22% of the pig samples (Fremaux et al., 2009).

Thus, one reason for these differences in our results and reported values of other studies may be because *Bacteroidales* species prevalent in the northern and western states may be different from those prevalent in the southeastern USA. We collected cattle fecal samples from Auburn University Beef Teaching Unit as well as pasture grazing cattle fields around Auburn. We do not have dietary records (using of antibiotics, supplementary etc.) about cows that we collected fecal samples. Because gut population variability may be due to different management practices, particularly with regard to the diets, antibiotics and supplements used in beef cattle and milking cows. The age of the herd and climatic conditions may also affect the diversity of bacterial populations (Klieve et al., 2003; Shanks et al., 2011). Further, a pyrosequencing study was conducted in four different geographic locations around the US—Ohio, Georgia, Nebraska and Colorado—with different feeding operations. This study revealed that the species in cattle gut populations show considerable regional variability in addition to differences related to feeding operations, although the choice of feeding operation was more important for determining the cattle microbiome than the geographical location (Shanks et al., 2011).

Evaluation of BoBac primer revealed that the BoBac primer combines a high sensitivity (100%) with a low specificity, with 50% of sewage DNA also being amplified by BoBac primers (Table 3.2). This does not agree with the results reported by the original study, which reported high specificity and sensitivity for this assay (Layton et al., 2006), although Shanks et al. (2010) also found a low specificity of BoBac of only 47%.

The bovine origin Bac1, Bac2 and Bac3 primers tested for this study all exhibited 100% specificity but low sensitivities of 30%, 20% and 0%, respectively. These specificity results are comparable with those of a previous study (Shanks et al., 2010) that found Bac2 and Bac3 to have specificities of 100% and 98.9%, respectively. However, the same study reported higher sensitivity (prevalence) results, with 54% of the Bac2 and 69% of the Bac3 primers able to amplify 247 bovine fecal samples from 11 herds. The prevalence of Bac3, in particular, does not seem to be consistent: the overall prevalence of Bac3 varied from 0% to 100%, with fecal samples collected from one cattle herd in Nebraska showing 100% prevalence while another herd also from Nebraska showing 0% prevalence. Both Bac3 and CowM3 have the same gene target of sialic acid-specific 9-*O*-acetylerase secretory protein homolog, and both primers are specific with regard to amplifying only *Bacteroidales* from bovine feces. However, CowM3 has a better sensitivity than Bac3 and targets a different primer site in the sialic acid-specific 9-*O*-acetylerase secretory protein homolog (Shanks et al., 2010).

In the present study, CowM3 showed 100% sensitivity and 100% specificity against sewage *Bacteroidales*. This primer was developed in Cincinnati, OH, but it was tested for sensitivity using 60 bovine fecal samples collected from six different populations in three different states, all of which were amplified by CowM3 primers.

Specificity was evaluated against 16 animal groups from six states and CowM3 did not amplify any of these non-target DNAs (Shanks et al., 2008). Another study found that CowM3 has high specificity, at more than 98.9%, but low sensitivity compared with other ruminant associated markers such as BoBac (Shanks et al., 2010). When tested against 22 cattle fecal samples collected from various locations in and around Auburn, AL, where the cattle were kept on pastures, all the samples were amplified with CowM3 primers. This primer was then evaluated for specificity using 10 sewage samples collected from several locations within a 50 mile radius of Auburn, AL, and none of these sewage samples cross reacted with the CowM3 marker. Therefore, CowM3 primer was selected for the source identification studies.

However, some formation of primer dimers was observed in the CowM3 assay, especially in those samples with low concentrations. Hot start PCR, touchdown PCR, different Mg^{2+} concentrations, different primer concentrations and different annealing temperatures were used to address this problem, but none were successful (data not shown). The original study (Shanks et al., 2008) did not report any problem with primer dimers because they used this primer with the qPCR Taqman assay. In our study, Cow M3 primer dimer formation was a common problem for both qPCR and end-point PCR. Primer dimers do not affect the target amplification at the initial stage of qPCR cycles, but at a later stage where sample amplification is taking place after C_q of 26, the formation of primer dimers significantly affected the reaction rate and efficiency and, ultimately, the estimated initial gene concentration (Mehra and Hu, 2005). Our study failed to find a significant amount of cattle markers in the water samples, which may be due to the masking of low cattle signatures by primer dimers, although it is also possible

that cattle fecal matter from the watershed simply did not reach the stream. Based on our results, CowM3 with end-point PCR or qPCR with SYBR Green assay does not seem to be a good option for the detection of cattle fecal contamination in water samples.

Furthermore, as the CowM3 marker has been patented by USEPA, there has been no published research using CowM3 other than that reported by Shanks' research group in Cincinnati, OH. The fact remains, however, that this is the only cattle associated primer that exhibited 100% specificity and was abundant in all the cattle fecal samples tested for this study. In a large-scale study, Boehm et al. (2013) compared the performance of CowM3 with those of 40 other microbial source tracking methods, reporting that CowM3 showed 100% sensitivity and specificity for known fecal samples but poor sensitivity for the challenge samples at low concentrations. However, CowM3 was evaluated in only one laboratory, namely the developer's lab. These results clearly demonstrate the necessity of testing this marker by other research laboratories. Based on our results, in future studies, one should be cautious before using CowM3 with the SYBR Green assay, but it is worthwhile further studying this marker using the Taqman assay.

Parkerson Mill Creek is on the Alabama 303(d) List due to its elevated levels of fecal bacteria, so knowing the sources of fecal contamination in this watershed would undoubtedly help when determining appropriate best management practices to mitigate the problem. The Auburn University Beef Teaching Unit is located near Site Q, which was a suspected source of fecal contamination in this watershed. Also, sewage lines are located along Parkerson Mill Creek, although detailed information on their precise location and status, for example a sewage line map, is not available. Although pets and

wildlife may also contribute to fecal pollution, this study mainly focused on detecting human and cattle contribution to the fecal pollution in this watershed.

Fecal contamination was evaluated using three parameters: *E. coli*, general *Bacteroidales* Bac32 end-point PCR assay and AllBac qPCR assay. These three criteria provided clear evidence that all the water samples collected from Parkerson Mill Creek contained fecal bacteria. This watershed is heavily contaminated with fecal matter, with 95% of the samples tested containing *E. coli* concentrations above the EPA's 5-day geometric mean criterion for recreational water. The ultimate goals of any bacterial source tracking study have to be to identify the sources of fecal contamination as well as quantifying the risks associated with the presence of pathogens in the water. A study conducted in Japan to determine the correlation between pathogens and *Bacteroidales* general and human-associated markers in a municipal wastewater treatment plant and in surface water found that if the human and general *Bacteroidales* gene marker concentrations were above 10^3 and 10^4 gene copies/100 ml, respectively, there was a possible risk of pathogens such as *E. coli* O157:H7 and *Salmonella* also being present (Savichtcheva et al., 2007). The results reported here show that most of the samples collected from Parkerson Mill Creek have high concentrations of general as well as human associated *Bacteroidales* gene markers, with 45.2% of water samples containing general *Bacteroidales* gene copy numbers above 10^4 /100 ml and 13.3% of water samples containing human associated markers above 10^3 /100 ml. These results strongly suggest the likelihood of pathogen contamination in Parkerson Mill Creek from fecal matter. To date, there has been no investigation of potential pathogens in this water body and it is crucial that such research should be undertaken as soon as possible.

Comparison of the end-point PCR and qPCR results revealed that end-point PCR identified only 14.2% of the samples as containing human markers while qPCR put this figure somewhat higher, detecting human markers in 87.5% of the samples. This difference was not unexpected because most qPCR assays run for 40 to 45 amplification cycles, while end-point PCR consists of only 30 cycles. Therefore, qPCR was better able to amplify the low marker concentrations present in the water samples, especially in the later stages of amplification. In addition, SYBR Green and Taqman assay chemistries are more sensitive than end-point PCR DNA amplification chemistries

(http://www6.appliedbiosystems.com/support/tutorials/pdf/rtqpcr_vs_tradpcr.pdf).

However, if the water samples have high concentrations of a particular marker, end-point PCR is sensitive enough to detect this fecal contamination. There was no significant difference in the CowM3 amplification between end-point PCR and qPCR, but this may be due to the formation of primer dimers in both end-point PCR and qPCR.

This study found a good correlation between *E. coli* and *Bacteroidales*. The comparison between culture-based and molecular-based methods is not common. A study conducted in the Great Lakes compared *Enterococcus* culture based and molecular based methods (Haugland et al., 2005) and found a significant correlation between these two methods ($r = 0.68$), while another study of California coastal beaches also found a relationship between fecal indicator bacterial abundance and the existence of human-associated *Bacteroidales* markers (Santoro and Boehm, 2007). Our study, as well as previously reported studies, suggests the utility of using general *Bacteroidales* as an alternative indicator.

In this study, there was a negative relationship between the prevalence of the human marker and rainfall patterns. These results are in direct contrast with the *E. coli* and cattle results, both of which showed a positive correlation with rainfall. It is reasonable to assume that *E. coli* and cattle fecal matter entered the stream with the surface runoff water, while human fecal sources are much less likely to do so but instead could originate from seepage from a sewage-carrying line close to the stream. This assumption is supported by the rainfall data. When rainfall is high, these human markers would be diluted and higher concentrations were indeed observed when there was no rainfall, especially in January when water samples showed high concentrations. Broken sewage lines or leaks in sewage-carrying clay pipes may contribute to the human marker concentrations in this creek. On the other hand, a strong correlation between *E. coli* geometric means and total P, as well as 'high level' turbidity in the stream water, suggests that *E. coli* should be positively correlated with rainfall. Consequently, *E. coli*, P and sediment particles are all likely to enter the stream as a result of water runoff after rain events.

3.5. Summary

E. coli concentration, general *Bacteroidales* 16S rRNA gene amplification with end-point PCR and qPCR can all successfully be used to detect fecal contamination in a watershed. qPCR is more sensitive than end-point PCR and also allows for the quantitative estimation of fecal contamination but is a more technically demanding and expensive approach. End-point PCR can also be useful in detecting sources of fecal contamination in certain situations, however. This study successfully utilized end-point PCR with touchdown thermocyclic conditions and reamplification of PCR products to

amplify stream water samples containing low concentrations of fecal matter. Source identification is a vital factor that must be considered by those charged with managing local and national waterways, especially in a mixed land-use watershed, so the selection of proper molecular markers for a particular watershed is imperative.

The dominant fecal contamination source can vary based on rainfall intensity and how that rainfall contributes to the runoff. This study clearly showed that human sewage made a significant contribution to fecal contamination during dry periods, although it showed lower concentrations during wet periods due to dilution. No cattle-associated markers were detected in January because there was no runoff after rain events during this time. The greatest number of water samples that tested positive for cattle markers occurred in November, when rain fell after a long dry period. The accumulation of cattle fecal matter on land can be swept into the stream with runoff water, thus contributing to the fecal contamination in the stream. *Bacteroidales* is a good alternative indicator for *E. coli*, and this study found a positive correlation between general *Bacteroidales* markers and *E. coli* concentrations. Finally, Parkerson Mill Creek is polluted with fecal matter, and given that humans seem to be a significant potential source, there is a serious risk that pathogens are also present in this water.

3.6 References

1. **ADEM.** 2011. Parkerson Mill Creek Watershed Management Plan http://www.auburn.edu/administration/facilities/organization/campus-planning-space-management/campus-planning/landscape-master-plan/documents/PMCPLAN_LR.pdf
2. **ADEM.** 2013. Water Quality Criteria. 1-335-6. Division 6. Volume 1. <http://www.adem.state.al.us/alEnviroReglaws/default.cnt>
3. **Ahmed, W., J. Stewart, T. Gardner, D. Powell, P. Brooks, D. Sullivan, and N. Tindale.** 2007. Sourcing faecal pollution: A combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments. *Water Res.* **41**:3771-3779.
4. **Bernhard, A. E., and K. G. Field.** 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**:1587-1594.
5. **Bernhard, A. E., and K. G. Field.** 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **66**:4571-4574.
6. **Boehm, A. B., L. C. Van De Werfhorst, J. F. Griffith, P. A. Holden, J. A. Jay, O. C. Shanks, D. Wang, and S. B. Weisberg.** 2013. Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res.* (in press)
7. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005a. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3179-3183.
8. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005b. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3184-3191.
9. **Dick, L. K., E. A. Stelzer, E. E. Bertke, D. L. Fong, and D. M. Stoeckel.** 2010. Relative decay of *Bacteroidales* microbial source tracking markers and cultivated *Escherichia coli* in freshwater microcosms. *Appl. Environ. Microbiol.* **76**:3255-3262.

10. **Fremaux, B., J. Gritzfeld, T. Boa, and C. K. Yost.** 2009. Evaluation of host-specific *Bacteroidales* 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Res.* **43**:4838-4849.
11. **Fremaux, B., T. Boa, and C. K. Yost.** 2010. Quantitative qPCR assays for sensitive detection of Canada Goose-specific fecal pollution in water sources. *Appl. Environ. Microbiol.* **76**:4886-4889.
12. **Gawler, A. H., J. E. Beecher, J. Brandao, N. M. Carroll, L. Falcao, M. Gourmelon, B. Masterson, B. Nunes, J. Porter, A. Rince, R. Rodrigues, M. Thorp, J. M. Walters, and W. G. Meijer.** 2007. Validation of host-specific *Bacteroidales* 16S rRNA genes as markers to determine the origin of faecal pollution in Atlantic Rim countries of the European Union. *Water Res.* **41**:3780-3784.
13. **Gourmelon, M., M. P. Caprais, C. Le Mennec, S. Mieszkin, C. Ponthoreau, and M. Gendronneau.** 2010. Application of library-independent microbial source tracking methods for identifying the sources of faecal contamination in coastal areas. *Environ. Sci. Technol.* **61**:1401-1409.
14. **Haugland, R. A., S. C. Siefring, L. J. Wymer, K. P. Brenner, and A. P. Dufour.** 2005. Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Res.* **39**:559-568.
15. **Kildare, B. J., C. M. Leutenegger, B. S. McSwain, D. G. Bambic, V. B. Rajal, and S. Wuertz.** 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: A Bayesian approach. *Water Res.* **41**:3701-3715.
16. **Klieve, A. V., D. Hennessy, D. Ouwkerk, R. J. Forster, R. I. Mackie, and G. T. Attwood.** 2003. Establishing populations of *Megasphaera elsdenii* YE 34 and *Butyrivibrio fibrisolvens* YE 44 in the rumen of cattle fed high grain diets. *J. Appl. Microbiol.* **95**:621-630.
17. **Lamendella, R., J. W. S. Domingo, D. B. Oerther, J. R. Vogel, and D. M. Stoeckel.** 2007. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of *Bacteroidetes* 16S rRNA gene. *FEMS Microbiol. Ecol.* **59**:651-660.
18. **Lamendella, R., N. VerBerkmoes, and J. K. Jansson.** 2012. 'Omics' of the mammalian gut - new insights into function. *Curr. Opin. Biotechnol.* **23**:491-500.
19. **Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Saylor.** 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based qPCR assays

for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* **72**:4214-4224.

20. **Lee, Y. J., M. Molina, J. W. S. Domingo, J. D. Willis, M. Cyterski, D. M. Endale, and O. C. Shanks.** 2008. Temporal assessment of the impact of exposure to cow feces in two watersheds by multiple host-specific PCR assays. *Appl. Environ. Microbiol.* **74**:6839-6847.
21. **Lu, J. R., J. S. Domingo, and O. C. Shanks.** 2007. Identification of chicken-specific fecal microbial sequences using a metagenomic approach. *Water Res.* **41**:3561-3574.
22. **Lu, J., J. W. Santo Domingo, S. Hill, and T. A. Edge.** 2009. Microbial diversity and host-specific sequences of Canada Goose feces. *Appl. Environ. Microbiol.* **75**:5919-5926.
23. **Mehra, S., and W. S. Hu.** 2005. A kinetic model of quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* **91**:848-860.
24. **Mieszkin, S., J. F. Yala, R. Joubrel, and M. Gourmelon.** 2010. Phylogenetic analysis of *Bacteroidales* 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by qPCR. *J. Appl. Microbiol.* **108**:974-984.
25. **Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* **74**:890-90.
26. **Peed, L. A., C. T. Nietch, C. A. Kelty, M. Meckes, T. Mooney, M. Sivaganesan, and O. C. Shanks.** 2011. Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. *Environ. Sci. Technol.* **45**:5652-5659.
27. **Salyers, A. A.** 1984. *Bacteroides* of the human lower intestine -track *Annu. Rev. Microbiol.* **38**:293-313.
28. **Santoro, A. E., and A. B. Boehm.** 2007. Frequent occurrence of the human-specific *Bacteroides* fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators. *Environ. Microbiol.* **9**:2038-2049.
29. **Sauer, E. P., J. L. VandeWalle, M. J. Bootsma, and S. L. McLellan.** 2011. Detection of the human associated *Bacteroides* genetic marker provides evidence

of widespread sewage contamination of storm water in the urban environment. *Water Res.* **45**:4081-4091.

30. **Savichtcheva, O., N. Okayama, and S. Okabe.** 2007. Relationships between *Bacteroides* 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Res.* **41**:3615-3628.
31. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. D. Siciliano.** 2005. Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with qPCR for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* **7**:249-259.
32. **Shanks, O. C., C. Nietch, M. Simonich, M. Younger, D. Reynolds, and K. G. Field.** 2006a. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Appl. Environ. Microbiol.* **72**:5537-5546.
33. **Shanks, O. C., J. W. Santo Domingo, R. Lamendella, C. A. Kelty, and J. E. Graham.** 2006b. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Appl. Environ. Microbiol.* **72**:4054-4060.
34. **Shanks, O. C., J. W. Santo Domingo, J. R. Lu, C. A. Kelty, and J. E. Graham.** 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Appl. Environ. Microbiol.* **73**:2416-2422
35. **Shanks, O. C., E. Atikovic, A. D. Blackwood, J. R. Lu, R. T. Noble, J. S. Domingo, S. Seifring, M. Sivaganesan, and R. A. Haugland.** 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **74**:745-752
36. **Shanks, O. C., C. A. Kelty, M. Sivaganesan, M. Varma, and R. A. Haugland.** 2009. Quantitative PCR for genetic markers of human fecal pollution. *Appl. Environ. Microbiol.* **75**:5507-5513.
37. **Shanks, O. C., K. White, C. A. Kelty, S. Hayes, M. Sivaganesan, M. Jenkins, M. Varma, and R. A. Haugland.** 2010. Performance assessment PCR-based assays targeting *Bacteroidales* genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **76**:1359-1366.
38. **Shanks, O. C., C. A. Kelty, S. Archibeque, M. Jenkins, R. J. Newton, S. L. McLellan, S. M. Huse, and M. L. Sogin.** 2011. Community structures of fecal

bacteria in cattle from different animal feeding operations. *Appl. Environ. Microbiol.* **77**:2992-3001.

39. **USEPA.** 1986. Ambient Water Quality Criteria EPA 440/5-84-002. http://water.epa.gov/action/advisories/drinking/upload/2009_0413_beaches_1986_crit.pdf
40. **USEPA.** 2012. Recreational Water Quality Criteria. 820-F-12-061. <http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/factsheet2012.pdf>
41. **Van De Werfhorst, L. C., B. Sercu, and P. A. Holden.** 2011. Comparison of the host specificities of two *Bacteroidales* quantitative PCR assays used for tracking human fecal Contamination. *Appl. Environ. Microbiol.* **77**:6258-6260.
42. **Walters, S. P., and K. G. Field.** 2009. Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. *Environ. Microbiol.* **11**:1410-1421
43. **Yang, X., L. Xie, Y. X. Li, and C. C. Wei.** 2009. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *Plos. One* **4**. e6074. doi:10.1371/journal.pone.0006074.

Table 3.1. Primer sets tested for specificity and sensitivity.

Assay	Primer Sequences (5'-3')	Target gene	Size	Reference
Bac32	32F: ACGCTAGCTACAGGCTT 708R: CAATCGGAGTTCTTCGTG	General <i>Bacteroidales</i> 16S rRNA	676 bp	Bernhard & Field, 2000a
AllBac	296F: GAGAGGAAGGTCCCCAC 412R: CGCTACTTGGCTGGTTCAG	General <i>Bacteroidales</i> 16S rRNA	106 bp	Layton et al., 2006
HF183	183F: ATCATGAGTTCACATGTCCG 708R: CAATCGGAGTTCTTCGTG	Human origin <i>Bacteroidales</i> 16S rRNA	525 bp	Bernhard & Field, 2000b
qHF183	183F: ATCATGAGTTCACATGTCCG 265R: TACCCCGCCTACTATCTAATG	Human origin <i>Bacteroidales</i> 16S rRNA	82 bp	Seurinck et al., 2005
Hum163	163F: CGTCAGGTTTGTTCGGTATTG 163R: AAGGTGAAGGTCTGGCTGATGTAA	Human origin Hypothetical protein BF3236	165 bp	Shanks et al., 2008
Hum 336	336F: CCAACGGCGTAACTTCTTCA 336R: ATTACCGGATTACAAACCTTATG	Human origin outer membrane efflux protein precursor	162 bp	Shanks et al., 2008
HuBac	566F: GGGTTTAAAGGGAGCGTAGG 692R: CTACACCACGAATTCCGCCT	Human origin <i>Bacteroidales</i> 16S rRNA	116 bp	Layton et al., 2006
CowM3	M3F: CCTCTAATGGAAAATGGATGGTATCT M3R: CCATACTTCGCCTGCTAATACCTT	Bovine origin <i>Bacteroidales</i> 9-O-acetylerase secretory protein gene	122 bp	Shanks et al., 2008
CF128	128F: CCAACYTTCCCGWTA 708R: CAATCGGAGTTCTTCGTG	Ruminant origin <i>Bacteroidales</i> 16S rRNA	580 bp	Bernhard & Field, 2000b
CF193	CF 193: TATGAAAGCTCCGGCC 708R: CAATCGGAGTTCTTCGTG	Ruminant origin <i>Bacteroidales</i> 16S rRNA	515 bp	Bernhard & Field, 2000b
BoBac	367F: GAAG(G/A)CTGAACCAGCCAAGTA 467R: GCTTATTCATACGGTACATACAAG	Bovine origin <i>Bacteroidales</i> 16S rRNA	100 bp	Layton et al., 2006
Bac1	Bac1F : TGCAATGTATCAGCCTCTTC Bac1R: AGGGCAAACCTCACGACAG	Bovine origin <i>Bacteroidales</i>	196 bp	Shanks et al., 2006b
Bac2	Bac2F: GCTTGTTGCGTTCCTTGAGATAAT Bac2R:ACAAGCCAGGTGATACAGAAAG	Bovine origin <i>Bacteroidales</i>	274 bp	Shanks et al., 2006b
Bac3	Bac3F:CTAATGGAAAATGGATGGTATCT Bac3R: GCCGCCAGCTCAAATAG	Bovine origin <i>Bacteroidales</i>	166 bp	Shanks et al., 2006b

Table 3.2. Specificity and sensitivity of selected primers.

Assay	Target source	No. of sewage samples	No. of cattle fecal samples	Specificity (%)	Sensitivity (%)
HF183	Human	30	19	100% (19)	100% (30)
Hum163	Human	10	5	60% (5)	100% (10)
Hum336	Human	8	8	100% (8)	0% (8)
HuBac	Human	2	2	0% (2)	100% (2)
Bac1	Bovine	10	10	100% (10)	30% (10)
Bac2	Bovine	10	10	100% (10)	20% (10)
Bac3	Bovine	6	6	100% (6)	0% (6)
CF128	Ruminant	20	41	15.0% (20)	100% (41)
CF193	Ruminant	6	10	100% (6)	0% (10)
BoBac	Bovine	2	2	50% (2)	100% (2)
CowM3	Bovine	22	22	100% (22)	100% (22)

No. of samples used for specificity and sensitivity tests are given in the parentheses.

Table 3.3. Performance information for qPCR.

Quality control parameter	AllBac	qHF183	CowM3
qPCR efficiency	106%	104%	94%
R ² for standard curves	0.995	0.996	0.997
LLOQ (gene copies/5 µl)	1000	10	10
Limit of detection (gene copies/5 µl)	168	6.5	4.8
Composite standard curve	Y=-3.19x+34.17	Y=-3.22x+33.27	Y=-3.48X+34.12

Table 3.4. Number of samples tested positive for general *Bacteroidales*, host specific molecular markers and *E. coli*.

Site	End-point PCR			qPCR*		<i>E. coli</i>	
	Bac32	HF183	CowM3	AllBac	qHF183		
November 2009							
H	10/10	2/10	1/10	10/10	10/10	1/10	10/10
B	10/10	2/10	1/10	10/10	7/10	1/10	10/10
D	10/10	2/10	1/10	10/10	9/10	3/10	10/10
Q	10/10	2/10	3/10	10/10	10/10	1/10	10/10
December 2009							
H	10/10	0/10	0/10	10/10	2/10	3/10	10/10
B	10/10	0/10	0/10	10/10	10/10	1/10	10/10
D	10/10	0/10	0/10	10/10	8/10	0/10	10/10
Q	10/10	0/10	0/10	10/10	9/10	0/10	10/10
January 2010							
H	10/10	0/10	0/10	10/10	10/10	0/10	10/10
B	10/10	0/10	0/10	10/10	10/10	0/10	10/10
D	10/10	4/10	0/10	10/10	10/10	0/10	10/10
Q	10/10	5/10	0/10	10/10	10/10	0/10	10/10

*For qPCR, gene copies above LOD were considered positive.

Table 3.5. Selected nutrient parameters and turbidity monitored during the study period.

Site	Ca (mg/l)	Mg (mg/l)	K (mg/l)	Total P (mg/l)	Organic C (mg/l)	Total N (mg/l)	C/N ratio	Turbidity (NTU)
Nov-2009								
H	23.5	12.1	2.94	0.083	22	0.885	24.9	10.8-102
B	35.1	14.8	7.36	0.041	29.3	2.52	11.6	8.50-55.3
D	31.2	15.1	5.22	0.01	29.6	1.92	15.4	4.60-84.2
Q	20.9	12.1	4.21	0.034	23.6	1.66	14.2	4.60-76.5
Dec-2009								
H	17.3	9.5	2.24	0	20.9	0.614	34.1	8.31-63.1
B	33.3	13	6.11	0.001	27.6	1.84	14.9	2.60-49.5
D	24.6	12.3	4.42	0.005	25.8	1.43	18	3.63-51.4
Q	38.6	10.4	3.43	0.007	18.7	1.31	14.3	4.59-49.8
Jan-2010								
H	23.4	13.6	1.94	0	26.4	0.585	45.1	8.35-9.15
B	27.1	14.2	3.77	0.003	25.2	2.51	10	1.50-12.3
D	27.7	14.5	3.3	0.003	25.9	1.73	15	2.86-15.8
Q	23.2	13.2	3.09	0	21.8	2.04	10.7	3.36-4.50

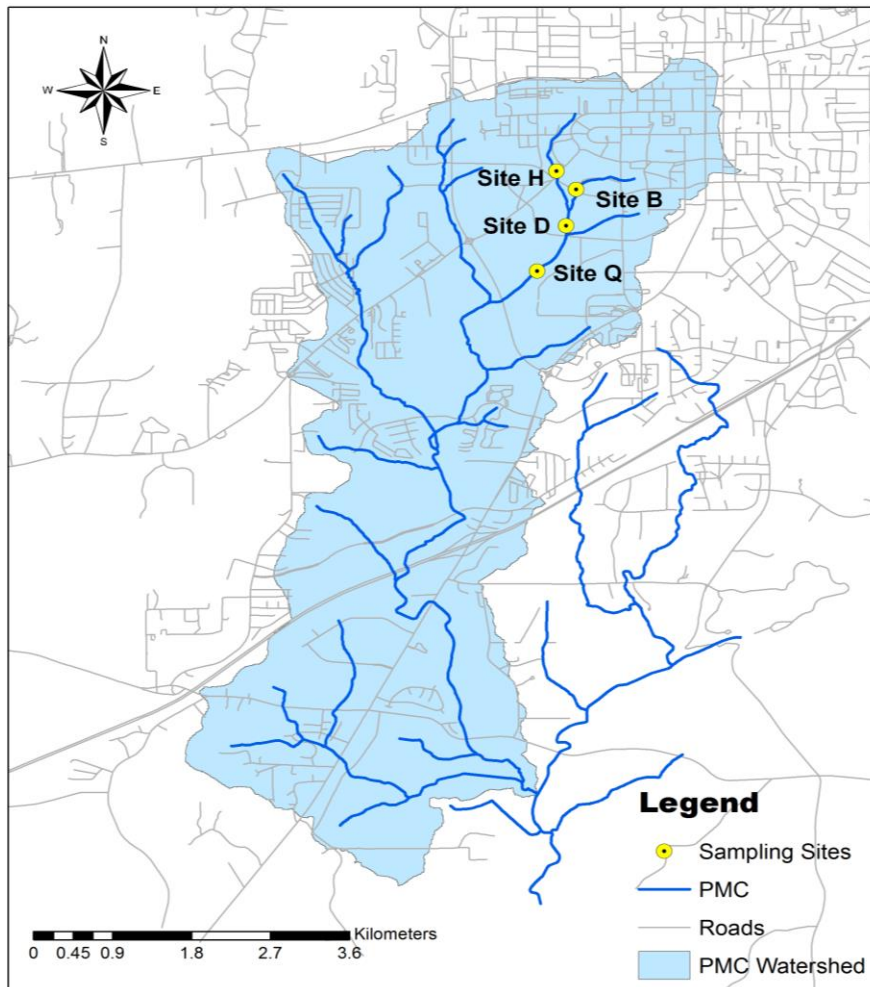


Fig. 3.1. Sampling sites at the Parkerson Mill Creek watershed.

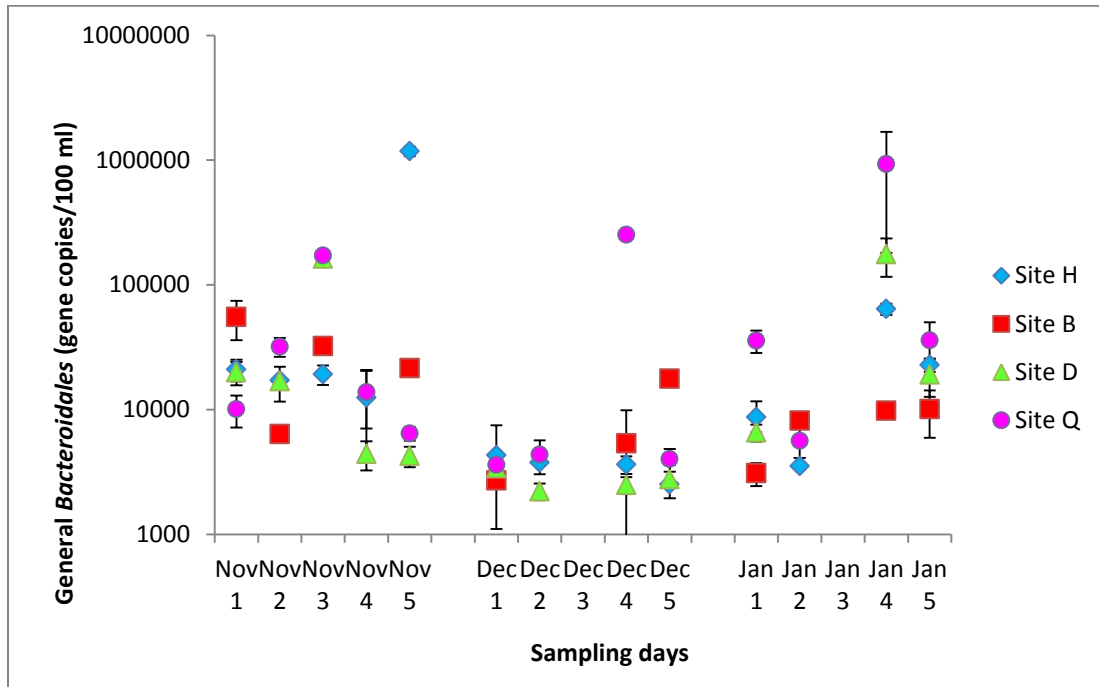


Fig. 3.2. General *Bacteroidales* marker concentrations found above the LLOQ at different sampling sites during the sampling period.

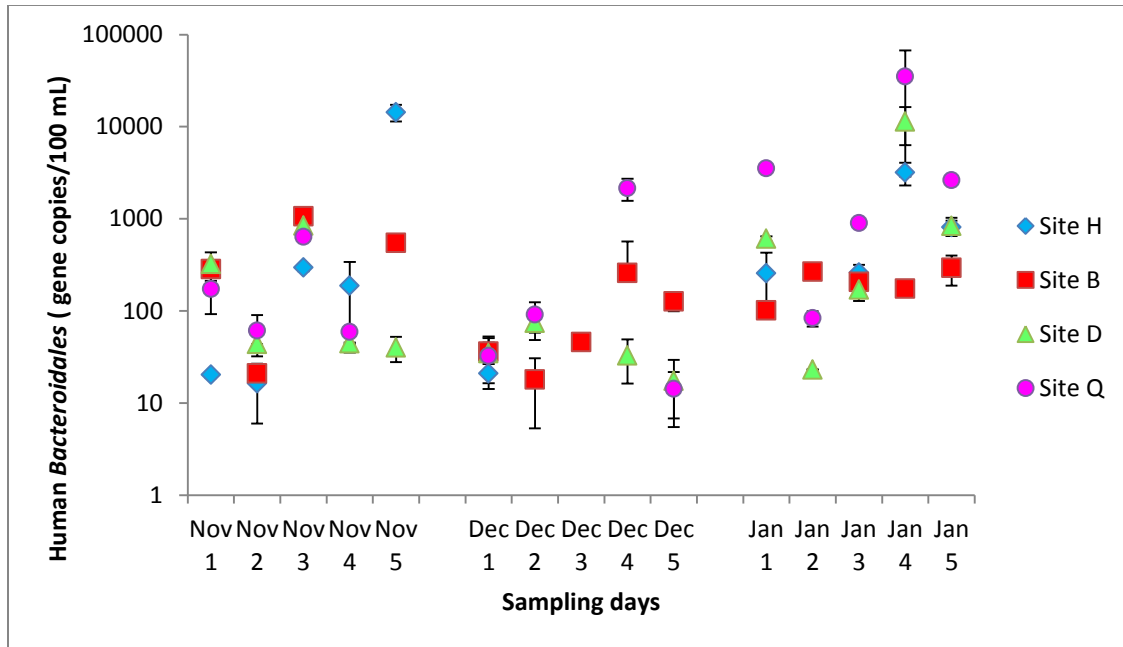


Fig. 3.3. Human-associated *Bacteroidales* marker concentrations found above the LLOQ in different sampling sites during the sampling period.

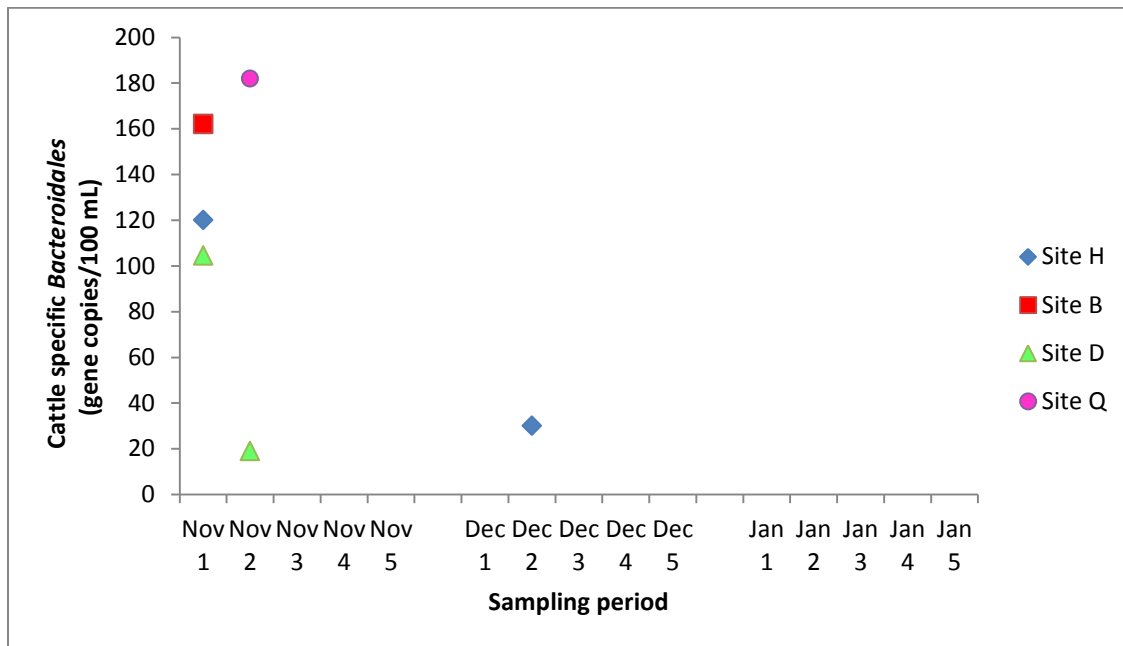


Fig. 3.4. Cattle-associated *Bacteroidales* marker concentrations found at different sampling sites during the sampling period.

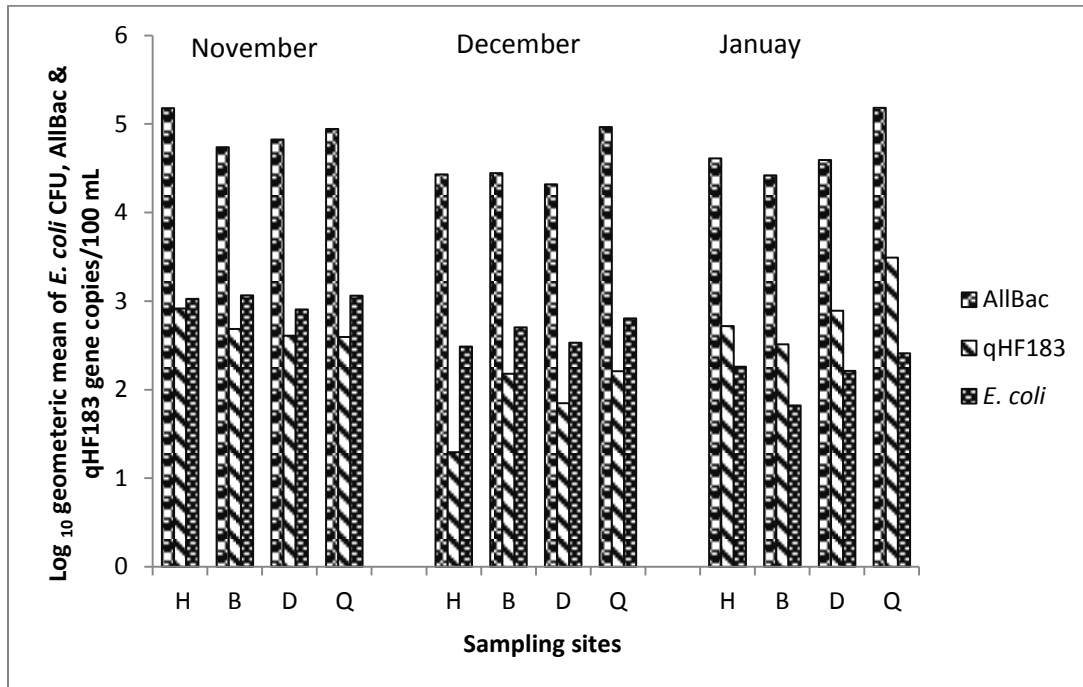


Fig. 3.5. Five-day geometric means of *E. coli* concentrations, general *Bacteroidales* and human-associated genetic markers found during the study period.

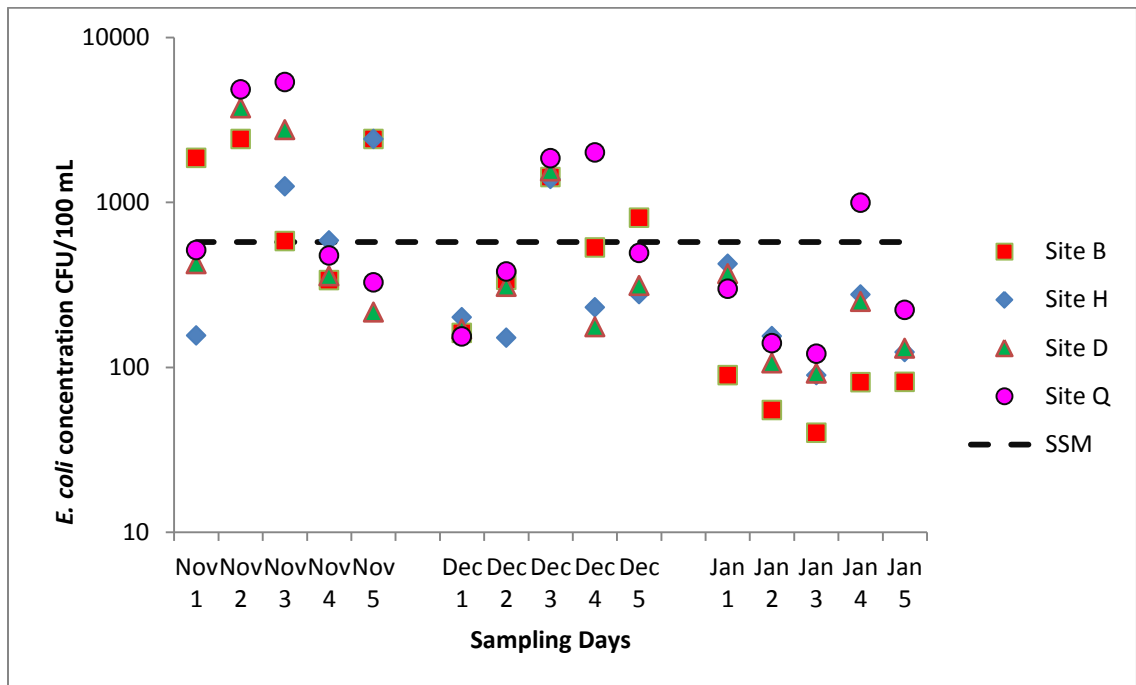


Fig. 3.6. *E. coli* concentrations found at different sampling sites during the sampling period. The dashed line shows the single sample maximum (SSM) criterion for fish and wildlife (576 CFU/100 mL).

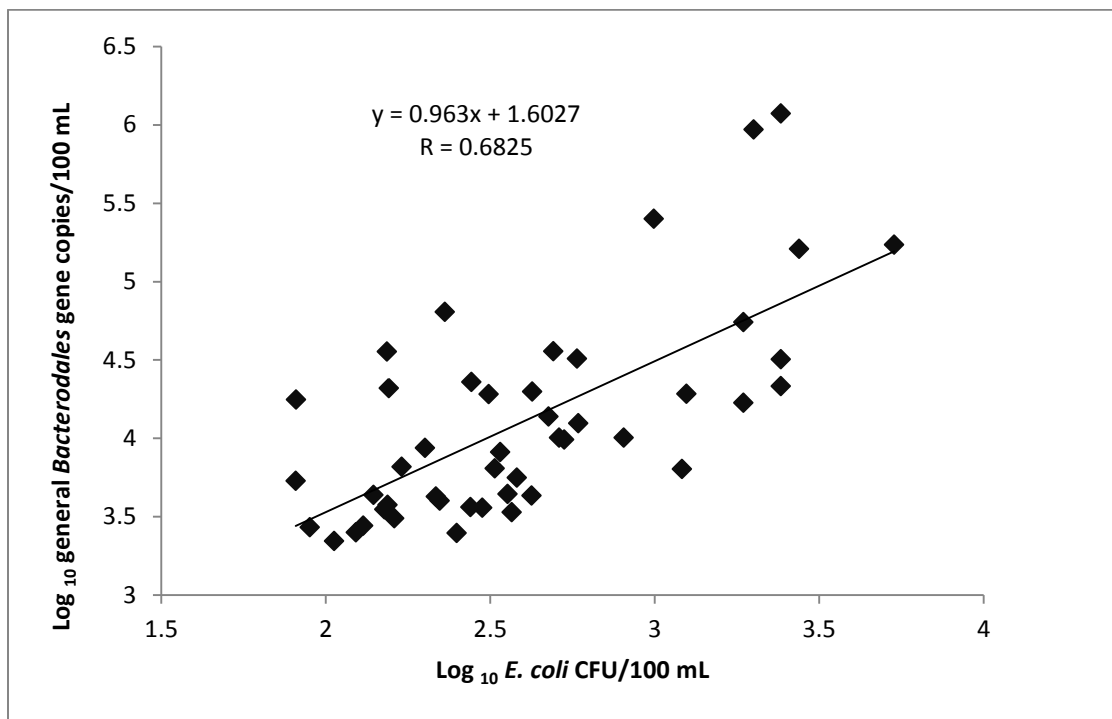


Fig. 3.7. The overall correlation between the general *Bacteroidales* marker and *E. coli*.

4. Differentiation of *Bacteroidales* 16S rRNA genetic markers from live and dead cells and their persistence in the secondary habitat

Abstract

Amplification of host-associated gene fragments of fecal *Bacteroidales* using quantitative real-time PCR (qPCR) is a popular approach for fecal source identification. Since PCR cannot discriminate between DNA from live and dead cells, PCR-based methods may overestimate fecal bacterial populations in the environment. This can be addressed by using propidium monoazide (PMA) as a DNA intercalating agent to distinguish between DNA from live and dead cells, and PMA was therefore utilized in this study to determine survival of viable *Bacteroidales* in their secondary habitat. Microcosm experiments revealed that viable *Bacteroidales* survived for 6 days in the stream water and 9 days in the sediment, while human-associated *Bacteroidales* remained in stream water and sediment microcosms for 3 and 4 days, respectively. *E. coli* survived much longer, with 2,500 CFU/100 ml remaining in stream water microcosms after 14 days and 137 CFU/100 ml in sediment microcosms after 75 days. The results of this study indicate that about 50% of total *Bacteroidales* detected by qPCR originated from dead cells or extracellular DNA. Detection of *Bacteroidales* by PMA assisted qPCR provides information about recent fecal pollution in surface waters.

4.1. Introduction

Currently, the order *Bacteroidales* is considered one of the most promising alternative indicator organisms for bacterial source tracking. This group is abundant in feces, and the phylum Bacteroidetes represents about 25%–30% of the human gut population (Salyers, 1984; Yang et al., 2009). Most members of the phylum are harmless and only few are opportunistic pathogens (Salyers, 1984). As *Bacteroidales* are obligate anaerobes, they have a shorter life span in their secondary habitat than other facultative anaerobes such as *E. coli*. This is a fundamental feature in the bacterial source tracking field because water samples that test positive for *Bacteroidales* indicate recent water pollution with fecal matter. The current fresh water fecal indicator, *E. coli*, has an extended survival period in its secondary habitat, persisting for up to 260 days under sterile conditions at 4°C in stream water (Flint, 1987), and is also capable of regrow in the secondary habitat (Ishii et al., 2006; Byappanahalli et al., 2006). Thus, the simple presence of *E. coli* does not necessarily indicate recent fecal pollution in water and the use of *Bacteroidales* as an alternative source-specific fecal indicator has therefore been suggested (Newton et al., 2011).

Recent findings suggest that fecal *Bacteroidales* bacteria may survive longer in oxygenated environments than other obligate anaerobes (Xu et al., 2003). Sequences of *Bacteroides thetaiotaomicron* revealed that its genome contains gene sequences for Complex I (NADH-quinone oxidoreductase) and Complex II (succinate dehydrogenase)

of the aerobic respiratory pathway, as well as other enzymes involved in oxygen-dependent respiration (Xu et al., 2003). Microcosm studies have also found that *Bacteroidales* 16S rRNA gene markers can persist in stream water for eight to 24 days and that their persistence depends mainly on the temperature and predators in the water (Seurinck et al., 2005; Okabe et al., 2007; Bell et al., 2009). However, these studies did not differentiate between DNA from live and dead *Bacteroidales* cells. Even after cell death, DNA can persist for two weeks in stream water, or more than 16 weeks under sterile conditions at 4°C (Josephson et al., 1993).

DNA has a tendency to adsorb to and settle with sediments. In addition, anaerobic pockets in sediments can provide niches for anaerobes. Thus, *Bacteroidales* survival in sediment may be different from their survival in water. This was supported by a survival study that found that extracellular DNA can persist for 55 days in seawater, 21 days in fresh water and 40 days in sediment (Nielsen et al., 2007). The prevalence of DNA from non-viable cells and bacteria in sediment is likely to adversely affect the results of bacterial source tracking because turbulent water currents during a storm period could resuspend these sediments (Eichmiller et al., 2013), thus elevating bacterial cell numbers and DNA concentrations in the water column and rendering efforts to provide information about recent water pollution with fecal matter problematic.

The inability of PCR to discriminate between DNA derived from live and dead cells and extra cellular DNA may result in significant overestimation of microbial populations in the environment. Various techniques have been used to differentiate between live cells and dead cells, with commonly used methods including culturable cell counts, fluorescence-based live/dead assays, and flow cytometry (Nebe-von-Caron 2000;

Kramer et al., 2009). However, most of these techniques suffer from limitations and hence cannot conveniently be used to separate DNA associated with live cells from that from dead cells or from free DNA. Detection utilizing mRNA is a precise way to determine live cells in environmental water (Walters and Field, 2009; Liang et al., 2012), but in addition to being a technically demanding and expensive approach, mRNA suffers from a short half-life and is unstable in the environment (Josephson et al., 1993).

Another possible approach is to use the dyes propidium monoazide (PMA) and ethidium monoazide (EMA) to selectively differentiate between DNA associated with live and dead cells (Nogva et al., 2003; Rudi et al., 2004; Nocker and Camper, 2006; Nocker et al., 2006). PMA and EMA are DNA intercalating agents; their azide groups intercalate with DNA molecules to produce strong covalent bonds in the presence of bright visible light, inhibiting PCR amplification. EMA/PMA has the ability to penetrate only the cell walls and cell membranes of dead cells, not those of live cells and any unbound EMA/PMA reacts with the water molecules in solution and is thus inactivated. Consequently, PMA/EMA treatment facilitates the selective amplification of live cells' DNA in the bacterial population (Kramer et al., 2009; Nogva et al., 2003; Rudi et al., 2004; Nocker and Camper 2006; Nocker et al., 2006; Vesper et al., 2007; Rawsthorne et al., 2009). However, studies using *E. coli*, *Campylobacter jejuni*, and *Listeria monocytogenes* revealed that EMA can also cause significant loss of live cells because intact cells of some bacterial species can take up more EMA than other organisms, resulting in cell damage (Nocker et al., 2006; Flekna et al., 2007). This is less of a problem with the more recently introduced PMA, which has a better selectivity and can only enter dead cells (Nielsen et al., 2007; Nocker et al., 2006). The use of EMA/PMA to

detect live pathogens in clinical samples is common for qPCR (Kramer et al., 2009; Kobayashi et al., 2009), but as yet this technique has not been widely applied for environmental samples. This may be due to insufficient light penetration through particulate matter and the high levels of suspended solids present in surface water, which may interfere with photo induced crosslinking of PMA/EMA to DNA (Varma et al., 2009; Wagner et al., 2008).

This technique has been successfully applied to quantify the live and dead *Bacteroidales* present in the effluent and influent of a sewage treatment plant and to identify the prevalence of *Bacteroidales* and pathogens in water microcosms (Bae and Wuertz, 2009, 2012). Additionally, PMA was successfully used to differentiate live *Enterococcus* and *Bacteroidales* from dead ones in wastewater (Varma et al., 2009). None of these studies have used general *Bacteroidales* 296F/412R and human associated 183F/265R primers, however, which are widely used to detect fecal contamination (Seurinck et al., 2005; Okabe et al., 2007; Bernhard and Field 2000a,b) and there appear to be no published reports of the use of PMA to discriminate the association of these two markers with DNA from live and dead *Bacteroidales* in microcosms containing stream water and sediments.

The objectives of this study were therefore to verify the ability of PMA to separate live and dead *Bacteroidales* in stream water and sediments, and determine the survival of viable *Bacteroidales* and *E. coli* in these environments.

4.2. Methods and Materials

Sewage sample collection and preparation. Sewage influent and effluent samples were collected in 1-L sterile high density polyethylene (HDPE) bottles from a sewage treatment plant in Auburn, AL, kept on ice and transported to the laboratory. Sewage samples were concentrated by centrifugation on the day of sample collection. Briefly, 10 centrifuge tubes each containing 45 ml samples were centrifuged at 2,750 g for 15 minutes at 4°C. The supernatant was decanted and pellets were combined and resuspended in 45 mL of phosphate buffered saline (PBS) at pH 7.4.

Validation of PMA treatment conditions for microcosms prepared with stream water. Optimization of the PMA concentration, the incubation time in the dark, and the time required to crosslink DNA and PMA under visible light were conducted using defined mixtures of viable and killed cells prepared by mixing fresh and boiled sewage samples. Briefly, half of the concentrated sewage was boiled for 15 minutes. Then boiled and fresh sewage were mixed in five 500-mL volumetric flasks, with the proportions adjusted to ensure that the flasks contained 0, 25, 50, 75 and 100% of diluted (1:10) fresh sewage, respectively. The final volumes were brought to 500 ml with autoclaved stream water. Each volumetric flask contained 10% sewage. Suspensions were mixed well, transferred to 1-L Erlenmeyer flasks, and mixed again. After incubation at room temperature (20°C) for 1 hour, three 30-ml samples were removed from each microcosm and transferred to three 50-mL centrifuge tubes and centrifuged at 2,750 g for 15 minutes. Supernatants were decanted and the pellets were resuspended in 1 mL of PBS. Each sample was divided into two equal portions: one half (0.5 mL) was stored at 4°C for

DNA extraction and the other half was used for PMA treatment as described by Bae and Wuertz (2009).

In the dark, 2.5 µl of 20 mM PMA (Biotium Inc., Hayward, CA) was added to the sewage suspension (0.5 mL) in a 1.8-ml clear micro centrifuge tube; the final concentration of PMA was 100 µM. Samples were mixed well and incubated in the dark for 5 minutes at room temperature. Samples were then exposed to light from a 650-W halogen bulb for 10 minutes. To prevent excessive heating of cells, the tubes were horizontally laid on ice 20 cm away from the light source. After the heat treatment, samples were centrifuged at 10,000 g for 5 minutes, the supernatant was discarded and pellets were washed twice with PBS. Finally, the pellets were resuspended in 100 µl of PBS and stored at 4°C until DNA extraction.

***Bacteroidales* survival in microcosms prepared with stream water.** Stream water samples were collected from Site Q of Parkerson Mill Creek in Auburn, Alabama. Microcosms were prepared in triplicates using 2-L Erlenmeyer flasks, each containing 990 mL of autoclaved fresh stream water, and 10 mL of concentrated (10X) sewage as inoculum. The final sewage concentration of the microcosms was 10%. All three microcosms were covered with aluminum foil and shaken for 1 hour at 72 rpm. The first two samples were collected one and eight hours after inoculation, respectively, and the remaining samples were collected at 24-hour intervals. At each sampling time, flasks were well mixed and 30 mL of the contents removed from each microcosm and concentrated to 1 mL as described above. One half (0.5 mL) of the concentrated sample was stored at 4°C for direct DNA extraction, and the other half was used for PMA treatment. The *E. coli* concentration in each microcosm was also enumerated using the

modified membrane thermotolerant *Escherichia coli* agar (m-TEC) medium (Difco, Detroit, MI) following USEPA Method 1603 (USEPA, 2002).

Optimization of PMA treatment conditions for microcosms prepared with sediment.

Preliminary experiments were conducted to determine the appropriate amount of sediments and PMA concentrations to be used in the study. Fresh sediment was collected from the bottom of Parkerson Mill Creek at Site Q. Three different concentrations of sediment (1%, 5% and 10%) were used to prepare microcosms in the initial experiment. Microcosms of different sediment concentrations along with different PMA concentrations (100 μ M, 200 μ M and 300 μ M) and exposure times (15, 20 and 30 minutes) were used to determine the optimal conditions required to determine *Bacteroidales* survival in the sediment. Microcosms with 5% and 10% sediment did not allow adequate light penetration to activate the crosslinking between DNA and PMA, so in the subsequent experiments microcosms were prepared with 1% sediment and 10% sewage. Sediments and fresh stream water were autoclaved for one hour and allowed to cool overnight. To determine the appropriate PMA concentration and light exposure time, one 500 mL sediment microcosm was prepared with autoclaved stream water and sewage containing 50% boiled sewage. The Erlenmeyer flask was well mixed and incubated for one hour at room temperature (20°C) after which duplicate samples (10 mL) were removed from the flask and centrifuged at 2,750 g for 15 minutes at 4°C. The supernatant was decanted and pellets were resuspended in 2 mL PBS. One milliliter of the sample was stored at 4°C for DNA extraction and 1 mL was used for PMA treatment.

Different PMA concentrations (100, 200 and 300 μ M) and light exposure times (10, 15, 20, and 30 min) were tested. DNA was extracted from both PMA treated and un-

treated samples, and the general *Bacteroidales* marker was determined by qPCR. These preliminary experiments indicated that 1% sediment and 100 μ M PMA with 10 minutes exposure time were the optimal conditions for PMA treatment. To confirm these initial findings, triplicate microcosms containing 1% sediment were set up and inoculated with 0, 25, 50, 75 and 100% of fresh sewage. The PMA treatment was carried out at 100 μ M and a light exposure time of 10 minutes.

***Bacteroidales* survival in the sediment microcosm.** Microcosms were prepared with autoclaved 1% sediment and stream water in triplicate with a final concentration of fresh sewage of 10%. The microcosms were mixed well to ensure a consistent distribution of the sewage. All three microcosms were covered with aluminum foil and shaken at 72 rpm for 1 hour at room temperature. Samples were removed as described in the water microcosm experiment. A 10 mL sample was removed from each flask, transferred to a 50-mL centrifuge tube and centrifuged at 2,750 g for 15 minutes at 4°C. The supernatant was decanted and the pellets resuspended in 2 mL of PBS. Each sample was divided into two portions, with one half (1 mL) of the sample stored for DNA extraction and the other half used for PMA treatment. PMA treatment was performed using the methods described in the previous section. The *E. coli* concentration in each microcosm was enumerated using the modified m-TEC medium (USEPA, 2002).

DNA extraction. DNA was extracted from 300 μ L of sample using the PowerSoil DNA Isolation Kit (MOBIO, Carlsbad, CA) following the manufacturer's instructions except for the following: 1) the centrifuge speed was increased from 10,000 g to 12,000 g in all steps; 2) the spin filter was air dried at room temperature for 10 minutes after ethanol elution, which allowed any remaining ethanol in the spin filter to evaporate; 3) the eluting

buffer (C6) was warmed to 37°C, added to the spin filter, and incubated for 2 minutes at room temperature. The DNA concentration was measured using a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Preparation of qPCR standards. DNA extracted from sewage samples was amplified with general *Bacteroidales* primers 32F/708R (Bernhard and Field, 2000a) and human associated primers HF183F/708R (Bernhard and Field, 2000b) and resolved on 1.5% agarose gel to confirm the presence of a single band. The product sizes for general and human associated *Bacteroidales* markers were 686 bp and 525 bp, respectively. The remaining PCR amplified products were cleaned with the DNA Clean & Concentrator Kit (Zymo Research, Orange, CA) and cloned into the TOPO 2.1 cloning vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Plasmid DNA was extracted from *E. coli* colonies and cleaned with the QIAprep Spin Miniprep Kit (Qiagen Sciences, Germantown, MD). The presence of plasmid inserts was verified by PCR amplification with 32F/708R and 183F/708R primers prior to sequencing with an ABI 3100 DNA Genetic Analyzer with M13 primers. Each sequence was checked to confirm the presence of primer sites for general *Bacteroides* primers 32F/708R and AllBac primers 296R/412F as well as the human associated primers 183F/708R and 183F/265R. Plasmid DNA concentration was measured with a NanoDrop ND-1000 spectrophotometer and the gene copy numbers were determined. All the primers used in this study are listed in Table 4.1.

qPCR assay. The 15 µL qPCR reaction mixtures contained 5 µL of template DNA (4.68 to 11.32 ng/µL), 1.4 µL of PCR grade water, 0.5 µL of 2% BSA (Sigma, St Louis, MO), 7.5 µL of 2X SYBR Green master mixture (Applied Biosystems, Carlsbad, CA), and 0.3 µL each of 10 µM forward and reverse primers. The general *Bacteroidales* 16S rRNA

genetic marker was amplified with AllBac 296F and 412R primers, and the human-associated *Bacteroidales* genetic marker was amplified with 183F and 265R primers. The amplification was performed using an Applied Biosystems StepOne Real-Time PCR instrument using the following thermocycling conditions: initial denaturation at 95°C for 10 min, then 40 cycles of denaturation at 95°C for 15 S, primer annealing at 60°C for 30 S and extension at 72°C for 30 S. Subsequently, the melting curve analysis was performed with a temperature gradient of 0.5°C per minute from 95°C to 60°C.

Amplification inhibition in qPCR was evaluated using a dilution series consisting of undiluted, 10x and 100x dilutes samples. Based on the results of the inhibition test, DNA extracted from the water samples was diluted 10-fold and DNA extracted from the sediment samples was used directly, without dilution. All the samples were run with at least five 10-fold dilution standards in duplicate; no template control and all the samples were run in duplicate. The limit of detection (LOD) was calculated as described by Dick et al. (2010), where the LOD was defined as the fifth percentile among the observed threshold cycles (C_q) across all blanks and negative-control reactions.

Chemical and physical parameters. The laboratory room temperature was recorded every day and found to range between 20 and 21°C. Chemical parameters such as pH (Accumet pH meter 25, Fisher Scientific), electrical conductivity, selected nutrients: total C, dissolved C, total N, and dissolved N (TOC-V Combustion Analyzer, Shimadzu, Columbia, MD), and turbidity (HACH 2100 Turbidity meter) of microcosms were measured on the first day and the last day of sampling. Sediment particle size was analyzed using the hydrometer method (American Society for Testing and Materials, 1985).

4.3. Results

qPCR quality control. The quantification range of the standard curve for the AllBac marker was from 10^2 to 10^7 gene copies and that of the human-associated marker was 10^1 to 10^6 gene copies per 5 μ l of DNA extract. The LOD for AllBac was 37 copies and qHF183 was 3.2 copies per 5 μ l of DNA extract. Some of the important quality control parameters of qPCR are listed in Table 4.2.

To assess the amplification inhibition, DNA samples were prepared without dilution and with 10-fold and 100-fold dilutions, and then amplified with AllBac primers. Based on the results of the Student's t Test, undiluted samples from three microcosms prepared with stream water were significantly different from those samples of the 10-fold dilution ($P=0.009$), but there were no differences between samples of the 10-fold and 100-fold dilutions ($P=0.667$). Thus, DNA samples from stream water microcosms were diluted 10 times to overcome inhibition. Microcosms prepared with 1% sediment did not show significant differences between undiluted samples and those diluted 10 and 100 times ($p > 0.05$). This may be due to the adsorption of the inhibitors onto clay particles in the sediments.

Evaluation of the removal of DNA from dead cells during the DNA extraction

process. If DNA from dead cells and extracellular DNA could be removed during the DNA extraction process, there would be no need to discriminate between the dead cells' DNA and extracellular DNA from DNA associated with live cells. Therefore, the first step in this evaluation was to determine if, and to what extent, the DNA extraction process removed dead cells and extracellular DNA. Cells in fresh sewage were killed by heat treatment, and mixtures containing 100, 50 and 0% heat-treated sewage were

prepared. DNA was extracted and amplified with AllBac *Bacteroidales* primers. The results revealed that the DNA extraction process did not adequately remove the DNA from killed cells: the AllBac marker concentration was 195,120 copies/5 μ L of DNA extract in the mixture with 100% fresh sewage, 130,039 copies/5 μ L of DNA (67% of the total) in the mixture containing 50% heat-treated sewage, and 121,610 copies/5 μ L of DNA (62% of the total) in the mixture with 100% heat-treated sewage.

Validation of PMA treatment conditions for the stream water microcosms. Samples were taken in triplicate from microcosms with defined percentages of fresh sewage (0, 25, 50, 75 and 100%). There is a linear relationship between the percentage of fresh sewage and viable *Bacteroidales* detected by the AllBac PMA-qPCR assay ($R^2=0.93$) (Fig. 4.1a). The linear relationship also existed between the proportion of fresh sewage and viable human-associated *Bacteroidales* ($R^2=0.92$) (Fig. 4.1b).

***Bacteroidales* survival in microcosms prepared with stream water.** Fig. 4.2 shows the persistence of general *Bacteroidales* 16S rRNA genetic markers in stream water microcosms with and without PMA treatment. The initial concentration of the AllBac marker without PMA treatment was 4.3×10^7 gene copies per 100 ml of water from the microcosm; after the PMA treatment, the gene copy numbers were reduced to 1.8×10^7 gene copies per 100 ml of microcosm water, or 43% of the untreated level (Table S 4.1). Overall, there was a difference of about 50% gene copies between the PMA-treated and untreated sewage. AllBac marker concentrations originating from live cells varied from 37% of those without PMA treatment at 8 hours to 88% at Day 5 (Table S4.1), but their decay rates followed the same pattern. A considerable reduction in the number of gene copies was observed between Day 0 and Day 1 (Table 4.3). After Day 1, only 22% and

26% of the PMA-treated and untreated AllBac markers remained, respectively, and by Day 2, these percentages had dropped to 1.2% and 1.3%, respectively. In both PMA treated and untreated samples, AllBac markers were detected up until Day 6. Fig. 4.3 illustrates the persistence of human-associated markers with and without PMA treatments. The initial concentration of human associated *Bacteroidales* markers without PMA treatment was 7.9×10^5 copies/100 ml; that with PMA treatment was 3.6×10^5 copies/100 ml. Human-associated *Bacteroidales* markers have a shorter prevalence period than general *Bacteroidales* markers, their signals had fallen below the limit of detection by Day 3. Initially, 46% of the qHF183 markers originated from live cells and this percentage dropped to 18% by Day 3.

Optimization of PMA treatment conditions for sediment microcosms. The next set of experiments consisted of 50% heat-treated sewage in 10, 5, and 1% sediments at different PMA concentrations (100 μ M, 200 μ M and 300 μ M) and light exposure times (15, 20 and 30 minutes). The ratio between untreated and PMA-treated samples from the mixture with 50% heat-treated sewage was expected to be 2:1; approximately 50% of AllBac markers should be amplified in qPCR. However, this was not the case for the samples from microcosms with 10% and 5% sediment. This could have been either due to PMA absorption onto clay sediment particles or to poor light penetration, which is essential for cross linkage between PMA and DNA (data not shown). For microcosms with 1% sediment, 100 μ M PMA treatments with different light exposure times (10, 15 and 20 minutes) were used to determine the optimum exposure time. Exposure times of 15 and 20 minutes resulted in the amplification of less than 50% of the live cells. Excessive exposure time may facilitate PMA entering the intact cells and crosslinking with

intracellular DNA, which may result in the underestimation of gene copies from live cells. A total of 52% of live cells were amplified when sewage was treated with 100 μM PMA and a 10-minute exposure. This experiment was repeated using triplicates and exposure times of 10- and 15-minutes. Fifteen-minute exposure times again underestimated the live cells, while a 10-minute exposure amplified 48% of the live cells (Fig. 4.4). These parameters were further evaluated using defined percentages of fresh and heat-treated sewage. As shown in Fig. 4.5, a linear relationship ($R^2=0.98$) was observed between the AllBac marker concentration and the percentage of fresh sewage in the mixture.

***Bacteroidales* survival in the microcosm prepared with 1% sediment.** The results of the sediment microcosm study indicate that the general *Bacteroidales* and human associated *Bacteroidales* markers persist longer in sediments, both with and without PMA treatments, than in stream water. The initial concentration of general *Bacteroidales* markers in the sediment microcosm samples without PMA treatment was 4.0×10^7 gene copies/100 ml, while that in the samples that did receive the PMA treatment was 2.0×10^7 gene copies/100 ml (Fig. 4.6). Samples treated with PMA showed an approximately 50% reduction in the gene copies versus those without PMA treatment (Table S4.2). By Day 1, 52% and 43% of the molecular markers in the samples without and with PMA remained, respectively (Table 4.4). By Day 4, these percentages had decreased to 1.2% and 1.9%, respectively. General *Bacteroidales* molecular markers were detected until Day 9 (Fig. 4.6). Initial *Bacteroidales* 16S rRNA human marker concentrations in the sediment microcosms with and without PMA treatment were 1.4×10^5 and 3.0×10^5

copies/100 ml, respectively. These markers had dropped below the LOD by Day 3 (Fig. 4.7).

Survival of *E. coli* in stream water and sediments. *E. coli* persisted much longer than *Bacteroidales* in both sediment and stream water microcosms. Fig. 4.8 shows *E. coli* survival in microcosms prepared with stream water and sediment. The initial concentration of *E. coli* in the stream water microcosm was 2.7×10^6 CFU/100 ml and in the sediment microcosm was 8.6×10^5 CFU/100 ml. The *E. coli* concentration in the original sewage sample used to prepare the sediment microcosm was 4.7×10^6 CFU/100 ml. This indicates that some portion of the *E. coli* was initially adsorbed into the clay fraction of the sediment. The *E. coli* in the stream water decayed more rapidly than in the sediment microcosm. The stream water microcosm study lasted 14 days, and by Day 14 the *E. coli* concentration had dropped to 2,500 CFU/100 ml, 0.09% of the initial concentration. During this 14-day period a 3-log reduction in *E. coli* concentration was observed (Table 4.3). The sediment microcosm study continued for 75 days; by Day 75 the average *E. coli* concentration had fallen to 137 CFU/100 ml, only 0.02% of the initial concentration, and a 3-log reduction was again observed during this period (Table 4.4).

The results of this study revealed a good correlation between *E. coli* and AllBac marker concentrations in both the PMA and no PMA treatments in stream water and sediment microcosms. Correlation coefficients (*r*) of 0.98 were observed for the plots of *E. coli* and general *Bacteroidales* 16S rRNA markers for both the stream water and sediment microcosms that were not treated with PMA (Fig. 4.9). PMA treated samples exhibited a similar correlation between *Bacteroidales* and *E. coli*, with *r* again being 0.98.

Application of PMA to separate live and dead cells in wastewater treatment plant

influent and effluent. Influent and effluent sewage samples were collected from a sewage treatment plant in Auburn, Alabama. The effects of PMA were studied by qPCR amplification using both general *Bacteroidales* and human-associated primers. Influent samples without and with PMA treatment contained 8.5×10^8 and 2.9×10^8 AllBac gene copies/100 ml, respectively; only 35% of the molecular markers originated from live cells. The human-associated *Bacteroidales* 16S rRNA marker found in the influent was 2.0×10^7 gene copies/100 ml without PMA; the sampled that had been treated with PMA contained 4.7×10^6 gene copies per 100 ml. Only 23% of the human-associated markers were derived from live cells. In the effluent, general *Bacteroidales* markers were almost at the LOD and human associated markers had fallen below it. *E. coli* concentrations in the influent and effluent were 3.0×10^6 and 2.3 CFU/100 ml, respectively (Fig. 4.10).

Change of nutrient parameters over time in stream water and sediment microcosms.

Several of the chemical parameters tested during the study period are shown in Table 4.5. Both the stream water and sediment microcosms showed a slight increase in pH on the final day of sampling compared with the beginning of the sampling period (0.5 and 0.1 pH increments in the water and sediment microcosm, respectively). pH in the sediment microcosm exhibited less fluctuation than in the water microcosm, which may be attributable to the buffering capacity of sediments. Both microcosms showed an increase in all nutrient levels, clear evidence of cell lyses. Only PO₄-P concentrations in the sediment microcosm were unchanged between the first and last samples. Dissolved organic carbon concentration decreased in the water microcosm, but no difference was observed in the sediment microcosm between the first and last samples.

4.4. Discussion

Effect of PMA treatment. AllBac markers in the microcosms with mixtures containing 100% and 50% heat-treated sewage were 62% and 67% of those containing fresh sewage, respectively. These results confirmed that the DNA extraction process did not adequately remove dead cells' DNA and that PMA treatment should therefore be used to discriminate between DNA from live and dead cells. This study found a good positive correlation between the percentage of live cells and respective gene copy numbers after PMA treatment, with correlation coefficients of 0.96 and 0.95 for the general *Bacteroidales* and human assays, respectively. This strong correlation reveals that not only does PMA have only a minimal effect on live cells, it also has the ability to discriminate effectively between DNA from live and dead cells. These results are consistent with those reported by Bae and Wurtz's (2009). The main disadvantage of using PMA for environmental samples is that sediments and suspended particulates in water samples can block the light penetration that is essential to activate the PMA. In addition, negatively charged clay particles can adsorb the positively charged PMA molecules. These are the two main constraints that need to be addressed before PMA treatment can be routinely applied to stream water samples to differentiate between live and dead cell DNA.

The results for the samples with different percentages of sediments (10, 5 and 1%) revealed that if the sediment percentage in a water sample exceeds 5%, this can reduce the performance of PMA. However, 1% sediment with 5-minute incubation in the dark and 10-minute light exposure did not affect the performance of the 100 μ M PMA treatment. The turbidity of the 1% sediment samples varied from 210 NTU to 235 NTU.

Another measurement conducted in Parkerson Mill Creek in Auburn, Alabama, found the maximum turbidity to be 102 NTU (Chapter 3). For this study, 60 water samples were collected in duplicate from each of four locations (Site H, Site B, Site D, Site Q) in Parkerson Mill Creek over a period of 3 months (November 2009 to January 2010), which included water samples collected during a dry period as well as just after a heavy storm (rainfall of 60.7 mm or 2.39 inches over 24 hr). The maximum measured turbidity was much less than the turbidity found in sediment samples. Most streams and creeks have low turbidity levels, generally less than 20 NTU (USEPA 1999). Large rivers may fall into the high turbidity category on occasion, but for most of the time their turbidity level is less than 100 NTU (USEPA, 1999). Therefore, PMA can effectively be used to differentiate between live and dead cell DNA in the majority of the United States' surface water bodies.

Persistence of *Bacteroidales* genetic markers in stream water microcosms. General *Bacteroidales* 16S rRNA gene markers were detected for 6 days by qPCR for samples after PMA treatment as well as without PMA treatment. The persistence of *Bacteroidales* human-associated markers was less than that of general *Bacteroidales* markers. Human-associated markers dropped below the LOD by Day 2 for samples that had received the PMA treatment; for those that had not, this took 3 days. Compared with most previous studies, *Bacteroidales* survival in our study was low. Microcosm temperature and the study locations may partially explain the short survival period of *Bacteroidales* observed here. Most of the previous survival studies conducted in the United States were for microcosm temperatures of less than 15°C. Walters and Field (2009) monitored the persistence of qHF183 and qHF134 markers at 13°C and were able to detect these

markers for 10 days with qPCR; a microcosm study conducted in Florida at an average temperature of 15°C found that qHF183 markers persisted for 14 days (Liang et al., 2012); and another study detected an even longer persistence of qHF183 of 24 days at 12°C (Seurinck et al., 2005). Our study was conducted at 20-21°C, considerably above those in the previous studies. The temperature is known to have a significant effect on the survival and persistence of molecular markers. For example, Kreader (1998) reported a strong effect of temperature on the survival of *B. distasonis*; the molecular markers were detected for at least 2 weeks at 4°C, 4 to 5 days at 14°C, 1 to 2 days at 24°C and only 1 day at 30°C. Several other studies have also suggested that physiochemical parameters such as high temperatures and sunlight enhance the decay of *Bacteroidales* molecular markers (Bae and Wuertz, 2009; Dick et al., 2010; Green et al., 2011; Schulz et al., 2011). Although our microcosms were not exposed to sunlight, they were kept at 20 to 21°C, 6 to 10°C higher than the earlier studies (Seurinck et al., 2005; Walters and Field 2009; Liang et al., 2012; Kreader 1998) so it is not unreasonable to suspect that the target genes may have deteriorated at a faster rate. Another survival study conducted by Bae and Wuertz (2012) using general *Bacteroidales* markers (BacUni-UCD) showed that this marker could be detected for up to 101 hours (4.2 days) at a temperature of 22.4°C. In our study AllBac markers persisted for 6 days in the stream water microcosm. Since both studies were conducted at temperatures above 20°C, this could explain the similar survival period of general *Bacteroidales*. However, human-associated BacHum-UCD persisted for far longer (95 hr) than qHF183 markers (2 days).

Only one of these studies was conducted in the southern region of the United States (Liang et al., 2012). In this Florida study, human fecal samples from only four

adults were used and hence this is unlikely to adequately represent the real diversity of *Bacteroidales* population. The order *Bacteroidales* has significant spatial variability, so the strains that exist in the southern region of the country may differ considerably from the strains found in other regions, and hence may employ different survival strategies.

Survival of *E. coli* in stream water and sediment microcosms. *E. coli* survived for a much longer period than either the general or human associated molecular markers. *E. coli* survival in the stream water microcosm was monitored until Day 14, by which point average *E. coli* concentrations had fallen to 2500 CFU/100 ml, a 3-log reduction from the initial concentration. In contrast, *E. coli* survival in the sediment microcosm was monitored for 75 days, after which time it had dropped to 137 CFU/100 mL. Sediment provides a particularly good habitat for *E. coli* survival as it offers an anaerobic environment, nutrients, suitable pH and protection from UV light and predators, thus permitting *E. coli* and most other anaerobic microorganisms to survive in the sediment longer than in the water column (Crabill et al., 1999; Davies et al., 1995). This prolonged survival in sediments can lead to problems during storm events, when turbulent water currents often resuspend sediments, thus adding settled microbes into the stream water, and leading to increased concentrations in storm water. These results showed that *E. coli* added to the stream water more than 75 days earlier could still result in violations of the USEPA recreational water quality criteria due to the resuspension of these settled organisms. The results of this study clearly demonstrate the limitations of using *E. coli* as an indicator organism with which to identify recent fecal contamination. However, no increase in cell numbers was observed during the period of our study so these results do

not support *E. coli* regrowth in the secondary habitat as reported by Ishii et al. (2006), Byappanahalli et al. (2006) and Brennan et al. (2010).

An analysis of the sediment showed its textural class to be a sandy loam, containing 70% sand and only 15% clay. The relatively low clay content may also affect the persistence and survival of both *Bacteroidales* genetic markers and *E. coli*. The presence of a high percentage of clay provides more anaerobic microsites where both *Bacteroidales* and *E. coli* can survive. Smaller particles also have a larger surface to volume ratio and so can adsorb more nutrients such as Ca^{2+} , Mg^{2+} , and NH_4^+ , all of which are essential for their growth and survival. Some studies have reported that large particle size and low organic matter content of the sediment have a negative effect on *E. coli* survival and persistence in the secondary habitat (Craig et al., 2003; Lee et al., 2006; Pote et al., 2009). Since our sediment microcosms contained a high percentage of sand, this may have had a negative effect on the survival of both *E. coli* and *Bacteroidales*. However, the nutrient analysis revealed that nutrient availability may not be the primary reason for the death of these organisms. Maximum bacterial growth occurs for a C/N ratio in the range 7:1 to 5:1. Here, the initial C/N ratio of the water microcosm was 9:1, dropping to 6:1 by the end of the experiment (Day 14). In the sediment microcosm experiment, the ratio of 6:1 did not change significantly throughout the experiment and the concentrations of other cations such as Ca, Mg, and K had increased by the end of the experiment. Hence, nutrients do not seem to be a limiting factor in this case.

Cell lyses may also add nutrients as well as some toxic compounds to the environment, and given that the experiments conducted for this study utilized a batch system, these accumulated toxins could have adversely affected their survival and

regrowth. This condition does not apply to natural environments such as stream water or a stream bed, where stream water continually circulates and dilutes the accumulating toxins. In addition, *Bacteroidales* and *E. coli* prefer anaerobic conditions, but these experiments were not conducted under strict anaerobic conditions. Although the microcosms were prepared in closed containers, the containers were opened at intervals to remove the samples, thus introducing air into the microcosm and preventing the development of anaerobic conditions in the microcosm. Therefore, it is reasonable to expect these microbes to survive longer under natural conditions than in the laboratory. On the other hand, sterilized stream water and sediment were used for this experiment, so it is safe to assume that the microcosms were free of predators and other competitors. Natural sewage samples are likely to contain both, so the use of sterile water and sediment may not have adversely affected this experiment.

4.5. Summary

The findings of this study revealed that PMA can successfully be used to discriminate between DNA from live and dead cells. When conducting research on environmental samples, however, the application of PMA to each sample may be not economically viable. Random testing of at least 10% of samples can provide preliminary data regarding the percentages of DNA present that are derived from dead and live cells. Survival studies clearly indicated that *E. coli* can survive in the secondary habitat for a longer period than *Bacteroidales*. Therefore, the use of *E coli* as an indicator organism does not provide sufficient information about recent water pollution. Human markers persisted less than 3 days in both the stream water and sediment microcosms. This suggests that the detection of qHF183 markers in a water body accurately reflects recent water pollution with human fecal matter.

4.6. References

1. **Bae, S., and S. Wuertz.** 2009. Discrimination of viable and dead fecal *Bacteroidales* bacteria by quantitative PCR with propidium monoazide. *Appl. Environ. Microbiol.* **75**:2940-2944.
2. **Bae, S., and S. Wuertz.** 2012. Survival of host-associated *Bacteroidales* cells and their relationship with *Enterococcus* spp., *Campylobacter jejuni*, *Salmonella enterica* Serovar Typhimurium, and adenovirus in freshwater microcosms as measured by propidium monoazide-quantitative PCR. *Appl. Environ. Microbiol.* **78**:922-932.
3. **Bell, A., A. C. Layton, L. McKay, D. Williams, R. Gentry, and G. S. Saylor.** 2009. Factors influencing the persistence of fecal *Bacteroides* in stream water. *J. Environ. Qual.* **38**:1224-1232.
4. **Bernhard, A. E., and K. G. Field.** 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**:1587-1594.
5. **Bernhard, A. E., and K. G. Field.** 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **66**:4571-4574.
6. **Brennan, F. P., F. Abram, F. A. Chinalia, K. G. Richards, and V. O'Flaherty.** 2010. Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil. *Appl. Environ. Microbiol.* **76**:2175-2180.
7. **Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky, and S. Ishii.** 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ. Microbiol.* **8**:504-513.
8. **Crabill, C., R. Donald, J. Snelling, R. Foust, and G. Southam.** 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. *Water Res.* **33**:2163-2171.
9. **Craig, D. L., H. J. Fallowfield, and N. J. Cromar.** 2004. Use of macrocosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with in situ measurements. *J. Appl. Microbiol.* **96**:922-930.
10. **Davies, C. M., J. A. H. Long, M. Donald, and N. J. Ashbolt.** 1995. Survival of fecal microorganisms in marine and fresh water sediments *Appl. Environ. Microbiol.* **61**:1888-1896.

11. **Dick, L. K., E. A. Stelzer, E. E. Bertke, D. L. Fong, and D. M. Stoeckel.** 2010. Relative decay of *Bacteroidales* microbial source tracking markers and cultivated *Escherichia coli* in freshwater microcosms. *Appl. Environ. Microbiol.* **76**:3255-3262
12. **Eichmiller, J. J., R. E. Hicks, and M. J. Sadowsky.** 2013. Distribution of genetic markers of fecal pollution on a freshwater sandy shoreline in proximity to wastewater effluent. *Environ. Sci. Technol.* **47**:3395-3402.
13. **Flekna, G., P. Stefanic, M. Wagner, F. J. M. Smulders, S. S. Mozina, and I. Hein.** 2007. Insufficient differentiation of live and dead *Campylobacter jejuni* and *Listeria monocytogenes* cells by ethidium monoazide (EMA) compromises EMA/real-time PCR. *Res. Microbiol.* **158**:405-412.
14. **Flint, K. P.** 1987. The long-term survival of *Escherichia coli* in river water. *J. Appl. Bacteriol.* **63**:261-270.
15. **Green, H. C., O. C. Shanks, M. Sivaganesan, R. A. Haugland, and K. G. Field.** 2011. Differential decay of human faecal *Bacteroides* in marine and freshwater. *Environmental Microbiology* **13**:3235-3249.
16. **Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl. Environ. Microbiol.* **72**:612-621.
17. **Josephson, K. L., C. P. Gerba, and I. L. Pepper.** 1993. Polymerase chain-reaction detection of nonviable bacterial pathogens. *Appl. Environ. Microbiol.* **59**:3513-3515.
18. **Kobayashi, H., M. Oethinger, M. J. Tuohy, G. S. Hall, and T. W. Bauer.** 2009. Improving clinical significance of PCR: use of propidium monoazide to distinguish viable from dead *Staphylococcus aureus* and *Staphylococcus epidermidis*. *J. Orthop. Res.* **27**:1243-1247.
19. **Kramer, M., N. Obermajer, B. B. Matijasic, I. Rogelj, and V. Kmetec.** 2009. Quantification of live and dead probiotic bacteria in lyophilised product by real-time PCR and by flow cytometry. *Appl. Microbiol. Biotechnol.* **84**:1137-1147.
20. **Kreder, C. A.** 1998. Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Appl. Environ. Microbiol.* **64**:4103-4105.
21. **Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Saylor.** 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR

- assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* **72**:4214-4224.
22. **Liang, Z. B., Z. L. He, X. X. Zhou, C. A. Powell, Y. E. Yang, M. G. Roberts, and P. J. Stoffella.** 2012. High diversity and differential persistence of fecal *Bacteroidales* population spiked into freshwater microcosm. *Water Res.* **46**:247-257.
 23. **Lee, C. M., T. Y. Lin, C. C. Lin, G. A. Kohbodi, A. Bhattl, R. Lee, and J. A. Jay.** 2006. Persistence of fecal indicator bacteria in Santa Monica Bay beach sediments. *Water Res.* **40**:2593-2602
 24. **Nebe-von-Caron, G., P. J. Stephens, C. J. Hewitt, J. R. Powell, and R. A. Badley.** 2000. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. *J. Microbiol. Methods* **42**:97-114.
 25. **Newton, R. J., J. L. VandeWalle, M. A. Borchardt, M. H. Gorelick, and S. L. McLellan.** 2011. *Lachnospiraceae* and *Bacteroidales* alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. *Appl. Environ. Microbiol.* **77**:6972-6981.
 26. **Nielsen, K. M., Johnsen, P. J., Bensasson, D. and D. Daffonchio.** 2007. Release and persistence of extracellular DNA in the environment. *Environ. Biosafety Res.* **6**:37-53.
 27. **Nocker, A., and A. K. Camper.** 2006. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Appl. Environ. Microbiol.* **72**:1997-2004.
 28. **Nocker, A., C. Y. Cheung, and A. K. Camper.** 2006. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *J. Microbiol. Methods* **67**:310-320.
 29. **Nogva, H. K., S. M. Dromtorp, H. Nissen, and K. Rudi.** 2003. Ethidium monoazide for DNA-based differentiation of viable and dead bacteria by 5' -nuclease PCR. *Biotechniques* **34**:804-810.
 30. **Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* **74**:890-901.
 31. **Pote, J., L. Haller, R. Kottelat, V. Sastre, P. Arpagaus, and W. Wildi.** 2009. Persistence and growth of faecal culturable bacterial indicators in water column and sediments of Vidy Bay, Lake Geneva, Switzerland. *J. Environ. Science-China* **21**:62-69.

32. **Rawsthorne, H., C. N. Dock, and L. A. Jaykus.** 2009. PCR-based method using propidium monoazide to distinguish viable from nonviable *Bacillus subtilis* spores. *Appl. Environ. Microbiol.* **75**:2936-2939.
33. **Rudi, K., H. K. Hoidal, T. Katla, B. K. Johansen, J. Nordal, and K. S. Jakobsen.** 2004. Direct real-time PCR quantification of *Campylobacter jejuni* in chicken fecal and cecal samples by integrated cell concentration and DNA purification. *Appl. Environ. Microbiol.* **70**:790-797.
34. **Salyers, A. A.** 1984. *Bacteroides* of the human lower intestine track. *Annu. Rev. Microbiol.* **38**:293-313.
35. **Schulz, C. J., and G. W. Childers.** 2011. Fecal *Bacteroidales* diversity and decay in response to variations in temperature and salinity. *Appl. Environ. Microbiol.* **77**:2563-2572.
36. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. D. Siciliano.** 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. *Environ. Microbiol.* **7**:249-259.
37. **USEPA.** 2002. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermo tolerant *Escherichia coli* agar (modified mTEC). **EPA 821-R-02-023.**
38. **USEPA.** 40 CFR part 136. **Appendix B. 314-315.**
http://www.epa.gov/region9/qa/pdfs/40cfr136_03.pdf
39. **USEPA.** 2012. Recreational water quality criteria. **820-F-12-061.**
<http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/factsheet2012.pdf>
40. **USEPA.** 1999. EPA guidance manual turbidity provisions. Chapter 9. Turbidity in source water. EPA-815-R-99-010.
<http://water.epa.gov/lawsregs/rulesregs/sdwa/mdbp/mdbptg.cfm>
41. **Varma, M., R. Field, M. Stinson, B. Rukovets, L. Wymer, and R. Haugland.** 2009. Quantitative real-time PCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Res.* **43**:4790-4801.
42. **Vesper, S. J., M. E. Rogers, A. N. Neely, and R. A. Haugland.** 2007. Opportunistic *Aspergillus* pathogens measured in home and hospital tap water by quantitative PCR (QPCR). *J. Water Health* **5**:427-431.

43. **Wagner, A. O., C. Malin, B. A. Knapp, and P. Illmer.** 2008. Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. *Appl. Environ. Microbiol.* **74**:2537-2539.
44. **Walters, S. P., and K. G. Field.** 2009. Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. *Environ. Microbiol.* **11**:1410-1421.
45. **Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. I. Gordon.** 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Sci.* **299**:2074-2076.
46. **Yang, L., X. Lu, C. W. Nossa, F. Francois, R. M. Peek, and Z. Pei.** 2009. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. *Gastroenterology* **137**:588-597.

Table 4.1. Primers, target genes and amplicon lengths for qPCR assays

Assay	Primer Sequences (5'-3')	Target gene	Size	Reference
Bac32	32F: ACGCTAGCTACAGGCTT 708R: CAATCGGAGTTCTTCGTG	General <i>Bacteroidales</i> 16S rRNA	676 bp	Bernhard & Field, 2000a
AllBac	296F: GAGAGGAAGGTCCCCCAC 412R: CGCTACTTGGCTGGTTCAG	General <i>Bacteroidales</i> 16S rRNA	106 bp	Layton et al., 2006
HF183	183F: ATCATGAGTTCACATGTCCG 708R: CAATCGGAGTTCTTCGTG	Human associated <i>Bacteroidales</i> 16S rRNA	525 bp	Bernhard & Field, 2000b
qHF183	183F: ATCATGAGTTCACATGTCCG 265R:TACCCCGCCTACTATCTAATG	Human associated <i>Bacteroidales</i> 16S rRNA	82 bp	Seurinck et al., 2005

Table 4.2. Quality control parameters for AllBac and qHF183 qPCR assays

Assay	Composite standard curve	Range of quantification (gene copies)	Limit of detection	Efficiency of composite std. curve	R ² of composite std. curve
AllBac	y= -3.208x + 34.37	100 to 10 ⁶	37	105.0%	0.999
qHF183	y= -3.258x +33.944	10 to 10 ⁶	3.2	102.8%	0.998

Table 4.3. Survival of *Bacteroidales* and *E. coli* (as percentages of the initial concentrations) in microcosms prepared with stream water.

	AllBac		qHF183		<i>E. coli</i>
	No PMA	PMA	No PMA	PMA	
Day 0	100%	100%	100%	100%	100%
8- hrs	72.80%	61.80%	71.80%	51.70%	
Day 1	26.10%	22.40%	28.10%	25.70%	71.60%
Day 2	1.23%	1.35%	0.64%	0.26%	11.80%
Day 3	0.39%	0.52%			7.28%
Day 4	0.16%	0.20%			2.52%
Day 5	0.09%	0.19%			1.74%
Day 6	0.07%	0.13%			1.16%
Day 8					0.19%
Day 11					0.12%
Day 14					0.09%

Table 4.4. Survival of *Bacteroidales* and *E. coli* (as percentages of the initial concentrations) in microcosms prepared with sediment.

	AllBac		qHF183		<i>E. coli</i>
	No PMA	PMA	No PMA	PMA	
Day 0	100%	100%	100%	100%	100%
8 hrs	90.9%	67.7%	76.9%	76.4%	80.5%
Day 1	52.2%	43.1%	52.0%	51.9%	79.0%
Day 2	15.5%	19.1%	11.1%	7.9%	39.7%
Day 3	3.3%	5.8%	0.31%	0.67%	32.3%
Day 4	1.2%	1.9%			7.74%
Day 5	0.44%	0.64%			5.60%
Day 6	0.21%	0.37%			3.62%
Day 7	0.14%	0.17%			2.30%
Day 9	0.02%	0.03%			1.61%
Day 16					1.15%
Day 23					0.86%
Day 33					0.34%
Day 45					0.13%
Day 60					0.07%
Day 75					0.02%

Table 4.5. Selected chemical parameters for microcosms during the study period

Parameter	Stream water microcosm		Sediment microcosm	
	First day (Day 0)	Last day (Day 14)	First day (Day 0)	Last day (day 75)
pH	7.28	7.63	7.46	7.56
EC (s/m)	0.02	0.03	0.03	0.03
Total org. C (mg/l)	20.84	18.09	25.49	25.74
Dissol.org. C (mg/l)	21.04	18.76	23.76	23.9
Total N (mg/l)	2.23	2.75	4.01	4.1
Dissolved N (mg/l)	2.21	2.87	4.06	4
Total C/N ratio	9.35	6.58	6.36	6.28
NH ₄ -N (mg/l)	0.991	1.957	2.045	2.953
PO ₄ -P (mg/l)	1.991	2.35	1.932	1.932
Ca (mg/l)	17.83	20.5	26.9	27.46
K (mg/l)	6.33	7.43	5.27	5.3
Mg (mg/l)	5.46	6.41	11.47	12.45
Turbidity (NTU)	N/A	N/A	210 (SD 4.7)	235 (SD 3.5)

SD: standard deviation

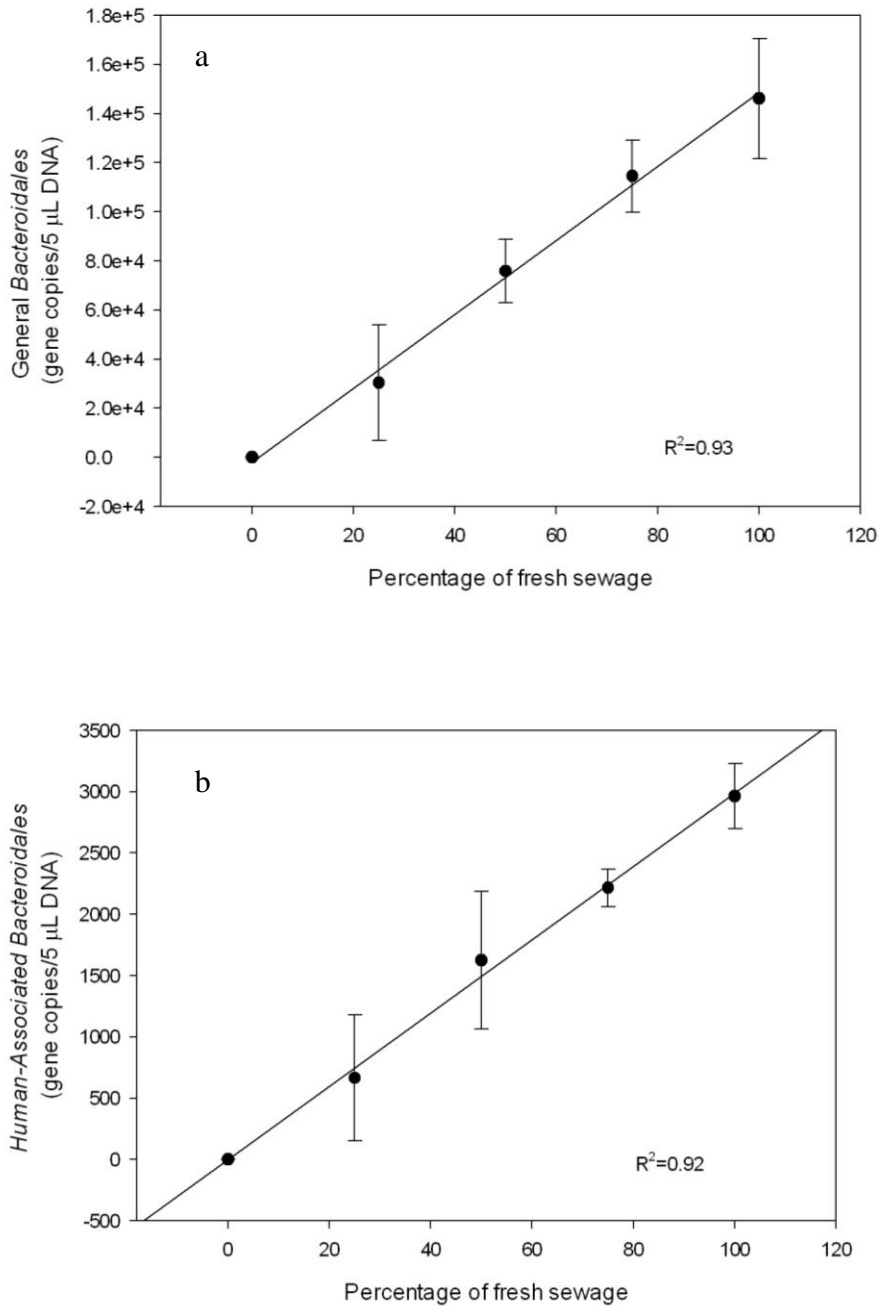


Fig. 4.1. Verification of the effectiveness of PMA (at 100 µM and 10-minute light exposure) to discriminate between live and dead *Bacteroidales* cells in stream water for PCR amplification of defined ratios of fresh and heat treated sewage for general *Bacteroidales* markers (a) and human associated *Bacteroidales* markers (b). The error bars represent standard deviations for three samples.

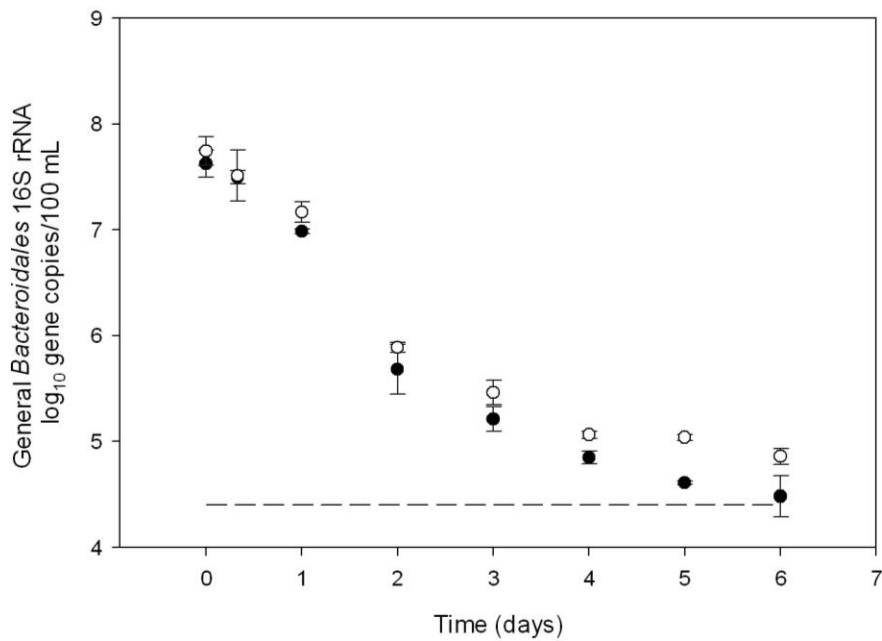


Fig. 4.2. Persistence and survival of general *Bacteroidales* 16S rRNA gene copies (AllBac) in microcosms prepared with stream water. Open circles denote the qPCR amplification of DNA extracted from stream water microcosms without PMA treatment, closed circles denote the qPCR amplification of DNA extracted from stream water microcosms with PMA treatment, and the dashed line represents the LOD. The error bars represent standard deviations for three samples.

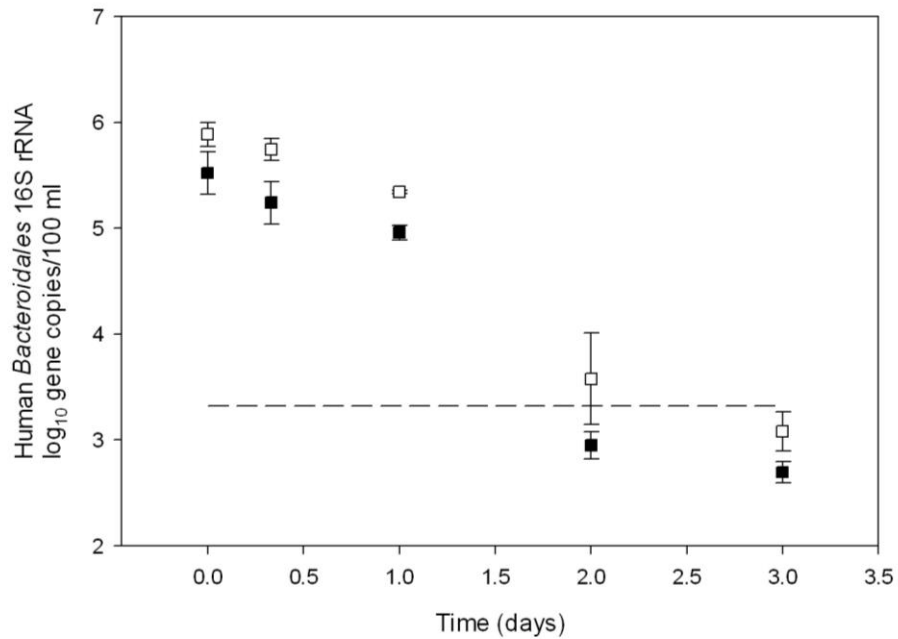


Fig. 4.3. Persistence and survival of human-associated *Bacteroidales* 16S rRNA (qHF183) gene copies in microcosms prepared with stream water. Open squares denote the qPCR amplification of DNA extracted from stream water microcosms without PMA treatment, closed squares denote the qPCR amplification of DNA extracted from stream water microcosms with PMA treatment, and the dashed line represents the LOD. The error bars represent standard deviations for three samples.

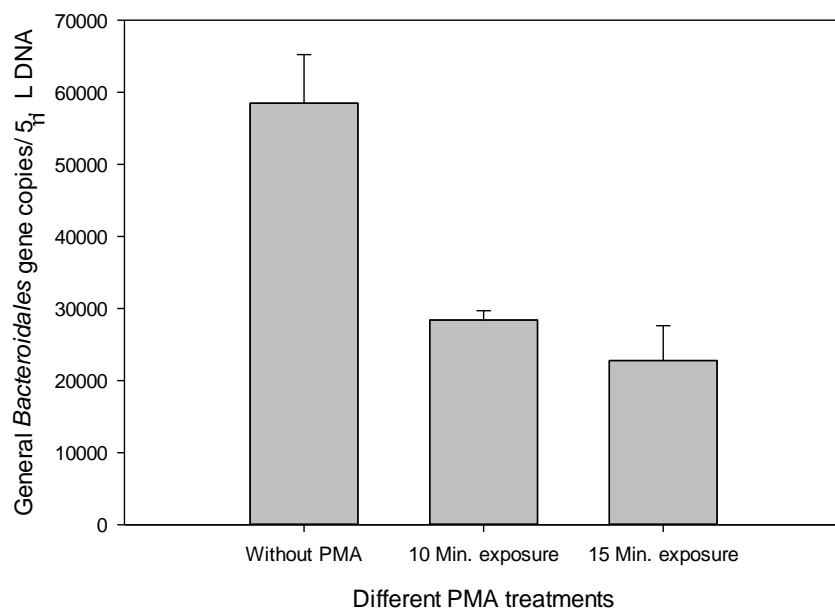


Fig. 4.4. Effects of light exposure during PMA treatment on general *Bacteroidales* 16S rRNA AllBac markers in sediment samples under conditions of 1% sediment and 100 μ M PMA.

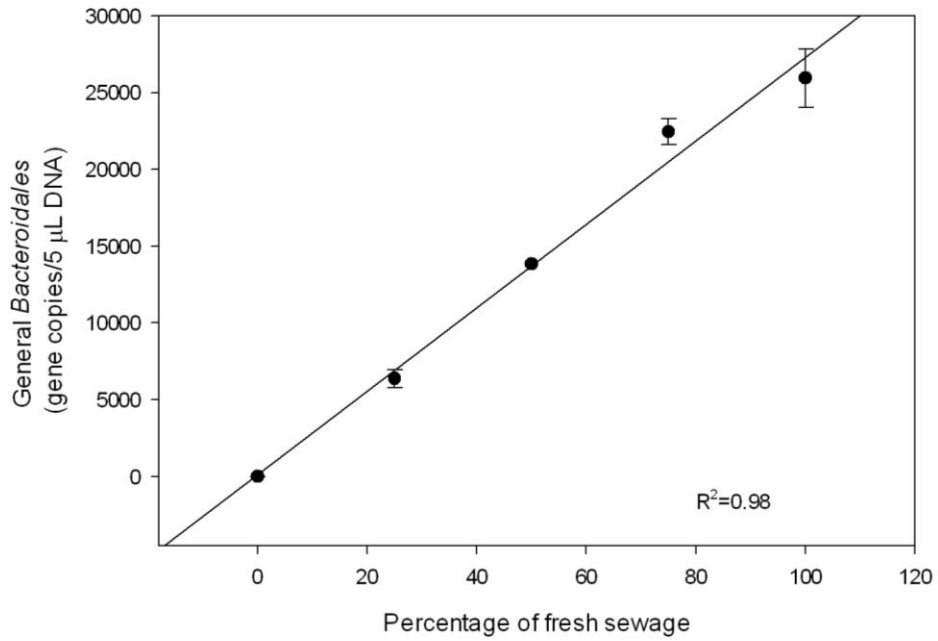


Fig. 4.5. Relationship between the percentages of *Bacteroidales* live cells and general *Bacteroidales* 16S rRNA markers under conditions of 1% sediment, 100 μM PMA and 10 minutes of light exposure using defined ratios of fresh and heat treated sewage samples. The error bars represent standard deviations for three samples.

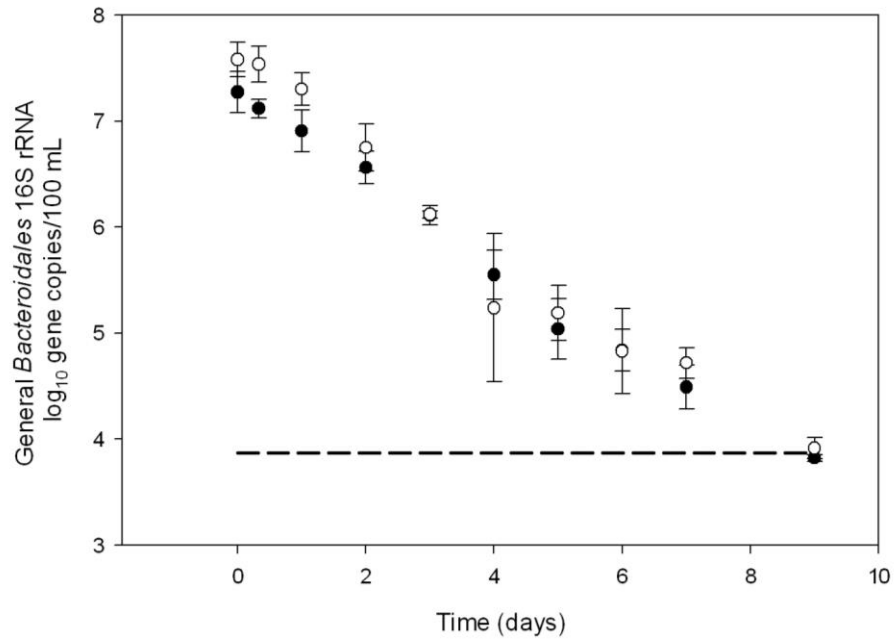


Fig. 4.6. Persistence and survival of general *Bacteroidales* 16S rRNA gene copies in microcosms prepared with 1% sediment. Open circles denote the qPCR amplification with DNA extracted from sediment microcosms without PMA treatment, closed circles denote the qPCR amplification with DNA extracted from sediment microcosm with PMA treatment, and the dashed line represents LOD. The error bars represent standard deviations for three samples.

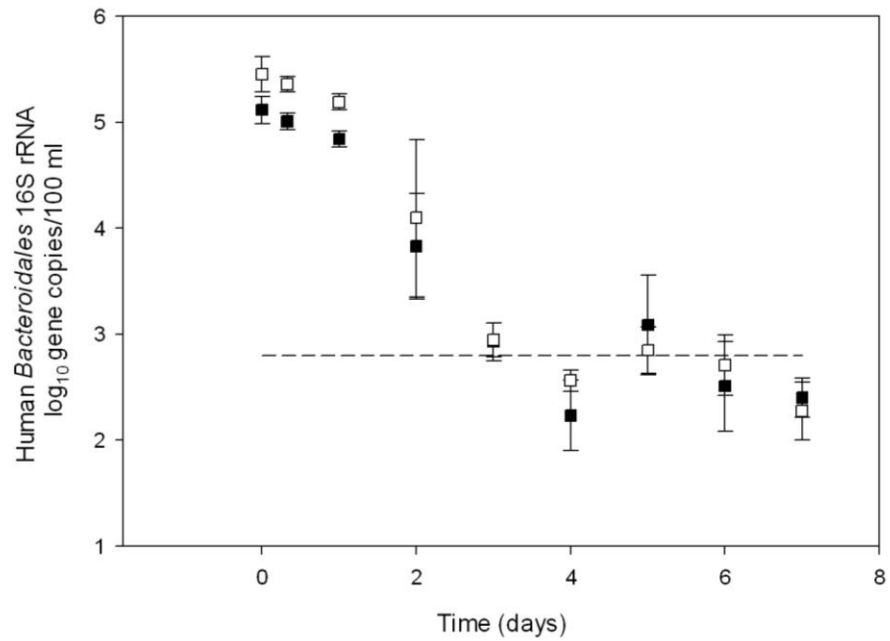


Fig. 4.7. Persistence and survival of human associated *Bacteroidales* 16S rRNA gene copies in microcosms prepared with 1% sediment. Open squares denote the qPCR amplification with DNA extracted from sediment microcosm without PMA treatment, closed squares denote the qPCR amplification with DNA extracted from sediment microcosms with PMA treatment, and the dashed line represents the LOD. The error bars represent standard deviations for three samples.

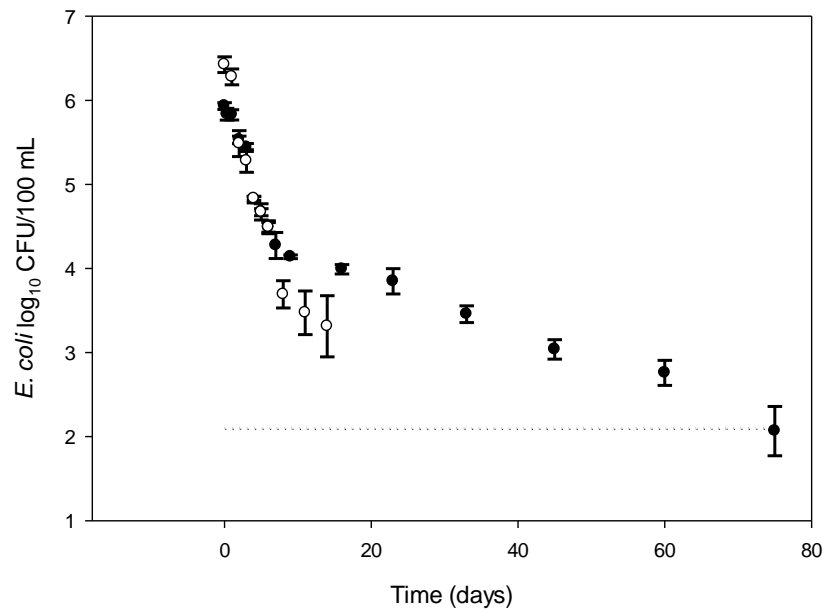


Fig. 4.8. Survival of *E. coli* in microcosms prepared with stream water and 1% sediment. Open circles denote the survival of *E. coli* in stream water microcosms and closed circles represent the survival of *E. coli* in 1% sediment microcosms. The dashed line represents the USEPA 2012 criterion for 30-day geometric means for recreational water (126 CFU/100 ml). The error bars represent standard deviations for three samples.

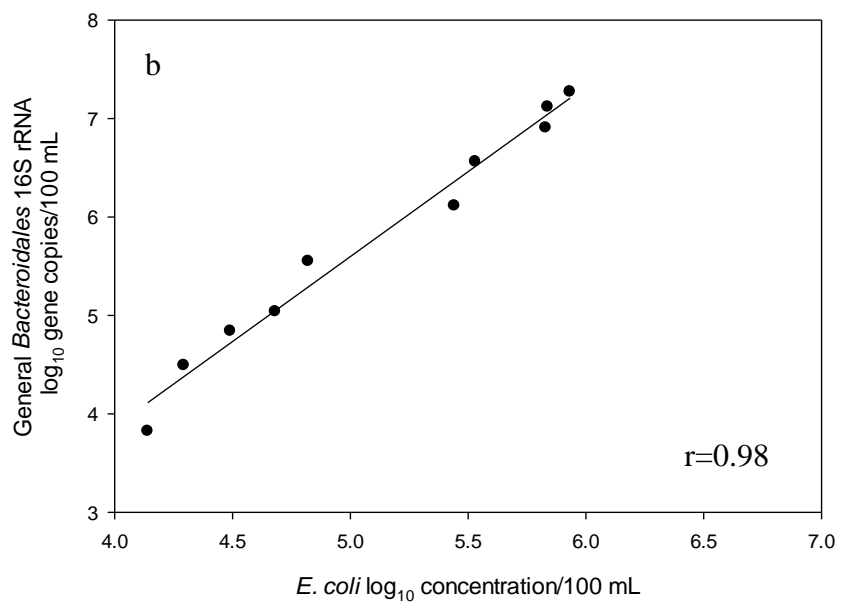
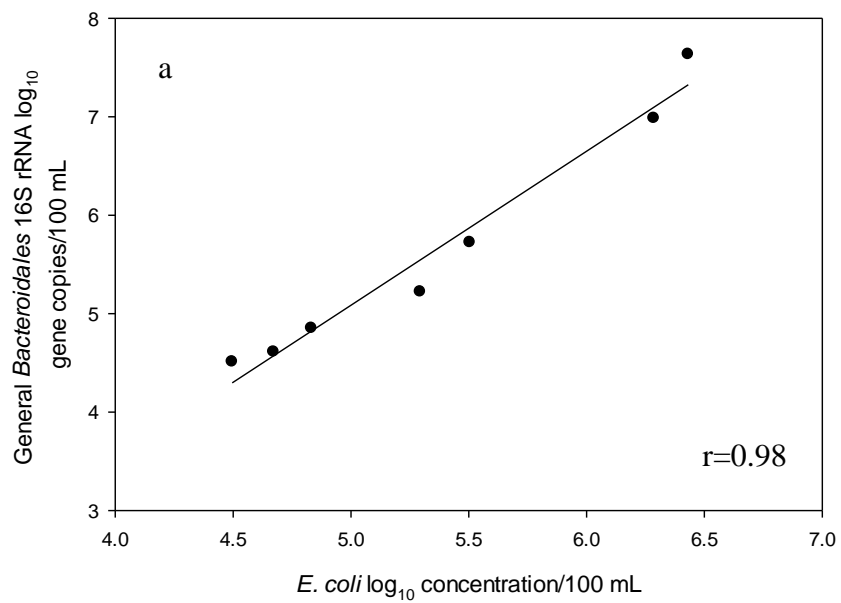


Fig. 4.9. The correlation between *E. coli* and general *Bacteroidales* 16S rRNA (AllBac) gene markers found in (a) stream water microcosms and (b) sediment microcosms without PMA treatment.

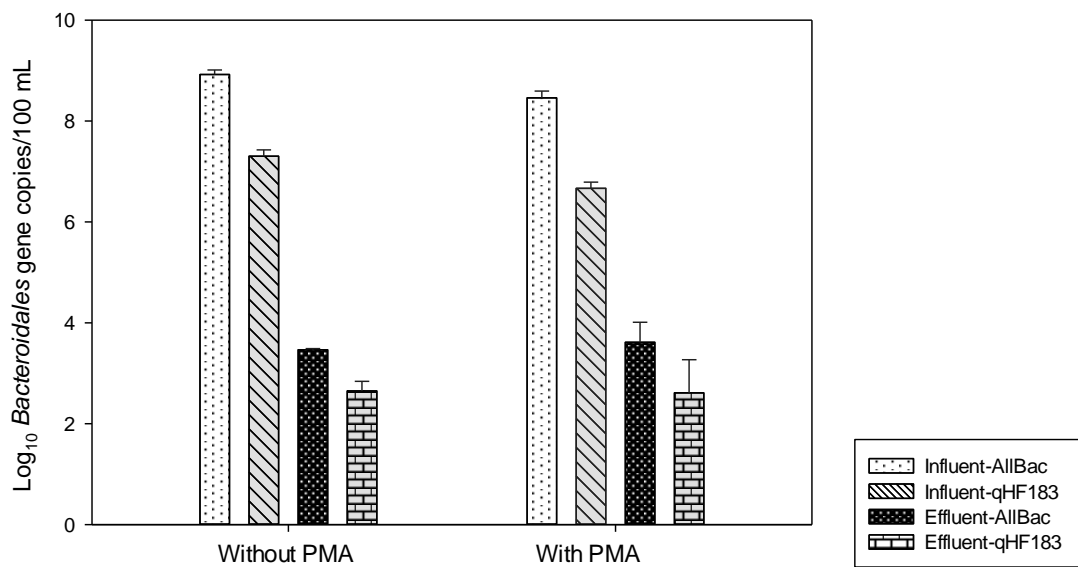


Fig. 4.10. The ability of PMA to discriminate between DNA from live and dead cells in the influent and effluent of a wastewater treatment plant in Auburn, AL.

5. Summary and future directions

This is a comprehensive study conducted to identify the suitability of utilizing *Bacteroidales* as an indicator bacterium to identify the source of fecal contamination in surface water. Library based method and non-library based methods were compared and potential human and cattle molecular markers for use in Alabama watersheds evaluated. The molecular markers selected as a result of this evaluation were then used to quantify the fecal contamination in two different watersheds, to identify the correlation between *E. coli* and *Bacteroidales*, to determine the survival of *Bacteroidales* and *E. coli* in stream water and sediment microcosms, and to differentiate between DNA associated with live and dead cells in samples.

The comparison of library based and non-library based methods indicated that both approaches have both advantages and disadvantages. The use of the library based rep-PCR DNA fingerprinting method for a watershed with mixed land used patterns is a useful approach for detecting the sources of fecal contamination from various source groups. This method may also be helpful for efforts to gather information on long-term patterns, especially those related to changing sources, of fecal contamination in a watershed and their temporal variability because it enables researchers to identify more source groups than non-library based methods. On the other hand, non-library based methods, especially gene amplification with qPCR, can provide faster results than library based approaches, which can be particularly important when quick decisions need to be made regarding designated beaches or shellfish farming areas. Therefore, the selection of the proper approach is crucial and will mainly depend on the research objectives, the

availability of funds and the time frame. If sufficient funds are available, the combination of two or more approaches would provide more robust data.

This study revealed that before adopting a molecular marker for a particular watershed, it is imperative to test the specificity and sensitivity of those markers because markers can have significant spatial variability. Also, in an MST study researchers should consider what percentage of gene copies originate from live cells and what percentage from dead cells.

Further, based on the designated use of the water source, researchers should establish clear research goals. If the study is conducted for a designated beach area and if beach closings happen frequently due to elevated fecal indicator bacteria (FIB) levels, their research should focus primarily on detecting human pathogens such as *Shigella*, *Campylobacter jejuni*, pathogenic *E. coli* (O157:H7) and *Salmonella*. These data can then be used to assess the risk of swimmers becoming ill using microbial risk assessment (QMRA) tools.

If any watershed has elevated FIB, researchers should use MST to identify the precise sources of fecal contamination in order to develop total maximum daily loads (TDML) for a particular watershed. Before selecting a MST method, a sanitary survey should be conducted in the watershed and an appropriate method selected based on survey results, funding availability, research goals and the relevant time frame.

Microbial community analysis is another approach that can be used to determine how microbial populations change spatially and temporally. If a city is planning a new wastewater treatment plant or deciding whether to shut down an old one, a thorough

community analysis could provide useful information on how these communities change overtime downstream.

Microbial source tracking is a powerful tool that can successfully be used to detect the sources of fecal pollution in a watershed. Using the correct method at the correct time is imperative in order to produce the accurate results needed for soundly-based future management decisions.

Appendices

S. 2.1 Enumeration of *E. coli* and verification

Water samples were collected from six sampling sites, kept on ice, transported to the laboratory and processed within 6 hours. From each sample, three dilutions were filtered through 0.45 µm membrane filters under vacuum. Membrane filters were placed on modified membrane-thermotolerant *Escherichia coli* agar (m-TEC) media (Difco, Detroit, MI) and incubated at 37°C for 2 hours, and then incubated at 44.5°C for 24 hours. Isolates giving a magenta/red color were selected (1 colony per plate) and streaked on MacConkey agar (Difco). After overnight incubation at 37°C, a single colony of dark pink color was selected from each plate. Half of that colony was streaked on Chrom agar (Chromagar Microbiology, Paris, France) and the other half was streaked on MacConkey agar. After overnight incubation at 37°C, colonies that were dark pink in color on the MacConkey agar and blue on the Chrom Agar were selected to inoculate citrate agar (BBL, Cockeysville, MD), EC broth with 4-methylumbelliferyl-D-glucuronide (EC-MUG) (Difco), 1% trypton (Fisher Biotech, Fair Lawn, NJ), and methyl red–Voges-Proskauer (Difco) broth. Isolates were identified as *E. coli* if they did not use citrate as a substrate, grew at 44.5°C, produced gas and fluorescence in EC-MUG broth, produced indole from tryptophan, and produced an acidic end product 68 when grown in methyl red-Voges-proskauer broth. *E. coli* isolates were suspended in 50% glycerol /nutrient broth (Difco) and stored at –80°C for subsequent analysis.

S. 2.2 rep-PCR DNA fingerprinting development

The rep-PCR DNA fingerprints of the *E. coli* isolates were obtained using the BOX A1R primer (5'-CTA CGG CAA GGC GAC GCT GAC G-3') (Rademaker and de Bruijn, 1997) and *E. coli* whole cells as the templates for PCR. PCR was performed according to a protocol modified after Rademaker and de Bruijn (1997) and Dombek et al. (2000). Briefly, the *E. coli* isolates were grown on Plate Count Agar for 18 hours. A portion of a single colony was then removed using a 1- μ l sterile inoculation loop and suspended in 100 μ l of PCR grade water in a microcentrifuge tube. The 25 μ l PCR mixture contained 2 μ l of whole cell suspension, 2.5 μ l of 10X Promega reaction buffer without MgCl₂ (Promega, Madison, WI); 3.0 μ l of 25 mM MgCl₂ (Promega); 0.2 μ l of 100 mM dNTP's (Promega); 0.2 μ l of BSA (2% stock) (Invitrogen, Carlsbad, CA); 1.0 μ l of 10 μ M BOX A1R primer (final conc 0.2 pmol/ μ l) (Invitrogen); 0.4 μ l (2 units) of Taq DNA polymerase (Promega), and 15.7 μ l of PCR grade water. PCR was performed using a Biometra T-Gradient thermocycler (Whatman, Göttingen, Germany) using the following conditions: initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 94°C for 1 minute, annealing at 60°C for 1 minute, and extension at 72°C for 1 minute, then a final extension at 72°C for 10 minutes. A negative control containing sterile water and a positive control containing *E. coli* ATCC 25922 were included in each PCR set.

PCR products (25 μ l) were mixed with 5 μ l of 6X loading dye (Promega) and 10 μ l of each reaction mixture was resolved using 1.5% agarose gel (25 cm x 20 cm) in 0.5X TBE buffer. One kb Plus DNA ladder (0.66 μ g/well; Invitrogen) was added to the 1st, 10th, 19th, 28th and 36th lane and a negative control was added to the 35th lane (Fig. 2). The gels were electrophoresed at room temperature for 7 hours at 130 V and stained with 0.5

$\mu\text{g/ml}$ ethidium bromide (Fisher Biotech) in 0.5X TBE buffer for 1 hour. Gel images were captured using the Gel Logic 200 imaging system (Eastman Kodak Co., Rochester, NY) and saved as bip files. Prior to analyzing images with the BioNumeric software, the bip files were converted to 8-bit TIFF format.

S. 2.3 Sequence of plasmid DNA containing sewage DNA insert that contains primer sites for HF 183 and qHF 183

5'-TCTCTGACACGTCTCGCGTGTCTGTTGCGCCAGCTTGGTACCGAGCTCGG
ATCTCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTT**CAATCGGAGT**
TCTTCGTGATATCTAAGCATTTACCGCTACACCACGAATTCGCCTGCC
TCAACTGCACTCAAGACATCCAGTATCAACTGCAATTTTACGGTTGAGCC
GCAAACTTTCACAACTGACTTAAACATCCATCTACGCTCCCTTTAAACCC
AATAAATCCGGATAACGCTCGGATCCTCCGTATTACCGCGGCTGCTGGCA
CGGAGTTAGCCGATCCTTATTCATAAAGTACATGCAAACGGGTATGCATA
CCCGACTTTATTCCTTTATAAAAGAAGTTTACAACCCATAGGGCAGTCAT
CCTTCACGCTACTCGGCTGGTTCAGGCCATCGCCATTGACCAATATTCC
TCACTGCTGCCTCCCGTAGGAGTTTGGACCGTGTCTCAGTTCCAATGTGG
GGGACCTTCCTCTCAGAACCCCTATCCATCGTTGACTAGGTGGGCCGT**TA**
CCCCGCCTACTATCTAATGGAACGCATCCCCATCGTCTACCGGAAAATAC
CTTTAATCATG**CGGACATGTGAACTCATGAT**AAGGGCGAATTCTGCAGAT
ATCCATCACACTGGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTC
GCCTATAGTGAGTCGATTACAATTCACTGGCCGTCGTTTACACGTCGTGA
CTGGGAAACCCTGGCGTTACCCAACCTAATCGGCTTGCAGACATCCCCTT
TCGCAGCTGGCGAATAGCGAAAGGCCGCACGATCGCCTTCCCACAGTGCC
CACTGAATGCGAATGGACCCCCCTGAAGGGCCATAAGGCGGGCGGGGTGG
TGGTAAGCGAGGGGACCTAATTGCAGGCCTAGGCCGTTCTTTGTTTCTCC
TTCTTTTCGCCGTCGGGGTTTCCGGTAAGTTTAATCGGGCTCTTAGGGTC
GATTAGGTAAAGCACTCGACCAAACTTGATAGGGTAGGGTCCGATTGGCC
CTCTTAAGCGTTCCTTGACTGGGACACTCTTATGGGACTGTCACGGAACC
TCACATTCGGTATTTTATAAGCATTGCATTCCTTTGTAAAGAGTATTTAA
ATCCGATAATTGGGCAGGTTGAGGGACATCC-3'

183F: 5'ATC ATG AGT TCA CAT GTC CG-3' → **CGG ACA TGT GAA CTC ATG AT**

708R: **5'CAA TCG GAG TTC TTC GTG-3'**

265R: **5' TAC CCC GCC TAC TAT CTA ATG-3'**

S. 2.4 Sequence of plasmid DNA containing sewage DNA insert that contains primer sites for Bac 32 and AllBac

5'TACATGATTACGCCAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCG
CCAGTGTGCTGGAATTCGCCCTT**CAATCGGAGTTCTTCGTG**ATATCTAAGCA
TTTCACCGCTACACCACGAATTCGCCCACTTTGTGCGTACTCAAGGAAACCA
GTTGCGCTGCAGTGCAGACGTTGAGCGTCTACATTTCACAAACACGCTTAATC
TCCGGCCTACGCTCCCTTTAAACCCAATAAATCCGGATAACGCCCGGACCTTC
CGTATTACCGCGGCTGCTGGCACGGAATTAGCCGGTCCTTATTCATAAGGTAC
ATGCAAAAAGCCTCACGAGACTCACTTTATCCCTTATAAAAGCAGTTTACAA
CCCATAGGGCCGTCATCCTGCA**CGCTACTTGGCTGGTTCAG**ACTCTCGTCC
ATTGACCAATATTCCTCACTGCTGCCTCCCGTAGGAGTTTGGACCGTGTCTCA
GTTCCAAT**GTGGGGGACCTTCCTCTC**AGAACCCTACTGATCGTTGCGTTG
GTGGGCCGTTACCCCGCCAACAAGCTAATCAGACGCATCCCCATCCATCACC
GATAAATCTTTAATCTCTTTTCAGATGTCTTCTAGAGATATCATTGGGTATTAG
TCTTACTTTCGCAAGGTTATCCCCAAGTGGTGGGCAGGTTGGATACGCGTTAC
TCACCCGTGCGCCGGTCGACGCCTATCGGAAGCAAGCTTCCAATATCGTTTCC
CTCGACTTGCATGTGTT**AAGCCTGTAGCTAGCGTT**AGGCGAGTTCTGCAAAT
ATCCATCACACTGGCGGCCGCTCGAGCATGCTCTAGAGGGCCCAATTCGCCT
ATGTGAGCCTATACATTCCTGGCCGTCGTTTAAACGTCTTACTGGAAACCTGG
CGTACCACTTAATCGCCTGCACAATCCCCTTCCCAGCTGGGTATACCAAAG
CCCGCCCCATC-3'

708R: 5'-CAA TCG GAG TTC TTC GTG-3'

32F: 5'-AAC GCT AGC TAC AGG CTT-3' →5'- AAG CCT GTA GCT AGC GTT

296F: 5'-GAG AGG AAG GTC CCC CAC-3' →GTG GGG GAC CTT CCT CTC

412R: 5'-CGCTACTTGGCTGGTTCAG-3'

Table S2.1. The correlation coefficients between *Bacteroidales* and *E. coli* by month

Month	Correlation coefficient
Mar-07	0.57
Apr-07	0.20
May-07	-0.32
Jun-07	-0.32
Jul-07	0.93
Aug-07	-0.27
Sep-07	-0.23
Oct-07	0.12
Nov-07	0.67
Dec-07	0.98
Jan-08	0.80
Feb-08	0.81

Table S 2.2. The correlation coefficients (r) between *Bacteroidales* and *E. coli* by sampling site

Site	Correlation coefficient
RSP	0.51*
RHM	0.24
BM	0.13
LT	0.09
WS	0.1
CW	0.53

*only 4 data points are available.

Table S 2.3. Daily rainfall in the Catoma Creek watershed on the day of sample collection and the monthly average

Sampling date	Avg. monthly rainfall (inch)	Daily rainfall (inch)	No. of dry days before collecting samples
26-Mar-07	0.25	0	10
18-Apr-07	1.92	0	3
22-May-07	0.41	0	9
11-Jun-07	1.53	0	3
18-Jul-07	2.92	0	1
13-Aug-07	1.89	0	17
11-Sep-07	3.92	1.18	8
10-Oct-07	2.77	0	1
14-Nov-07	1.63	0.28	21
12-Dec-07	4.06	0	9
15-Jan-08	5.64	0	4
12-Feb-08	3.14	0.66	12

Source: NOAA (<http://www.ncdc.noaa.gov/>)

Table S 3.1. Pearson product moment correlation between *E. coli*, AllBac markers and physiochemical parameters.

	K	Mg	Total P	Org. C	Total N	Turbidity low	Turbidity high	AllBac GM	<i>E. coli</i> GM
Ca	0.62 0.0316 12	0.301 0.341 12	-0.0433 0.894 12	0.267 0.402 12	0.471 0.122 12	-0.238 0.457 12	0.00158 0.996 12	0.105 0.744 12	0.246 0.441 12
K		0.465 0.128 12	0.15 0.641 12	0.658 0.02 12	0.702 0.0109 12	-0.182 0.571 12	0.262 0.41 12	0.0219 0.946 12	0.467 0.126 12
Mg			-0.0113 0.972 12	0.823 0.00101 12	0.646 0.0232 12	-0.257 0.419 12	-0.28 0.378 12	0.428 0.165 12	-0.155 0.631 12
Total P				-0.0622 0.848 12	-0.0745 0.818 12	0.64 0.0251 12	0.679 0.0151 12	0.46 0.132 12	0.64 0.0249 12
Org. C					0.463 0.129 12	-0.125 0.699 12	-0.0191 0.953 12	0.0383 0.906 12	0.0363 0.911 12
Total N						-0.576 0.0501 12	-0.223 0.486 12	0.198 0.537 12	-0.0551 0.865 12
Turbidity low							0.479 0.115 12	0.198 0.537 12	0.465 0.128 12
Turbidity high								-0.0385 0.905 12	0.813 0.00129 12
AllBac GM									0.282 0.375 12
<i>E. coli</i> GM									

Cell content: Correlation coefficient; P values and number of samples

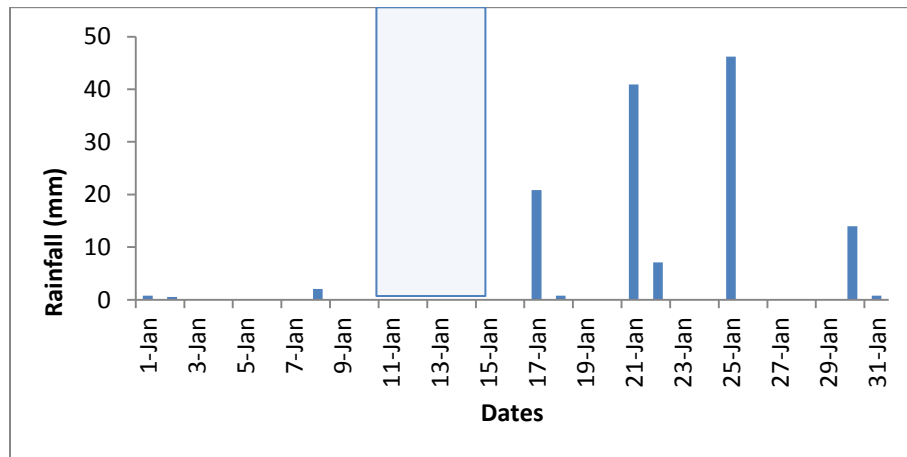
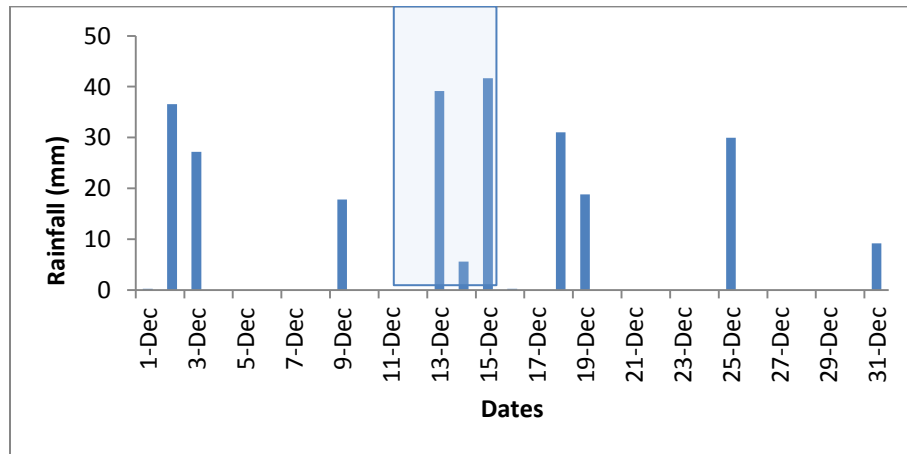
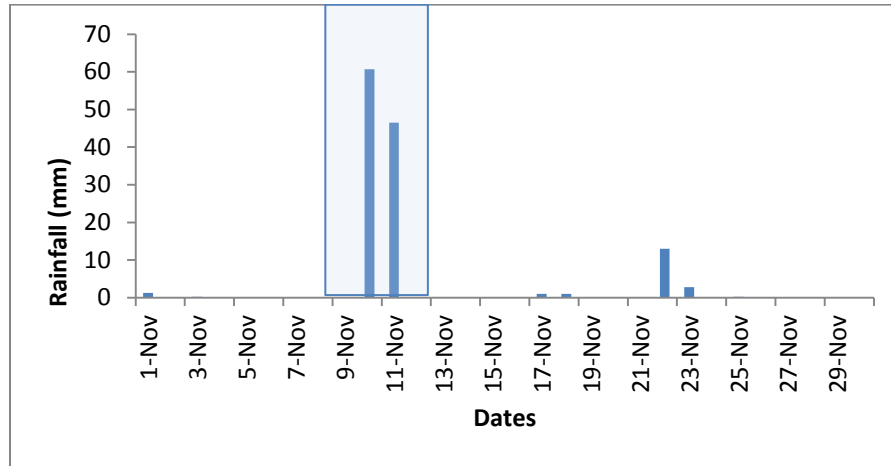


Fig. S 3.1. Rainfall distribution during the sampling months. Gray boxes represent the sampling periods in each month (Source: <http://www.awis.com/mesonet>).

Table S.4.1. Survival of *Bacteroidales* and *E. coli* in their secondary habitat: microcosm prepared with stream water

Days	AllBac		qHF183		<i>E. coli</i>	
	Gene copies/ 100 mL	As a % of initial conc	Gene copies/ 100 mL	As a % of initial conc	CFU/ 100 mL	As a % of initial conc
Day 0	43,162,567 18,761,611 (43.4%)		786,975 358,031 (45.5%)		2,700,000	
8- hrs	31449109 11592865 (36.8%)	72.8% 61.8%	565,210 185,267 (32.7%)	71.8% 51.7%	N/A	
Day 1	9,672,309 4,896,251 (50.6%)	26.1% 22.4%	221,421 91,980 (41.5%)	28.1% 25.7%	1,933,333	71.6%
Day 2	528,827 254,132 (48.0%)	1.23% 1.35%	5,037 919 (18.2%)	0.64% 0.26%	320,000	11.8%
Day 3	166,927 97,020 (58.1%)	0.39% 0.52%	Below LOD		196,667	7.28%
Day 4	70,813 37,856 (53.5%)	0.16% 0.20%			68,000	2.52%

Day 5	40,806 <i>35,854</i> (87.8%)	0.09% <i>0.19%</i>	47,000	1.74%
Day 6	32,299 <i>23,892</i> (73.9%)	0.07% <i>0.13%</i>	31,333	1.16%
Day 8	Below LOD		5,133	0.19%
Day 11			3,300	0.12%
Day 14			2,500	0.09%

The first number in the AllBac and qHF183 gene copies represents the number of gene copies amplified by qPCR without PMA treatment, numbers in italic form represent the number of gene copies amplified by qPCR with PMA treatment and the numbers within parenthesis denote the percentage of gene copies derived from live cells.

Table S.4.2. Survival of *Bacteroidales* and *E. coli* in their secondary habitat: microcosm prepared with sediment

Days	AllBac		qHF183		<i>E. coli</i>	
	Gene copies/100 ml	As a % of initial conc.	Gene copies/100 ml	As a % of initial conc.	CFU/100 ml	As a % of initial conc.
Day 0	39,882,518 19,929,615 (49.9%)		298,616 134,718 (45.1%)		856,667	
8-hrs.	36,270,236 13,283,230 (36.6%)	90.9% 67.7%	229,630 10,2879 (44.4%)	76.9% 76.4%	690,000	80.5%
Day 1	20,826,686 8,591,865 (41.2%)	52.2% 43.1%	156,701 69931 (44.6%)	52.0% 51.9%	676,667	79.0%
Day 2	6,164,879 3,800,693 (61.6%)	15.5% 19.1%	33,013 10,591 (32.1%)	11.1% 7.9%	340,000	39.7%
Day 3	1,319,740 1,152,256 (87.7%)	3.3% 5.8%	923 898 (97%)	0.31% 0.67%	276,667	32.3%
Day 4	461,426 388,337 (84.2%)	1.2% 1.9%	Below LOD		66,333	7.74%
Day 5	176,253 126,695 (71.8%)	0.44% 0.64%			48,000	5.60%
Day 6	86,975 73,756 (84.8%)	0.21% 0.37%			31,000	3.62%
Day 7	54,015 33,359 (61.8%)	0.14% 0.17%			19,667	2.30%
Day 9	8,356 6,672 (79.8%)	0.02% 0.03%			13,800	1.61%
Day 11	Below LOD					
Day 16					9,867	1.15%
Day 23					7,333	0.86%
Day 33					2,900	0.34%
Day 45					1,113	0.13

Day 60	597	0.07%
Day 75	137	0.02%

The first number in the AllBac and qHF183 gene copies represents the number of gene copies amplified by qPCR without PMA treatment, numbers in italic form represent the number of gene copies amplified by qPCR with PMA treatment and the numbers within parenthesis denote the percentage of gene copies derived from live cells.