# Microbial Source Tracking of Human and Animal Waste Pollution of Diverse Watersheds and of Urban Drainage Systems using Molecular Methods.

A Thesis

Submitted to the Faculty

of

Drexel University

by

Michael O'Neil Ryan

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

December 2012

## DEDICATION

I would like to dedicate this Doctoral dissertation to the three most influential women in my life. To my late mother who passed on to me the insatiable appetite for scientific knowledge. To my late grandmother who instilled in me the aspiration of always aiming for the highest levels in anything I do. To my aunt (my surrogate mother, my sister, and my friend) who instilled in me a respect for education and the fortitude that goes with hard work; who has always been motivating and supportive of my ambitions; and whose counsel I highly regard.

## ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my teachers, colleagues and friends that made my thesis worthwhile.

I would like to extend a special thanks to my advisor Dr. Charles N. Haas. I am grateful for the opportunities he gave to me in pursuing my doctorate, and making my time at Drexel University academically stimulating and rewarding. I greatly appreciate his encouragement and support. I am honored to have been chosen as his student for such a worthwhile and interesting research project.

I would like to thank my committee members for their advice throughout my research. Gary Burlingame provided invaluable support and kept my research relevant to environmental issues concerning watersheds in the Philadelphia area. Dr. Jacob Russell guided and helped to develop my microbiological knowledge and expertise. Dr. Patrick Gurian ensured my statistical concepts and applications were pertinent. Dr. Mira Stone Olson guided my understanding of microbiology and hydrology. Dr. Franco Montalto encouraged the broader implications of my research in terms of the engineering applicability including towards Green Infrastructures. It has been an honor to have had such a diverse and knowledgeable committee.

I would like to thank the faculty and staff in Civil, Architectural and Environmental Engineering at Drexel University for their continued help in all matters pertaining to my degree, especially Barbara Interlandi, Sharon Stokes, Kenneth Holmes and Gregory Hilley. Finally, thanks to my fellow graduate student peers, past and present. My time at Drexel University would have been substantially less rewarding and enjoyable without you.

This project is funded through the Philadelphia Water Department and the LD Betz Environmental Engineering Endowment.

**TABLE OF CONTENTS**

## LIST OF TABLES

**LIST OF FIGURES**

**ABSTRACT**

Microbial Source Tracking of Human and Animal Waste Pollution of Diverse Watersheds and of Urban Drainage Systems using Molecular Methods.

Michael O'Neil Ryan

Dissertation Advisor: Prof. Charles N. Haas. L.D. Betz Professor

The central hypothesis of microbial source tracking (MST) is that there exists a distinguishable relationship between microbes excreted in feces and their particular hosts. These associations could be due to adaptations of microbes to their host's gut environment and may involve coevolution between microbes and their hosts. In this study, we performed phylogenetic analyses on 16S rRNA gene sequences associated with different host groups to determine if *Bacteroides dorei* and *Enterococcus faecium* show co-evolutionary characteristics, thus, host specificity. Bioinformatic analyses were also conducted on *E. faecium* surface protein gene (*esp*$_{fm}$) for host specificity. The results of our analyses provide good support for *B. dorei* human host specificity, and very little support for such specificity in *E. faecium*.

Given these findings, a redesigned primer and probe set was developed for a probe based qPCR assay that specifically targets bacterial sequences from a human-specific *B. dorei* lineage (HF68-HR183rc) and another targeting the *E.*

*faecium esp$_{fm}$* gene. The results from comparative assays of HF68-HR183rc against a HF183 protocol obtained from the literature showed similar sensitivity, but an improvement in the specificity of our newly designed protocol. However, assays of rabbit samples showed a high number of positive assays with both the redesigned (58%) and the HF183 (67%) protocol. The newly designed HF68-HR183rc assay could be of considerable use to screen a watershed for human fecal sources of pollution if it is known not to be impacted by rabbits. Assays of the redesigned *esp$_{fm}$* protocol were not as successful as a quantitative assay, requiring an enrichment process.

The HF68-HR183rc qPCR MST protocol was used to help characterize the sources of pollution of sub-watersheds of the temperate environs of Philadelphia/Delaware watershed; watersheds of the tropical island of Puerto Rico; and different types of green infrastructures in Philadelphia and New York.

Keywords: microbial source tracking; *Bacteroides dorei*; *Enterococcus faecium*; qPCR; molecular (genetic) markers; 16S rRNA gene; bioinformatics; phylogenetic

# CHAPTER 1: BACKGROUND AND LITERATURE REVIEW

## 1.1 Introduction

The most generally accepted methods for monitoring or predicting the occurrence of potential pathogenic microorganisms in aquatic environs are based on cultivation and enumeration of fecal indicator bacteria (FIB), which include total and fecal coliforms, *Escherichia coli*, and fecal enterococci. However, conventional methodologies used for the enumeration of FIBs provide no information as to the origins of the pollution or to the different types of host sources (human vs. animal, cow vs. deer, etc.) that contribute to the pollution (Reischer et al. 2008; Scott et al. 2003; Walters and Field 2006). Additionally, FIBs have been shown to survive and grow after discharge into receiving water (Bucci et al. 2011; Desmarais et al. 2002; Solo-Gabriele et al. 2000). Compliance with the Clean Water Act (1997, original title: Federal Water Pollution Control Amendments of 1972) and the federal requirements to develop and implement total maximum daily loads (TMDLs) have made fecal source determination an imperative issue in the United States (Layton et al. 2006; Santo Domingo et al. 2007; Simpson et al. 2002; USEPA 2005b). Microbial source tracking (MST) employs methods that attempt to identify specific fecal source groups with the objectives of developing best management practices (BMPs) to control fecal contamination from relevant human/animal sources, protecting recreational-water users from water-borne pathogens, and preserving the integrity of drinking water supplies (USEPA 2005b).

This research will be applied to the Delaware Valley watershed containing natural water resources used by Philadelphia Water Department (PWD) as sources of potable water. This watershed serves Philadelphia County with a population exceeding 1.5 million (U.S.CensusBureau 2010). Unfortunately, this watershed is the recipient of wastes produced by its inhabitants and those from suburban communities. In 2004 Drexel University's Department of Civil, Architectural and Environmental Engineering (CAEE) commenced a PWD-supported research project to develop methods to determine the sources of bacterial pollution in this watershed, focusing on the sub-watershed of the Wissahickon Creek in southeastern Pennsylvania. The research methodology then employed an enrichment/cultivation library dependent (ELD) MST approach using both genotypic and phenotypic components (Pope 2009). The genotypic element involved the use of repetitive polymerase chain reaction (rep-PCR) to amplify, visualize, and fingerprint the DNA of target microbes and build a data library. The phenotypic element used antibiotic resistance analysis (ARA) in determining the antibiotic resistance of target microbes as acquired from their hosts. The microbes targeted for that research were *Escherichia coli* and *Enterococcus* spp. Isolates were obtained from environmental water bodies, wastewater treatment plant (WWTP) effluents, and fecal samples at predetermined sites throughout the watershed. The host species included human (WWTP effluent), livestock, wildlife and domestic animals.

In 2009, I recommended that the focus of the MST protocol change to a non-enrichment library independent (NELI) approach. This strategy is the current trend in the field of MST with the appeal that these assays can be completely performed in a matter of hours, are sensitive, inexpensive (excluding capital cost), quantitative, free from the need for cultivation, and amenable to automation in comparison to ELD MST (Field and Samadpour 2007; Santo Domingo et al. 2007). Thus, the objective of this proposal involves using a NELI MST protocol to track and quantify the FIB prevalence in some of the PWD's resources by using quantitative real-time polymerase chain reaction (qPCR) techniques. The proposed methodology has employed human specific genetic markers for the organisms *Enterococcus faecium* (*E. faecium*) and *Bacteroides* spp. to differentiate human vs. non-human sources of pollution, and ultimately appraise the potential health risk attributed to waterborne pathogens.

**1.2 Ideal Source Performance Criteria**

The ideal indicator of a pathogen's occurrence in water is stipulated to be the pathogen itself (Berg 1978). However, it is highly infeasible to use microbiological assays to screen for every waterborne pathogen due to differences in their incubation times and growth requirements, their intermittent presence in environmental waters, their low population densities when present, and the economic limitations of testing involved. Thus, traditional regulations used to determine microbial health related water quality are based on FIB enumeration. The ideal FIB was originally based on Bonde's (1966) criteria, and is stipulated to have the following attributes: be present whenever a pathogen is present and absent when the pathogen is absent; not proliferate to any greater extent in the aqueous environment than the pathogen; occur in much greater numbers than the pathogen; be easily detected; be more resistant to disinfectants and environmental stresses; and be randomly distributed in samples to be examined. These attributes were refined to continue FIB's relevance to public health protection based primarily on the development and increasing availability of new measurement and detection methods (NRC 2004). The criteria of an ideal FIB now include: the correlation of the indicator with health risks, and its ability to indicate the origin of fecal pollution (NRC 2004). How fecal indicators correlate to pathogens, fecal microbial flora, and aquatic environments requires assessment since different scenarios may lead to different results and interpretations (Figure 1).

Establishing performance standards is essential to evaluate the accuracy of developing and existing MST methods (Stoeckel and Harwood 2007). Methods that fail a certain performance criterion should be discouraged from being used for MST utilization although they may prove useful for research applications (Santo Domingo et al. 2007). There are many criteria to consider for MST studies (Figure 2).

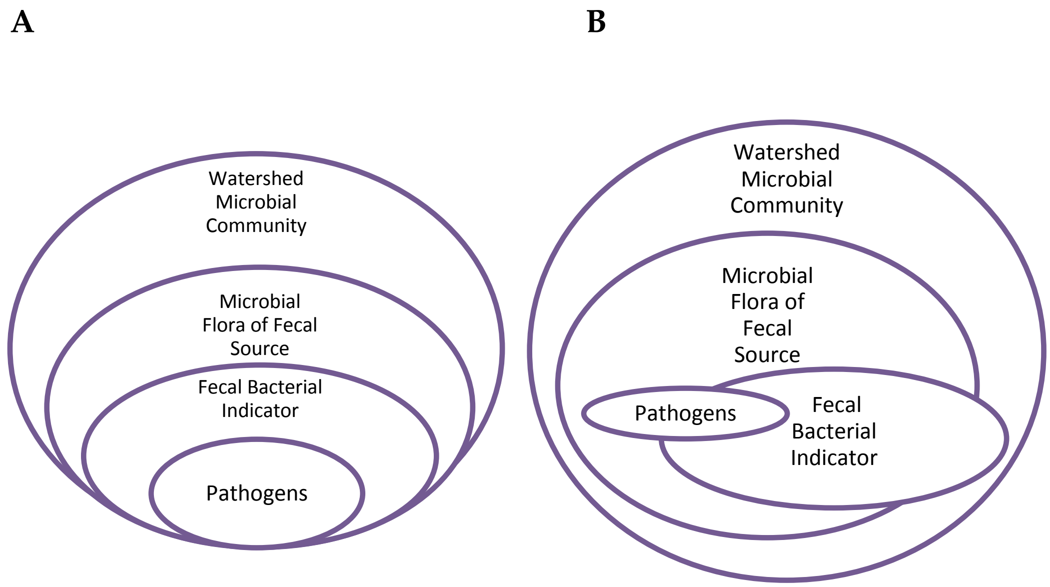



**Figure 1: Venn diagram graphically depicting the distribution of pathogens to fecal distribution in a watershed excerpted from (Santo Domingo et al. 2007).**

**"A" indicates the ideal scenario with all pathogens being a subset of the fecal indicator bacteria (FIB). "B" depicts the scenario that is typical of current assays and leads to problems in correlating FIBs to pathogen.**

Some criteria (e.g. sensitivity, specificity) could be universally applicable and others would depend on the MST method being used. Ultimately, criterion results would determine the relative success of MST methods and their appropriateness as standard methods for utilization for field application. State agencies and other stakeholders with analogous objectives can easily assess the applicability of a MST method if all relevant protocols were evaluated by standardized performance criterions.

The appropriate criteria for a given MST method would be dependent on the objective of a particular study (Santo Domingo et al. 2007; Stoeckel and Harwood 2007). A presence/absence (binary) PCR protocol is adequate to determine the source of fecal microbial pollution, however, inadequate to determine the levels of the pollution. In the latter case a qPCR or metagenomic protocol is required especially if complex watersheds with temporally variable sources are being evaluated.

Many current MST studies fail to include adequate validation, which has led to exaggerated expectations for method performance. The following are sets of performance criterions deemed essential for MST studies (Santo Domingo et al. 2007; Stoeckel and Harwood 2007; USEPA 2005b). The computation involved for these criterions will be discussed in more detail in the statistical chapter.

### 1.2.1 Host Sensitivity

Sensitivity is a diagnostic statistical measure defined as the proportion of true positives that are correctly identified by a test (Altman and Bland 1994). In MST, it is the percentage of positive results of assays to determine the presence of a host-specific target using samples known to contain the host target in question.

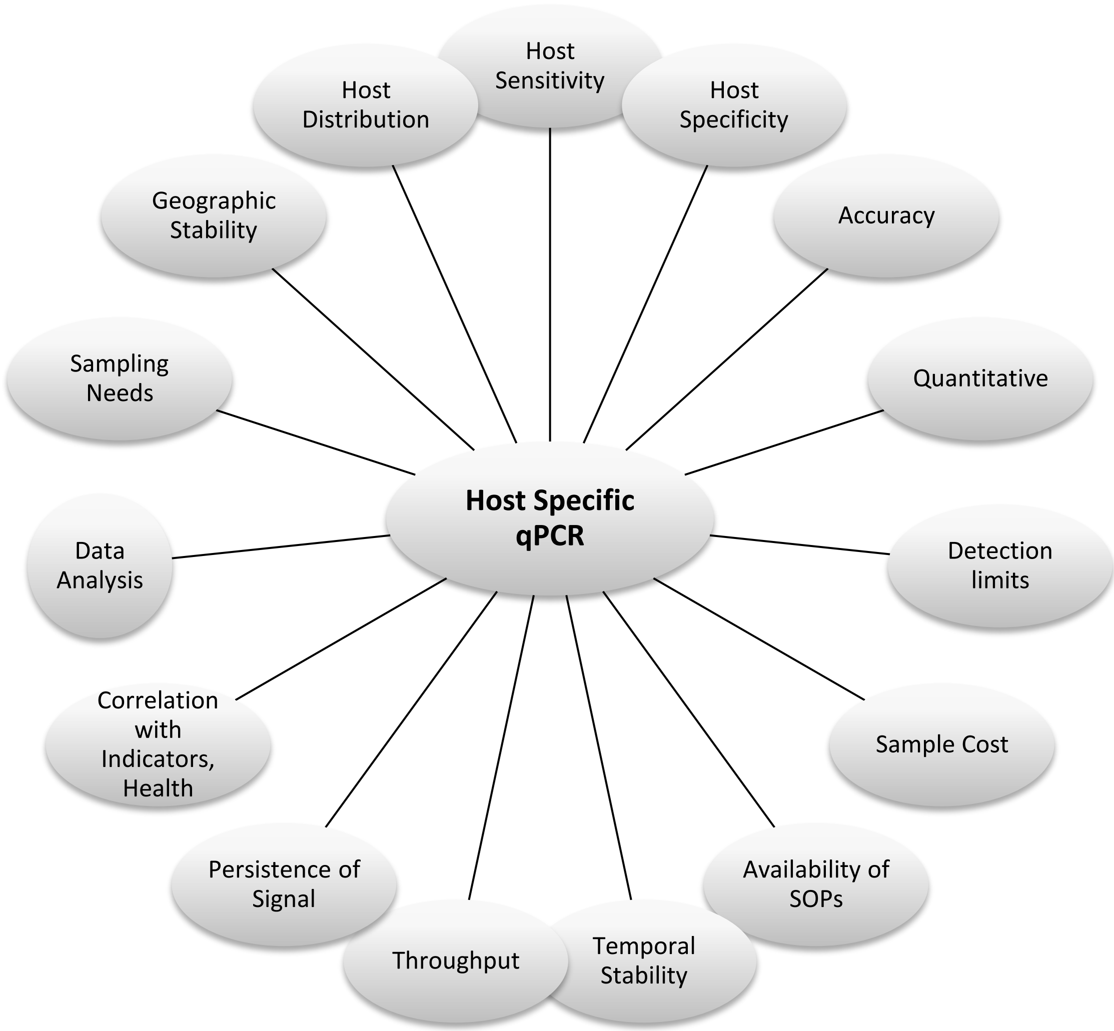



**Figure 2: Diagram showing areas for consideration for qPCR MST research - excerpted from (Santo Domingo et al. 2007).**

**The importance assigned to any issue will depend on method, end user and research perspective.**

It gives an indication of the robustness of an assay providing that the targets are present at levels at or above the detection levels. Direct testing of fecal or spiked samples is used to challenge an assay and to determine its rate of false negatives. An ideal MST assay is anticipated to give results of 100% host sensitivity.

### 1.2.2 Host Specificity

Specificity is a diagnostic statistical measure defined as the proportion of true negatives that are correctly identified by a test (Altman and Bland 1994). In MST, it is the percentage of negative results of assays to determine the presence of a host-specific target using samples known to not contain the host target in question. Ideally, MST assays should not give positive signals (cross react) with non-specific targets. Unfortunately, all MST protocols to date have shown non-specific reactions in some form. This makes it difficult to precisely characterize the source of fecal pollution in environmental samples. Realistically, there may be no genetic marker that is 100% specific or sensitive, however, this is still a worthwhile aspiration.

### 1.2.3 Accuracy

The accuracy of a given host-specific protocol is defined as the proportion of all samples (specific and non-specific) that are correctly classified or identified to have (or not have) the host-specific target (Santo Domingo et al. 2007). Calculated specificities and sensitivities change as MST assays are tested against additional individuals, host species, watersheds, geography and other environs. Thus, conditional probabilistic assessment is important to determine the predictive accuracy of an assay given its sensitivity and specificity values. Bayes' theorem has been used on occasions to assess positive and negative predictive values for MST assays (Kildare et al. 2007; Weidhaas et al. 2011).

### 1.2.4 Host Distribution

Ideally, the genetic markers targeted by a qPCR MST assay should be present in all members of all populations and at approximately similar levels of abundance in all individuals of the target host group (Simpson et al. 2002; USEPA 2005b). The probability of correctly assigning the source of fecal pollution of environmental samples to the actual host source is reduced if the host-specific targets are sparingly and infrequently found in host fecal samples. In cases of low abundance, the assay's detection limit for environmental samples needs to be very low or the samples require concentration of large volumes to increase the odds of detection. In cases where there are inconsistencies in detecting the target in its host fecal sample then it will be difficult to establish relationships between fecal load and the assay's signals.

### 1.2.5 Sample and Method Limits of Detection

The method detection limit (MDL) is the theoretical limit under ideal sample conditions (PCR grade water with no PCR inhibitors) (Santo Domingo et al. 2007).

Theoretically, qPCR assays should be able to detect a few to a single copy of a gene per reaction. Known quantities of a targeted molecular marker are used to estimate the MDL. Plasmids containing the target genetic strand are normally used for this to construct standard curves for MDL determination. Sample limit of detection (SLOD) is an estimate of the detection limit for a sample. In practice, SLOD tends to be higher than MDL due to the carryover of PCR inhibitory contaminants in DNA extracted from complex environmental samples, to a mixture of competing targets, to loss of targets during sample concentration, and to DNA extraction inefficiency. Figure 3 shows a flow chart of a typical qPCR MST protocol. At each of these stages, there is the progressive chance of reduced detection limits.

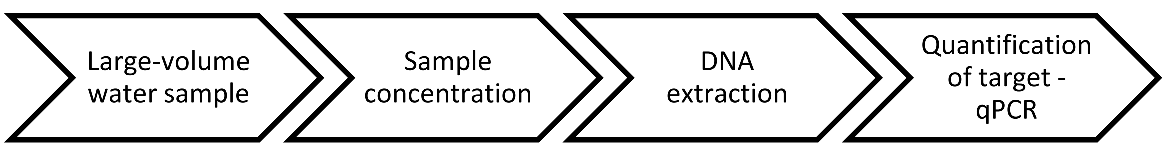



**Figure 3: Flow chart showing overview of sample processing for qPCR MST assay of large water samples.**

### 1.2.6 Temporal and Geographical Stability

Ideally, a host-specific target should be found in its host fecal samples irrespective of temporal, seasonal, regional or geographical differences. This would allow for global applicability of the MST assays. However, forces influenced by a host's diet, health and other intestinal environmental factors may lead to divergent evolution of the targeted microbes that colonizes the host's alimentary system. These impacts are exacerbated by geographic isolation of the hosts (and their microbial inhabitants) and temporal differences in sample collection. Targeting the 16S rDNA gene that is highly conserved in bacteria, in comparison to the rest of the genome, is hoped to minimize the impact of these forces. It is thus appropriate to validate the stability of MST protocols using a large number of samples (known and unknown sources) collected over time from different geographic regions.

## 1.3 Microbial Source Tracking Paradigm

The fundamental assumption of MST is that there exists a traceable association between microbes excreted in feces and their particular hosts, and that the relative proportion of these identifying characteristics/traits are temporally and geographically stable (Field and Samadpour 2007). These host associations could be related to differences in host diet, intestinal environment, metabolic characteristics, symbiotic coevolution, or a combination of these. The USEPA proposed that the ideal MST identifier should: be found only in one host species; be found in all populations of that host species; not be subject to mutation or methodological variability; be stable temporally and geographically; have representative samples; not proliferate and have relative decay rates to pathogens; and be correlated to regulatory water quality parameters and health risks (USEPA 2005b).

Several methods have been proposed that attempt to differentiate sources of fecal pollution in surface waters using different microorganisms (e.g., *Bifidobacterium* spp., *Bacteroides fragilis* phages, *Clostridium perfringens, Escherichia coli*, F-specific RNA coliphages, fecal *streptococci, Rhodococcus coprophilus*, and enterovirus) (Kirs and Smith 2007). However, similar to the fact that no "ideal" indicator organism for the assessment of water quality has been identified, no identifier has yet been recognized as ideal for MST. Thus, there is an active body of research seeking better identifiers for fecal pollution of environmental waters (USEPA 2005b).

### 1.3.1 Enrichment Library Dependent (ELD) MST

Initially MST techniques focused on traditional FIBs with the hope of finding host associations. ELD MST involved using traditional microbiological techniques of cultivation to grow and maintain a collection of FIBs isolated from different known fecal sources. Phenotypic, and to an increasing extent, genotypic techniques are used to differentiate host groups and the data compiled as a library of known associations.

Unknown samples are screened against this library with the use of different statistical computations in an attempt to determine their origin.

Phenotypic methods include antibiotic resistance analysis (ARA): a biochemical technique using antibiotic resistance patterns to identify fecal sources. This protocol is based on the premise that the microbial inhabitants of a particular host acquire antibiotic resistance because of being exposed to antibiotics that are administered to this host. One study used ARA to analyze 1,435 fecal streptococcus isolates obtained from five host species (human, beef cattle, dairy cattle, poultry, turkey) (Wiggins 1996). Samples collected from two pristine streams and two polluted streams were included in the library of isolates. Based on widespread use, five antibiotics (chlortetracycline, halofuginone, oxytetracycline, salinomycin, streptomycin) were then used to evaluate the isolates. Discriminant analyses were performed using different pooled and separated source and antibiotic combinations. In Wiggins (1993), a high average rate of correct classification (ARCC) was obtained when the human group was compared to the pool of all animals (95%) and was highest when the human group was compared solely against the pooled wild animals group (98%).

Wiggins later expanded on his research by creating a merged multi-watershed library of 6,587 enterococci isolates comprising six different Virginia watersheds (Wiggins et al. 2003). Representativeness, the standard type of discriminant analysis, indicated how well the merged library could classify non-library isolates and was determined to be good by resubstitution. Non-library isolates were classified (on average) as well as the merged library isolates by Jackknife statistical analysis, which was used to cross-validate the ARCC obtained from the resubstitution analysis. Only the largest of these six watersheds (2,931 isolates) was determined to be of good representativeness. However, the highest ARCCs were obtained when the smaller sub-libraries were used (65 to 85% - lower than the initial study), but they were much less able to correctly classify non-library isolates. The entire library was temporally stable for at least one year, a finding supported by another ARA study (Ebdon and Taylor 2006).

Ebdon and Taylor attempted to use a library of 2,739 enterococci isolates from 12 fecal source types in the U.K.to predict the origin of 456 enterococci isolates from France, Spain and Sweden. The ARCC for U.K. isolates was 72%; however, it fell to 43% when non-U.K. isolates were used. The authors used Shannon's diversity index to determine that the ideal MST library should contain at least 400-500 isolates per source, an immense size requirement. They concluded that patterns of resistance amongst isolates contained in their U.K. library were not representative of those found in the isolates obtained from the other locations, and that it may be possible that libraries from diverse geographical regions do not share common attributes. This implies the requirement of watershed specific libraries for antibiotic MST studies. The time required for microbial isolation, purification and susceptibility testing, the geographic dissimilarity in antibiotic resistance patterns, and the large library size requirement are clear limitations of the ARA protocol.

One promising ELD MST genotypic technique is repetitive extragenic palindromic sequence based polymerase chain reaction (rep-PCR), a DNA fingerprinting technique that can be used to differentiate microbes of known animals and human sources. One study used *Escherichia coli* (*E. coli*) strains isolated from human and six different animal fecal sources (duck, geese, chicken, pig, sheep and cow), to construct a library comprised of 208 isolates (Dombek et al. 2000). BOX A1R and REP 1R primers were used to amplify the DNA between adjacent repetitive extragenic elements in order to obtain strain-specific DNA fingerprints of these host groups. Jaccard similarity coefficients using fuzzy logic and area-sensitive options were used to analyze the band patterns obtained by gel electrophoresis by dividing the number of bands that occurred in pairs of fingerprints by the total number of bands common and unique to both. Band matching values were gradually decreased with the distance between bands using the fuzzy logic option and differences in area between matching bands were taken into account using the area-sensitivity options. DNA fingerprint relatedness and the successful assignment of isolates to correct source groups were statistically determined by Jackknife analysis.

Dombek et al.'s (2000) study showed host specificity, and was one of the first encouraging signs that molecular techniques could be used for MST. A dendrogram based on the BOX–derived fingerprint data was constructed by using the neighbor-joining clustering method to determine strain relatedness. Discriminant analysis, a statistical clustering technique using multivariate analysis of variance (MANOVA), was used to analyze binary band matching characteristics of the BOX–derived fingerprints assigned to their corresponding host group. It was observed that strains from chickens formed a single cluster, although this group also showed similarity to a pig fingerprint profile. Strains from humans formed four clusters, however, a goose and duck profile clustered with one of the human groups. Most of the members of the other animal groups clustered together, but relationships between groups were also detected.

In 2009, Joanna Pope concluded her thesis based on the PWD-Drexel ELD MST study (Pope 2009). She used both ARA and PCR fingerprinting approaches. ARA was used to analyze 82 *Enterococcus* spp. and 50 *E. coli* isolates obtained from WWTP effluent and animal fecal sample groups (domestic, livestock, waterfowl, and wildlife). Based on local use, four antibiotics (amoxicillin, ampicillin, erythromycin, and tetracycline) were used, with applications at two dosage levels. K-nearest neighbor (KNN) analysis showed the livestock group had the highest RCC values (100% for *E. coli* and 81.8% for enterococci). However, cluster analysis showed no clearly defined clusters/clades and discriminant analysis generated rate of correct classification (RCC) values below the USEPA (2005) recommended 80%. The second aspect of the Pope 2009 protocol used BOX A1R primers to fingerprint 375 enterococci and 315 *E. coli* isolates from WWTP effluent and animal fecal samples (domestic, livestock, wildlife, and zoo). In most cases Jaccard similarity coefficients, cluster analysis, and KNN showed low RCC (<80%) for *E. coli* fingerprints. However, KNN and jackknifing showed promising results for the enterococci fingerprints with all groups generating RCC values greater than 80%.

This Pope (2009) ELD MST method was used to characterize the source of fecal pollution of two small creeks located in Philadelphia at the Schuylkill Center for Environmental Education (SCEE) and the Fox Chase Farm (FCF). The SCEE creek was expected to show fecal impacted primarily from wildlife and/or waterfowl and the FCF creek from livestock animals. However, the results indicated that the predominant sources of fecal pollution at these sites were septic systems with 72% allocation for SCEE and 50% for FCF samples (Figure 4 & 5).

Although the ELD MST protocol using FIBs shows promise, MST comparative studies have indicated that geographical and temporal fluctuations limit the reproducibility of the technique (Simpson et al. 2002; USEPA 2005b). The ability of ELD methods to correctly predict fecal sources is dramatically influenced by the statistical methods used to classify environmental samples (Lasalde et al. 2005; Ritter et al. 2003;

Santo Domingo et al. 2007). Also, a large library size (>1000) , which takes considerable time to construct, is now considered a requirement to improve the rates of correct classification (Johnson et al. 2004).

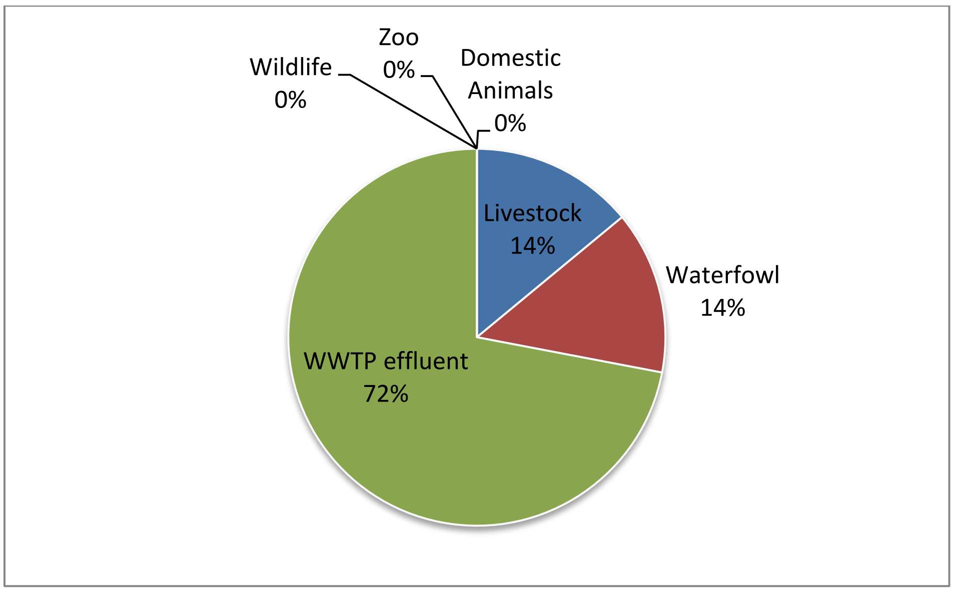



**Figure 4: Pie chart depicting the fecal source allocation derived from the ELD BOX-PCR fingerprints and knn analysis (*k*=1) of enterococci isolates from the Schuylkill Center for Environmental Education (SCEE) water samples (Pope 2009).**

### 1.3.2 Enrichment Library Independent (ELI) MST

In order to increase the specificity of MST protocols it has been recommended to look for genes that code for microbial structures facilitating attachment to cells of host gastrointestinal tract or toxin/virulence specific genes (Scott et al. 2002; Simpson et al. 2002). Improvements in molecular techniques have facilitated this approach with the added advantage of increasing the speed of the protocol and decreasing the expense and tediousness involved with constructing libraries. Focus was also directed towards the easily cultivated FIB, *Enterococcus*, as an alternative to *E. coli*. This bacterial genus comprises anaerobes that are Gram-positive cocci occurring in pairs or short chains with two members, *E. faecium* and *E. faecalis*, most commonly found in human excrement (Gilmore et al. 2002).

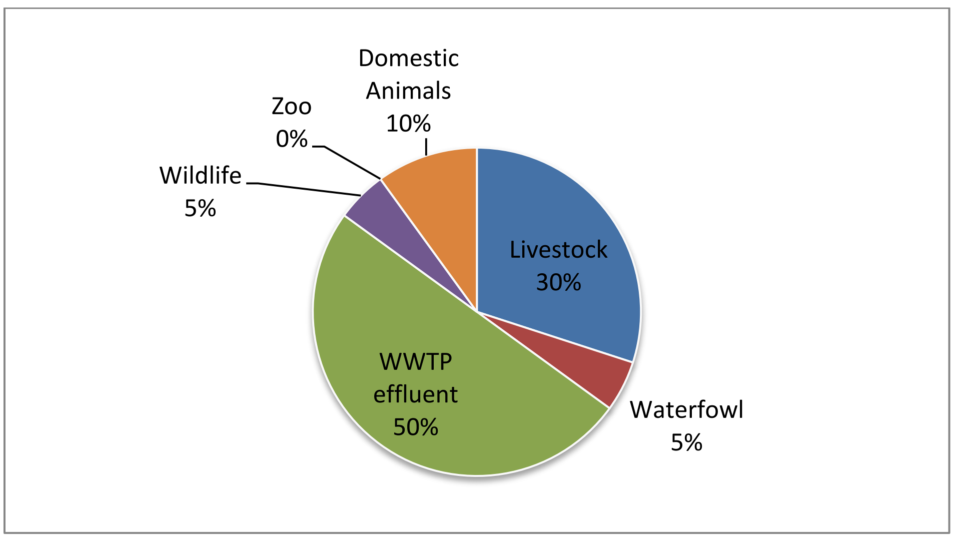



**Figure 5: Pie chart depicting the fecal source allocation derived from the ELD BOX-PCR fingerprints and knn analysis (*k*=1) of enterococci isolates from the Fox Chase Farm (FCF) water samples (Pope (2009)**

One of the more promising ELI MST protocols uses PCR and gel electrophoresis (binary PCR) to screen for the presence of a human specific putative virulence factor, the enterococcal surface protein (*esp*) gene, in *E. faecium* (Scott et al. 2005). A membrane filter cultivation technique is used as an enrichment step to minimize PCR inhibition which tends to occur when environmental samples are concentrated, to increase sensitivity of the PCR assays, to promote growth of stressed or injured cells, and to increase detectability of cells in low concentrations. Scott et al. (2005) analyzed 102 *E. faecium* isolates from 8 animal hosts (poultry, swine, bovine, geese, seagull, pelican, and wild birds) and 65 human (sewage and septic) samples by this PCR approach and showed sensitivity results of 97% and specificity of 100%. On average 58±24 cfu of cultivable *Enterococci* per membrane filter was necessary to detect the $esp_{fm}$ gene. The marker was also shown to be detectable in freshwater samples for a shorter duration (5 days) as compared to the cultivable organism (10 days). The marker and cultivable organisms were found to persist longer in saline water compared to freshwater.

The Scott et al. (2005) protocol has been applied in a number of subsequent studies with varying degrees of success. One study (Whitman et al. 2007) tested 64 human and 233 animal fecal samples and showed 93.1% of the human raw sewage influent samples identified positively. However, no pit toilet waste samples and only 30% of septic truck waste samples were identified positively as of human origin. Specificity values showed that 4.7% of the animal samples (dog and gulls) were identified incorrectly as human sources. Horizontal transfer of fecal bacteria between humans and their companion pets may explain the presence of this marker in dogs (Field and Samadpour 2007; van den Bogaard et al. 2002). Conversely, another study (Ahmed et al. 2009a) screened 32 human and 50 animal fecal samples and classified all samples accurately.

Another study, using biochemical phenotypic fingerprinting with the "PhPlate" system (PhPlate AB, Stockholm, Sweden), analyzed 4,057 enterococci isolates from 10 host groups, supporting the hypothesis that *Enterococcus* spp. exhibits host association (Ahmed et al. 2005). This PhPlate system creates a biochemical fingerprint of each isolate comprised of quantitative data for each bacterial isolate by using microtiter plates to measure the kinetics of cell metabolism. The PhPlate software then calculates the level of similarity between the tested isolates. Simpson's diversity index was used to determine phenotypic diversity, and the similarities between the isolates were calculated as correlation coefficients. Unweighted-pair group method with arithmetic averages (UPGMA) was used to construct clusters. This protocol was able to differentiate human from non-human sources and differentiate different animal sources.

### 1.3.3 Non-Enrichment Library Independent (NELI) MST

Building on the findings of Scott et al. (2005), the *E. faecium* $esp_{fm}$ human specific marker was used to assay 197 fecal samples from 13 host groups in Australia (Ahmed et al. 2008d). The specificity and sensitivity of the maker was determined to be 100% and 90.5% respectively. Ahmed expanded on this study to develop a qPCR SYBR Green protocol to quantitatively determine human sources of *E. faecium* pollution and then applied the protocol to 16 sewage and 16 environmental water samples (Ahmed et al. 2008b). They obtained comparative results to Scott *et.al.* (2005) with cultivable *Enterococci* to $esp_{fm}$ ratios of 16:1 and 21:1 for raw sewage and secondary effluent respectively. Since the *esp* gene is considered a single copy gene, direct conversion of fluorescent signals into gene copies or cell counts is possible. The concentration of the $esp_{fm}$ human specific marker found in 100 mL of the environmental samples was calculated to be in the range of $1.1\times10^{2}$– $5.3\times10^{2}$. The $esp_{fm}$ gene copy concentration computed for sewage ranged between $9.8\times10^{3}$– $3.8\times10^{4}$ per 100 mL.

The need to find microbes that show improved host correlation motivated MST researchers to investigate other fecal-associated microbes for MST. Since most of the members of the fecal microbial flora are anaerobic and difficult to cultivate, the focus on a culture independent approach became necessary. The 16S ribosomal RNA gene of microbes was explored for use as a host specific marker since it is considered to be highly conserved yet contains regions that are variable enough to allow for genus and species identification (Chakravorty et al. 2007). *Bacteroides* spp., enteric anaerobes which are considered to be restricted to warm-blooded animals and belongs to the order Bacteroidales that make up approximately 10 to 60% of human fecal microbial flora, became prime MST candidates (Costello et al. 2009; Franks et al. 1998; Layton et al. 2006; USEPA 2005b).

Kreader (1995) performed one of the first encouraging studies that tested the assertion of *Bacteroides* spp.'s association with human waste. PCR assays using 16S rDNA primers specific for *B. distasonis, B. thetaiotaomicron*, and *B. vulgatus* were used to analyze fecal extracts from 9 humans and 70 animals. This study showed the targeted *Bacteroides* species to be detected rarely in animal excrement, and when detected the levels were very low. Another study built on this research by utilizing length heterogeneity PCR (LH-PCR) and terminal restriction fragment length polymorphism (T-RFLP) to determine *Bacteroides-Prevotella* host associated community profiles of fecal samples (Bernhard and Field 2000a; Bernhard and Field 2000b). These techniques were used to analyze differences in the lengths of gene fragments due to insertions and deletions, with the main objective of designing 16S ribosomal DNA host specific markers of *Bacteroides-Prevotella* to distinguish human from cow fecal pollution. It should be noted that *Prevotella* is a recent taxonomical reclassification of the members of the *Bacteroides* genus that are indigenous to oral cavities (Shah and Collins 1990). The close similarity/relationship of these two groups may have been the reason this assay simultaneously targeted both. Bernhard and Field (2000) used PCR to amplify *Bacteroides-Prevotella* 16S rDNA fragments from 16 fecal samples of healthy human volunteers and 16 cow samples with Bac32F and Bac708R primers. The products for 10 individuals of each host species were then pooled and cloned. These clones were PCR amplified with 6-FAM labeled Bac32F and either Bac303R or Bac708R. The Bac303R amplicons were LH-PCR analyzed. The Bac708R amplicons were first digested with *Hae*III or *Aci*I restriction enzymes before T-RFLP analysis. Five cow specific peaks were observed by both the LH-PCR (2) and T-RFLP (3) techniques, and two human specific peaks were found with the T-RFLP technique.

Bernhard and Field (2000) sequenced clones of the DNA fragments that showed promising host specific LH-PCR and T-RFLP patterns and used Basic Local Alignment Search Tool (BLAST) analysis and DNA sequence alignment to analyze these sequences against 16S rRNA sequences from other members of the *Bacteroides* genus obtained from GenBank. BLAST is one of the most widely applied bioinformatics tools, and is used to find similarities between DNA or protein sequences (Altschul et al. 1990; Zvelebil and Baum 2007). It performs pairwise sequence alignments between the query sequence and a sequence database. The GenBank database at the National Center for Biotechnology Information (NCBI) is publicly available (Benson et al. 2012) and was used by Bernhard and Field (2000) to obtain reference sequences from the *Cytophaga, Flavobacter* and *Bacteroides* groups for manual alignment. Bernhard and Field (2000) then used DNA distance matrix (DNADIST) analysis to calculate evolutionary distance using the Kimura two-parameter model for nucleotide change and a transition/transversion ratio of two. Phylogenetic analysis of the sequence fragments was done using the neighbor-joining method and showed clustering based on host isolates. The Ribosomal Database Project (RDP), which houses one of the largest collections of aligned and annotated 16S rRNA sequences (Cole et al. 2009; Olsen et al. 1992), was then used to compute similarities between sequences and to preliminarily check the specificity of the primers. *Bacteroides-Prevotella* isolates from both human and cows showed clusters that were almost distinct. Interestingly, four sequences (HF8, HF102, HF117 and HF145) were found to be >98.9% similar to each other forming the HF8 cluster depicted by the phylogenetic tree (Figure 6). The HF74 sequence was found to be 93.9% to 94.9% similar to the members of the HF8 cluster.

Three human and two cow specific primers were later developed by Bernhard and Field (2000). Of particular interest is the HF183 human specific primer targeting both the HF74 sequence and the HF8 cluster that have been used with high degrees of success in other MST studies. The HF183 marker has subsequently been determined to occur in the V2 hypervariable region of the 16S rDNA. This region shows one of the highest degrees of sequence diversity of the 16S gene among bacteria and was determined to be the most appropriate site to target in this gene to differentiate bacteria (Chakravorty et al. 2007).

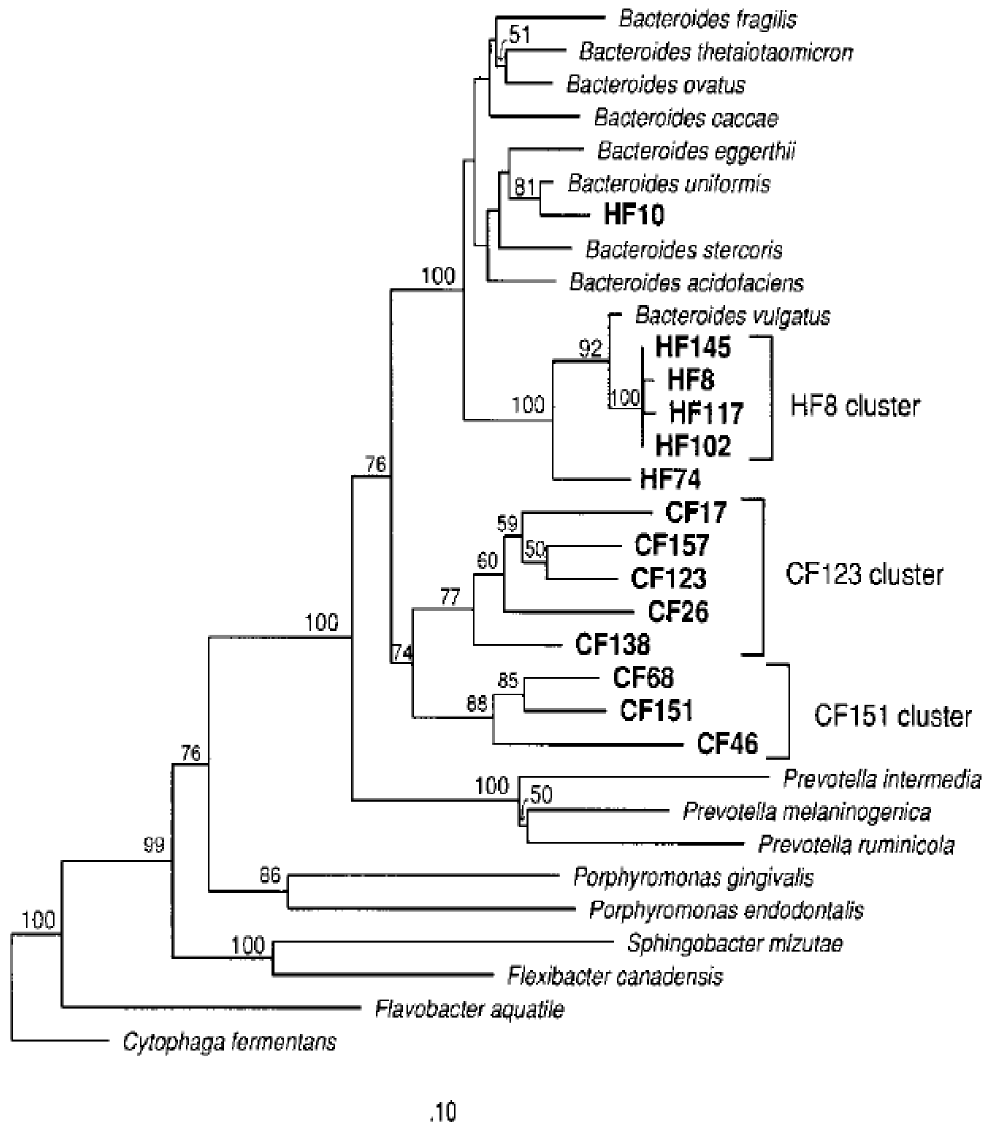



**Figure 6: Phylogenetic tree depicting a human specific (HF8) and two cow specific (CF123 and CF151) clusters of *Bacteroides* 16S rRNA sequences (Bernhard and Field 2000a)**

Building on Bernhard and Field (2000), Fogarty and Voytek (2005) collected fecal samples from nine different animal host species from multiple locations in Virginia, West Virginia and Indiana. *Bacteroides-Prevotella* spp. were isolated and analyzed to determine if unique populations could be obtained by the T-RFLP technique using three enzymes (AciI, MspI, and HaeIII). They successfully amplified 16S rDNA from *Bacteroides-Prevotella* of 48 cows, 20 dogs, 21 deer, 23 horses, 20 humans, and 19 pigs. Amplification of the 16S rDNA from *Bacteroides-Prevotella* of avian host species proved problematic with 14 of the 36 geese, 5 of the 22 chickens, and none of the seagull fecal samples generating amplicons. Jaccard similarity coefficients were calculated and discriminant analysis was performed using the T-RFLP profiles obtained, and a dendrogram was created using UPGMA cluster analysis. The study found that sequences from samples clustered primarily by host groups, and that location did not influence clustering. Isolates from all 20 humans and 3 of the dogs formed one cluster. Most of the cows (41), pigs (15), deer (16) and horses (20) formed different clusters. More similarity was observed within host groups than between groups. Three-dimensional discriminant analysis was also used to illustrate the relationship of the profiles.

In 2009 a comparative study used a qPCR SYBR Green protocol to evaluate different human-specific genetic markers targeting different microbial species against the concentrations of traditional FIBs (Ahmed et al. 2009b). Human-specific *Bacteroides*

(HF183), *E. faecium* (*esp*$_{fm}$), adenovirus (HS-AVs) and polyomavirus (HS-PVs) assays and cultivated *E. coli*, *Enterococci*, and *Clostridium perfringens* were used to compare fecal samples collected from 5 different animal hosts (cattle, pigs, sheep, dogs and ducks). All markers showed 100% sensitivity and specificity except for the HF183, which showed a specificity of 98% due to one dog sample showing a positive result. Serial dilution of samples from freshwater, seawater, and distilled water, all spiked with fresh sewage, were used to determine detection sensitivity of each primer set. The HF183 marker was proven the most sensitive with fecal detection capability 3-4 orders of magnitude better than the other markers, and was detected at dilutions ($10^{-7}$ – $10^{-8}$) that showed no cultivable FIB. Similar to previous *esp*$_{fm}$ studies, this marker was detected in samples that had cultivable enterococci. Detection was observed when the cultivable enumeration was a minimum 71±6 and 40±1 cfu per membrane filter for enterococci and *E. coli* respectively.

In 2009, Ahmed published another study evaluating four different *Bacteroides* genetic markers: HF183, BacHum, HuBac, BacH and Human-Bac. They assayed 186 samples from 11 animal species, including human, and the HF183 marker again proved the most successful by displaying the highest specificity (99%).

A recent comparative study assessed five presence-absence (binary) PCR and 10 qPCR assays that target human specific *Bacteroidales* (Shanks et al. 2010b). The study examined DNA extracted from 54 primary influent sewage samples from different geographical locations in the United States, and 174 samples from 23 different animal species. Of the human specific markers evaluated, only the HF183 and HF134 showed 100% sensitivity when binary PCR was used. None of the binary PCR assays showed 100% specificity. HF183 and HumM19 showed the highest binary PCR specificity of 95%. For the qPCR evaluation, the HF183 earned top grades based on the criteria of sensitivity, specificity, sample limit of detection, and calibration curve correlation coefficient ($R^2$) values.

The literature strongly indicates the possibility of a host association of *E. faecium* and *Bacteroides*, and the potential for their use in MST. However, additional research is required to confirm these relationships and MST suitability (USEPA 2005b). A review of the literature highlights the fact that no single marker or approach, to date, can be considered an ideal MST method, with current recommendations arguing for the application of multiple markers or approaches in a toolbox approach (Schriewer et al. 2010; USEPA 2005b). Sensitivity and specificity analyses of each genetic marker to the target hosts of an investigated region are recommended before using markers for MST (Santo Domingo et al. 2007).

### 1.4 Hypotheses and Objective

Based on the literature, *Bacteroides* 16S ribosomal DNA genetic markers, particularly the HF183 human specific marker, showed very good MST predictive capabilities for human fecal pollution. The *E. faecium* surface protein genetic marker (*esp$_{fm}$*) also showed good results. Thus, these two organisms and molecular markers are good candidates for MST research and a toolbox approach to human vs. non-human differentiation.

My research hypotheses are:

- Human specific genetic markers found in *Bacteroides* and *Enterococcus faecium* will differentiate human from non-human sources of fecal pollution (Hypothesis 1).
- Human specific genetic markers found in *Bacteroides* are geographically stable (Hypothesis 2).
- A toolbox approach, using our human specific molecular qPCR assays and current cultivable FIB tests, will provide better predictive information on the source of pollution than the use of a single protocol (Hypothesis 3).
- A toolbox approach, using multiple molecular qPCR markers targeting multiple hosts, will provide better predictive information of the source of pollution than the use of a single molecular marker or cultivable FIB tests (Hypothesis 4).

My research objective is to develop qPCR MST assays to identify and quantify human microbial pollution in watersheds using genetic markers targeting *Bacteroides* and/or *Enterococcus*. Figure 7 outlines the workflow of the analytical aspect of the research undertaken to achieve the proposed criteria.

**1.5 Scope of research work**

Outlined below are the steps that broadly define the scope of the work in the presented research:

- Sampling strategy – determine the types, locations, frequencies, and collection method for water, WWTP effluent and fecal samples
- Method development - develop a standard operating procedure (SOP) for a qPCR MST protocol using *E. faecium* and/or *Bacteroides*
- Performance criteria assessment - analyze the generated data in a systematic and appropriate manner to determine specificity and sensitivity of the protocol
- Comparative evaluation – in-house assay comparison of the protocol's performance against one assessed to be the most promising in the literature
- Application evaluation – apply MST protocols against environmental water bodies with known and uncertain sources of pollution to assess their applicability

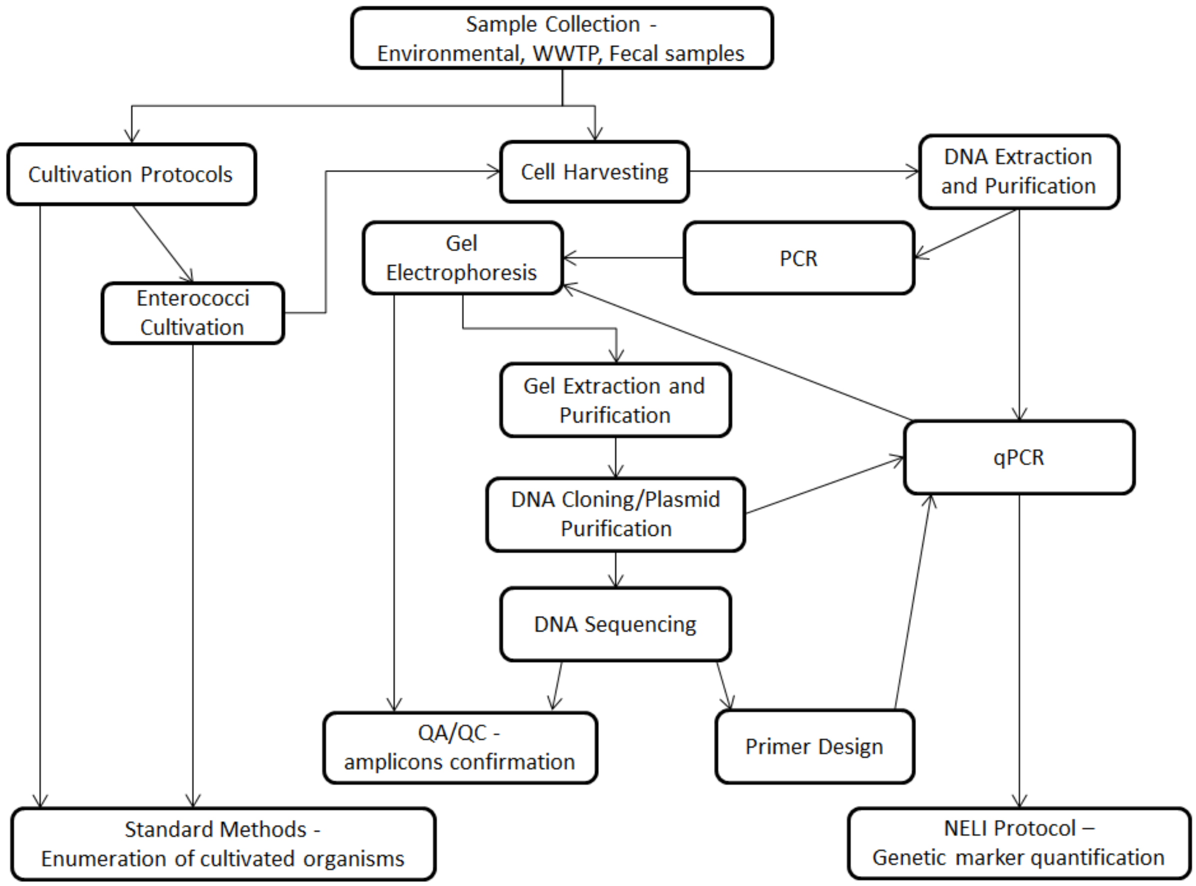



**Figure 7: Flow chart of MST analytical work plan.**

# CHAPTER 2: EVALUATING THE MICROBIAL SOURCE TRACKING ATTRIBUTES OF *BACTEROIDES DOREI*


## Abstract

The central supposition of microbial source tracking is that there exists a distinguishable relationship between microbes excreted in feces and their particular hosts. These associations could be due to adaptations of microbes to their host's gut environment and may involve coevolution between microbes and their hosts. In this study, we performed phylogenetic analyses on 16S rRNA gene sequences associated with different host groups to determine if *Bacteroides dorei* shows co-evolutionary characteristics, thus, host specificity. The results of our analyses provide good support for *B. dorei* human host specificity. Given this finding, a redesigned primer and probe set was developed for a qPCR assay that specifically targets bacterial sequences from a human-specific *B. dorei* lineage. The results from comparative assays against a HF183 protocol obtained from the literature showed similar sensitivity, but an improvement in the specificity of our newly designed protocol. However, assays of rabbit samples showed a high number of positive assays with both the redesigned (58%) and the HF183 (67%) protocol. The newly designed assay could be of considerable use to screen a watershed for human fecal sources of pollution if it is known that rabbit enteric pollution is not impacting it.

**2.1 Introduction**

Microbial source tracking (MST) attempts to identify specific fecal source groups with the objectives of developing best management practices (BMPs) to control fecal contamination from relevant human/animal sources (USEPA 2005b). MST also aims to protect recreational-water users from water-borne pathogens and to preserve the integrity of drinking water supplies. The central supposition of MST is that there exists a traceable association between microbes excreted in feces and their particular hosts, and that the relative proportions of these identifying traits are temporally and geographically stable (Field and Samadpour 2007; USEPA 2005b). These microbe-host associations may arise due to adaptations of the microbe to differences in host diet or the intestinal environment, and further, could represent long-term, coevolved relationships. Such relationships should facilitate MST efforts due to the presence of unique DNA sequences among microbes from particular hosts.

Potential microbial indicators investigated by MST researchers include *Bifidobacterium spp*. (Lynch et al. 2002; Mushi et al. 2010), *Escherichia coli* (Johnson et al. 2004; Scott et al. 2003), *Bacteroides fragilis* phages (Puig et al. 1999; Tartera et al. 1989), *Clostridium perfringens* (Fujioka and Shizumura 1985), F-specific RNA coliphages (Kirs and Smith 2007; Stewart-Pullaro et al. 2006), *Enterococcus* spp. (Scott et al. 2005; Wiggins et al. 2003), enteric viruses (Jiang et al. 2001; Lee and Kim 2002), and *Bacteroides* spp.

(Bernhard and Field 2000a). Also, several laboratory methods have been proposed that attempt to differentiate sources of fecal pollution. Methods investigated include cultivation phenotypic techniques (e.g. antibiotic resistance) (Wiggins et al. 2003), genotypic fingerprinting (Johnson et al. 2004), presence/absence polymerase chain reaction (PCR) (Ahmed et al. 2007; Bernhard and Field 2000b) and quantitative PCR (qPCR) (Haugland et al. 2010; Shanks et al. 2009).

Some of the various techniques previously mentioned are dependent on the construction of extensive libraries of cultivable isolates and others show varying success/failure rates in their specificity and sensitivity dependent on geographical, temporal and/or assay technique differences (Gourmelon et al. 2007; USEPA 2005b). Generally, the more consistent and promising techniques have explored the 16S ribosomal RNA gene of microbes since it contains both highly conserved regions and regions that are variable enough to allow for genus and species identification (Chakravorty et al. 2007). *Bacteroides* spp. have become frequently targeted using culture-independent, 16S based methods (Ahmed et al. 2008b; Bae and Wuertz 2012; Bernhard and Field 2000b; Shanks et al. 2007). These enteric anaerobes are thought to be restricted to warm-blooded animals and make up approximately 30 to 40% of human fecal microbial flora (Layton et al. 2006; USEPA 2005b). *Bacteroides* MST PCR protocols have been relatively effective and protocols using the associated human-specific forward primer, HF183 (Bernhard and Field 2000b), were considered to be better than most (Shanks et al. 2010b). The amplicons produced by different HF183-associated primer sets have been shown to be highly linked with the recently classified species, *Bacteroides dorei* (Bakir et al. 2006; Haugland et al. 2010). Although HF183 protocols have been relatively successful as human specific MST assays, they have also yielded PCR (and qPCR) amplification of *Bacteroides* from other animal hosts (e.g. dog, chicken and sheep) (Ahmed et al. 2009a; Ahmed et al. 2009b; Shanks et al. 2010b). This none-specificity could be exacerbated by the universality of the reverse primers normally paired with the HF183 forward primer, so clearly, there is still room for improvement.

This study seeks to improve the human host-specificity of *Bacteroides* MST protocols. To achieve this, we downloaded the 16S rRNA gene sequences of a human-associated *B. dorei* from the National Center for Biotechnology Information (NCBI) database and used them for BLASTn queries for specific host types. Sequence alignments and phylogenetic analyses were then used to examine trends of host association. Different sets of primers and probes that targeted human specific clades of *Bacteroides* were designed for qPCR assays. The sensitivity and specificity of these protocols were then examined against known host samples, and compared against a published HF183 qPCR assay to determine improvements.

## 2.2 Materials and Methods

### 2.2.1 Sequence identification and phylogenetic analyses

The NCBI database was queried to obtain a *B. dorei* 16S rRNA gene sequences from human feces. One full-length 16S rRNA gene sequence (accession # EU722737.1) was used to perform BLASTn queries that were limited to microbes from particular animal hosts. Host animals were primarily selected based on their high likelihood to contribute to fecal pollution in watersheds located in the Philadelphia area. As such, the host groups selected for evaluation included human, domesticated animals (cat and dogs), livestock (chicken, cow, horse, pig and sheep) and wildlife (deer and goose). To limit our BLASTn searches to microbes from these hosts, each was performed by specifying the host name in the Entrez query option. The Entrez queries were further limited to include (fecal OR feces) requirements. All GenBank entries were examined to ensure that they were isolated from the intended host. To reduce the possibility of outlier sequences, the Ribosomal Database Project (RDP) online tool, which is ideally suited for 16S rRNA gene sequence analysis (Cole et al. 2009) was used to classify each sequence to ensure that they were of the order Bacteroidales with bootstrap support values ≥80%.

The final pooled dataset encompassed 497 different sequences from 46 different studies. The host dataset that was comprised with the least number of discrete studies was that of cow with sequences from two different studies. The human, pig and sheep host datasets were the most diverse, with each comprised of sequences from eight different studies. The final dataset included the top 108 (8 identified as B. dorei) BLASTn results of Bacteroidales classified fecal associated 16S rRNA gene sequences from human hosts and the top 50 for all but two of the other animal hosts. The horse (27 sequences) and cat (11 sequences) datasets provided the exceptions, as RDP classification revealed that many of their top BLASTn hits were not from the genus *Bacteroides*.

The maximum identities and query coverage results for the top BLASTn hits for each host are listed in Table 4. Positively identified *B. dorei* 16S rRNA gene sequences from the NCBI database (all human associated sequences included in our dataset) were named "B_dorei" to facilitate their identification on the generated phylogenies. The 16S rRNA gene sequence of a *Clostridium botulinum* was added to the dataset as an out-group sequence, to root the eventual phylogenetic tree with an organism that was distantly related to all the sequences in the dataset.

After aligning sequences on the RDP website, SeaView4 was used to construct maximum likelihood phylogenies for aligned datasets using the program PhyML (Gouy et al. 2010). The approximate likelihood-ratio test (aLRT) of the SeaView4's GRT model was used to generate each tree, and the proportion of invariable sites was estimated during the run. The starting tree topology was obtained by using the neighborhood joining algorithm for biological dataset (BioNJ) option that takes into account evolutionary distances obtained from aligned sequences (Gascuel 1997). A best of two reasonable tree topology was determined by selecting the option for the "BEST" of Nearest Neighbor Interchanges (NNI) or Subtree Pruning and Regrafting (SPR) topological moves (Guindon et al. 2010). Branch support for the tree topology was determined with the approximate likelihood-ratio test (Anisimova and Gascuel 2006). Bootstrapping with 100 replicate was used to support the nodes of each generated tree.

The resulting phylogenies were uploaded to the interactive tree of life (iTOL) website for annotation (Letunic and Bork 2007; Letunic and Bork 2011). Rooted phylogenetic tree files generated with SeaView4 were used to perform principal coordinate analysis (PCoA) and pairwise UniFrac P-test significance tests assessing the similarity of Bacteroidales microbes from the different host GI environments (Lozupone et al. 2006). The online suite of tools provided by Fast UniFrac enabled the estimation of phylogenetic overlap for gut microbes of different hosts. This, in turn, allowed us to identify phylogenetic patterns consistent with host specificity (i.e. clustering of microbes from the same host species). To accomplish this, all sequences were assigned to a host environment based on their species of origin. PCoA analyses were then performed using UniFrac distances calculated between each pair of hosts. The results were presented as scatter plots. UniFrac P-test statistics were also computed to determine whether bacteria from different hosts shared significantly less phylogenetic overlap than expected by chance.

### 2.2.2 Primer design

The Molecular Evolutionary Genetics Analysis version 5 (MEGA5) software (Tamura et al. 2011) was used to examine the aligned dataset to identify segments that were conserved among human-associated Bacteroidales, but divergent from homologous sites in Bacteroidales from the other animal hosts. These sections of the 16S rRNA gene were designated regions of interest for the design of potential human specific primer sets. The sequences of these regions (~300bp in length) were then uploaded to Roche's universal primer library (UPL) online primer design tool to design potential primer sets (amplicons <100bp in length). The option for pasting sequences for organisms was selected and all other recommended (default) primer 3 settings were used for the Roche UPL tool.

All potential primers and probes were searched against the sequence alignment to determine identical matches using the MEGA5 software motif search option. Primers that predominantly matched human associated bacterial sequences were considered highly favorable for further human specific assay development. PrimerBLAST, RDP's Probe Match (Cole et al. 2003) and NCBI BLASTn analyses were subsequently used to predict the sensitivity and specificity of the primers. PrimerBLAST was also used to test the specificity of the primer oligos against sequences of the targeted organism (*Bacteroides*) in the NCBI nr database. Probe Match analyses allowed us to search for matches between candidate primers and all bacterial 16S rRNA fragments with sequence at the particular gene coordinates (Cole et al. 2005). NCBI's BLASTn queries allowed for the general screening of the primer sequences against all sequences in the NCBI nucleotide collection (nr/nt) database. Primers that match (and thus, that should amplify) the targeted human host bacterial sequences were considered human sensitive; higher numbers of matches to human host sequences were inferred to reveal higher sensitivity. Primers that did not match (and, thus, that should not amplify) sequences associated with non-human hosts were considered human specific: the smaller the number of non-human associated sequence matches, the higher the specificity.

### 2.2.3 Sample collection and DNA extraction

Human-associated influent samples were collected from a wastewater treatment plant (WWTP) that serves a portion of the Philadelphia population. These were used to determine the sensitivity of the human specific assays. Specificity assays were conducted with fecal samples that were collected from different animal hosts, including include cow, sheep, pig, horse, dog and rabbit. These samples were collected over a one year period at different locations throughout Philadelphia. Table 5 shows the number of WWTP influent and different distinct types of animals assayed. All attempts were made to collect fresh samples and DNA extraction was conducted within 6 hours of collection.

All DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's protocol. All extracted DNA samples were stored at -20°C until analyzed by qPCR.

### 2.2.4 qPCR Assays

A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. The protocol for all assays used the Roche LC480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed pre-incubation for 5 min at 95°C; 40 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, and each included no template (negative) controls along with positive controls. Qiagen PCR Cloning Plus kits were used to produce plasmids with inserts of amplicons from assays of WWTP influent samples, which were then used as positive controls.

Our most promising human specific qPCR assay was evaluated with known fecal sources and compared against a human specific HF183 protocol (Shanks et al. 2010b) obtained from the literature that was considered by the authors to be one of the more specific (95%) and sensitive (100%).

### 2.2.5 Sequencing

Amplicons obtained from several of the qPCR assays were also cloned, purified and then submitted to the University of Pennsylvania department of genetics for DNA sequencing (see more below).

### 2.2.6 Specificity and Sensitivity

The sensitivity of a human specific marker is the probability of a positive test result given that the assay is of known human fecal origin (true positive rate):

$$Sensitivity = \quad P(T^{+}|H) = \quad \frac{\alpha^{+}}{\alpha^{+} + \alpha^{-}}$$

Where $\alpha^{+}$ is the number of positive assays of fecal samples of known human origin, $\alpha^{-}$ are the number of negative assays of fecal samples of known human origin, $T^{+}|H$ is a test considered positive given that the samples are all of known human origins.

The specificity of a human specific marker is the probability of a negative test result given that the assay is of a known animal (non-human) fecal source:

$$Specificity = \quad = \quad P(T^{-}|A) = \quad \frac{\beta^{-}}{\beta^{+} + \beta^{-}}$$

Where $\beta^{+}$ is the number of positive assays of fecal samples from known animal origin, $\beta^{-}$is the number of negative assays of fecal samples of known animal origin,

$T^{-}|A$ is a test considered negative given that the samples are all of known animal origins.

**Table 1: Top BLASTn hits to the 16S rRNA sequence of human-associated *B. dorei* (accession# EU722737.1).**

| Animal Host | Organism | Accession # | Query coverage (%) | Max identity (%) |
|---|---|---|---|---|
| **human** | *Bacteroides dorei* | NR_041351.1 | 100 | 100.0 |
| **cat** | uncultured Bacteroidales bacterium | AY695707.1 | 46 | 88.8 |
| **chicken** | uncultured Bacteroides sp. | AY597130.1 | 35 | 89.9 |
| **cow** | Uncultured bacterium | FJ680698.1 | 93 | 97.1 |
| **deer** | Uncultured bacterium | GU198354.1 | 99 | 83.9 |
| **dog** | Uncultured bacterium | FJ978666.1 | 94 | 96.9 |
| **goose** | Uncultured bacterium | FJ390684.1 | 59 | 89.7 |
| **horse** | Uncultured bacterium | FJ493112.1 | 61 | 85.7 |
| **pig** | swine fecal bacterium RF3G-Cel1 | FJ753838.1 | 94 | 96.6 |
| **sheep** | uncultured Bacteroidales bacterium | EU573910.1 | 71 | 88.8 |

## 2.3 Results

### 2.3.1 BLAST Results

The BLASTn results using the *B. dorei* 16S rRNA gene query sequence EU722737.1 showed the bacterial sequences associated with the human host to have very high percentage similarity. All 108 top hits of the human associated group had over 99.0% maximum identity to the query, with a minimum of 99% query coverage. The top cow-associated sequence was the most similar non-human microbe to the *B. dorei* query (97.1% maximum identity). However, similarly high sequence identity was found for the top associated sequences from dog (96.9%) and pig (96.6%), which had similar query coverage compared to the cow-associate (Table 1). The sequences from all other host groups were more distantly related to the *B. dorei* query sequence, with <95.0% max identity, irrespective of query coverage.

### 2.3.2 Phylogenetics

Figure 8 show the phylogenetic tree generated from the 16S rRNA alignment. All of the human-associated sequences related to the focal *B. dorei* strain, grouped into a monophyletic clade (human-exclusive cluster) with 99% bootstrap support. The other animal host sequences of the *B. dorei* dataset also tended to form host-associated clusters. However, humans were unique in that all of their sequences grouped into one host-specific clade.

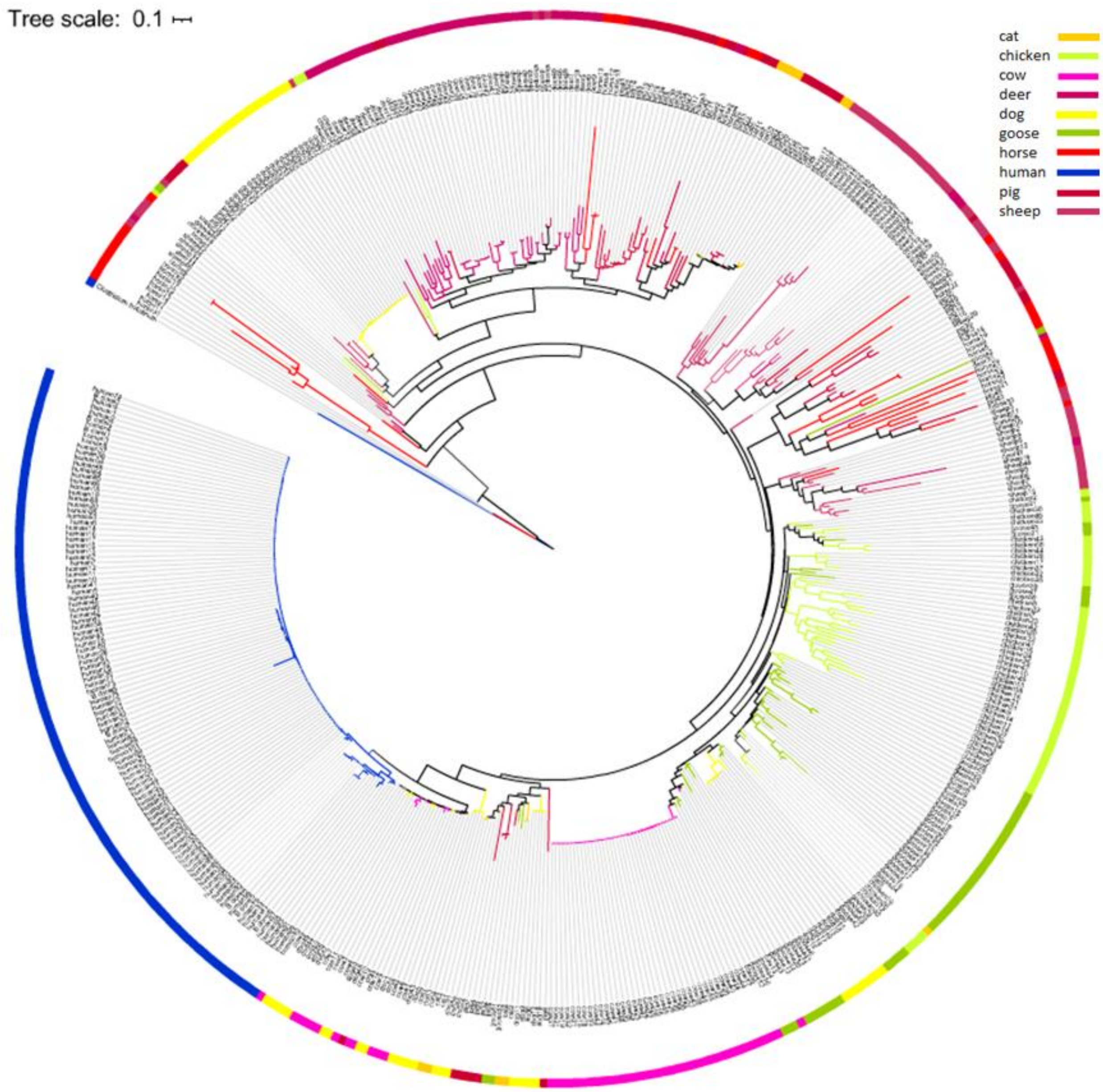



**Figure 8: 16S rRNA maximum likelihood phylogeny of *Bacteroides dorei* and relatives.**

**The tree was constructed from an alignment of the top BLASTn hits of *B. dorei* 16S rRNA sequence EU722737.1. Sequences used to construct the tree included ten**

**identified *B. dorei*, the top 108 hits from human hosts (8 identified as *B. dorei*) and the top 50 hits from each chicken, cow, deer, dog, goose, pig and sheep hosts. Only the top 27 horse and 11 cat BLASTn hits were used since most BLASTn hits from these organisms did not classify to the order Bacteroidales. *Clostridium botulinum* was used as the out-group to root the tree. Branches and color strips illustrate the host affiliation of each of the analyzed sequences, revealing a clear trend of host specificity for human associates (blue). Branch lengths of the tree are drawn to scale.**

UniFrac P-test significance analyses (1000 permutations) revealed highly significant differences (p-value <0.001) between human and all non-human-associated sequences, indicating that the clustering seen for human-associated *B. dorei* relatives was non-random. A similar result was obtained when the p-values were adjusted for multiple comparisons using Bonferroni correction. Unweighted UniFrac PCoA of the ten different host environments showed dissimilarity between sequences from the human and those of other non-human communities using the first three principal coordinate axes (Figure 9). However, collections of *B. dorei* relatives from dogs and cows clustered with those of humans, reflecting the fact that several of the sequences from these animals were immediate outgroups to the human-associated *B. dorei* clade (Figure 8).

### 2.3.3 Primer Design

The *B. dorei* 16S rRNA alignment showed two regions that were conserved for the human host group and variable for the other animal host groups. These regions were designated regions of interest for human specific primer design. The HF183 human specific forward primer from the Bernhard and Field (2000) study was found to be located in one of these regions. The other region was located upstream of the HF183 forward primer. A ~300bp section of the query sequence that contained these two regions of interest was selected for human specific primer design.

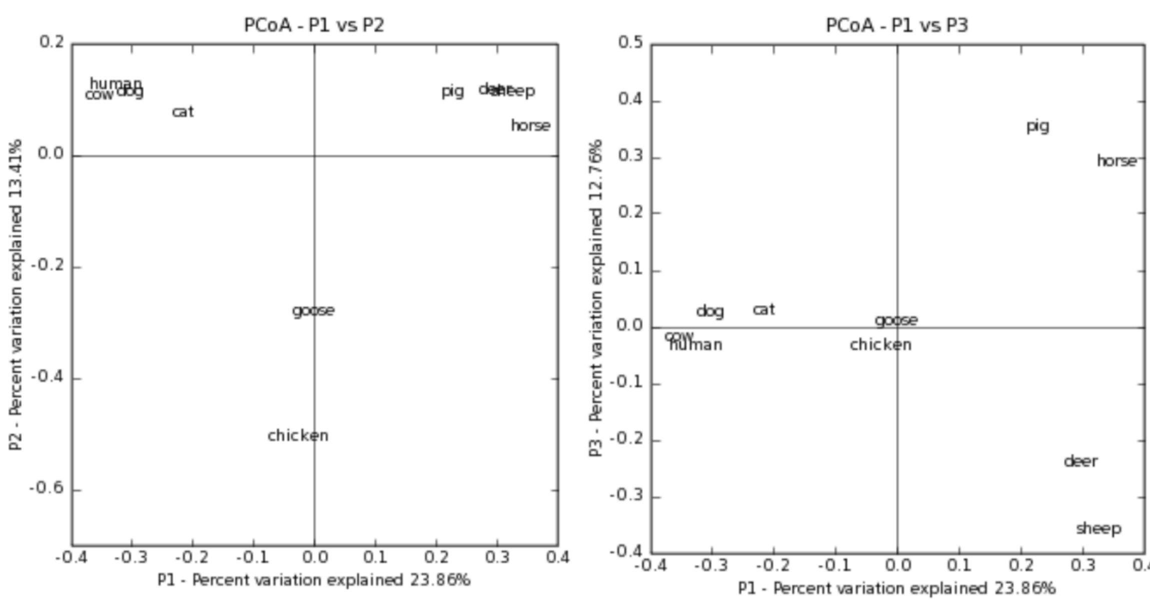



**Figure 9: Unweighted UniFrac principal coordinates analysis (PCoA) of Bacteroidales microbes from various animal hosts.**

**Unweighted UniFrac was used to generate a matrix of pairwise distances between collections of top BLASTn hits to *B. dorei* from different vertebrate hosts. A scatterplot was then generated from the matrix of distances using Principal Coordinate Analysis. The plots show the first three principal coordinates axes. Human associated sequences with relatedness to *B. dorei* show little similarity (no overlapping) to most non-human associated sequences related to this microbe, except for those from cow and, to a lesser extent, dog and cat.**

The online Roche UPL primer tool generated a number of potential human host specific qPCR primer-probe sets. Of interest, the tool returned a primer-probe set that produced amplicons of ~ 72bp in length. The sequences of the oligos in the primer-probe set included the forward primer HF68 (5′-GGC AGC ATG GTC TTA GCT TG-3′), the reverse primer of HF119 (5′-GGC AGG TTG GAT ACG TGT TAC-3′), and an internal probe (designated as Roche UPL probe #156) of 5′-GCT GAT GG-3′. The numbers in the primer names referred to nucleotide coordinates of the 5′ ends of our primers based on the reference *E. coli* sequence (accession # J01695.2). The MEGA5 motif search of the aligned *B. dorei* 16S rRNA dataset using HF68 revealed all but two of the 108 human-associated sequences of the *B. dorei* dataset to have exact matches. However, homologous segments from microbes of dogs (7/50), chickens (7/50) and cows (12/50), were also identified as perfect matches for this forward primer, indicating possible non-specificity. HR119 showed less host-specificity, as it matched with at least one sequence from seven of the host groups investigated. RDP's probe match and NCBI's BLASTn queries both supported this non-specificity (data not shown). However, incorporation of the UPL probe # 156 increased the specificity of the HF68-HR119 PCR assay. Motif searching showed all of the human-associated sequences had homologous segments to the UPL probe #156 (except for two of the identified *B. dorei* sequences, which did not contain sequence information in this region). The probe was also an exact match to one dog-, pig- and chicken-associated sequence.

The HF183 primer (5'-ATC ATG AGT TCA CAT GTC CG-3') obtained for the Shanks et al. 2010 study also showed all but two human-associated sequences were exact matches, while no sequences from microbes of other hosts were exact matches to this primer. The reverse primer BthetR1 (5'-CGT AGG AGT TTG GAC CGT GT-3') and the probe [6~FAM] CTG AGA GGA AGG TCC CCC ACA TTG GA [TAMRA~6~FAM] for this protocol both showed exact matches to most of the sequences from all host groups, suggesting potential limits to the specificity of this assay.

**2.3.4 qPCR**

Based on our bioinformatics analyses suggesting the specificity of the HF68 and HF183 primers, we paired the HF68 forward primer with the reverse complement of the

HF183 primer (HR183rc) in a redesigned protocol. This HF68-HR183rc primer set was used with the UPL probe#156 for a qPCR assay. This protocol was compared with the HF183-BthetR1 qPCR protocol and the results are shown in Table 2. Both protocols showed 100% sensitivity for WWTP influent samples with gene copies/mL of sample averaging an order of magnitude ~1.0E8. Overall, the redesigned protocol showed better specificity (72%) than that of the HF183-BthetR1 (55%). The specificity results were dramatically impacted by assays of rabbit fecal samples. Assay of rabbit samples showed high occurrences of the markers with the HF68-HR183rc protocol showing 58% of the samples positive and the HF183-BthetR1 protocol showing 67% positive. The maximum concentration of the markers in the rabbit samples reached an order of magnitude of ~1.0E9 gene copies/g of sample using both protocols.

To follow up on cross-amplification of rabbit-associated microbes, six of the extracted rabbit DNA amplicons were sequenced (GenBank Accession #'s JX889727-JX889732). Bioinformatics tools were used to further analyze these sequences to determine similarity with the reference B. dorei sequence and relatives obtained from NCBI. A smaller phylogenetic tree (Figure 10) was constructed using the top 5 BLASTn hits for each animal host as previously outlined. This new tree included the six rabbit sequences from two distinct animals. The results showed the isolated rabbit sequences to be highly similar to those of the human sequences, grouping into the same clade.

**Table 2: Results of two human-targeted *Bacteroides* qPCR assays on Waste Water Treatment Plant and animal fecal samples.**

| Animal type | # of distinct samples assayed | # of distinct samples positive with HF68-HR183rc | # of distinct samples positive with HF183-BthetR1 |
|---|---|---|---|
| **WWTP influent** | 20 | 20 | 20 |
| **cow** | 20 | 0 | 2 |
| **sheep** | 12 | 0 | 0 |
| **pig** | 10 | 0 | 6 |
| **horse** | 12 | 0 | 0 |
| **rabbit** | 24 | 14 | 16 |
| **dog** | 10 | 0 | 0 |

The tree was constructed from an alignment of the top BLASTn hits of *B. dorei* 16S rRNA sequence EU722737.1. Sequences used to construct the tree included ten identified *B. dorei,* the top 10 hits from human hosts (8 identified as *B. dorei*) and the top 5 hits from each chicken, cow, deer, dog, goose, pig, horse, cat and sheep hosts. Also include in the tree were 6 rabbit-associated sequences isolated during this study. *Clostridium botulinum* was used as the out-group to root the tree. Branches and color strips illustrate the host affiliation of each of the analyzed sequences, revealing relatedness of human and rabbit associates. Branch lengths of the tree are drawn to scale and bootstrap support values >50 are depicted.

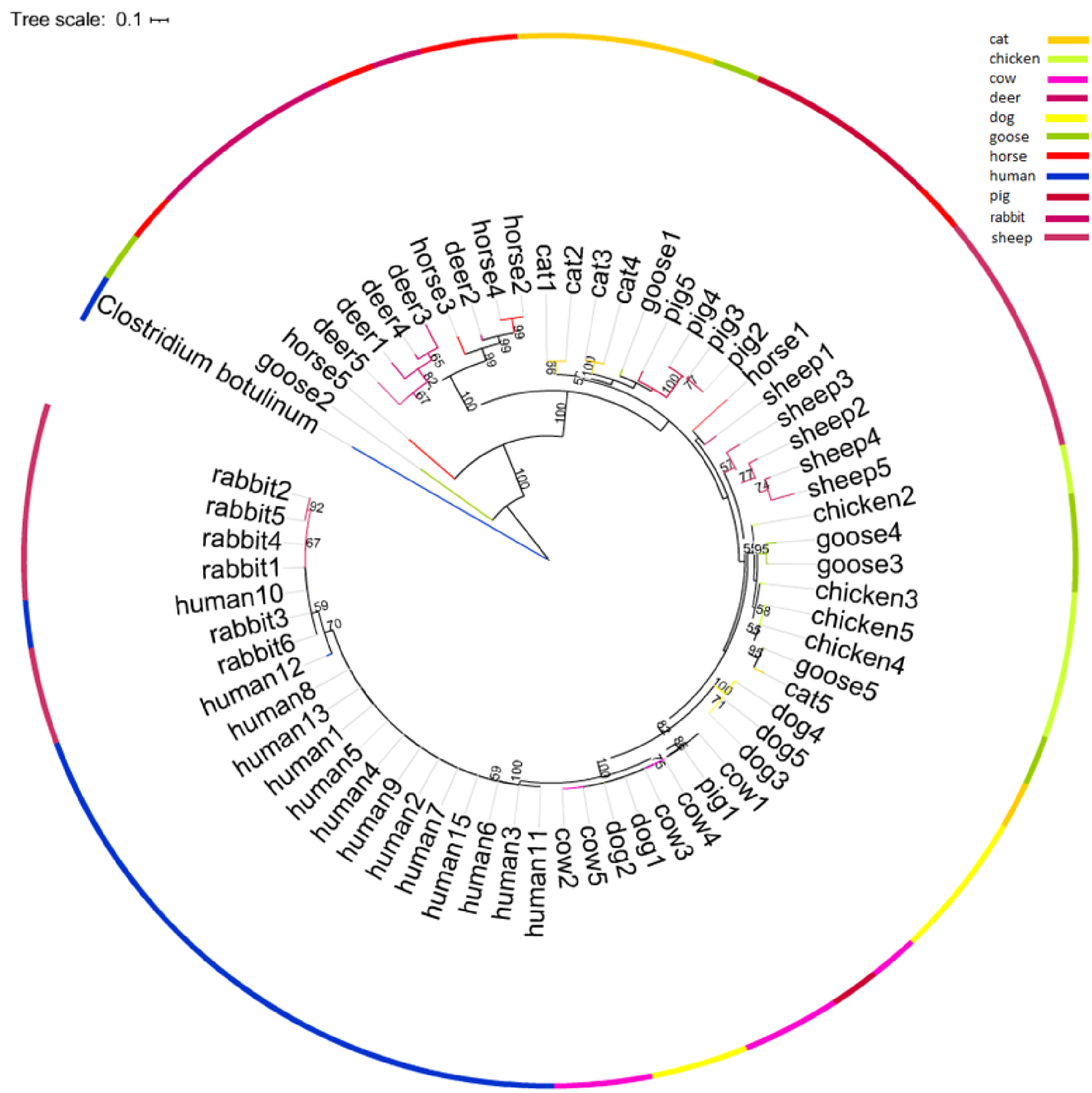



**Figure 10: 16S rRNA maximum likelihood phylogeny of *Bacteroides dorei* and relatives, including sequenced rabbit DNA samples.**

**The tree was constructed from an alignment of the top BLASTn hits of *B. dorei* 16S rRNA sequence EU722737.1. Sequences used to construct the tree included ten**

**identified *B. dorei*, the top 10 hits from human hosts (8 identified as *B. dorei*) and the top 5 hits from each chicken, cow, deer, dog, goose, pig, horse, cat and sheep hosts. Also include in the tree were 6 rabbit-associated sequences isolated during this study. *Clostridium botulinum* was used as the out-group to root the tree. Branches and color strips illustrate the host affiliation of each of the analyzed sequences, revealing relatedness of human and rabbit associates. Branch lengths of the tree are drawn to scale and bootstrap support values >50 are depicted.**

### 2.4 Discussion

Numerous MST studies have investigated *Bacteroides* spp. as prime candidates for differentiating host groups using PCR-based techniques (Ahmed et al. 2009a; Ahmed et al. 2009b; Bernhard and Field 2000a; Bernhard and Field 2000b; Haugland et al. 2005; Haugland et al. 2010; Santo Domingo et al. 2003; Scott et al. 2005; Shanks et al. 2010b; USEPA 2005b). Yet some of these studies have shown cross-reaction of the human-specific assays with known non-human associated samples. These “false positive” results could lead to unnecessary mitigating actions and thus indicate the need to find improved MST protocols. The main aim of this study was to investigate phylogenetic relationships between sequences associated with different host animals, and to use the results to design a molecular survey for human fecal bacteria. Through focusing on clades of human-specific bacteria, we aimed to design a human-MST assay with improved specificity.

The BLASTn analysis of the *B. dorei* dataset indicated relatively good differentiation between the human and most non-human host associated bacterial sequences (max identity <98.0%; Table 4) with similar findings in our phylogenetic analysis. The existence of a human-specific clade, inclusive of all identified *B. dorei,* supports the assertion that this organism is primarily human-associated (Bakir et al. 2006; Haugland et al. 2010). Most sequences in our dataset tended to form host-specific clusters, although none was as exclusive as the trend seen for humans. UniFrac P-test significance analysis supported the inference that the sequences associated with the human GI environment are different from those of non-humans. It therefore appears that this trend is non-random and indicative of host-specificity.

Studies of different animal hosts have shown factors such as geography, diet, and relatedness appear to shape the composition of symbiotic gut communities (Hooper and Gordon 2001; Ley et al. 2008; Russell et al. 2009). Other studies have indicated potential coevolution between mammals and their gut microbes (Ochman et al. 2010). Our findings support the hypothesis that there exist host specific relationships between gut-associated *Bacteroides spp.* and their animal hosts, with *B. dorei*, in particular, showing host specific relationships with humans (Haugland et al. 2010). *Bacteroides* spp.

are obligate anaerobes and so it is possible that *B. dorei* co-evolved with humans due, in part, to limited possibilities for transfer between human and non-human hosts.

The resulting qPCR protocol of this study could prove beneficial in attempting to differentiate human from non-human hosts. Bioinformatics analyses of the primer and probe oligos associated with reported *Bacteroides* HF183 protocols, including that of Shanks et al (2010), showed only the forward HF183 primer to be human specific. The universal nature of the reverse primers and the probes of reported HF183 protocols could be contributing to the non-human qPCR signal observed, as has already been shown in some studies (Haugland et al. 2010; Shanks et al. 2010b). Combining the HF68 primer with the reverse complement of HF183 showed improved human specificity with no reduction in sensitivity when compared to the HF183-BthetR1 protocol. This is postulated to be due to the non-specific, almost universal, nature of the BthetR1 primer and the companion probe used in that protocol. The HF68 primer and the probe#156 both indicated increased selectivity via bioinformatics analysis. These theoretical trends were supported by the empirical findings in the assays conducted in this study.

The bioinformatics results indicated that bacteria species highly similar to *B. dorei* could also be human host specific (Haugland et al. 2010). However, this postulate is questioned since assays using both of the protocols showed very high concentrations and occurrences of the human-like *Bacteroides* in fecal samples from rabbit.

The fact that both qPCR protocols used in this study showed the propensity to amplify *Bacteroides* associated with both human and rabbit fecal pollution does not eliminate their use for MST applications. The newly designed assay could be used to screen a watershed for human fecal sources of pollution if it is known to not be impacted by rabbits, and vice versa. Alternatively, a rabbit specific marker could be explored to augment the process of host/source differentiation. The newly designed assays could be used as part of a multi-tiered molecular approach to pinpoint sources of fecal contamination.

Future work will include the continued evaluation of the new qPCR protocol for its potential applicability. It should be noted that no single primer set may ever be ideal to determine host pollution, thus a toolbox approach is supported. We advocate that the *B. dorei* human specific markers be one of the instruments in this toolbox to more accurately determine human sources of pollution.

# CHAPTER 3: INVESTIGATING CO-EVOLUTIONARY ATTRIBUTES *ENTEROCOCCUS FAECIUM* AND THEIR MICROBIAL SOURCE TRACKING POTENTIAL

## Abstract

Several methods have been proposed to differentiate sources of fecal pollution in surface waters using different microorganisms. One promising protocol used the enterococcal surface protein ($esp_{fm}$) gene, from *Enterococcus faecium*. Both the *esp* and the 16S rRNA genes of *E. faecium* were evaluated as possible targets in this study. A very low sensitivity was obtained for assays conducted on *E. faecium* isolates, which improved only when an enrichment protocol was included. *E. faecium* 16S rRNA phlologenetic and bioinformatic analyses indicated human and non-human host associated bacterial sequences had little evolutionary divergence. This study showed limitations in using *E. faecium* 16S rRNA and *esp* genes in differentiating the hosts of fecal pollution.

**3.1 Introduction**

Agencies and utilities were required to assess point and non-point source pollution impacting watersheds by the amendment of the Clean Water Act (CWA) 33 USC § 1251 et seq. (1972), the Water Quality Act (1987), and with the implementation of total maximum daily load (TMDL) standards (CWA § 303 (d)). Sources of fecal pollution are often contested, thus, microbial source tracking (MST) attempts to identify and differentiate fecal source groups. MST primarily investigates microbe that inhabits the gastrointestinal tract of animals and excreted in feces. The aim is to identify traits of these microbes that enable the differentiation of their host animals (Field and Samadpour 2007; USEPA 2005b). These microbe-host traits/associations could be related to adaptations of the microbe to differences in host diet and/or intestinal environment and further, could represent long-term, coevolved relationships.

Several methods have been proposed to differentiate sources of fecal pollution in surface waters using different microorganisms (e.g., *Bifidobacterium spp.*, *Bacteroides fragilis* phages, *Clostridium perfringens*, *Escherichia coli*, F-specific RNA coliphages, fecal streptococci, *Rhodococcus coprophilus*, and enterovirus) (Kirs and Smith 2007). One study proposed a promising protocol which used polymerase chain reaction (PCR) and gel electrophoresis to screen for the presence of a human-specific putative virulence factor, the enterococcal surface protein (*esp*$_{fm}$) gene found in *Enterococcus faecium* (Scott et al.

2005). The Scott et al. (2005) study used a cultivation method to first enrich for *Enterococcus* spp., and then used the human-specific $esp_{fm}$ marker to differentiate human from animal fecal pollution. This protocol suggested the $esp_{fm}$ marker to have high specificity and sensitivity to its human DNA target. One study successfully modified the Scott et al (2005) protocol to a SYBR Green qPCR protocol (Ahmed et al. 2009a). However, other studies had varied success, with one study obtaining positive results (false positives) with a number of other animal fecal samples, and failed detection (false negatives) of human fecal pollution in septic truck and pit toilet samples (Whitman et al. 2007). Another study has shown the $esp_{fm}$ gene to not be present in all human fecal associated *E. faecium* organisms (Kim et al. 2010), which could explain the reduced sensitivity.

This study used two different approaches to determine the potential of using *E. faecium* for MST purposes. Both the *esp* and the 16S rRNA genes of *E. faecium* were evaluated as possible targets. Firstly, the Ahmed et al. (2009) SYBR Green qPCR protocol using primers targeting the $esp_{fm}$ gene was evaluated and compared with a probe-based qPCR assay developed during this study. Secondly, this study tested the postulate that there exist host-specific relationships of *Enterococcus* spp. and their animal hosts. To achieve this, we downloaded 16S rRNA gene sequences of human-associated *E. faecium* from the National Center for Biotechnology Information (NCBI) database and used them for BLASTn queries for specific host types. Sequence alignments and phylogenetic analyses were then used to examine trends of host association.

## 3.2 Materials and Methods

### 3.2.1 Gene Analysis - $esp_{fm}$

#### *3.2.1.1 Library of Isolates*

The *esp*$_{fm}$ protocol was first tested against a library of *E. faecium* isolates stored at the Drexel MST facility, yielding a measure of the specificity and sensitivity of this marker. The library was comprised of 15% glycerol stocks of 297 *E. coli* and 353 enterococci isolates in 1 mL Tryptic Soy Broth (TSB) stored at -80$^{o}$C in 2ml cryogenic storage vials (Fisher). From the library, 157 isolates were randomly selected for both PCR-gel electrophoresis and SYBR Green qPCR MST analyses. This set of isolates included 73 human-associated WWTP, 56 animal fecal, 9 environmental water and 13 ATCC isolates. The animal fecal set was comprised of 1 gull, 6 woodchuck, 30 goose, 5 dog and 14 bovine isolates.

Isolates were inoculated into 1.0 mL of TSB in 1.5 mL microcentrifuge tubes and incubated at 37$^{o}$C for 18-24 hours @ 500 rpm in a Thermomixer (Eppendorf). The cells were harvested by centrifugation at 20,000 rpm (Eppendorf centrifuge 5430 R). DNA extractions were conducted with DNeasy Blood & Tissue kits (Qiagen) in accordance with the manufacturer's Gram positive bacteria spin-column protocol.

***3.2.1.2 Sample collection***

Human-associated samples were collected from the Northeast wastewater treatment plant (WWTP) on February 24, 2010 and were used to determine the sensitivity of human specific *esp$_{fm}$* assays. Duplicate 500 ml samples were collected from influent, secondary treatment (mixed liquor effluent) and effluent (pre and post chlorination) points on both occasions. All samples were transported in coolers on ice at ~ 4$^{o}$C until processing (≤ 6 hr.).

***3.2.1.3 DNA extraction***

An enrichment and a non-enrichment step were evaluated to determine the necessity of cultivation in the MST protocol used to analyze collected samples. The original *esp$_{fm}$* protocol required an enrichment step before DNA extraction, which increased assay time and prevented accurate quantification. For the no-enrichment protocol, DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's instructions. A modification to the manufacturer's instructions was used for the enrichment protocol. In brief, the conventional mEI enterococci enumeration method (USEPA 2005a) was used for enrichment. MFs were then placed in 2.0 mL safe lock microcentrifuge tubes (Qiagen) and 1.8 mL ASL buffer added. Samples were vortexed for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90$^{o}$C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAcube QIAamp DNA Stool protocol was followed is as is outlined in the kit's manual. All extracted DNA samples were stored at -20$^{o}$C until analyzed by qPCR.

#### *3.2.1.4 SYBR Green qPCR Protocol*

A SYBR Green *E. faecium* $esp_{fm}$ protocol was initially used for assays in the developmental phase of the MST project (Ahmed et al. 2009a). This prtocol was compared against a probe based version designed by this study. The original primer set included the *E. faecium* $esp_{fm}$ 5′– TAT GAA AGC AAC AGC ACA AGT T– 3′ (forward) and 5′– ACG TCG AAA GTT CGA TTT CC– 3′ (reverse) primers that generate amplicons of ~680 base pairs in length.

Reaction mixtures of 20µL consisted of Roche LightCycler 480 SYBR Green I Master mix; 300 nM of each forward and reverse primers. Reducing the volume of the DNA template from 5 µL to 2 µL assisted in the reduction of PCR inhibitory substances found in the complex environmental samples being investigated. The assays were conducted using a LightCycler 480 (Roche) and the PCR program was optimized to increase reaction specificity and reduce procedure time.

The modified amplification program employed a touchdown protocol of: pre-incubation for 10 min at 95$^{o}$C; 40 amplification cycles of 10 s of melting at 94$^{o}$C, 10 s of annealing starting at 65$^{o}$C (with a decrease of 0.5$^{o}$C at each cycle), extension of 30 s at

72$^{o}$C; melt curve analysis from 53$^{o}$C to 97$^{o}$C; and finally cooling for 10 s at 40 $^{o}$C. This protocol reduced the time required from 3 hours in the original assay to less than 1 hr. and increased the specificity of the primers. Nuclease-free water was used for no-template controls and DNA samples from clones of human specific *E. faecium* isolates obtained from Michigan State University (MSU) were used as positive controls for this PCR assay.

Samples were considered positive when the melt curve analysis profile of samples matched that of the positive control and had a similar melting temperature ($T_m$). The $T_m$ is the temperature at which half of the double stranded DNA in the SYBR Green PCR reaction well disassociates, becoming single stranded. A maximum crossover point (Cp) was also used to indicate positive results within the dynamic range of the standard curve. The Cp is the point at which the fluorescence of a sample rises above the background fluorescence and is dependent on the amount of target that is present at the beginning of the reaction. The higher the initial target concentration, the lower the Cp value.

#### *3.2.1.5 Additional QA/QC*

In addition to having positive and negative controls, Gel electrophoresis (2% agarose gels) was used to confirm the identity of amplicons generated by the qPCR protocols. Representative amplicons (DNA fragments) of the appropriate band lengths for each primer set were extracted and the DNA was purified using Qiagen's QIAquick Gel extraction kit. Purified amplicons were submitted to the University of Pennsylvania (UPenn) Department of Genetics DNA Sequencing Facility for sequencing as another means of quality control. The ribosomal database project (RDP) web tool was used to positively classify (confidence threshold 80%) the 16S rRNA gene sequences of the amplicons as *E. faecium*.

#### *3.2.1.6 Probe qPCR Protocol*

A probe based qPCR protocol was developed to replace that of the SYBR Green. This modification increased the selectivity and specificity of the MST protocol and reduced background, non-specific PCR products.

The original $esp_{fm}$ primer set was redesigned in an attempt to simultaneously improve the efficiency of the qPCR assay (i.e. by shortening the amplicons produced), while reducing non-specific amplification of the markers. ClustalW, the general purpose multiple sequence alignment web program for DNA or proteins (Thompson et al. 2002; Thompson et al. 1994), was used to align the original $esp_{fm}$ gene sequence (accession # AF444000.1) with a set of those sequenced at UPenn to determine similarities/differences. This gene sequence was also uploaded to the online Roche UPL primer tool that generated a number of potential human host specific qPCR primer-probe sets. Of interest, the tool returned a primer-probe set that produced amplicons of

~ 187bp in length. This redesigned primer set included the original forward *esp<sub>fm</sub>* primer and a redesigned reverse primer 5′– TGG ATT CGT GTC TCC GCT CTC T – 3 and an internal probe (designated as Roche UPL probe #9) of 5′- CAT CAC CA-3′. BLASTn analysis was conducted of these redesigned primers and the probe against the NCBI nucleotide collection (nr/nt) database to determine if there were improvements in the specificity of the redesigned primer set in comparison to the original.

#### *3.2.1.8 Assay*

A Roche LightCycler® 480 (LC480) Real-Time PCR 1.5 System was used to conduct qPCR assays using the redesigned probe-based protocol. This protocol used the Roche LightCycler® 480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed pre-incubation for 5 min at 95°C; 45 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate and includes nuclease free water for negative controls. DNA samples extracted from the MSU human specific *E. faecium* isolates were used as positive controls for assay.

### 3.2.2 Gene Analysis - 16S rRNA

The NCBI database was queried to obtain *E. faecium* 16S rRNA gene sequences from human feces. One full-length 16S rRNA gene sequence was selected for *E. faecium* (accession # EU722747.1) and used to perform BLASTn queries against the nr/nt database that were limited to microbes from particular animal hosts. Host animals were primarily selected based on their high likelihood to contribute to fecal pollution in watersheds located in the Philadelphia area. As such, the host groups selected for evaluation included human, domesticated animals (cat and dogs), livestock (chicken, cow, horse, pig and sheep) and wildlife (deer and goose). To limit our BLASTn searches to microbes from these hosts, each was performed by specifying the host name in the Entrez query option. The Entrez queries were further limited to include the terms fecal OR feces. All GenBank entries were examined to ensure that they were isolated from the intended host. The maximum identities and query coverage results for the top BLASTn hits for each host are listed in Table 10.

The 16S rRNA sequences of the top 50 BLASTn results for the human host and the top 10 for each of the other animal hosts were aligned, along with that of a selected outgroup (i.e. *Nitrospira marina*). Sequences were specifically aligned on the Ribosomal Database Project (RDP) website (Cole et al. 2009). SeaView4, a freely available software package, was then used to construct maximum likelihood phylogenetic trees (PhyML) for the aligned datasets (Gouy et al. 2010). The approximate likelihood-ratio test (aLRT) of the SeaView4's GRT model was used to generate each tree, and the proportion of invariable sites was estimated during the run. The starting tree topology was obtained by using the neighborhood joining algorithm for biological dataset (BioNJ) option that takes into account evolutionary distances obtained from aligned sequences (Gascuel 1997). The best of two reasonable tree topology was determined by selecting the option for the "BEST" of Nearest Neighbor Interchanges (NNI) or Subtree Pruning and Regrafting (SPR) topological moves (Guindon et al. 2010). Branch support for the tree topology was determined with the approximate likelihood-ratio test (Anisimova and Gascuel 2006).

The resulting phylogenies were uploaded to the interactive tree of life (iTOL) website for annotation (Figures 12) (Letunic and Bork 2007; Letunic and Bork 2011). Rooted phylogenetic tree files generated with SeaView4 were used to perform statistical analyses in FastUniFrac (Hamady et al. 2010). The online suite of tools provided by Fast UniFrac enabled the estimation of phylogenetic overlap for gut microbes of different hosts, which, in turn, allowed us to identify phylogenetic patterns consistent with host specificity (i.e. clustering of microbes from the same host species). To accomplish this, all sequences were assigned to a host environment based on their species of origin. PCoA analyses were then performed using UniFrac distances calculated between each pair of hosts. The results were presented as scatter plots. UniFrac P-test statistics were also computed to determine whether bacteria from different hosts shared significantly less phylogenetic overlap than expected by chance.

## 3.3 Results

### 3.3.1 Gene Analysis - $esp_{fm}$

#### *3.3.1.1 qPCR Results*

The PCR-gel electrophoresis results were comparable to those obtained by SYBR Green qPCR (Table 3). Both protocols yielded a specificity value of 96.4% for the same sample set although there was not 100% agreement of the samples identified as positive. The computed sensitivity for the SYBR Green qPCR protocol was 16.4% and 20.5% for the gel based PCR. Two of the nine environmental water isolates analyzed were confirmed positive by the SYBR Green qPCR protocol and none by the gel based protocol.

**Table 3: Results of *esp$_{fm}$* molecular marker tested against *E. faecium* isolates from the MST library by both gel PCR and SYBR Green qPCR.**

| Sample Type | # of isolates | # gel PCR positives | # SYBR Green qPCR positives |
|---|---|---|---|
| WWTP | 73 | 15 | 12 |
| Animal fecal | 56 | 2 | 2 |
| Water | 9 | 0 | 2 |

Conventional enumeration showed high levels of enterococci in the collected influent and the secondary treated WWTP samples and low levels in the final effluent samples (Table 4). All WWTP samples that were processed with an enrichment step before DNA extraction showed positive results for *esp*$_{fm}$ with both the SYBR Green and the probe based qPCR assays.

**Table 4: mEI and SYBR Green qPCR results for collected WWTP samples.**

| **Sample Type** | **mEI (CFU/100mL)** | **qPCR (gene copy/100mL)** |
|---|---|---|
| Set 1 PST Influent 1A | 20,000 | 15,000 |
| Set 1 PST Influent 1B | 20,000 | 6,000 |
| Secondary treatment 2A (after activated sludge treatment) | 410,000 | 17,000 |
| Secondary treatment 2B (after activated sludge treatment) | 610,000 | 14,000 |
| Effluent 2X (Post Chlorination) 3A | 3 | <1,000 |
| Effluent 2X (Post Chlorination) 3B | 8 | <1,000 |
| Negative controls (DI water) | 0 | <1,000 |

Conversely, on removal of the enrichment step the SYBR Green protocol showed positive signals for the influent and the secondary treated samples and none for the final effluent samples. However, some amplicons exhibited background products represented by multiple peaks of the melting curves (Figure 11) and multiple gel electrophoresis bands for the same sample. The probe based protocol showed no positive signals for the samples that lacked enrichment.

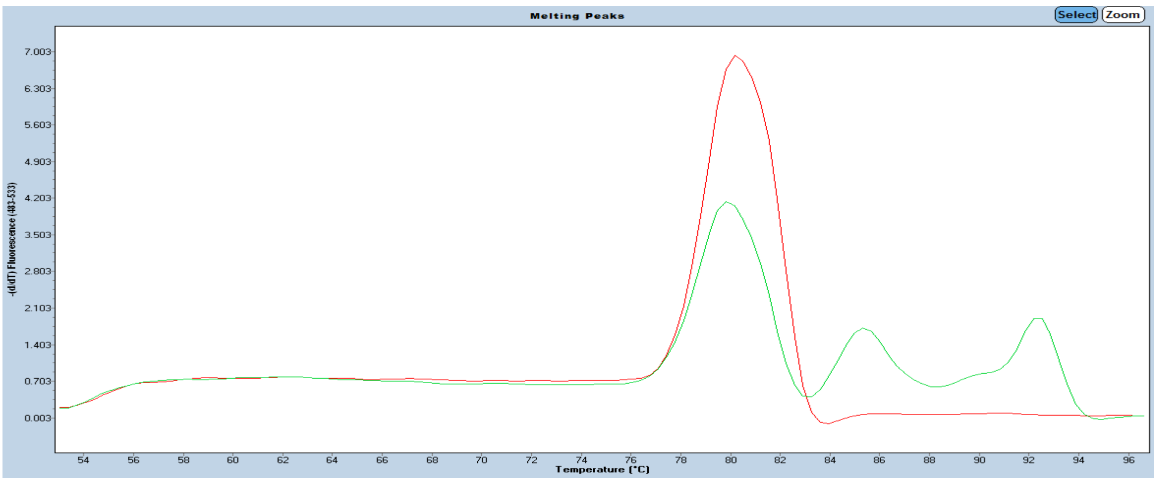



**Figure 11: Melting curve profile for a positive *esp$_{fm}$* control (red) and a sample showing multiple peaks (green).**

#### *3.3.1.2 BLASTn Results*

BLASTn analysis showed improvement in the specificity of the redesigned reverse primer in comparison to the original. The redesign potentially reduced cross reactivity of the marker set with terrestrial animals. However, there still exists the possibility for some non-specific amplification of other animals, in particular those (e.g.

seal and sealion) found in marine habitats (Table 5). For the origina forward $esp_{fm}$ primer analysis, the top seven BLASTn hits were of dog-associated $esp_{fm}$ gene (Table 6).

**Table 5: Top ten hits of the BLAST result for the redesigned reverse $esp_{fm}$ primer**

| Accession | Description | Max score | Total score | Query coverage | E value | Max identity |
|---|---|---|---|---|---|---|
| EU280890.1 | Enterococcus avium isolate eve 27d enterococcal surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU280887.1 | Enterococcus canintestini isolate eve 23b enterococcal surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815405.1 | Enterococcus sp. enrichment culture clone SeaLion_5_13 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815404.1 | Enterococcus sp. enrichment culture clone SeaLion_5_16 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815403.1 | Enterococcus sp. enrichment culture clone SeaLion_5_18 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815402.1 | Enterococcus sp. enrichment culture clone SeaLion_5_19 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815401.1 | Enterococcus sp. enrichment culture clone Seal_4_17 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815400.1 | Enterococcus sp. enrichment culture clone Seal_4_16 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815399.1 | Enterococcus sp. enrichment culture clone Horse_2_16 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815398.1 | Enterococcus sp. enrichment culture clone Horse_2_17 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |

### 3.3.2 Gene Analysis - 16S rRNA

#### *3.3.2.1 BLASTn Results*

The BLASTn results using the *E. faecium* 16S rRNA gene query sequence (EU722747.1) showed the top 10 hits of the human associated group averaged 93.9%

max identity with a minimum of 94% query coverage. The 50th sequence in this group showed 88.0% max identity with 99% query coverage. The top hit for the chicken-, deer- and pig-associated bacterial sequences all showed close relationships (max identity ≥99.5%) to the human-associated *E. faecium* query sequence (Table 7).

**Table 6: Top ten hits of the BLASTn result using the original forward $esp_{fm}$ primer**

| Accession | Description | Max score | Total score | Query coverage | E value | Max identity |
|---|---|---|---|---|---|---|
| EU815343.1 | Enterococcus sp. enrichment culture clone Dog_13_6 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815342.1 | Enterococcus sp. enrichment culture clone Dog_13_5 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815338.1 | Enterococcus sp. enrichment culture clone Dog_13_11 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815334.1 | Enterococcus sp. enrichment culture clone Dog_13_10 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815333.1 | Enterococcus sp. enrichment culture clone Dog_13_9 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815331.1 | Enterococcus sp. enrichment culture clone Dog_13_3 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU815329.1 | Enterococcus sp. enrichment culture clone Dog_13_1 surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU621692.1 | Enterococcus sp. enrichment culture clone ES-18 enterococcal surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU394213.1 | Enterococcus sp. enrichment culture clone ES-19 enterococcal surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |
| EU394212.1 | Enterococcus sp. enrichment culture clone ES-17 enterococcal surface protein (esp) gene, partial cds | 44.1 | 44.1 | 100% | 0.007 | 100% |

#### *3.3.2.2 Phylogenetics*

The phylogenetic tree generated from the *E. faecium* 16S rRNA alignment showed some clustering based on host association (Figure 12). However, there were instances of high relatedness between human-associated and non-human-associated sequences.

**Table 7: Top BLASTn hits to the 16S rRNA sequence of human-associated E. faecium (accession# EU722747.1).**

| Animal Host | Organism | Accession # | Query coverage (%) | Max identity (%) |
|---|---|---|---|---|
| **human** | *Enterococcus hirae* | EU722743.1 | 97 | 99.5 |
| **cat** | Uncultured bacterium | EU359835.1 | 46 | 75.5 |
| **chicken** | *Enterococcus faecium* | AB481104.1 | 99 | 99.9 |
| **cow** | *Weissella* sp. | FJ695507.1 | 97 | 88.3 |
| **deer** | *Enterococcus faecium* | JF690891.1 | 97 | 99.5 |
| **dog** | *Lactobacillus reuteri* | AY324629.1 | 92 | 88.7 |
| **goose** | Bacterium | GU592671.1 | 90 | 85.5 |
| **horse** | *Enterococcus gallinaium* | FN821376.1 | 92 | 97.0 |
| **pig** | Uncultured bacterium | AF371531.1 | 97 | 99.9 |
| **sheep** | *Caldicoprobacter oshimai* | AB450762.1 | 98 | 80.4 |

Unweighted UniFrac PCoA of the different host environments showed human-associated sequences to differ from those of the other animal groups (Figure 13). However, P-test results showed marginally significant differences between the human- and chicken-associated sequences (p-value =0.045), and no significant difference between the human- and horse-associated sequences (p-value =0.405) after Bonferroni correction (Table 8). Horse-associated enterococci showed no significant difference from those of three other host-associated groups (cow, deer and goose) after Bonferroni correction. Also, only suggestive significance was shown between chicken-deer associates and for comparisons of bacteria from cows and geese.

The aligned sequences of the *E. faecium* 16S rRNA dataset showed few host-specific segments. No clear regions could be identified for the design of human host specific 16S rRNA primers. This observation, coupled with the results of the phylogenetic analysis and the BLASTn results, did not encourage attempts to design human host specific primers for the *E. faecium* group based on the 16S rRNA gene.

### 3.4 Discussion

Numerous studies have been proposed that attempt to differentiate sources of fecal pollution in surface waters. Initially, MST studies utilized conventional FIBs such as *E. coli* (Khatib et al. 2002; Whitman et al. 2008) and *Enterococcus* spp. (Ahmed et al. 2009a; Haugland et al. 2005; Santo Domingo et al. 2003; Scott et al. 2005; Wiggins et al. 2003) as prime candidates for differentiating host groups using molecular techniques. Most of these studies show cross-reaction of the human-specific assays with known non-human associated samples. Since unnecessary mitigating actions could result from incorrect source identification, the need for improved MST protocols is clear.

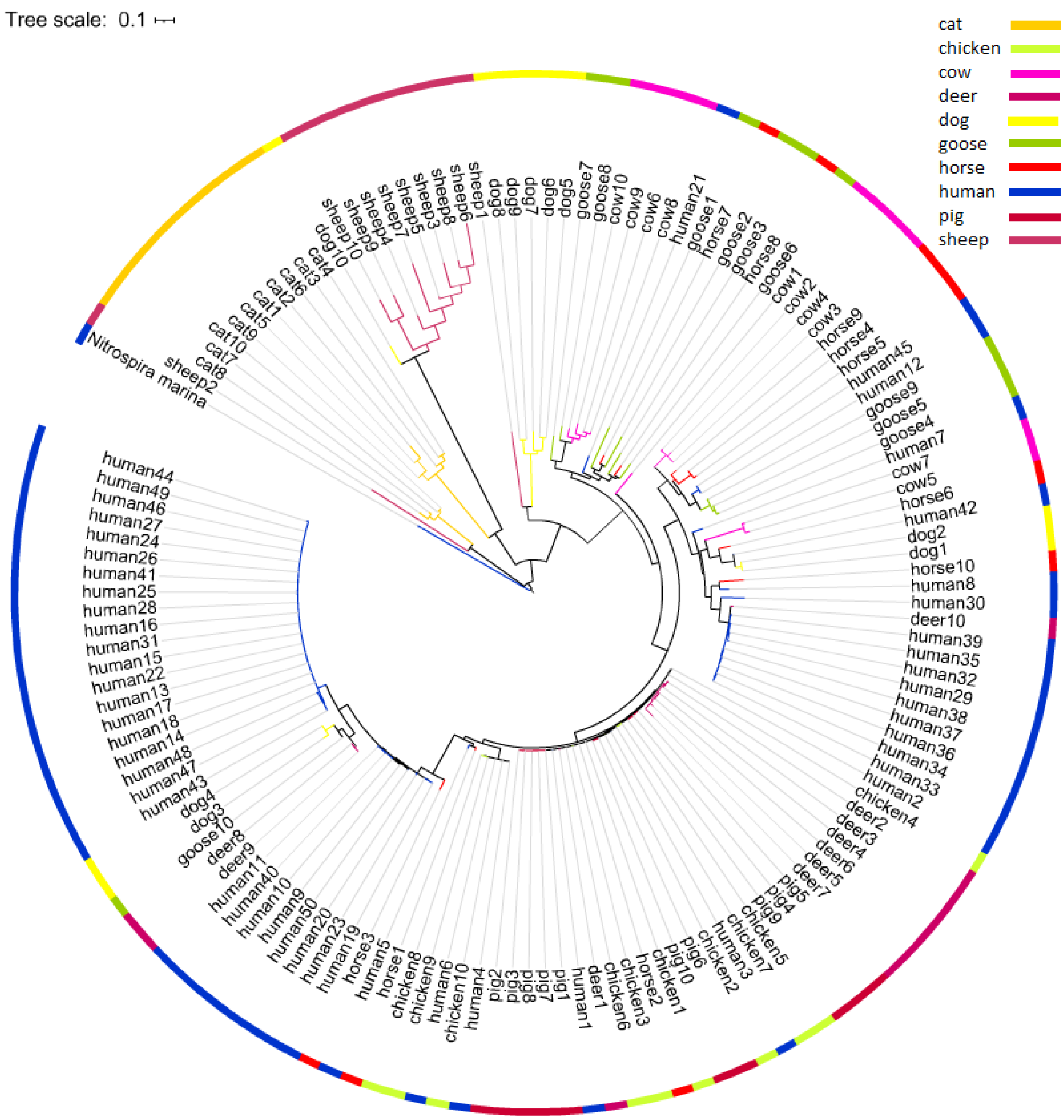



**Figure 12: 16S rRNA maximum likelihood phylogeny of *Enterococcus faecium* and relatives.**

**Sequences used to construct the tree were the ten identified *E. faecium*, the top 50 hits from human hosts and the top 10 hits from each cat, chicken, cow, deer, dog, goose, horse, pig and sheep host. *Nitrospira marina* was used as the out-group to root the**

**tree. Branches and color strips illustrate the host affiliation of each of the analyzed sequences, revealing a weak trend of host specificity for some human associates. Branch lengths of the tree are drawn to scale.**

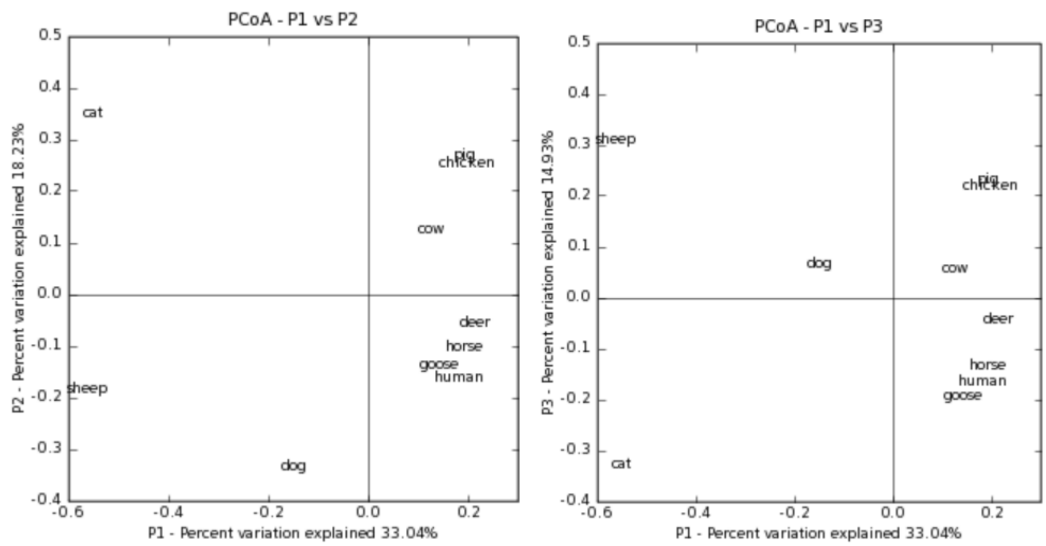



**Figure 13: Unweighted UniFrac PCA of the different host groups for *E. faecium*.**

**Unweighted UniFrac was used to generate a matrix of pairwise distances between communities. A scatterplot was then generated from the matrix of distances using Principal Coordinate Analysis. The plots show pair-wise combinations of the first three principal coordinates axes as visualized in 2D. Human associated sequences show dissimilarity (no overlapping) to non-human associated sequences, although they are close to those from the geese, horse and deer hosts.**

**Table 8: UniFrac P-test significance for E. faecium and relatives.**

**UniFrac P-test significance was used to estimate similarity between each pair of host groups for E. faecium and relatives using 1000 permutations. The lower diagonal shows p-values that were adjusted for multiple comparisons using the Bonferroni correction. The upper diagonal shows raw p-values without correction. No significant differences were observed between bacteria from the following host pairs when the Bonferroni corrections were used: human-horse, horse-goose, horse-deer and horse-cow.**

| | **cat** | **chicken** | **cow** | **deer** | **dog** | **goose** | **horse** | **human** | **pig** | **sheep** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Cat** | | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| **chicken** | <0.001 | | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| **cow** | <0.001 | <0.001 | | <0.001 | <0.001 | 0.002 | 0.005 | <0.001 | <0.001 | <0.001 |
| **deer** | <0.001 | 0.090 | <0.001 | | <0.001 | <0.001 | 0.003 | <0.001 | <0.001 | <0.001 |
| **dog** | <0.001 | <0.001 | <0.001 | <0.001 | | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| **goose** | <0.001 | <0.001 | 0.090 | <0.001 | <0.001 | | 0.035 | <0.001 | <0.001 | <0.001 |
| **horse** | <0.001 | <0.001 | 0.225 | 0.135 | <0.001 | 1.000 | | 0.009 | <0.001 | <0.001 |
| **human** | <0.001 | 0.045 | <0.001 | <0.001 | <0.001 | <0.001 | 0.405 | | <0.001 | <0.001 |
| **pig** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | | <0.001 |
| **Sheep** | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | |

## Description

| |
|---|
| (< 0.001) Highly significant |
| (0.001-0.01) Significant |
| (0.01-0.05) Marginally significant |
| (0.05-0.1) Suggestive |
| (> 0.1) Not significant |

The main aim of this study was to investigate co-evolutionary relationships between *E. faecium* sequences associated with different host animals and to determine potential human specific qPCR-based assays that target human-associated members of this species. We reasoned that groups showing host-specificity would be better candidates for MST protocols due to the likely presence of shared-derived nucleotides (synapomorphies) required for probe specificity.

The *E. faecium* organism was postulated to have adapted to the gut of different hosts and the *esp* gene specialized to enable this microbe to adhere to the surface of the gastrointestinal tract of specific host's. This is one of the reasons this gene was considered to be a possible human specific marker. Assays conducted on *E. faecium* isolates from the MST library showed a specificity of 96.4% with both the PCR-gel electrophoresis and the SYBR Green assays. Other studies showed false positive results for the human specific $esp_{fm}$ (Byappanahalli et al. 2008; Whitman et al. 2007). It is possible that *E. faecium* shifting from host to host; or horizontal transfer of the *esp* gene contributing to the reported wide range of host specificity values for $esp_{fm}$.

A very low sensitivity was obtained for assays conducted on *E. faecium* isolates from the MST library, even for the more sensitive SYBR Green qPCR protocol (20.5%). The low sensitivity was supported by the Kim et al (2010) study that posited that the $esp_{fm}$ gene was not present in all human fecal associated *E. faecium* organisms (<10%

sensitivity for their study). It was assumed that assays of DNA from a composite of *Enterococcus* spp. that comprised the WWTP samples analyzed would improve the sensitivity of the protocols. However, this was the case only after an enrichment process. The sensitivity of both protocols was reduced dramatically when the enrichment step was removed. The SYBR Green qPCR protocol showed lower gene copies per 100 mL of sample than the CFU per 100 mL of the conventional enumeration method. In addition, the multiple peaks of the SYBR Green qPCR melting curve profiles indicated the presence of non-specific amplicons. This implies inaccurate quantification of the $esp_{fm}$ targets. The probe-based qPCR protocol implies there may not be any of the $esp_{fm}$ targets in the assayed WWTP samples since it detected none.

The 16S rRNA gene of *E. faecium* was investigated as a possible alternative to the $esp_{fm}$. The 16S rRNA gene is considered one of the most conserved in bacteria (Isenbarger et al. 2008) and was promisingly used for *Bacteroides* related MST (Green et al. 2011; Haugland et al. 2010; Shanks et al. 2010b). The BLASTn analysis of the *E. faecium* 16S rRNA dataset indicates poor differentiation between the human and the non-human host associated bacterial sequences. The high similarity of bacteria from different animal hosts to the *E. faecium* 16S rRNA query sequence indicates that these sequences had little evolutionary divergence. This is supported by the unweighted UniFrac PCA analysis that showed close similarity between human-associated 16S rRNA gene sequences and those from goose, horse and deer hosts. Overall, there may be little human host specificity that can be detected for *E. faecium* with the slowly evolving 16S rRNA gene. The phylogenetic tree for the *E. faecium* dataset also showed dispersion of the different host-associated sequences across the tree. Twenty of the human-associated sequences formed a human specific clade; however, we did not attempt to design a primer-probe set for this clade because it was believed the sensitivity of this assay would be less than desirable.

It is possible that because *E. faecium* is a facultatively anaerobic organism it has had opportunities to be interchanged between different animal hosts. This could have prevented *E. faecium* from exclusively coevolving with a particular host group. The resulting limitation—the high similarity between *E. faecium* 16S rRNA gene sequences in different hosts—did not encorage us to develop human specific markers targeting this gene.

Overall, this study showed limitations in using *E. faecium* 16S rRNA and *esp* genes in differentiating the hosts of fecal pollution. This organism may still prove beneficial to the MST community, should other genes show a stronger signal of host specificity. Since enterococci are still used for conventional water quality assessment of recreational waters, the use of general/universal molecular assays for this group of organisms may be more advantageous. This general molecular assay could be used in a toolbox to help determine total fecal pollution levels.

# CHAPTER 4: VALIDATION OF THE QPCR MST PROTOCOL DEVELOPED THROUGHOUT THIS STUDY.

## 4.1 Introduction

State and local agencies are currently required to monitor fecal pollution levels based on methodology established in the early 1900s (Bonde 1966; NRC 2004). These methods do not allow identification of the specific sources of microbial pollution. For decades, critics have highlighted issues associated with the poor correlation between bacterial indicator densities and pathogen levels (Santo Domingo et al. 2007). Compliance with the Clean Water Act (1997, original title: Federal Water Pollution Control Amendments of 1972) and the federal requirements to develop and implement total maximum daily loads (TMDLs) made fecal source determination an imperative issue in the United States (Layton et al. 2006; Santo Domingo et al. 2007; Simpson et al. 2002; USEPA 2005b).

Microbial source tracking (MST) employs methods that aim to identify specific fecal source groups so that best management practices (BMPs) can be developed to control fecal contamination from relevant human/animal sources, protect recreational-water users from water-borne pathogens, and preserve the integrity of drinking water supplies (USEPA 2005b). Based on the literature, *Bacteroides* 16S ribosomal DNA genetic markers, particularly those that were associated with the HF183 human specific marker, showed very good MST predictive capabilities for human fecal pollution. Our bioinformatics analyses of *Bacteroides* 16S rRNA sequences obtained from both the National Center for Biotechnology Information (NCBI) database and our DNA sequencing efforts indicated of that we developed a potentially improved *Bacteroides* quantitative polymerase chain reaction (qPCR) MST protocol. However, before we could apply our protocol to real world situations, we needed to confirm our bioinformatics assessment by conducting sensitivity and specificity analyses.

Sensitivity and specificity evaluation requires the assay of samples obtained from known host sources. A variety of sources of pollution might influence water quality of complex watersheds, and in most cases, these are temporally variable. Thus, determining the level of host specificity of a MST protocol is essential (Santo Domingo et al. 2007; USEPA 2005b), particularly of hosts located within the watershed of concern (Field and Samadpour 2007). However, unlike clinical samples which are normally comprised of isolates, environmental water and animal enteric samples are comprised of a diverse cocktail of microorganisms with high concentrations of inhibitory PCR substances (humic acids, proteins, etc.). Taking these issues into consideration, an evaluation of DNA extraction protocols from different manufacturers was conducted to determine which was the most efficient at producing the purest DNA eluate at its highest concentration.

One of the limitations of DNA extraction protocols are the very small sample volume/mass recommended by manufacturers (as low as 200 µL by volume). This affects the limit of detection (LOD) of a protocol, especially in cases where the abundance of molecular targets is low. *Bacteroides* are enteric anaerobes that are restricted to warm-blooded animals (Layton et al. 2006; USEPA 2005b), thus, their concentrations in environmental waters may not be high enough to be detected. In order to improve our LOD we explored different methods of concentrating environmental water sample used in the DNA extraction process. Our evaluation took into consideration the possibility of increasing qPCR inhibitory substances during the concentration process.

## 4.2 Materials and Methods

### 4.2.1 Sample collection

A total of 58 human-associated samples were collected from a wastewater treatment plant (WWTP) that serves the Philadelphia populace on 10 separate occasions and used to assess the sensitivity of the qPCR protocol. On each occasion 120 – 500 mL grab samples were collected from the primary settling tank (PST) influent (duplicate), the aeration/mixed liquor effluent (MLE) (duplicate), the pre-chlorination (PreCl) final effluent and the post-chlorination (PostCl) final effluent sampling points by the plant's personnel. All samples were transported to the MST laboratory in igloos on ice (<8$^{o}$C) for expeditious processing (<6hr after collection). At the laboratory, samples were stored in the refrigerator (~4°C) until processed.

A total of 256 distinct animal fecal samples were collected from 21 animal types to assess the specificity of the qPCR protocol used in this study (Table 3). From this collection, 2486 replicates were prepared in 180-220 mg proportions and stored in 2 mL microcentrifuge tubes at -80 °C until required for DNA extraction. Most of the animal fecal samples were collected from: the Fox Chase Farm (FCF), a working educational farm located in Northeast Philadelphia; the Schuylkill Center for Environmental Education (SCEE), located in the northwest corner of Philadelphia, in the neighborhood of Roxborough; and the W. B. Saul High School of Agricultural Sciences (SHS), located in the upper Roxborough section of Philadelphia. Dog fecal samples were collected from SHS, the Schuylkill River Park Dog Run located at 25th and Spruce Street in Center City Philadelphia, and from the homes of colleagues. Cat fecal samples were collected from the homes of colleagues. All samples were transported to the MST laboratory in igloos with ice packs (<8°C) and processed within 6 hours of collection.

#### 4.2.2 FIB Enumeration

Most of the WWTP samples collected were enumerated for *Enterococcus* by membrane filtration (MF) and mEI cultivation in accordance with the EPA's standardized Method 1600 (USEPA 2005a). All samples were simultaneously processed for DNA extraction.

### 4.2.3 DNA Extraction

Four different DNA extraction kits were evaluated using the first three set of WWTP samples collected to determine the most efficient protocol to process WWTP and fecal samples. DNeasy Blood and Tissue Kit (Qiagen), MagMax (Invitrogen), PowerWater DNA (MoBio) and QIAamp DNA Stool Kit (Qiagen) were each used to extract DNA from WWTP samples in accordance to the manufacturers' protocol. The concentrations of the total DNA extraction yields for all samples were determined with a NanoDrop ND-000 UV spectrophotometer (Thermo Fisher Scientific) with absorbance readings at 260 nm, and the purity of the DNA preparations determined by A260/A280 ratios.

DNA extractions were initially performed by the manual protocol and eventually automated using a QIAcube instrument (Qiagen). QIAamp DNA Stool kits (Qiagen) were determined by our evaluation to be the best option to process samples. The method was followed in accordance with the manufacturer's protocol except for liquid environmental samples. The protocol was modified for these samples by a concentration process in order to increase the detection of the genetic markers during the qPCR process. Three different sample concentration processes were evaluated. These included:

- increasing the volume processed for DNA extraction from 0.2 mL to 0.6 mL
- centrifuging up to 50 mL of sample for 15 min at 20,000g and extracting DNA from the resulting pellets
- membrane filtering up to 1 L of sample using mixed cellulose ester 0.45 μm membranes and extracting DNA from the membrane

In all cases involving MF the maximum possible volume of the sample that could be filtered depended on its turbidity. Typically, the volume filtered ranged from 1 mL (MLE) to 100 mL (post-chlorination final effluent) of WWTP sample and 500 to 1000 mL environmental water samples.

The QIAamp manufacturer's instructions were then followed for the first two options. For the MF option, MFs were placed in 2.0 mL safe lock microcentrifuge tubes (Qiagen) and 1.8 mL ASL buffer added. Samples were vortexed for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90$^{o}$C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAcube QIAamp DNA Stool protocol was followed precisely. All extracted DNA samples were stored at -20°C until qPCR analysis.

**4.2.4 SYBR Green qPCR Protocol**

A SYBR Green *Bacteroides* HF183 qPCR protocol was initially used for assays in the developmental phase of the MST project (Ahmed et al. 2009a). The original HF183 prtocol was compared against one redesigned for this study. This was our first attempt to redesign the HF183 assay in order to improve its efficiency by shortening the amplicons produced, while simultaneously reducing non-specific amplification of the markers. When paired with the reverse primer Bac708R (5′ – CAA TCG GAG TTC TTC GTG – 3′), the HF183 primer (5′ – ATC ATG AGT TCA CAT GTC CG – 3′) generates an amplicon ~520 base pairs (bp) in length. The redesigned *Bacteroides* assay utilizes the original HF183 forward primer and a new reverse primer Bac368R (5′ – CAG GCC ATC GCC CAT TGA CCA – 3′), generating amplicons of ~211 bp in length.

Reaction mixtures of 20µL cosisisted of Roche LightCycler 480 SYBR Green I Master mix; 300 nM of each forward and reverse primers. Reducing the volume of the DNA template from 5 µL to 2 µL assisted in the reduction of PCR inhibitory substances found in the complex environmental samples being investigated. The assays were conducted using a LightCycler 480 (Roche) and the PCR program was optimized to increase reaction specificity and to reduce procedure time.

The modified amplification program employed a touchdown protocol of: pre-incubation for 10 min at 95°C; 40 amplification cycles of 10 s of melting at 94°C, 10 s of annealing starting at 65°C (reduced by 0.5°C at each cycle), and 30s of extension at 72°C; melt curve analysis from 53°C to 97°C; and finally cooling for 10 s at 40 C (Table 1). This protocol reduced the original assays from 3 hr. to less than 1 hr. and increased the specificity of the primers. Nuclease-free water was used as no-template controls (NTC) (negative controls) and WWTP samples as positive controls in all assays.

Samples were considered positive when the melt curve analysis profile of samples matched that of the positive control and had a similar melting temperature ($T_m$). The $T_m$ is the temperature at which half of the double stranded DNA in the SYBR Green PCR reaction well separates or melts. A maximum crossover point (Cp) was also used to indicate positive results within the dynamic range of the standard curve. The Cp is the point at which the fluorescence of a sample rises above the background fluorescence and is dependent on the amount of target that is present at the beginning of the reaction. The higher the initial target concentration, the lower the Cp value.

#### 4.2.5 Agarose Gel Electrophoresis protocol

In addition to having positive and negative controls, agarose gel electrophoresis was used to confirm the identity of amplicons generated by the qPCR protocols. The protocol was optimized to achieve maximum separation in the shortest time. Agarose gels, 2% (w/v), were prepared by weighing out the appropriate amount of agarose and diluting in 120ml of 1.0X Tris-Acetate-EDTA buffer. This mixture was heated to a boil with mixing until the agarose was completely melted and then cooled to approximately 60°C. The agarose gel was then casted using a 24-tooth comb and allowed to cool and to solidify. Gels were stored in 1.0X Tris-Acetate-EDTA buffer in the refrigerator if not immediately used.

For each gel, 5μL PCR product was loaded into the wells and then electrophoresis was performed at 100V for ~2 hour. Post staining of the gels were done with a 3x GelRed (Biotium) nucleic acid gel staining solution for 1 hour and then viewed using the UV transilluminator and photographed. Images were captured with a Canon PowerShot G6 digital camera.

#### 4.2.6 DNA Sequencing

Representative amplicons (DNA fragments) of the appropriate band lengths for each primer set were extracted from gels for DNA sequencing. DNA was extracted and purified using Qiagen's QIAquick Gel extraction kits. Purified amplicons were submitted to the University of Pennsylvania (UPenn) Department of Genetics DNA Sequencing Facility for sequencing as another means of quality control. The ribosomal database project (RDP) web tool was used to positively classify the 16S rRNA gene sequences of the HF183 amplicons as *Bacteroides*.

### 4.2.7 Probe qPCR Protocol

A probe based qPCR protocol was developed to increase the selectivity and specificity of the MST protocol and reduced background non-specific PCR products.

#### *4.2.7.1 Redesigned Primer-Probe Set-pMR*

The finalized human specific non-enrichment library independent (NELI) qPCR MST redesigned protocol targeted *Bacteroides dorei* and included a novel forward primer HF68 (5′– GGC AGC ATG GTC TTA GCT TG – 3′) and HF183rc, the reverse complement of the original forward primer HF183 (5′ – CGG ACA TGT GAA CTC ATG AT – 3′). This primer set generated amplicons of ~135 base pairs in length. Roche UPL hydrolysis probe #156 (5′-GCT GAT GG-3′) was used with this primer set to complete this qPCR MST assay. This primer-probe set was designated pMR.

#### *4.2.7.2 Shanks' HF183 Primer-Probe Set- pOS*

Shanks et al's (2010) HF183-based primer-probe set (pOS) was used in this study for comparative purposes. The pOS primer sets included the *Bacteroides* HF183 forward primer 5′– ATC ATG AGT TCA CAT GTC CG – 3′ and the BthetR1reverse primer 5′ – CGT AGG AGT TTG GAC CGT GT – 3′. This primer set generated amplicons of ~179 base pairs in length. The fluorescently labeled probe used for this qPCR assay was [6~FAM] CTG AGA GGA AGG TCC CCC ACA TTG GA [TAMRA~6~FAM].

A Roche LightCycler® 480 (LC480) Real-Time PCR 1.5 System was used to conduct all qPCR assays. The protocols for both assays used the Roche LightCycler® 480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed pre-incubation for 5 min at 95°C; 45 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, with standard curves, nuclease free water for no-template-controls (NTC) and purified plasmids derived from HF183 WWTP amplicons for positive controls.

#### 4.2.8 Standard Curves

A plasmid DNA construct was developed to function as a plasmid DNA standard for calculation of assay calibration curves. Amplicons obtained from the NE-WWTP influent were purified with QIAquick Gel extraction kits (Qiagen) and then cloned, before plasmid purification with Qiagen PCR cloning plus kits. The eluate was quantified and diluted to generate samples ranging from approximately 10 to 1.0E9 molecules of template DNA per μL for standard curves for each qPCR assay.

**4.2.9 Statistics**

The sensitivity of a human specific marker is the probability of a positive test result given that the assay is of known human fecal origin (true positive rate):

$$Sensitivity = \quad P(T^{+}|H) = \quad \frac{\alpha^{+}}{\alpha^{+} + \alpha^{-}}$$

Where $\alpha^{+}$ is the number of positive assays of fecal samples of known human origin, $\alpha^{-}$ is the number of negative assays of fecal samples of known human origin, $T^{+}|H$ is a test considered positive given that the samples are all of known human origins.

The specificity of a human specific marker is the probability of a negative test result given that the assay is of a known non-human-fecal source:

$$Specificity = \quad = \quad P(T^{-}|A) = \quad \frac{\beta^{-}}{\beta^{+} + \beta^{-}}$$

Where $\beta^{+}$ is the number of positive assays of fecal samples from known non-human animal origin, $\beta^{-}$is the number of negative assays of fecal samples of known non-human animal origin, $T^{-}|A$ is a test considered negative given that the samples are all of known non-human animal origins.

Applying Bayes' Theorem, the positive predictive values (PPV) of a human specific marker is the probability that an assay of a sample of unknown fecal source is positive for human fecal pollution given a positive test result.

$$PPV \quad = \quad P(H|T^{+}) \quad = \quad \frac{\sum True\ Positive}{\sum Test\ Outcome\ Positive}$$

$$= \frac{(sensitivity)(prevalence)}{(sensitivity)(prevalence) + \ (1 - specificity)(1 - prevalence)}$$

$$= \frac{P(T^{+}|H)P(H)}{P(T^{+}|H)P(H) + P(T^{+}|A)P(A)}$$

Where $P(H)$ is the probability that the assay/test is of known human fecal source (prevalence) and $P(A)$ is the probability that the test is of known animal fecal source (1-prevalence).

The negative predictive value (NPV) of a human specific marker is the probability that an assay of a sample of unknown fecal source is negative for human fecal pollution given a negative test result.

$$NPV \quad = \quad P(A|T^{-}) \quad = \quad \frac{\sum True\ Negative}{\sum Test\ Outcome\ Negative}$$

$$= \frac{(specificity)(1 - prevalence)}{(specificity)(1 - prevalence) + \ (1 - sensitivity)(prevalence)}$$

$$= \frac{P(T^{-}|A)P(A)}{P(T^{-}|A)P(A) + P(T^{-}|H)P(H)}$$

**4.3 Results**

As expected, the enterococci enumeration showed high counts of these microbes in the WWTP influent and mixed liquor sample (Table 9). The higher counts observer of the mixed liquor samples may be due to the concentration process that takes place in an aeration tanks during the activated sludge process. The pre and post chlorination effluent samples showed very low or no enterococci counts.

**Table 9: Number of Enterococci (CFU/100 mL) of NE-WWTP sample based on mEI cultivation**

| **Sampling Point** | **1/19/2010** | **2/2/2010** | **6/09/2010** | **3/1/2011** | **3/7/2011** | **3/14/2011** |
|---|---|---|---|---|---|---|
| PST Influent 1 | 22,000 | 20,000 | 40,000 | 40,000 | 32,000 | 52,000 |
| PST Influent 2 | 26,000 | 20,000 | 40,000 | 46,000 | 36,000 | 58,000 |
| Aeration/Mixed liquor effluent 1 | 100,000 | 470,000 | >200,000 | 76,000 | 68,000 | 88,000 |
| Aeration/Mixed liquor effluent 2 | 220,000 | 610,000 | >400,000 | 93,000 | 70,000 | 90,000 |
| Pre-chlorinated final effluent | <100 | 800 | - | <1000 | no sample | turbid |
| Post chlorinated final effluent | <100 | 300 | - | <1000 | 2 | 1 |
| DI controls - Pre, Post | <0.4 | <0.5 | <1 | <0.4 | <0.4 | <0.4 |

### 4.3.1 DNA Extraction Evaluation

DNA concentrations determined by NanoDrop readings (normalized to ng per µL of sample) showed the four extraction protocols to have similar trends in extracted

DNA levels for each sample type (Figure 14). For each sample type, the MagMax method produced the highest concentrations of DNA followed by the DNeasy, QIAamp and PowerWater methods, in that order. Extraction of DNA from the mixed liquor samples proved the most challenging given the high levels of suspended solids they contained. However, these samples produced the highest DNA concentrations by all four extraction kits. The PostCl effluent samples produced the lowest concentrations of DNA per μL of sample.

The SYBR Green qPCR protocol indicated that the DNeasy, MagMax and PowerWater protocols generated PCR inhibitor carryover, resulting in little to no DNA amplification by qPCR for most samples. An increase in amplification occurred at greater sample dilution confirming PCR inhibitory contamination. Unlike the other three DNA extraction protocols, PCR amplification was obtained from DNA samples generated with the QIAamp. Serial dilutions of some of these samples resulted in expected corresponding qPCR concentrations with efficiency values close to two, indicating that this protocol solved or at least minimized the PCR inhibitory problem. Thus, the method of choice for further assays in this study was the QIAamp protocol since it generated less background PCR products. Modification of the qPCR amplification's time and temperature settings, and adjustment in the concentration of the primers and templates further reduced background amplification signals, including those of possible primer dimers. The Methods section outlines the finalized program.

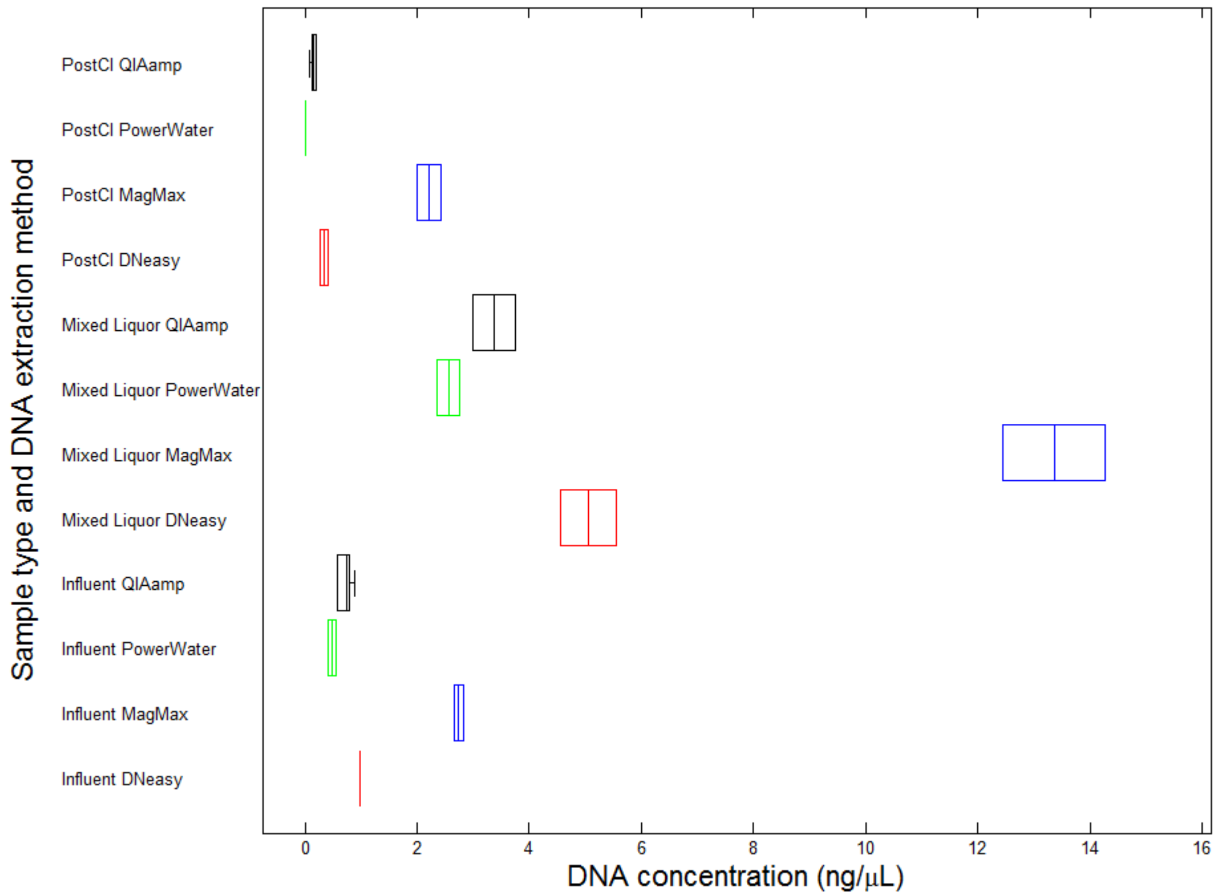



**Figure 14: Boxplot showing DNA concentration (ng/μL) before normalization of WWTP samples extracted by different methods.**

**The boxes represent the 25th, 50th and 75th percentile and whiskers, when present, represent the minimum and maximum NanoDrop© determined DNA concentrations of each extraction protocol per sample type. The DNA concentrations are normalized to ng per μL of sample. Colors represent DNA extraction methods. Key: black-QIAmp, green-PowerWater, blue-MagMax, red-DNeasy.**

The $T_m$ for the original SYBR Green HF183 assay was determined to be ~ 86 ± 1.3°C (Figure 15). However, shoulder peaks appeared at ~83 ± 1°C and were postulated to be indicative of hybrid genotypes due to variations of the GC content of the targeted DNA fragment sequence or amplification of non-specific targets (LightCycler® 480 Instrument Operator's Manual). These postulates were supported by gel electrophoresis where non-specific bands were observed in some of the samples.

A.


B.

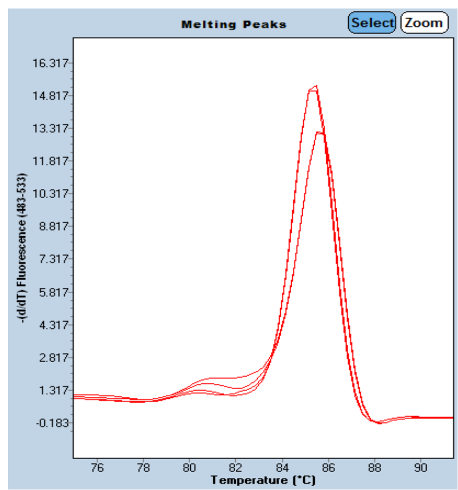



**Figure 15: Melting curves profiles of the original (left) versus the redesigned (right) HF183 SYBR Green qPCR assay. Peak: ~ 86°C.**

**A. Red and green curves of two different WWTP samples done in triplicate. B. Red curves of a WWTP samples done in triplicate.**

Positive HF183 qPCR products were determined by matching their $T_m$ profiles to that of gel electrophoresis bands confirmed to be HF183 amplicons, examples of which are shown in Figure 16.

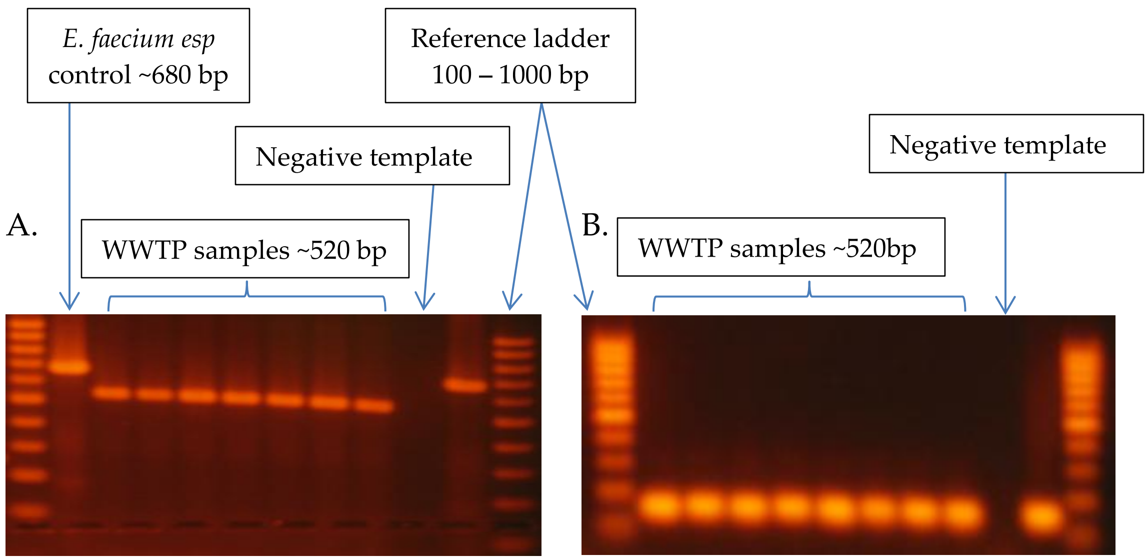



**Figure 16: Gel electrophoresis of amplicons from the Ahmed et al (2008) HF183 protocol (~520 bp) (left) versus our redesigned HF68-HR183rc (~179 bp) (right) SYBR Green qPCR assay.**

Preliminary assays of samples collected from cats and dogs showed good presence/absence differentiation, with no human fecal DNA melting curve profiles for the cat and dog samples (Figure 17). However, non-human amplification or primer dimers were observed at ~76°C, which prevented quantification applicability of the protocol.

The HF68-HR183rc protocol showed much improved results in comparison to the SYBR Green versions. Assays of the initial cat and dog samples showed no qPCR signals, suggesting improved specificity on the HF68-HR183rc assays.

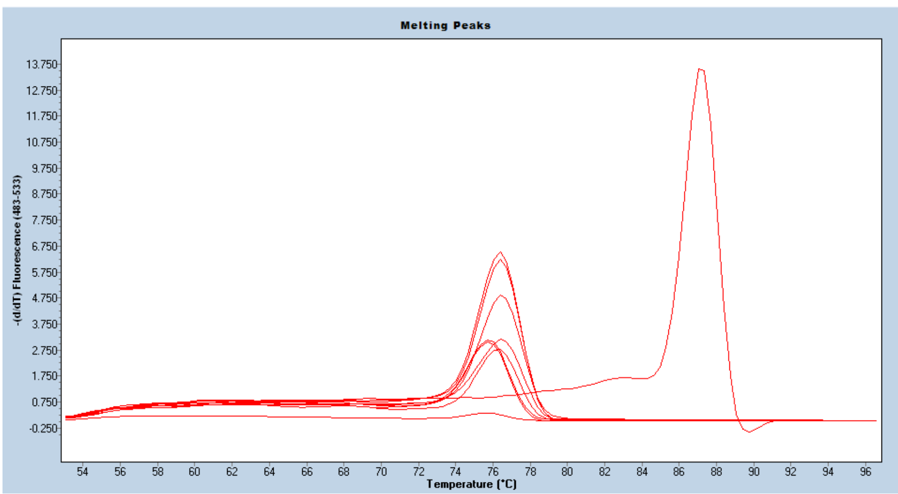



**Figure 17: Assay of cat and dog samples. Positive control HF183 ~ 86°C; non-human amplification or primer dimers – 76°C**

The qPCR results showed that the three different DNA concentration processes had similar values (Figure 18). Membrane filtration (MF) concentration with small volumes (<1 mL) of samples tended to produce slightly lower results in comparison to the direct method. The centrifugation process proved time consuming and an evaluation of the supernatant showed it to have remnants of the molecular marker. The MF protocol proved to be the best method for samples with low concentrations of the molecular targets that could be processed via MF of large volumes of the sample. In these cases, the direct and the centrifugation methods sometimes showed no qPCR results.

#### 4.3.2 Statistics

The sensitivities of the MST protocols were shown to be dependent on the DNA extraction protocol and the sample processing/concentration method used. As previously highlighted, different DNA extraction protocols were associated with varying presence/absence/levels of PCR inhibitory substances.

All WWTP samples assayed with the pMR and the pOS protocols showed qPCR amplification (100% sensitivity). The qPCR results showed slightly higher concentration of the pMR targets in the influent as compared to the MLE samples, with the number of gene copies per mL of sample in the order of magnitude of 1E+05 – 1E+06 (three of the set of collected WWTP samples shown in Table 10). The post-chlorinated samples showed the lowest levels of the human specific molecular markets. However, the levels may be enough to compensate for dilution effects and allow detection of WWTP effluent pollution of receiving environmental waters.

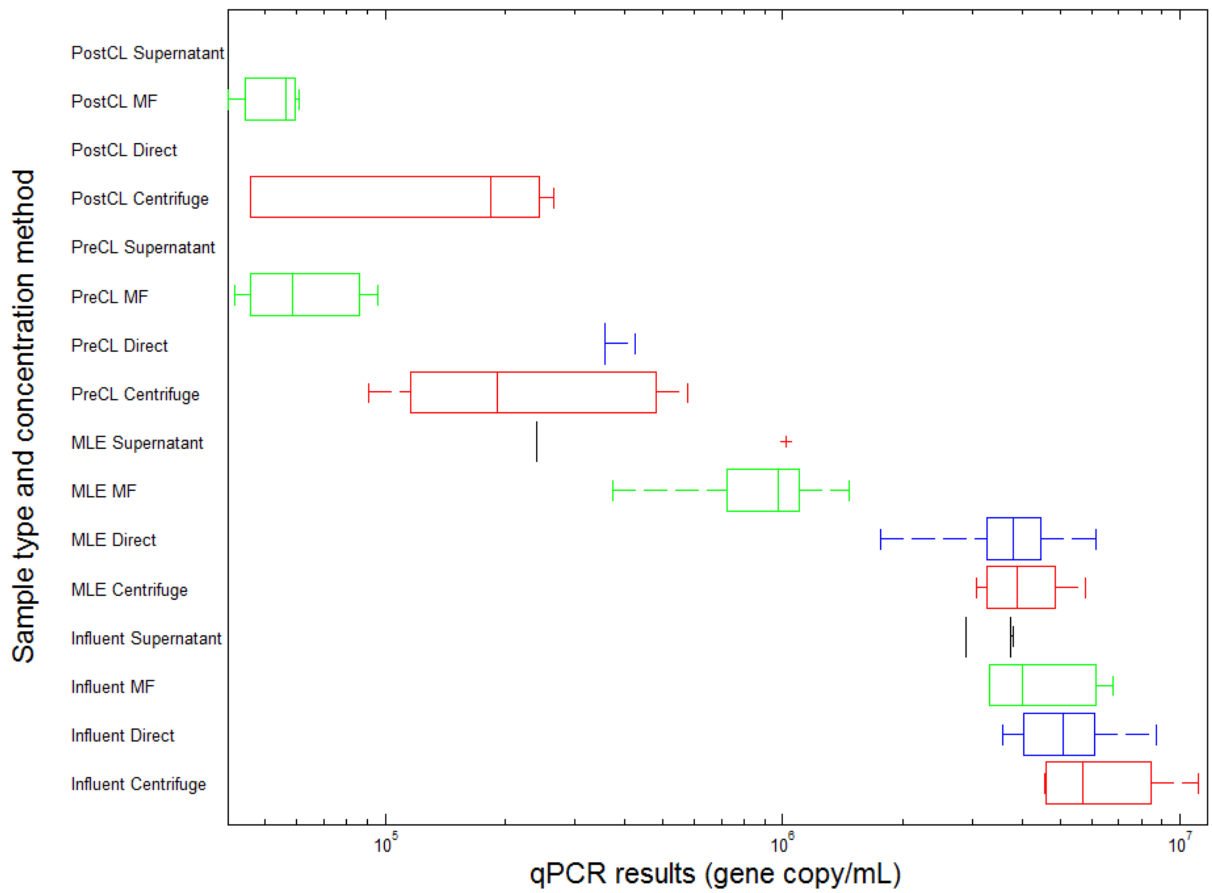



**Figure 18: Boxplot showing qPCR results (gene copies/mL – log transformed) of WWTP samples concentrated by different methods.**

**The boxes represent the 25th, 50th and 75th percentile and the whiskers represent the minimum and maximum concentrations of the pMR molecular markers for each sample concentration method per sample type. The molecular marker concentrations are normalized to gene copies per µL of sample. Colors represent sample concentration methods. Key: black-supernatant, green-membrane filtration, blue-direct, red-centrifuge pellet.**

The detected pMR target concentrations gradually decreased throughout the WWTP treatment train. The pre- and post-chlorination final effluent both showed similar pMR target concentrations.

Overall, the pMR protocol showed specificity of 93% with 58% of the rabbit fecal samples showing positive signals (Table 11). The concentration of the markers in the rabbit samples reached orders of magnitude of 1E+09 gene copies/g of sample. The pOS protocol showed 81% specificity with 66% of the rabbit samples showing positive signals.

**Table 10: Number of *Bacteroides* 16S rRNA gene copies in 100 mL of NE-WWTP sample**

| **Sampling Point** | **3/14/2011 gene copies / mL** | **4/21/2011 gene copies / mL** | **4/25/2011 gene copies / mL** |
|---|---|---|---|
| PST Influent 1 | 3.89E+05 | 2.56E+05 | 7.77E+06 |
| PST Influent 2 | 2.55E+05 | 2.96E+05 | 4.90E+06 |
| Aeration/Mixed liquor effluent 1 | 2.61E+05 | 4.81E+04 | 3.17E+06 |
| Aeration/Mixed liquor effluent 2 | 2.80E+05 | 5.88E+04 | 3.17E+06 |
| Pre-chlorinated final effluent | 2.33E+03 | 3.92E+03 | 1.93E+05 |
| Post chlorinated final effluent | 1.33E+03 | 3.15E+03 | 2.43E+05 |
| DI controls - Pre, Post | <3.60E+00 | <3.60E+00 | <3.60E+00 |

**Table 11: Table showing the number of samples of each animal type assayed, and the number of samples positive for the pMR and the pOS human specific *Bacteroides* primer sets.**

| Animal type | # of distinct samples assayed | # of replicates | # of distinct samples positive for pMR | # of distinct samples positive for pOS |
|---|---|---|---|---|
| cow | 34 | 408 | 0/34 | 2/14 |
| sheep | 24 | 288 | 1/24 | 1/6 |
| pig | 21 | 252 | 0/21 | 8/10 |
| horse | 16 | 192 | 0/16 | 0/10 |
| goat | 25 | 300 | 0/24 | 0/4 |
| chicken | 21 | 252 | 0/21 | 0/21 |
| rabbit | 24 | 288 | 14/24 | 16/24 |
| dog | 16 | 192 | 1/16 | 5/16 |
| cat | 8 | 24 | 0/8 | 0/8 |
| duck | 1 | 1 | 0/1 | 0/1 |
| deer | 1 | 3 | 0/1 | 0/1 |
| opossum | 1 | 1 | 0/1 | 0/1 |
| squirrel | 1 | 1 | 1/1 | 1/1 |
| red tailed hawk | 1 | 4 | 0/1 | 0/1 |
| raccoon | 1 | 12 | 0/1 | 0/1 |
| ground hog | 1 | 12 | 0/1 | 0/1 |
| goose | 20 | 240 | 0/20 | 0/15 |
| mice | 22 | 4 | 0/22 | 0/22 |
| rat | 7 | 7 | 0/7 | 0/7 |
| Gerbil | 8 | 1 | 0/8 | 0/8 |
| guinea pig | 4 | 4 | 0/4 | 0/4 |
| Total animal samples | 257 | 2486 | 17/256 | 32/169 |

.

Taking into consideration the sensitivity and specificity values the positive predictive value (PPV) of the pMR and pOS assays were 77% and 54%, respectively. The negative predictive value (NPV) of both assays was 100%. These results indicate both assays would equally predict the absence of human enteric pollution. However, of the two assays, the pMR assay would more accurately predict its presence of human enteric pollution.

**4.4 Discussion**

*Bacteroides* spp. are found in much higher concentrations in human enteric samples in comparison to the conventional FIB Enterococci. This is one of the reasons that *Bacteroides* spp. were investigated as potential MST FIBs. The use of using PCR-based techniques targeting *Bacteroides* spp. have been shown to be relatively successful in MST in differentiating host groups (Ahmed et al. 2009a; Ahmed et al. 2009b; Bernhard and Field 2000a; Bernhard and Field 2000b; Haugland et al. 2005; Haugland et al. 2010; Santo Domingo et al. 2003; Scott et al. 2005; Shanks et al. 2010b; USEPA 2005b). Throughout this study two human specific qPCR protocols obtained from the literature were evaluated.

The first protocol evaluated, a SYBR Green qPCR protocol (Ahmed et al. 2009a), showed limitations as a MST assay. Analysis of the SYBR Green $T_m$ profiles of positive samples could determine presence/absence of the targeted amplicons, however, the amplification of non-specific DNA or primer dimers negated this protocol's use for quantitative purposes. Further literature review was conducted to identify a better human specific protocol, however, all those available showed some qPCR signals of DNA of enteric microbes associated with non-human hosts. Of the protocols evaluated, the probe based qPCR protocol - pOS (Shanks et al. 2010b) showed the best sensitivity and specificity values.

Bioinformatics analyses of the primer and probe oligos associated with reported *Bacteroides* HF183 protocols, including the pOS, showed only the forward HF183 primer to be human specific. The human non-specificity of the reverse primers and the probes could be contributing to the non-human qPCR signal observed. Bioinformatics inferred that our redesigned protocol, pMR, had the potential to be more specific than the pOS since the HF68 primer showed equal human specificity as the HF183. The qPCR assays showed the specificity of the pMR protocol to be 93% (PPV = 77%) in comparison to 81% for the pOS (PPV = 54%), supporting our predicitons. The PPV values for both protocols are expected to increase with an increase in the ratio of assay of human-associated WWTP to animal samples.

This study showed very high concentrations and occurrences of the human specific marker in fecal samples from rabbit. An evaluation of these findings is discussed in Chapter 3. The non-specific human results do not eliminate the applicability of the protocols as long as the limitations are considered. Small rural watersheds might be impacted by only a few host sources. Thus, if an evaluation of a watershed indicated that the cross-amplified hosts are not present in that given environment then the protocol could still provide useful information to end users (Santo Domingo et al. 2007).

The qPCR assays of both the pOS and the pMR protocols showed 100% sensitivity for WWTP samples with gene copies/100 mL of sample ranging from 1.0E5 to 1.0E8. These concentrations are approximately one order of magnitude lower than that reported in the literature (Converse et al. 2009). This reduction in concentration could be due to the improved selectivity of the pMR protocol, however, we cannot overlook the possibility that it is due to the difference between the WWTP samples of the literature and that of our study. The standard *Enterococcus* enumeration showed CFU/100 mL four orders of magnitude lower (maximum of 5.2 x 1.0E4) than the pMR results. This highlights the much higher concentrations of *Bacteroides* spp. to Enterococci in human enteric samples.

The evaluation of the DNA extraction methods showed the QIAamp stool kit from Qiagen to produce the least PCR inhibitory substances. This protocol was specifically designed to extract DNA from complex samples such as WWTP, environmental and solid enteric samples; however, a sample concentration method was needed to improve its LOD. The postCl WWTP effluent samples showed no qPCR signals when the QIAamp protocol was used as recommended with only 200 μL of sample (LOD~1600 gene copies/mL). Using a MF protocol to concentrate 100 mL of sample improved the LOD to ~3.2 gene copies/mL for the QIAamp protocol. Overall, the QIAamp stool kit was considered to be the most suitable for our MST purposes.

The MF method tended to produce lower DNA concentrations than the centrifugation and the direct methods. This could be because lysed DNA passed through the pores of the MF and was thus not present during the amplification stage of the qPCR protocol. Human specific *Bacteroides* DNA markers were found to persist for eight days in sewage microcosms exposed to light and 28 days in similar microcosms in darkness (Tambalo et al. 2012; Walters et al. 2009). Thus the exclusion of lysed DNA is viewed as an advantage to the MST protocol since the trapped intact and more likely viable anaerobic *Bacteroides* microbial cells would indicate a more recent pollution event. Another advantage to the MF sample concentration process is that PCR inhibitory substances would also pass through the pores thus limiting their impact in the amplification stage.

We acknowledge the protocols' limitations and that no single MST assay is or may ever be ideal. However, our assays showed the pMR protocol that used a MF sample concentration aspect for liquid samples to be an improvement over at least two qPCR MST protocols obtained via literature review. Our assessment indicates the pMR protocol may be the best human specific *Bacteroides* 16S rRNA assay available for MST use. This would be highly beneficial to the MST community, utilities and regulatory entities.

# CHAPTER 5: EVALUATING THE POTENTIAL EXPOSURE TO WATERBORNE PATHOGENS BY USING QPCR MST

## Abstract

Preserving the integrity of drinking source water supplies and protecting recreational-water users from water-borne pathogens are essential in reducing waterborne health risks. Microbial source tracking attempts to make the process more effective by employing methods that seek to identify specific fecal source groups. A human specific *Bacteroides* qPCR protocol developed in this study showed improved specificity (99%) in comparison to another obtained from the literature (94%). Both assays showed sensitivities of 100%. Considering this finding, the protocol was then deployed to determine if the source of pollution impacting two creeks/tributaries was of human or non-human origin. The results of qPCR assays supported our expectations that these two creeks were not impacted by human sources of enteric pollution.

### 5.1 Introduction

Many pathogens are associated with feces and thus the contamination of water resources by fecal pollution represents a significant risk to the health of humans and animals. It is neither easy nor feasible to attempt to screen for each pathogen due to differences in their incubation times, growth requirements and economic constraints. Thus, fecal indicator bacterial (FIB) or coliforms are used to assess microbial health related water quality. Unfortunately, conventional methodologies of FIB enumeration provide no information as to the origins of the pollution or to the different types of host sources that contribute to the pollution.

Microbial source tracking (MST) employs methods that attempt to identify specific fecal source groups. The fundamental assumption of MST is that there exists a traceable association between microbes excreted in feces and their particular hosts. These associations could be due to adaptations of microbes to their host's gut environment and could include the coevolution of the microbes with their hosts. Thus, the objective of MST includes exploiting these associations to develop best management practices (BMPs) to control fecal contamination from relevant animal/human sources.

Preserving the integrity of drinking source water supplies and protecting recreational-water users from water-borne pathogens are essential in reducing waterborne health risks. In order to improve the effectiveness of mitigation procedures used by utilities and regulatory entities, there arose the need to find methods that differentiate human from non-human sources of fecal pollution. The 16S ribosomal RNA gene of *Bacteroides* spp. showed relative promise in this respect. These microbes are enteric anaerobes which are considered to be restricted to warm-blooded animals, and belongs to a phylum that makes up approximately 30 to 40% of human fecal microbial flora (Costello et al. 2009; Layton et al. 2006; USEPA 2005b). Although the HF183 human-specific primer set that targets *Bacteroides* spp. 16S rRNA sequences (Bernhard and Field 2000b) has been relatively successful (Ahmed et al. 2009b; Shanks et al. 2010b) there is still room for improvement since these studies show positive qPCR signals of the markers with other animals (Reischer et al. 2007; Shanks et al. 2010b).

Previously, we performed phylogenetic and bioinformatics analyses on 16S rRNA gene sequences associated with different animal host groups to determine if *Bacteroides* show co-evolutionary attributes. The results of these analyses provided good support for *B. dorei* human host specificity. Given our findings, a primer and probe set was designed for quantitative real-time polymerase chain reaction (qPCR) assays that specifically target bacterial sequences highly similar to *B. dorei* from human hosts. The main aim of this study was to evaluate the specificity and sensitivity of the qPCR protocol developed and to determine its applicability to determine human pollution in environmental water bodies.

## 5.2 Materials and Methods

### 5.2.1 Sample Collection

A total of 126 distinct animal fecal samples were collected from 11 animal types to assess the specificity of the qPCR protocol used in this study (Table 12). Dog fecal samples were collected from the Schuylkill River Park Dog Run and from the homes of colleagues. Cat fecal samples were collected from the homes of colleagues. All other animal fecal samples were collected from Fox Chase Farm (FCF) and the Schuylkill Center for Environmental Education (SCEE). All samples were transported to the MST laboratory in igloos with ice packs (<8$^{o}$C) for expeditious processing (<6hr after collection). FCF and SCEE are in close proximity to the environmental water bodies investigated. Thus, animal fecal samples were collected from these sites since they posed the most likely source of pollution at these creeks.

Human-associated influent samples collected from the Northeast wastewater treatment plant (NE-WWTP) were used to assess the sensitivity of the qPCR protocol. Samples were collected in 500 mL polypropylene (Fisher) bottles and transported to the lab in igloos with ice (<8$^{o}$C) and processes within 6 hours of collection.

Environmental water samples were collected from creeks/tributaries at SCEE and FCF. These sites were chosen primarily because they were not expected to be impacted by human enteric pollution. The SCEE sampling points were a pond at the Smith's Run

Creek and another point further downstream (Figures 19). The FCF sampling points were the headwaters and a pond along a tributary at FCF, and before and after the confluence of this tributary and the Pennypack Creek (Figures 20). Samples were collected less than one day after precipitation (wet event) and another period with precipitation greater than four days (dry event). Samples were collected in 1L polypropylene bottles at each site and transported to the lab in igloos with ice packs (<8°C), before processing within 6 hours of collection.

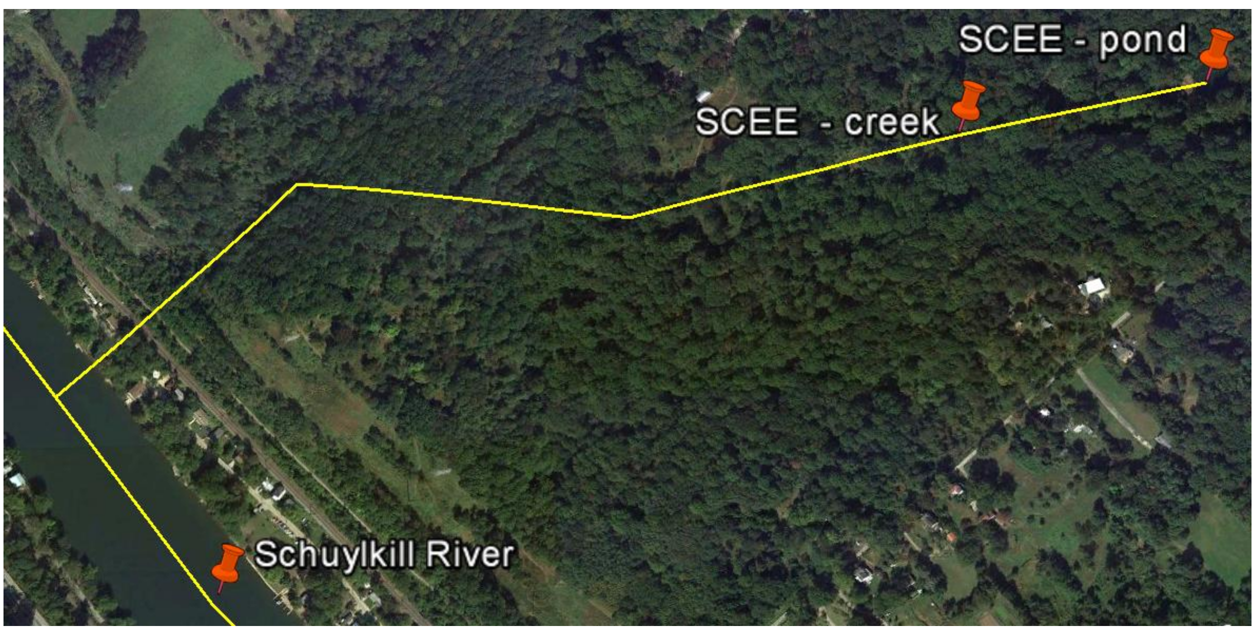



**Figure 19: Site A – Schuylkill Center for Environmental Education (SCEE) sampling points – wildlife impacted (image obtained using Google Earth)**

### 5.2.2 FIB Enumeration

*Enterococcus* enumeration that entailed membrane filtration and mEI cultivation in accordance with the EPA's standardized Method 1600 (USEPA 2005a) were conducted on each sample. All samples were simultaneously processed for DNA extraction.

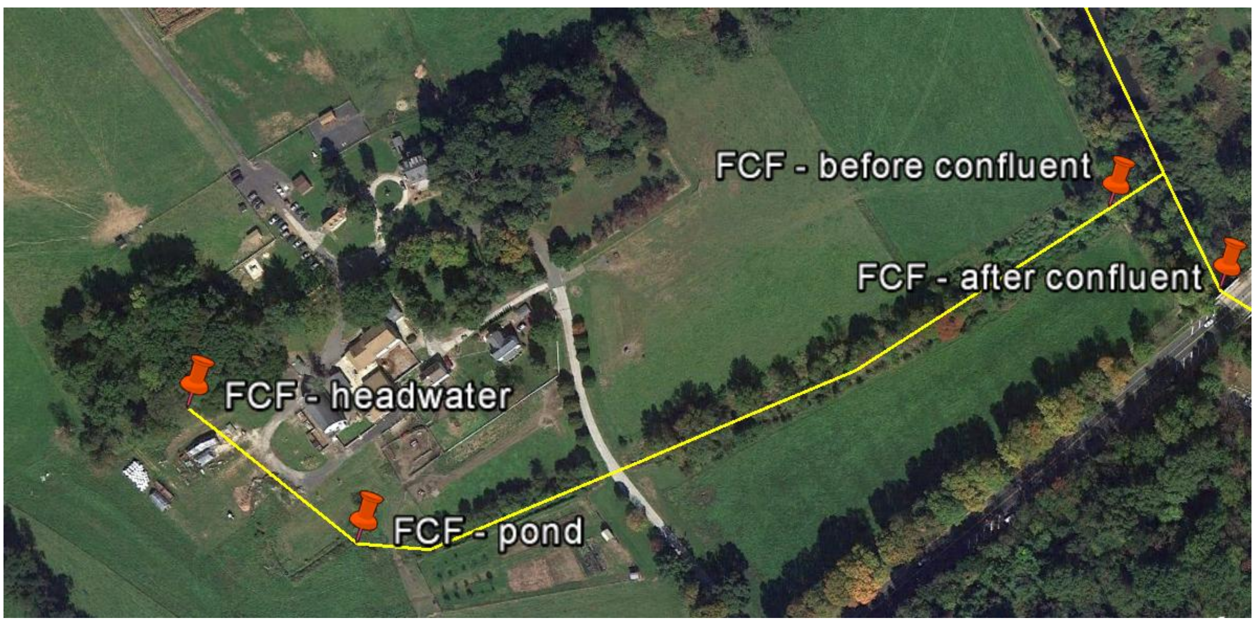



**Figure 20: Site B – Fox Chase Farm (FCF) sampling points – livestock impacted (image obtained using Google Earth)**

### 5.2.3 DNA Extraction

All DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's protocol, except for environmental water samples. In brief, mixed cellulose esters 0.45 μm membranes were used to vacuum filter environmental water samples. In all cases, dependent on turbidity, the maximum possible volume of the collected environmental water sample that could be filtered was attempted. The membranes were then placed in 2.0 mL safe lock microcentrifuge tubes (Qiagen) and 1.8 mL ASL buffer added. Samples were vortexed for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90$^{o}$C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples were added. The rest of the QIAcube QIAamp DNA Stool protocol was followed as outlined in the kit's manual. The concentration of the total DNA extraction yields for all samples were determined with a NanoDrop ND-000 UV spectrophotometer (Thermo Fisher Scientific) with absorbance readings at 260 nm, while the purity of the DNA preparations was determined by A260/A280 ratios. All extracted DNA were stored at -20$^{o}$C until analyzed by qPCR.

### 5.2.4 Standard Curves

A plasmid DNA construct was developed to function as a plasmid DNA standard for calculation of assay calibration curves. Amplicons obtained from the NE-

WWTP influent were purified with QIAquick Gel extraction kits (Qiagen) and then cloned; plasmids were then purified with Qiagen PCR cloning plus kits. The eluate was quantified and diluted to generate samples ranging from approximately 10 to 10e9 molecules of template DNA.

### 5.2.5 qPCR Protocol

The qPCR protocol from this study used a primer and probe set previously developed in this study. The forward primer, HF68, (5′-GGC AGC ATG GTC TTA GCT TG-3′) and the reverse complement of the HF183 primer (HF183rc) - (5′-CGG ACA TGT GAA CTC ATG AT-3′) were combined with a Roche UPL probe #156 (5′-GCT GAT GG-3′) for human specific qPCR assays. The HF183 protocol used for comparative analysis of the sensitivity and specificity tests was obtained from a previous study (Shanks et al. 2010). This protocol included the forward primer (5′-ATC ATG AGT TCA CAT GTC CG-3′), the reverse primer BthetR1 (5′-CGT AGG AGT TTG GAC CGT GT-3′) and the probe [6~FAM] CTG AGA GGA AGG TCC CCC ACA TTG GA [TAMRA~6~FAM]. All primers and probes, except for the Roche UPL probe, were obtained from Eurofins MWG Operon.

A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. The protocol for all assays used the Roche LC480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA and 0.5 μM of the forward and reverse primers. The program employed pre-incubation for 5 min at 95°C; 45 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, and each run included both no template controls (PCR-grade water) and positive controls (WWTP extracted DNA).

## 5.3 Results

### 5.3.1 Specificity/Sensitivity

Both MST qPCR protocols showed 100% sensitivity for the NE-WWTP influent samples with gene copies/100 mL of sample averaging orders of magnitude of 10e8. However, the protocol developed in this study (HF68-HF183rc) showed specificity of 99%, while the HF183- BthetR1 protocol showed 94% (Table 12). One dog sample was positive for both assays and six pigs for the HF183- BthetR1 assay

### 5.3.2 Environmental Samples

All of the FCF and SCEE samples showed *Enterococcus* presence (Table 13). The SCEE sites were expected to be impacted by a small population of wildlife (deer, Canadian goose etc.), as was indicated by the low enterococci counts. As expected, the FCF samples showed high enterococci concentrations. The FCF sites were highly likely to be impacted by livestock (cow, goat, sheep and horse) from the farm and to a lesser extent wild avian species (Canadian goose and duck). The site located after the confluence of the FCF tributary and the Pennypack was likely to be impacted by any run-off, storm water outfalls (SWO) located upstream, and by the FCF tributary. Samples collected during the wet event showed greater enterococci counts for the FCF samples than those collected during the dry event.

**Table 12: Table showing the number of samples assayed of the NE-WWTP, each animal type, and the number positive for the HF68-HR183 human specific *Bacteroides* primer sets.**

| Sample/Animal type | # of distinct samples assayed | # of distinct samples positive for HF68-HR183 |
|---|---|---|
| **WWTP influent** | 20/20 | 20/20 |
| **Cow** | 0/28 | 0/28 |
| **Sheep** | 0/18 | 0/18 |
| **Pig** | 0/15 | 6/15 |
| **Horse** | 0/16 | 0/16 |
| **Dog** | 1/13 | 1/13 |
| **Cat** | 0/8 | 0/8 |
| **Goose** | 0/20 | 0/20 |
| **Goat** | 0/25 | 0/25 |
| **Deer** | 0/1 | 0/1 |
| **Duck** | 0/1 | 0/1 |

**Table 13: Enterococcus enumeration (CFU/100mL) and average number of *Bacteroides* 16S rRNA gene (copies/100 mL) of the Schuylkill Center for Environmental Education (SCEE) and the Fox Chase Farm (FCF) environmental samples.**

| Sampling Point | Day 1 - dry event | | Day 2 - wet event | |
|---|---|---|---|---|
| | mEI - Average CFU/100mL | Average gene copies / 100 mL | mEI - Average CFU/100mL | Average gene copies / 100 mL |
| **Smith's Run Creek (SCEE) – stream** | 66 | <4 | 32 | <4 |
| **Smith's Run Creek (SCEE) – pond** | Not collected | <4 | Algal crowding | <4 |
| **Fox Chase Farm (FCF) – head water** | 146 | <4 | 81 | <4 |
| **Fox Chase Farm (FCF) – pond** | 208 | <4 | 262 | <4 |
| **Fox Chase Farm (FCF) – before confluence** | 324 | <4 | >800 | <4 |
| **Pennypack after Fox Chase Farm confluence** | 67 | 54 | >400 | 358 |
| **Pre and Post MF controls** | <1 | <4 | <1 | <4 |

The qPCR results show no HF68-HR183 markers in the samples from the SCEE or the FCF creeks/tributaries (Table 13). However, Pennypack Creek showed the markers in the samples from the site that was located after its confluence with the Fox

Chase Farm tributary. In addition, samples collected during the wet event showed greater numbers of the HF68-HR183 marker for this site than those collected during the dry event. Higher enterococci enumeration levels did not imply human pollution as was shown with the FCF before and after confluence results. The samples taken before the confluence had higher enterococci counts than those taken after; however, only the latter showed the human markers.

**5.4 Discussion**

Numerous MST studies have investigated *Bacteroides* spp. as prime candidates for differentiating host groups using PCR-based techniques (Ahmed et al. 2009a; Ahmed et al. 2009b; Bernhard and Field 2000a; Bernhard and Field 2000b; Haugland et al. 2005; Haugland et al. 2010; Santo Domingo et al. 2003; Scott et al. 2005; Shanks et al. 2010b; USEPA 2005b). Some of these studies show cross-reaction of the human-specific assays with known non-human associated samples. An MST qPCR assay performed with the HF68-HR183rc qPCR protocol developed during this study showed improved human specificity (99%) in comparison to using the HF183- BthetR1 protocol (94%). The protocols showed 100% sensitivity for the WWTP influent samples with gene copies/100 mL of sample averaging orders of magnitudes of 10e8.

The standard *Enterococcus* enumeration shows fecal pollution at all of the environmental sites. The SCEE sites were expected to be minimally impacted by a small population of wildlife (deer, Canadian goose etc.) as was indicated by the low enterococci counts. On the other hand, the FCF samples showed high enterococci concentrations. The FCF sites were highly likely to be impacted by livestock (cow, goat, sheep and horse) from the farm and to a lesser extent wild avian species (Canadian goose and duck).

None of the samples from the two sites showed the presence of the HF68-HR183rc markers except for the samples collected after the confluence of the FCF and the Pennypack creek. The absence of the human molecular marker was expected since these sites were not likely to be impacted by human enteric pollution. On the other hand, the Pennypack site was considered to be impacted by run-off and permitted suburban wastewater discharges located upstream of the sampling site, in addition to pollution of the FCF tributary itself. The samples collected during the dry weather event indicate a possible base level of human enteric source of pollution of the Pennypack, or the possibility of dry weather sources. The wet weather samples showed an elevation in the HF68-HR183rc markers, which indicate that storm water outfalls (SWO) could be the likely contributor to the elevation observed. Of note is that higher enterococci enumeration levels did not imply human pollution as was shown with the FCF before and after confluence results. This finding highlighted a limitation in using the conventional enterococci test to indicate human enteric pollution.

Overall, the results of qPCR assays supported our expectations that these two creeks/tributaries were not impacted by human sources of enteric pollution. This is the first study to conduct MST of these creaks using *Bacteroides* genetic markers. Future work will include the continued evaluation of the HF68-HR183 qPCR protocol for its general MST applicability and its particular suitability for use in the Philadelphia watershed. Additional evaluation of the Pennypack is required to confirm human enteric pollution of this creek and to ascertain the point/s of the intrusion/s. It should be noted that no single primer set may ever be ideal to determine host pollution, thus a toolbox approach is supported. The use of the *B. dorei* human specific marker as one of the instruments in this toolbox is supported by this work.

# CHAPTER 6: USING THE QPCR MST PROTOCOL TO DETERMINE LEVELS OF HUMAN FECAL POLLUTION IN CREEKS WITH UNCERTAIN SOURCES.


## Abstract

Numerous studies have shown *Bacteroides* spp. human specific MST protocols to be able to differentiate human from non-human enteric pollution. This study's *Bacteroides* qPCR MST protocol was used investigated the impact of human enteric pollution of five waterbodies in the Philadelphia watershed. The results of assays for the human specific *Bacteroides* markers indicated that all these urban watersheds were possibly impacted by human enteric pollution. Elevated enterococci and *Bacteroides* presence found at the environmental sites during wet weather event samples compared to that during corresponding dry weather event samples highlight the possible contribution of fecal pollution from run-off, CSOs and/or SWOs during wet weather events. The qPCR MST protocol provided better information than conventional mEI method; however, both used together give better inferences.

**6.1 Introduction**

Many waterborne diseases arise from exposure to waters contaminated with enteric waste and pose a significant risk to public health. The ability to identify the source of enteric pollution impacting a watershed is necessary for regulatory organizations and utilities to evaluate human health risks associated with recreational and potable source waters and to determine the appropriate remedial actions. Of particular interest to officials of cities similar to Philadelphia are combined sewer overflows (CSOs) and sanitary sewer overflows (SSOs) which were among the earliest constructed in the US cities. It is estimated that the annual discharge of CSOs in the US is 850 billion gallons (USEPA 2004). These combined sewer systems (CSSs) runoff contain raw sewage capable of harboring human pathogens (*Cryptosporidium*, *Vibrio cholerae*, noroviruses, etc.) and toxic pollutants (hormones, antibiotics, etc.) (USEPA 2004) and are thus of concern.

Current USEPA regulations recommend the used of *Escherichia coli* and *Enterococcus* spp. as standard indicators of fecal pollution for freshwater and marine water, respectively, and total coliform as indicators of the integrity of distribution systems (USEPA 2010b; USEPA 2012). However, these methods are unable to differentiate human from other sources of pollution. Thus, a number of microbial source tracking (MST) studies have developed molecular methods to identify and quantify human sources of pollution (Ahmed et al. 2008b; Kildare et al. 2007; Layton et al. 2006; Okabe et al. 2007; Seurinck et al. 2005). The more successful of these methods are based on quantitative polymerase chain reaction (qPCR) targeting the 16S rRNA gene of *Bacteroides* spp. These are enteric anaerobes that are thought to be restricted to warm-blooded animals and make up approximately 30 to 40% of human fecal microbial flora (Costello et al. 2009; Layton et al. 2006; USEPA 2005b). Although one particular protocol, known as the HF183 assay, has been relatively successful as a human specific MST assay, it has also shown PCR amplification of samples of other animal host origin (e.g. dog, cat, pig, chicken and sheep) (Ahmed et al. 2009a; Ahmed et al. 2009b; Layton et al. 2006; Okabe et al. 2007; Shanks et al. 2010b), indicating that there is still room for improvement in the specificity of human specific MST assays.

Our study has developed a *Bacteroides* qPCR MST protocol that has shown improved success of differentiating human vs. non-human sources of enteric pollution. The method was used to assay five waterbodies in the Philadelphia watershed that were of concern to the Philadelphia Water Department (PWD). These sites were impacted by enteric pollution of unknown origin and the PWD needed to determine if CSO, compromised sanitary laterals, storm sewers with illegal sanitary connections or other potential human sources were contributing to the pollution.

**6.2 Method**

### 6.2.1 Sampling Sites

The majority of the sites investigated were from the Wissahickon Creek (Figure 21) watershed located in southeastern Pennsylvania, falling within Montgomery and Philadelphia counties. The creek's headwater is located in Montgomery Township and the watershed drains approximately 64 square miles (PWD 2007b). It is a major tributary of the Schuylkill River and is comprised of approximately 134 linear miles of stream.

The area comprises a population of ~160,000 residents and spans 15 municipalities and the City of Philadelphia (PWD 2007b). More than 50% of the land area is comprised of residential development and over 30% of the watershed is impervious cover (PWD 2007b). This watershed is particularly important because the Queen Lane Water Treatment Plant (QLWTP) is located ~0.5 miles downstream of the Wissahickon Creek's confluence with the Schuylkill River.

The remaining samples were taken from the Poquessing Creek watershed (Figure 22) which is comprised of ~22 square miles and is the smallest of Philadelphia's major watersheds and is a tributary of the Delaware River (PWD 2007a).

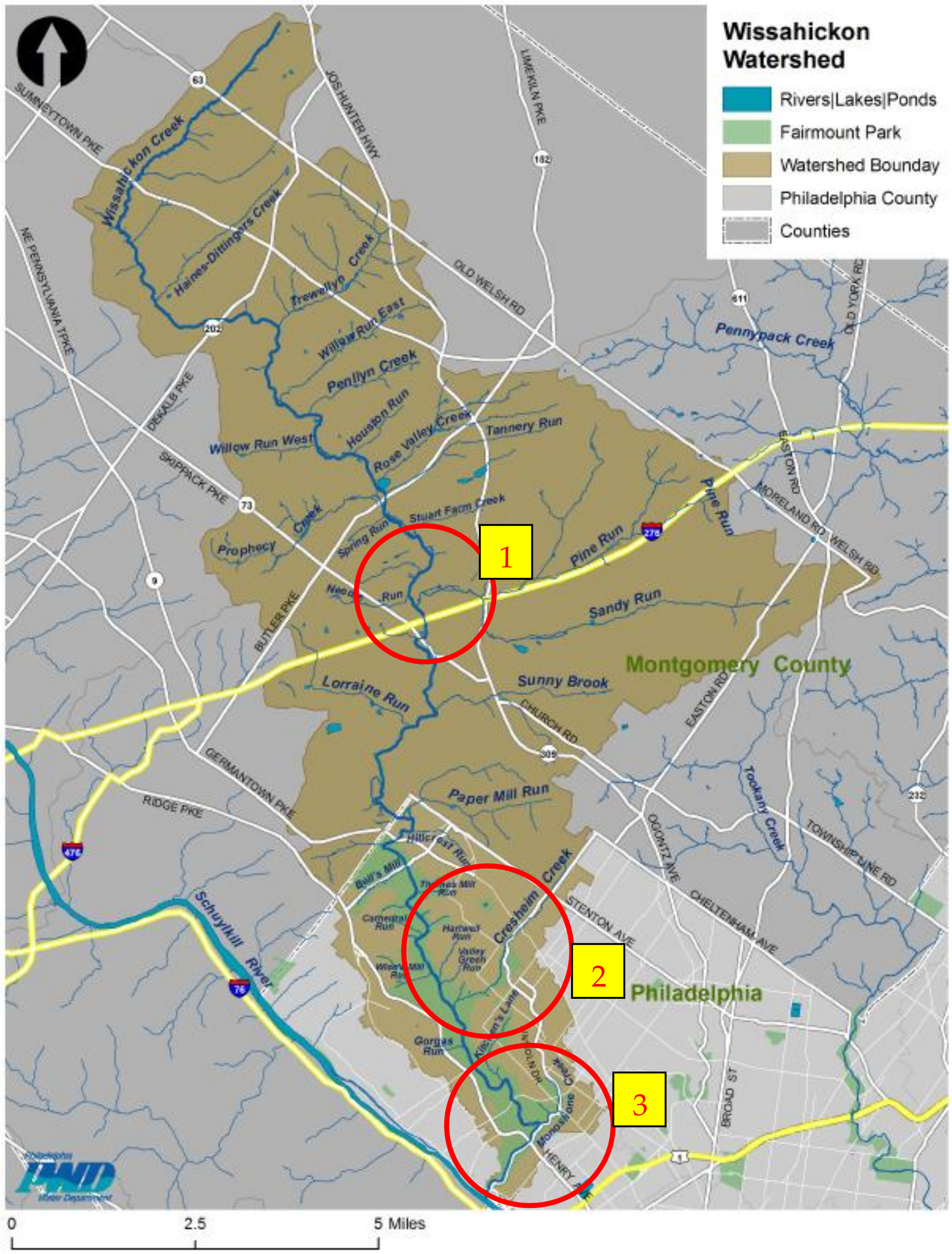



**Figure 21: Map showing the outlines of the waterbodies and counties comprising the Wissahickon Watershed (PWD 2012b). Sample sites circled red for; Sandy Run creek & the city limit (1), Valley Green & Cresheim creeks (2), and Monoshone creek and Ridge Avenue (3).**

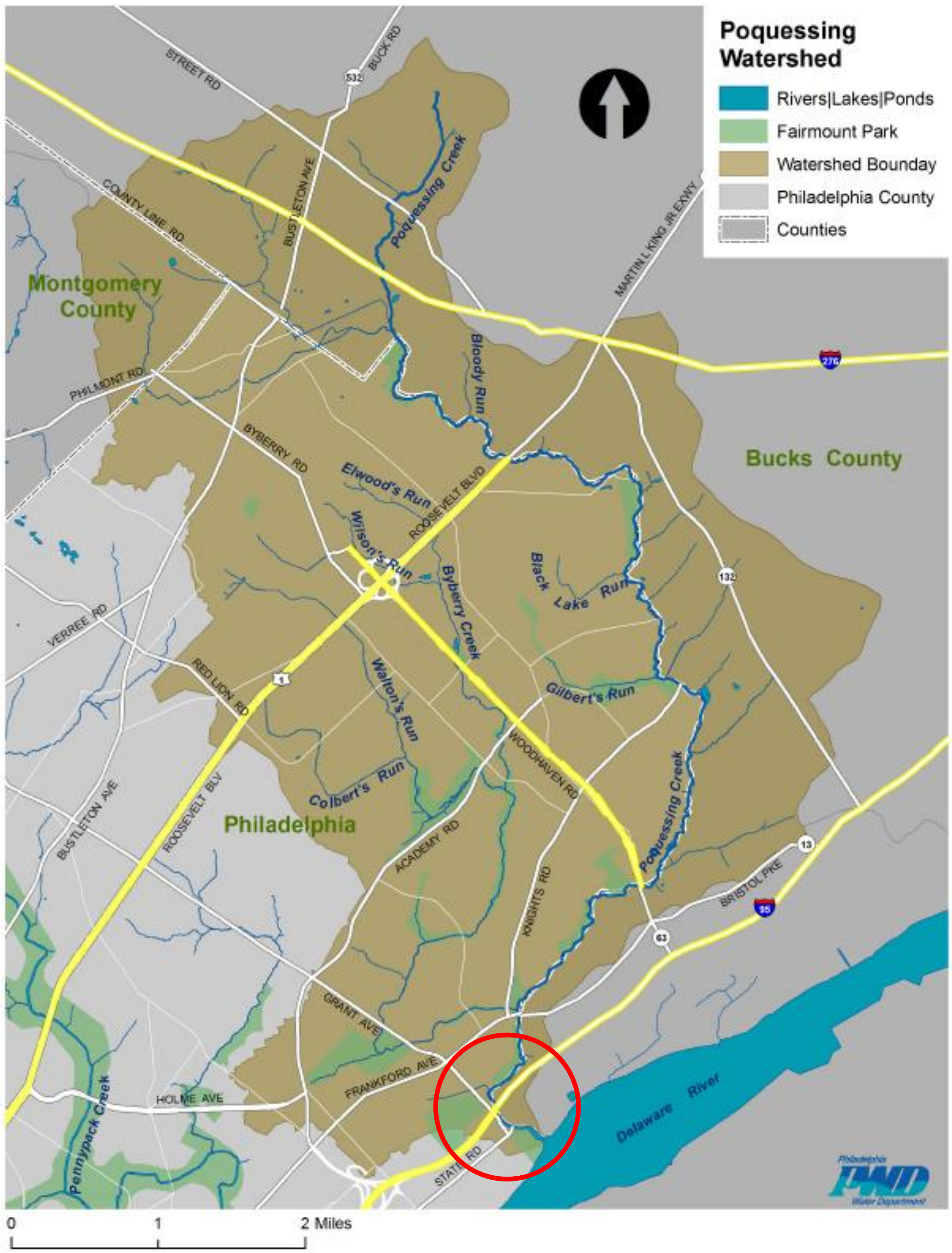



**Figure 22: Map showing the outlines of the waterbodies and counties comprising the Poquessing Watershed (PWD 2012a). Sample sites for the Poquessing circled red.**

A) B)

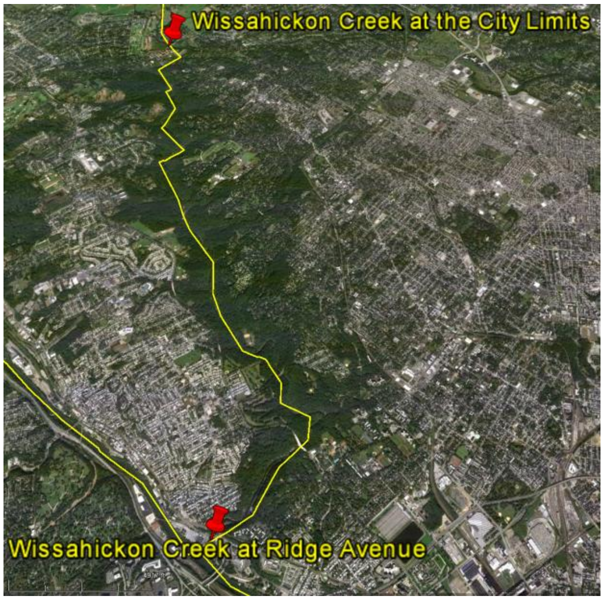



**Figure 23: A) Google Earth map of the Ridge Avenue and City Limits sties. B) Wissahickon Creek at Ridge Avenue Right (source: the author).**

The area houses a population of ~105,000 and spans Montgomery, Bucks and Philadelphia counties. Approximately 50% of the land area is comprised of residential development and over 38% of this watershed is impervious cover (PWD 2007a).

The six waterbodies investigated were; the Sandy Run, the Monoshone, the Valley Green Run, the Cresheim, the Wissahickon and the Poquessing Creeks (Figures 23-28). Sandy Run (8.1 stream miles), Valley Green Run (0.1 stream miles), Cresheim (3.1 stream miles) and Monoshone (1.3 stream miles) are tributaries of the Wissahickon.

The sampling points from the Wissahickon Creek watershed (Figures 23 to 26) were: Wissahickon Creek at Ridge Avenue (WS005) (GPS coordinates - 40.014623, -75.206871), Wissahickon Creek at the City Limits (Northwestern Avenue bridge - WISS400) (40.08884, -75.22781), Wissahickon Creek above the Sandy Run Creek (Germantown Academy school on Morris Rd - WS1210) (40.13887, -75.21687) , Sandy Run Creek – WSSR058 (40.13358, -75.21359), Monoshone headwaters along W. Johnson Street by Lincoln Drive (40.036872, -75.191422), Monoshone stormwater outfall (SWO) along Lincoln Drive (40.034897, -75.188717), Monoshone Creek before it enters the Wissahickon Creek (WSMC001) (40.027611, -75.192721), Wissahickon Creek above the Monoshone Creek confluence (40.027697, -75.193042), Valley Green Road Creek (40.056722, -75.214003), the Wissahickon Creek above and below its confluence with the Valley Green Road creek (40.055649, -75.218056), Cresheim Creek at Cresheim Valley Drive (40.062725, -75.200328), the Wissahickon Creek at Livezey Lane below its confluence with Cresheim Creek (40.049230, -75.213605).

The sampling points from the Poquessing Creek watershed (Figures 27 & 28) were: the Poquessing Creek at Bayswater Drive a few meters from the Delaware River confluence (40.053621, -74.979860), under the I95 overpass along Grant Avenue (40.055873, -74.984403), and at Red Lion Road just downstream of the confluence of two tributaries of the Poquessing (40.065336, -74.981019).

### 6.2.2 Sample Collection

Samples were collected from the environmental sites on different days as depicted in Tables 14 & 15. Five sets of samples were collected from the Monoshone-associated sites, four sets from the Ridge Avenue- and the Sandy Run Creek-associated sites, and two sets from the Cresheim Creek- and the Poquessing Creek-associated sites. Most of the samples were collected within a 48 hour duration that had a precipitation event except for those collected on June 27$^{th}$, 2011. Precipitation events were confirmed by the Pennsylvania State Climatologist (KPNE) website.

Samples were collected in 500 mL or 1L polypropylene bottles at each site and transported to the lab in igloos with ice packs (<8$^{o}$C) and processed for DNA extraction and enterococci enumeration within 6 hours of collection.

### 6.2.3 FIB Enumeration

*Enterococcus* enumeration by membrane filtration and mEI cultivation was performed in accordance with the EPA's standardized Method 1600 (USEPA 2005a) for each sample. The volume of sample filtered was dependent on the turbidity of the samples, which dictates the ease and time for filtration and the expected counts based previous results. Typically, volumes of 1, 10 and 100 mL of environmental water sample were membrane filtered for enumeration. All samples were simultaneously processed for DNA extraction.

A)


B)

C)

D)

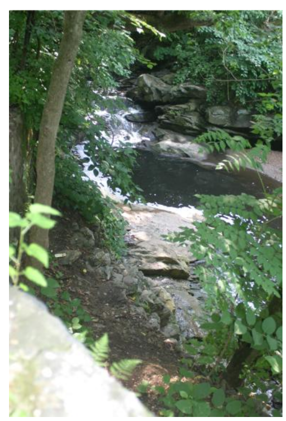
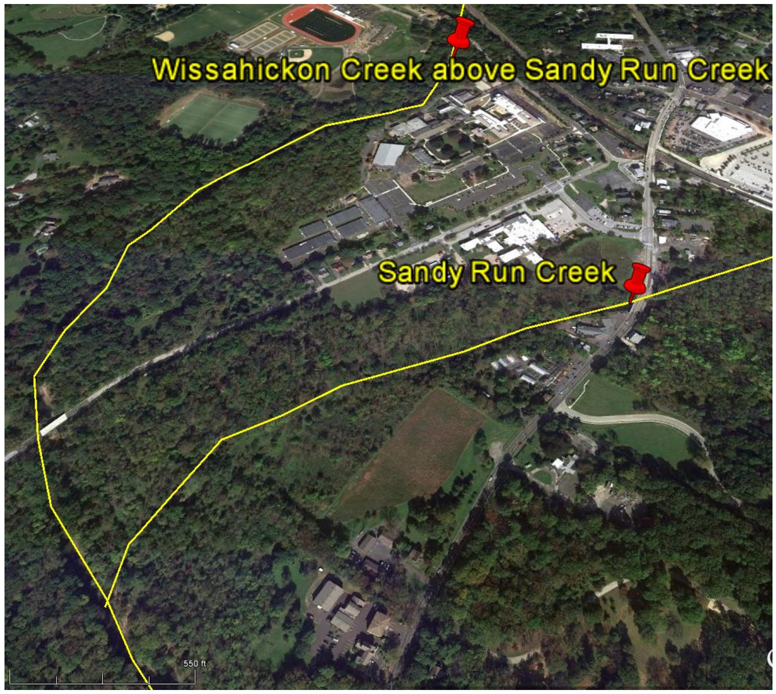


**Figure 24: - A) Google Earth map of the Monoshone sites. B) Monoshone Creek headwaters (source: the author). C) Storm water outfall (SWO) at the Monoshone Creek along Lincoln Drive (source: Gary Burlingame). D) Monoshone Creek before it enters the Wissahickon (source: the author).**

A)




B)

C)

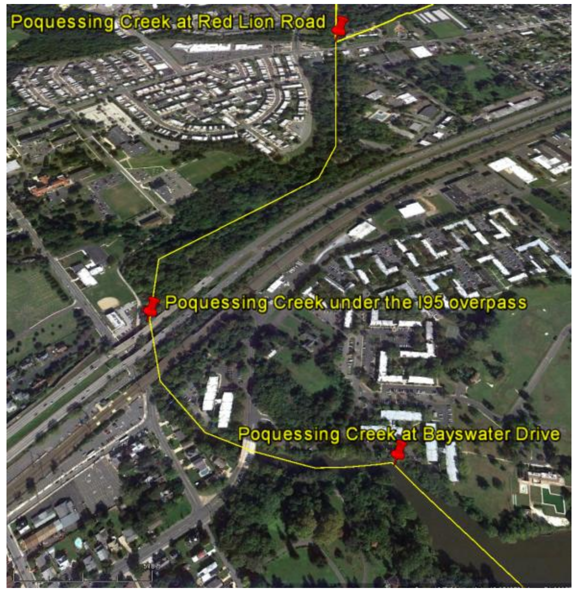
**Figure 25: A) Google Earth map of the Sandy Run sites. B) Sandy Run Creek (source: the author). C) Wissahickon Creek above Sandy Run (source: the author).**

A)


B) C)

**Figure 26: A) Google Earth map of the Valley Green and Cresheim Creek sites. B) Valley Green Road Creek (source: the author). C) Wissahickon Creek and Valley Green Road confluence (source: the author).**

A)




B)

C)

**Figure 27: A) Google Earth map of the Poquessing sites. B) Low tide at the Bayswater Drive Poquessing Creek site a few meters from the Delaware River (source: the author). C) Poquessing Creek site along Grant Avenue under the I95 highway overpass (source: the author).**

A) B)

**Figure 28: At the confluence of two creeks of the Poquessing at Red Lion Road (source: the author).**

### 6.2.4 DNA Extraction

All DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's protocol except for environmental water samples. In brief, 0.45 μm mixed cellulose ester membranes were used to vacuum filter environmental water samples. In all cases, dependent on turbidity, the maximum possible volume of the collected environmental water sample that could be filtered was attempted. Typically, 50 mL to 500 mL of environmental water sample was filtered for DNA extraction. The membranes were then placed in 2.0 mL microcentrifuge tubes (Qiagen) and 1.6 mL ASL buffer added.

Samples were vortex for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90°C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAcube QIAamp DNA Stool protocol was followed as outlined in the kit's manual. The concentrations of the total DNA extraction yields for all samples were determined with a NanoDrop ND-000 UV spectrophotometer (Thermo Fisher Scientific) with absorbance readings at 260 nm, while the purity of the DNA preparations was determined by A260/A280 ratios. All DNA extractions were stored at -20°C until qPCR analysis.

### 6.2.5 Standard Curves

A plasmid DNA construct was developed to function as a plasmid DNA standard for calculation of assay calibration curves. Amplicons obtained from WWTP influent were purified with QIAquick Gel extraction kits (Qiagen); cloned and plasmids purified with Qiagen PCR cloning plus kits. The eluate was quantified and diluted to generate samples ranging from approximately 1.0E+01 to 1.0E+09 molecules of template DNA.

### 6.2.6 qPCR Protocol

The human specific qPCR MST protocol (designated pMR) used a primer and probe set previously developed in this dissertation. The forward primer HF68 (5′-GGC

AGC ATG GTC TTA GCT TG-3’) and HF183rc, the reverse complement of the original forward primer HF183 (5′ – CGG ACA TGT GAA CTC ATG AT – 3′), were combined with a Roche UPL probe #156 (5’-GCT GAT GG-3’) for human specific qPCR assays. A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. The protocol for all assays used the Roche LC480 Probe Master reagents in concentrations in accordance with the manufacturer’s instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA and 0.5 μM of the forward and reverse primers. The program employed: pre-incubation for 5 min at 95°C; 45 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, including standard curves, no template controls and plasmids as positive controls.

**6.3 Results**

*Enterococcus* was detected at most of the sites. Enumeration results (Table 14) for the wet weather events showed statistically higher enterococci presence than that of the corresponding dry weather event samples (Wilcoxon Rank Sum test, p-values <0.05). The Monoshone Creek headwaters (Figure 24 A) showed low enterococci presence for both the wet and dry weather events. No *Enterococcus* enumeration results were obtained for the Cresheim Creek- and the Poquessing Creek-associated samples due to defective mEI media.

The human specific *Bacteroides* qPCR results for the Wissahickon, Monoshone and Sandy Run Creek samples are shown in Table 15. The samples obtained from the Wissahickon Creek at the City Limits (Northwestern Avenue Bridge) and the Monoshone headwaters showed no detectable human pollution during wet or dry weather events. The samples from Ridge Avenue and the Wissahickon Creek above the Monoshone Creek confluence showed high concentrations of the pMR targets during wet weather events and none during the dry weather event.

The samples from the Monoshone Creek before it enters the Wissahickon Creek, the Wissahickon Creek above Sandy Run Creek, and the Sandy Run Creek showed amplification of the human-specific *Bacteroides* targets during both the wet and the dry weather events. The numbers of the *Bacteroides* gene copies were higher for the wet than the dry weather event (Wilcoxon Rank Sum test, p-values $<0.10$). Given the small sample size and the exploratory nature of this comparison, a $\alpha$-value of 0.10 was used for this statistical evaluation.

**Table 14: Table showing sample location, enterococci CFU/100mL of sample for mEI membranes and the corresponding gene copy #/mL of sample for qPCR of DNA extracted directly from filtered samples.**

| | 2011/06/10 (wet) | | 2011/06/27 (dry) | | 2011/07/25 (wet) | | 2011/08/10 (wet) | | 2011/08/18 (wet) | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sample location** | CFU/100 mL | gene copy #/mL | CFU/100 mL | gene copy #/mL | CFU/100 mL | gene copy #/mL | CFU/100 mL | gene copy #/mL | CFU/100 mL | gene copy #/mL |
| **WS005 - Wissahickon Creek at Ridge Avenue** | >200 | 2.18E+02 | 10 | <3.60E+00 | >400 | <1 | >400 | 3.03E+01 | - | - |
| | - | - | - | - | - | - | >400 | 9.89E+01 | - | - |
| **WISS400 - Wissahickon Creek at the City Limits (Northwestern Avenue bridge)** | >200 | <3.60E+00 | 140 | <3.60E+00 | >400 | <1 | 178 | <1 | - | - |
| | - | - | - | - | - | - | 200 | <1 | - | - |
| **WS1210 - Wissahickon Creek sample above Sandy Run (Germantown Academy school on Morris Rd)** | >400 | 3.13E+01 | 50 | 1.33E+00 | >400 | 1.35E+03 | >400 | 7.11E+02 | - | - |
| | - | - | - | - | - | - | >400 | 3.61E+02 | - | - |
| **WSSR 058 - Sandy Run Creek** | >400 | 1.11E+03 | 160 | 0.647E+00 | >400 | 1.95E+02 | >400 | 3.05E+03 | - | - |
| | - | - | - | - | - | - | >400 | 8.43E+03 | - | - |
| **Monoshone headwaters along W. Johnson Street by Lincoln Drive** | 48 | <3.60E+00 | 12 | <3.60E+00 | >400 | <1 | 82 | <1 | 190 | <1 |
| | - | - | - | - | - | - | 82 | <1 | 196 | <1 |
| **Monoshone before SWO** | - | - | - | - | >400 | 4.00E+02 | >400 | <1 | >20,000 | <1 |
| | - | - | - | - | - | - | >400 | <1 | >20,000 | <1 |
| **Monoshone after SWO** | - | - | - | - | >400 | 1.83E+05 | >400 | 2.25E+05 | >20,000 | 4.44E+05 |
| | - | - | - | - | | | >400 | 2.20E+05 | >20,000 | 6.18E+05 |
| **WSMC001 - Monoshone Creek before it enters the Wissahickon Creek** | >400 | 3.18E+03 | 140 | 0.319E+00 | >400 | 3.53E+02 | >400 | 1.37E+03 | >20,000 | 1.35E+05 |
| | - | - | - | - | - | - | - | - | >20,000 | 1.31E+05 |
| **Wissahickon Creek above Monoshone Creek confluence** | >400 | 5.41E+01 | 40 | <3.60E+00 | >400 | <1 | >400 | 2.50E+02 | 2300 | 3.65E+02 |
| | - | - | - | - | | | - | - | 5000 | 9.48E+02 |
| **Pre & Post MF Control** | <1 | <3.60E+00 | <1 | <3.60E+00 | <1, <1 | <1, <1 | <1, <1 | <1, <1 | <1, <1 | <1, <1 |

**Note: Duplicate samples were taken at each site. SWO – storm water outflow**

Samples taken a few yards upstream of the Monoshone SWO along Lincoln Drive (Figure 24 A) showed no detectable human pollution except on 2011/07/25 (wet event) when it showed low concentrations (average ~400 gene copies/mL). In contrast, samples taken a few yards downstream of the SWO showed very high concentrations of the pMR targets on all occasions (order of magnitude 1.0E+05). The presence of the human specific *Bacteroides* targets just upstream of the SWO may be due to backflow during the wet event.

**Table 15: Table showing the gene copy/mL for the environmental samples assayed with the human specific qPCR MST protocol.**

| Sampling Sites | 2012/02/26 (wet) gene copy #/mL | 2012/03/14 (wet) gene copy #/mL |
|---|---|---|
| Cresheim Creek at Cresheim Valley Drive | 1.37E+04 | 5.06E+04 |
| Valley Green Road Creek 1 | - | 3.66E+05 |
| Valley Green Road Creek 2 | - | 7.65E+04 |
| Valley Green Road Creek 1 & 2 mix | 4.73E+01 | - |
| Wissahickon above Valley Green Road Creek confluence | 7.20E+01 | 4.12E+01 |
| Wissahickon Creek below VG confluence | 7.73E+01 | 5.32E+01 |
| Wissahickon Creek at Livezey Lane | 1.08E+02 | 4.31E+01 |
| Poquessing Creek at Red Lion Road | 2.19E+03 | 1.14E+03 |
| Poquessing under I95 Bridge | 1.81E+03 | 9.87E+02 |

| | | |
|---|---|---|
| Poquessing at Bayswater Drive | 1.20E+03 | 1.55E+03 |
| Sample Blank 1L PBS | <1E+00 | <1E+00 |

**Note: "-" implies not collected. All values are averages of triplicates.**

The Wissahickon Creek samples indicate a background level of the human specific markers (orders of magnitude 10s to 100s). The pMR protocol showed the Cresheim Creek at Cresheim Valley Drive and the Valley Green Road sites with higher concentrations (orders of magnitude 1000s to 10000s) of the human specific targets than that of the Wissahickon Creek samples. The Poquessing Creek associated sites all showed relatively similar concentrations of the human specific targets in the samples collected.

**6.4 Discussion**

This study investigated the impact of human enteric pollution on five different environmental waterbodies within the Philadelphia watershed. These creeks were of concern to the PWD and were reported to show elevated levels of enteric pollution via enterococci enumeration. The conventional method of enterococci enumeration (mEI plating) does not identify particular hosts/sources of the pollution, thus, MST protocols were needed. Numerous studies have shown *Bacteroides* spp. human specific MST protocols to be able to differentiate human from non-human enteric pollution (Ahmed et al. 2009a; Ahmed et al. 2009b; Bernhard and Field 2000a; Bernhard and Field 2000b; Haugland et al.

2005; Haugland et al. 2010; Santo Domingo et al. 2003; Shanks et al. 2010b). Here, we used the improved human specific *Bacteroides* qPCR MST protocol (pMR) developed during this dissertation to assay samples.

The elevated enterococci and *Bacteroides* presence found at the environmental sites during wet weather event samples compared to that during corresponding dry weather event samples highlight the possible contribution of fecal pollution from run-off, CSOs and/or SWOs during wet weather events. However, assays for the human specific *Bacteroides* markers suggest that there were no sources of human enteric pollution at the sites located at Wissahickon Creek at the City Limits (Northwestern Avenue bridge), or sites above the Monoshone SWO. This finding highlights one of the limitations of the conventional enterococci protocol where it indicated enteric pollution at the Monoshone headwater when none is expected to be there. Enterococci presence may be a result of persistence and/or growth in sediments and the water column from past exposure and not necessarily indicative of recent pollution (Bae and Wuertz 2012; Bergholz et al. 2011).

The results of assays for the human specific *Bacteroides* markers indicated that the Wissahickon Creek was possibly impacted by human enteric pollution from the Monoshone, especially during rain events. After such an event, samples collected from the Monoshone downstream of the SWO showed enterococci levels >20,000 CFU/100 mL and *Bacteroides* 16S rRNA gene copies of the order of magnitude above 1.0E+05. These levels indicate a potentially serious problem with the Monoshone SWO because these levels were similar to those obtained from the Northeastern Waste Water Treatment Plant influent samples. These findings indicate the Monoshone SWO is possibly impacted by compromised sewage laterals or other human sources located along this watercourse.

To a reduced extent, the pMR results for the Sandy Run (order of magnitude above 1.0E+03), the Valley Green Road Creek (order of magnitude above 1.0E+04) and the Cresheim Creek (order of magnitude above 1.0E+04) indicated these sites could possibly be sources of human enteric pollution of the Wissahickon Creek during rain events. All the Poquessing Creek samples indicate this Creek was a possible source of human enteric pollution to the Delaware. The presence of human enteric pollution in the Philadelphia urban watershed was similar to another study conducted on four watersheds in Milwaukee (Sauer et al. 2011). Our results, combined with those of the Sauer et al. (2011) study, illustrate stormwater as sources of pollution of creeks and rivers in urban watersheds. The results also indicate a possible background level of the human specific *Bacteroides* targets in the Wissahickon Creek during dry conditions. This indicates some persistence of the *Bacteroides* organism or genetic material in this water body, or a source of human pollution during dry conditions.

Overall, the qPCR MST protocol provided better information than the conventional mEI method. Results of assays of human specific *Bacteroides* imply a direct relationship between exposure to human pollution and the potential associated elevation of health risk. However, the use of cultivable and molecular methods together gives better inferences. Although an elevated human health risk is associated with human enteric pollution (Scott et al. 2002), the risks attributed to animal enteric pollution (Ahmed et al. 2012b) cannot be overlooked. In turn, both methods used together in a toolbox would enable PWD to make better decisions on its remediation and mitigation programs. The best approach to solving pollution issues in the Philadelphia watershed may include a long term systematic monitoring of all its watersheds and the inclusion of the human specific *Bacteroides* protocol as a part of a toolbox of methodologies.

# CHAPTER 7: DETERMINING THE POTENTIAL IMPACT OF HUMAN ENTERIC POLLUTION IN URBAN DRAINAGE SYSTEMS

## Abstract

Microbial and chemical contaminants are entrapped by stormwater as it runs over roads, rooftops, and compacted land. Thus, there is the possible unintentional increased risk of exposure of urban drainage system handlers to microbial pathogens. This preliminary study assessed samples collected from five urban drainage systems to determine if these sites were impacted by human enteric pollution. Our findings indicate the possibility that certain urban drainage systems could pose a possible health risk due to human enteric pollution. In addition, the perception of pollution at each site based on responders of a survey varied. Thus, the correct awareness by urban drainage system handlers/workers as to the human enteric pollution levels of water samples from different types of these systems could assist in minimizing potential risk of illness.

## 7.1 Introduction

Many developed and developing countries, including the United States, are faced with the difficulty of deriving growth and development while maintaining or improving the quality of their natural water resources. Buildings, roadways and driveways are common practices of conventional development, particularly of metropolitan areas, and result in the coverage of large areas with impervious surfaces. The increased impervious coverage results in changes in the conveyance and storage of water and can lead to an increase in the rate and volume of surface runoff, higher peak storm flows, and increased severity and frequency of flooding (USEPA 2009). These effects are particularly drastic when these land alterations occur in previously undeveloped areas (USEPA 2009). These increased flows negatively affect existing combined sewer overflows (CSO) and sanitary sewer overflows (SSO) by consuming conveyance capacity resulting in discharge of waste to natural waterbodies.

It is estimated that 850 billion gallons of CSO is discharged annually, and between three and 10 billion gallons of sewage are discharged annually in the United States as a result of between 23,000 and 75,000 SSO events (USEPA 2004). These CSO and SSO runoff contain raw sewage capable of harboring human pathogens (*Cryptosporidium*, *Vibrio cholerae*, noroviruses, etc.) and toxic pollutants (hormones, antibiotics, etc.), thus, contributing to an elevation in the level of gastrointestinal illness from recreational activities at beaches post-storm event (Colford et al. 2012; Curriero et al. 2001; Gaffield 2003; USEPA 2004; USEPA 2009).

One comprehensive approach to water quality protection in areas with increased urbanization is the implementation of Green Infrastructures (GI), which is defined by a range of natural and built systems that can occur at the regional, community, and site scales (USEPA 2009). GI may involve the detention or retention of stormwater runoff on the landscape to alleviate flooding and generation of CSO discharge during wet weather. However, captured runoff may pose an inadvertent human health concern due to the entrapment of chemical and microbial contaminants as stormwater runs over roads, rooftops, and compacted land (NRC 2008). Studies have shown the presence of zoonotic pathogens and toxin genes associated with intestinal and extraintestinal pathotypes of *Escherichia coli* in roof-rainwater harvesting tanks (Ahmed et al. 2012a; Ahmed et al. 2008a; Ahmed et al. 2012b; Ahmed et al. 2010). However, these studies attributed the pollution to animal enteric waste.

In general, it is posited that there is an elevated human health risk associated with human enteric pollution (Scott et al. 2002). Thus, of concern is the possible unintentional increased risk of public exposure to microbial pathogens in runoff via urban drainage systems. Our study assessed samples of urban runoff, sewer flow, and stored water inside GI facilities to determine if these sources were impacted by human enteric pollution. A human specific *Bacteroides* qPCR MST protocol developed throughout our study was used to assay the samples collected in New Jersey and New York. The results of the MST assays were compared against survey results of personnel associated with the sites. The survay ranked personnel's risk perception to human enteric pollution exposure at the sites.

## 7.2 Method

### 7.2.1 Sampling Sites and Sample Collection

Five sites were investigated, each with a different urban drainage characteristic. Site 1 was a combined sewer manhole at the intersection of Stratford Ave and 174th Street in the Bronx, New York City (Figure 29A). The manhole possesses domestic sewer and urban runoff generated on one residential block. Samples were also collected from one of the two catch basins that convey street runoff into the combined sewer. Site 2, Meadow Lake in Flushing Meadows/Corona Park (FMCP), Queens, is New York City's largest fresh water body. A total of about 38 acres of untreated highways, parking lots and roads drain directly to the Lake. The landscape surrounding Meadow Lake within the Park is comprised of several parking lots, playgrounds, mowed lawns and trails that are heavily used in the summer. The area is frequently flooded and difficult to transverse after large rain events. Samples were collected from the lake and from a puddle in a parking lot that will eventually discharge into a new stormwater treatment wetland (outlined in red) (Figure 29B).

A.


B.


**Figure 29: A) The 2 hectare Stratford Ave Catchment Area (Jeffers 2012). B) Aerial photo of Meadow Lake (Unisphere Inc. et al. 2010).**

Site 3 is a bioretention area (Greenstreet) located at Nashville and 116th Street, Queens, New York City. This Greenstreet has two infiltration beds comprised of mulch, Greenstreet soil mix and native sandy soil to allow water from street runoff to infiltrate into the ground. Runoff enters the beds primarily through a curb cut inlet on the west side of the site. Samples were collected from a flow diverter box that conveys urban runoff to the beds near the inlet (Figure 30A). Site 4 is located at the Tacony-Palmyra Bridge (Figure 30B), New Jersey. A pipe conveys runoff from one of the bridge's scuppers to a cistern, the sampling point, from which runoff is directed to a rain garden installed underneath the bridge's eastern span.

Site 5 is located at the ABC carpet stormwater treatment wetland, Bronx, New York City. Runoff from the parking lot is conveyed to a small, lined, engineered wetland for natural treatment before discharge into the Bronx River. At this location, samples were collected from the parking lot runoff catch basin, the wetland, and the Bronx River.

Samples were collected in 500 mL polypropylene bottles at each site and transported to the lab in igloos with ice packs (<$8^{o}$C); they were then processed for DNA extraction within 6 hours of collection.

A. B.

**Figure 30: A) The Nashville Greenstreet inlet, flume apparatus and catchment (Kimberly DiGiovanni). B) Collage of the Tacony-Palmyra Bridge site (Bita Alizadeh).**

**7.2.2 Survey**

A short survey was conducted of four students and two professors familiar with the sites. The survey required each individual to rank each site's likelihood of pollution from human enteric sources. The results of the survey were used to determine if there existed a correlation between risk perception and the qPCR results obtained.

### 7.2.3 DNA Extraction

All DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's protocol except for environmental water samples. In brief, mixed cellulose esters 0.45 μm membranes were used to vacuum filter environmental water samples. In all cases, dependent on turbidity, the maximum possible volume of the collected environmental water sample that could be filtered was processed. Typically, 50 mL or 100 mL of sample was filtered for DNA extraction. The membranes were then placed in 2.0 mL microcentrifuge tubes (Qiagen) and 1.6 mL ASL buffer added. Samples were vortexed for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90$^{o}$C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAcube QIAamp DNA Stool protocol was followed as outlined in the kit's manual. The concentration of the total DNA extraction yields for all samples were determined with a NanoDrop ND-000 UV spectrophotometer (Thermo Fisher Scientific) with absorbance readings at 260 nm, and the purity of the DNA preparations determined by A260/A280 ratios. All extracted DNA were stored at -20$^{o}$C until analyzed by qPCR.

### 7.2.4 Standard Curves

A plasmid DNA construct was developed to function as a plasmid DNA standard for calculation of assay calibration curves. Amplicons obtained from WWTP influent were purified with QIAquick Gel extraction kits (Qiagen). Cloneing and plasmid purification were done with Qiagen PCR cloning plus kits. The eluate was quantified and diluted to generate samples ranging from approximately 1.0E+01 to 1.0E+09 molecules of template DNA for standard curves for qPCR assays.

### 7.2.5 qPCR Protocol

The human specific qPCR MST protocol used a primer and probe set previously developed in this study. The forward primer, HF68, (5′-GGC AGC ATG GTC TTA GCT TG-3′) and HF183rc, the reverse complement of the original forward primer HF183 (5′ – CGG ACA TGT GAA CTC ATG AT – 3′), were combined with a Roche UPL probe #156 (5′-GCT GAT GG-3′) for human specific qPCR assays. A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. The protocol for all assays used the Roche LC480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed was as follows: pre-incubation for 5 min at 95°C; 40 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, including standard curves, no template controls and plasmids as positive controls.

### 7.3 Results

The results of the human specific *Bacteroides* assay for the GI sites showed the Stratford Avenue manhole site to be the most highly impacted by human enteric waste, followed by the Bronx River and then road runoff at both the ABC carpet wetland parking lot catchment and Stratford Avenue catch basin (Table 16). None of the samples from the Tacony, Meadow Lake and Nashville sites showed human impact.

The survey showed the Stratford Avenue manhole to be perceived by most of the responders as the site most affected by human enteric pollution (Table 17). On the other hand, the Tacony–Palmyra Bridge, the Meadow Lake parking lot, and all of the ABC Carpet sites were perceived to be influenced the least by human enteric pollution.

**Table 16: Average number of Bacteroides 16S rRNA gene copies per mL of sample.**

| GI Sampling Site | qPCR results gene copies/mL |
|---|---|
| Stratford Avenue - manhole | 1892500 |
| Bronx River – outfall of ABC Carpet wetland | 768 |
| ABC Carpet - parking lot runoff catch basin | 574 |
| Stratford Avenue - road runoff catchment basin | 266 |
| ABC Carpet - stormwater treatment wetland | <34 |
| Tacony–Palmyra Bridge - road runoff catchment tank | <17 |
| Meadow Lake - lake | <17 |
| Meadow Lake - parking lot puddle | <17 |
| Nashville - road runoff catch basin | <17 |
| Pre & Post MF Control | <3.6 |

A semi-log plot of the average gene copies/mL of sample determined by the human specific qPCR assays versus the average ranked perception of responders of the survey showed a correlation (Spearman Rank = 0.68) (Figure 31). The average ranks ordered the site from highest impacted to least impacted by human enteric pollution as; Stratford Avenue manhole, Bronx River, Stratford Avenue catch basin, Meadow Lake, ABC wetland, Nashville, ABC Carpet parking lot catch basin and wetland, Meadow Lake  parking lot, and Tacony. The perception that of all the sites sampled, human enteric pollution impacted the manhole the most followed by the Bronx River and then the, Stratford Avenue catch basin matched the qPCR results. The other sites were perceived to not be greatly impacted, which the qPCR results supported. One outlier in the findings was the ABC parking lot catch basin, which was perceived to not be impacted by human enteric pollution but showed human specific *Bacteroides* qPCR targets.

**Table 17: Survey results showing the number of responders that assigned a particular rank to a site.**

**The ranks were based on the responders' perception of the possible pollution of each site by human enteric sources with '5' being the worst and '1' the least impact. Six responders participated in the survey.**

| | # of times each rank was assigned | | | | |
|---|---|---|---|---|---|
| **GI Sampling Site** | **1** | **2** | **3** | **4** | **5** |
| Stratford Avenue - manhole | | | 1 | | 5 |
| Stratford Avenue - road runoff catch basin | 1 | 2 | 2 | 1 | |
| ABC Carpet - parking lot runoff catchment basin | 4 | 1 | 1 | | |
| ABC Carpet - stormwater treatment wetland | 4 | 1 | 2 | | |
| Bronx River – outfall of ABC Carpet wetland | | 1 | 3 | 1 | 1 |
| Tacony–Palmyra Bridge - road runoff catchment tank | 5 | 1 | | | |
| Meadow Lake - lake | | 5 | 1 | | |
| Meadow Lake - parking lot puddle | 4 | 2 | | | |
| Nashville - road runoff catch basin | 3 | 2 | 1 | | |

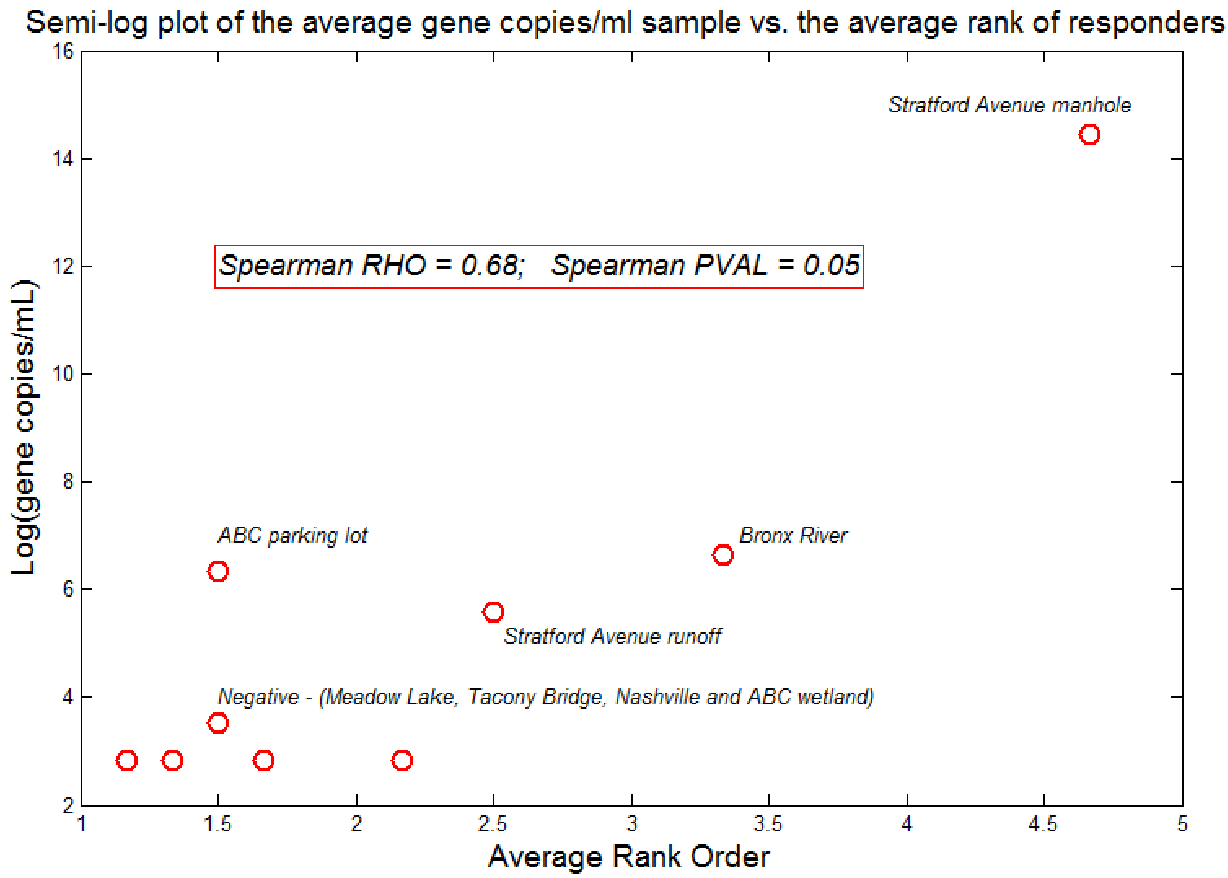



**Figure 31: Semi-log plot of the average gene copies/mL of sample vs. the average rank of responders.**

### 7.4 Discussion

Microbial and chemical contaminants are entrapped by stormwater as it runs over roads, rooftops, and compacted land (NRC 2008). Studies have shown the presence of zoonotic pathogens and toxin genes associated with pathogenic *Escherichia coli* in roof-harvested rainwater tanks (Ahmed et al. 2012a; Ahmed et al. 2008a; Ahmed et al. 2012b; Ahmed et al. 2010). Thus, the possible inadvertent increased risk of public exposure to microbial pathogens in runoff via urban drainage systems is of concern. In this study, five urban drainage systems were investigated to determine actual and perceived pollution by human enteric sources.

As expected, of all the samples collected those obtained from the manhole contained the highest level of human associated *Bacteroides* molecular markers (~1.89E+06 gene copies/mL). These samples were collected at a time and point when they were likely to be primarily comprised of sanitary waste from the residence on the Stratford Avenue block.

The Stratford Avenue catch basin and the ABC Carpet wetland parking lot runoff both showed human associated *Bacteroides* molecular markers (order of magnitude - 1.00E+02 gene copies/mL). It is postulated that backflow from the Stratford Avenue sanitary main may have resulted in the human signals of that catch basin. It is uncertain what contributed to that of the ABC Carpet parking lot runoff qPCR signals; however, it may be from truck cabin discharge of delivery haulers. The ABC Carpet wetland showed no human targets, which indicated a reduction in levels inputted from the parking lot runoff. The Bronx river samples, which were collected at the ABC Carpet wetland discharge point, indicate possible baseline pollution by human enteric sources, albeit at low levels. This could be a result of CSO discharge into this river upstream of the collection point. The implication of human enteric pollution in urban streams and rivers is similar to other MST studies (Newton et al. 2011; Sauer et al. 2011).

None of the other GI sites showed the presence of the human specific markers, which implies a reduced likelihood that these sites to pose a significant human health risk due to human enteric pollution. This is our expectation, however, only a limited number of samples were assayed for this study and further investigation (samples and tests) is required to confirm this finding. Although an elevated human health risk is associated with human enteric pollution (Scott et al. 2002), the risks attributed to animal enteric pollution (Ahmed et al. 2012b) should not be overlooked.

Our survey indicated that our respondents' perception correlated with the qPCR determined levels of human enteric pollution of the sites. Our responders had working knowledge of the sites sampled prior to the survey, yet not all correctly perceived the levels obtained. The correct awareness by urban drainage system handlers/workers as to the human enteric pollution levels of water samples from different types of these systems could assist in minimizing their risk of illness.

The inferences made by this study about urban drainage systems are based on a limited number of samples. More samples need to be assayed in order to make assertions that are more definitive. However, our findings indicate the possibility that certain urban drainage systems could pose a possible health risk due to human enteric pollution.

# CHAPTER 8: INVESTIGATING THE WATER QUALITY OF TROPICAL INLAND WATERS

## Abstract

Fecal indicator bacteria (FIBs) are used globally to assess the microbiological safety of drinking water, recreational waters, and shellfish aquatic habitats. However, a number of studies suggested FIBs may proliferate in sediments, warm, tropical and subtropical aquatic habitats. Environmental water samples collected from watersheds located on the tropical island of Puerto Rico were evaluated for their quality using conventional and our qPCR MST molecular protocols. The findings support the use of the protocol in both tropical and temperate regions. In addition, a toolbox approach using our molecular MST protocol along with conventional FIBs and pathogenic tests lead to a more comprehensive understanding of the results.

## 8.1 Introduction

Fecal indicator bacteria (FIBs) are used globally to assess the microbiological safety of drinking water, recreational waters, and shellfish aquatic habitats. For recreational waters the U.S. Environmental Protection Agency (USEPA) recommended *Enterococcus* spp. for marine and fresh waters and *Escherichia coli* (*E. coli*) for fresh waters only (USEPA 2000; USEPA 2012). However, a number of studies suggested these two FIBs may proliferate in warm, tropical and subtropical aquatic habitats (Byappanahalli and Fujioka 2004; Byappanahalli and Fujioka 1998; Byappanahalli et al. 2003; Carrillo et al. 1985; Desmarais et al. 2002; Hardina and Fujioka 1991; Rivera et al. 1988; Solo-Gabriele et al. 2000; Stewart et al. 2008). Other studies have revealed the persistence of these FIBs in the sediments of environmental waters (Davies et al. 1995; Fish and Pettibone 1995; Sherer et al. 1992) and of beach sands (Alm et al. 2004; Whitman et al. 2003; Yamahara et al. 2012; Yamahara et al. 2009).

*Bacteroides* spp., anaerobes, are members of the phylum Bacteroidetes which is one of the most abundant bacterial group within the human colon microbiota (Bakir et al. 2006; Benno et al. 1989; Ley et al. 2005). These organisms have been proposed as an alternative FIB by researchers (Allsop and Stickler 1985; Fiksdal 1985). *Bacteroides* spp. are not expected to proliferate in the environment and were shown to have a shorter survival time in aquatic environs than FIBs (Balleste and Blanch 2010; Fiksdal et al. 1985; Kreader 1998). Detection of their presence in aquatic environments would indicate recent enteric pollution. Although, one study indicated members of this order may survive for up to six days in oxygen stressed aquatic conditions (Avelar et al. 1998). Another study showed the persistence of detectable *Bacteroides* spp. genetic markers is dependent on predation and temperature (Kreader 1998).

Microbial source tracking (MST) attempts to differentiate and identify the source of enteric pollution and assumes that there exists a traceable association between microbes excreted in feces and their particular hosts (Field and Samadpour 2007; USEPA 2005b). Several studies have shown *Bacteroides* molecular assays using presence/absence polymerase chain reaction (PCR) (Ahmed et al. 2007; Bernhard and Field 2000b) and quantitative PCR (qPCR) (Haugland et al. 2010; Shanks et al. 2009) to be relatively successful for MST purposes. Our study has developed a *Bacteroides* qPCR assay that was shown to be highly human specific.

This study was a collaboration between Drexel University and the Center for Environmental Education, Conservation and Research of Inter American University of Puerto Rico (CECIA). Environmental water samples collected from watersheds located on the tropical island of Puerto Rico were evaluated for their quality using conventional and our qPCR MST molecular protocols. Previous aspects of our research assayed environmental and animal fecal samples collected from the temperate regions of the Philadelphia watershed of Pennsylvania. Thus, this study gave us the opportunity to assess our protocol in tropical environments to determine its geographical transferability. Fecal sample were collected of animal that were inhabitants of the watersheds investigated to assess the specificity of our MST protocol.

## 8.2 Method

### 8.2.1 Study Area

The water quality of four small streams that comprised four distinct watersheds of the tropical island of Puerto Rico was evaluated over a three-week period. These watersheds, El Real (Muñoz Rivera), Mulas Jagual, Quebrada Arriba and Apeadero, are located at the southeastern section of the island (Figure 32). They provided the only source of water for distribution to residences of small isolated rural districts located in the corresponding watersheds.

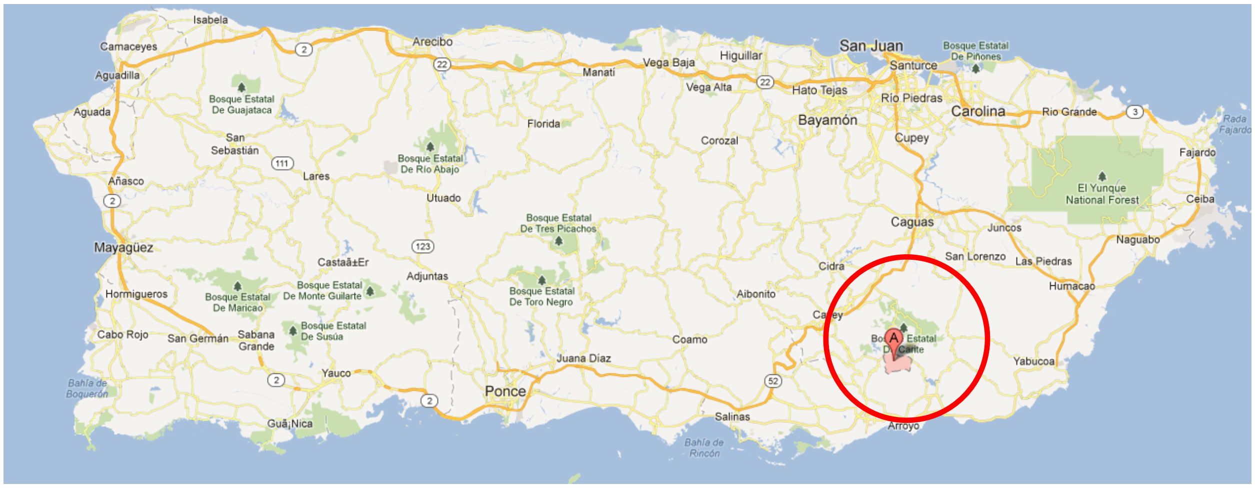



**Figure 32: Map of Puerto Rico. All five watersheds investigated are located at the South Eastern section of the island (encircled)(Google Maps).**

Source water was conveyed via PVC pipes from small concrete catchments located near the headwater of each stream to its respective concrete tank (located a few meters away) for storage and disinfection treatment before distribution to the public (Figure 33 & 34). Unlike the other watersheds, the Apeadero watershed serves two different distribution systems: Tanque Arriba and Tanque Abajo. The intake points for both these systems are located approximately a mile away from the storage/treatment tanks. Tanque Arriba's intake is sited close to the headwater of the source stream and that of Tanque Abajo is situated further downstream of this location. Disinfection was done by dosing with chlorine (calcium hypochlorite) using either a tablet or powder drip feed system. A local resident was trained by CECIA representatives to administer proper doses of disinfectant and maintain the system/s in their respective watershed.

A. B.

**Figure 33: Pictures of the source water catchment.**

**A) El Real. B) Quebrada Arriba (Michael Ryan)**

A. B.

**Figure 34: Pictures of the tanks used for storage, treatment and distribution of water.**

**A) Apeadero (Tanque Arriba) - disinfection by chlorine powder drip feed system. B) Quebrada Arriba - disinfection by chlorine tablet system.**

### 8.2.2 Non-Molecular Water Quality Protocols

The water quality of four of the watersheds was assessed using non-molecular based protocols. The fifth site was not assessed due to time limitations. Environmental water samples were collected from El Real (December 5th), Quebrada Arriba (December 8th) and Apeadero (Tanque Arriba) (December 14th).

#### *8.2.2.1 Standard Indicators*

Duplicate samples were enumerated for thermotolerant (total) coliforms (TC), fecal coliforms (FC), *Escherichia coli* (*E. coli*), fecal streptococci, and Enterococci. Samples were collected in 1 L polypropylene bottles and transported to the lab in coolers with ice packs (<8$^{o}$C) and processes within 6 hours of collection. The enumeration was based on inoculating culture media (in single tubes sets) with serial dilutions of the samples and computing log densities based on presence-absence of growth. Dilutions were from 100 mL for TC, FC and *E. coli* and from 10 mL for fecal streptococci. The procedure followed were in accordance with those of the Standard Method (APHA et al. 2005). Table 18 lists the different media and incubation conditions for each test.

#### *8.2.2.2 Salmonella*

Separate samples were also collected for presence/absence screening of *Salmonella*. Ten samples (1 L each) were collected for *Salmonella* assays following the procedure previously described (a total of 10 L each site). The samples were membrane filtered through a layer comprised of a cellulose absorbant pad (Millipore) placed on top of a mixed cellulose ester 0.45 μm membranes (Millipore). A filtration manifold apparatus was used to speed up the filtration process. On clogging, the pad and membrane were placed in the *Salmonella* liquid broth media and a new pair was added to the manifold. The procedure was repeated until the approximate 10 L of sample was processed. This protocol is an adaptation of Standard Methods techniques described elsewhere (Herson et al. 2005; Hunter et al. 2010) and allowed detection of densities of at least 1 CFU/10 L. Table 18 lists the different media and incubation conditions for the test.

**Table 18: List of media and cultivation conditions used to determine organisms based on standard methods.**

| **Media** | **Target organism** | **Incubation Duration** | **Incubation Temperature** |
|---|---|---|---|
| Presence-absence broth | Total or thermotolerant coliforms - presumptive | 24 ± 2 - 48 ± 3 hours | 35 +/-0.5°C. |
| Brilliant Green Lactose Bile Broth (BGLB) | Total or thermotolerant coliforms - confirmative | 48 ± 3 hours | 35 +/-0.5°C. |
| EC medium | Fecal coliform | 24 ± 2 hours | 44.5 ± 0.2 °C |
| Lauryl Tryptose Broth with MUG (LTB- mug) | Escherichia coli | 24 ± 2 hours | 35 +/-2°C. |
| Azide Dextrose | Fecal streptococci - presumptive | 24 ± 2 - 48 ± 3 hours | 35 +/-5°C. |
| Brain-Heart Infusion Agar (BHIA) | Fecal streptococci – validation 1 | 24 ± 2 - 48 ± 3 hours | 35 +/-5°C. |
| Brain-Heart Infusion Broth (BHIB) | Fecal streptococci – validation 2 | 24 ± 2 | 35 +/-5°C. |
| Bile esculine agar (BEA) | Fecal streptococci – validation 3 | 48 ± 3 hours | 35 +/-5°C. |
| Brain-Heart Infusion Broth (BHIB) + NaCl | Enterococci - validation4 | 48 ± 3 hours | 35 +/-5°C. |
| Tetrathionate broth | Salmonella enrichment | 24 ± 2 | 35 +/-5°C. |
| Xylose lysine desoxycholate | Salmonella selective growth | 24 ± 2 | 35 +/-5°C. |

#### *8.2.2.3 Cryptosporidium and Cyclospora*

Five 10 L carboys were used to collect samples for Cryptosporidium and *Cyclospora*. Samples were first coagulated as a concentration step. Cryptosporidium was determined by immunomagnetic separation (IMS) and immunofluorescence assay (IFA) staining. Cryptosporidium identification was by examining the stained slides using fluorescence microscopy. *Cyclospora* oocysts were determined by modified Ziehl Neelsen (mZN) acid-fast stain smears and identified by microscopic examination.

#### *8.2.2.4 Physicochemical parameters*

A separate 1 L sample was collected for tests of physicochemical parameters. A Corning® 450 meter and combination electrode was used for pH tests. Turbidity was measured with a Nephelometer. Each instrument was calibrated for use according to manufacturer instructions.

### 8.2.3 Animal Sample Collection

The specificity of the markers was assessed against animals that were indigenous to the watersheds studied. Peafowl (3), chicken (3), cow (3), pig (4), dog (4), rabbit (5) and horse (1) fecal samples were collected on an occasion when no environmental water samples were being collected. Samples were collected in sterile Kapak SealPak® pouches and transported to the lab in cooler with ice packs (<8°C) and processes for DNA extraction within 6 hours of collection.

### 8.2.4 Molecular Water Quality Protocols

Environmental (raw water) samples were collected in 1 L polypropylene bottles at each site on different occasions (Table 19). Samples collected from El Real (Muñoz Rivera), Mulas Jagual and Quebrada Arriba were grab samples taken directly from the streams. Samples from Apeadero (Tanque Arriba) and Apeadero (Tanque Abajo) were collected from raw water taps located at the respective tanks immediately before entering. Taps were swabbed with ethanol and the water then allowed to run freely for five minutes before sample collection. All environmental water samples were transported to the lab in cooler with ice packs (<8°C) and processes for DNA extraction within 6 hours of collection.

#### *8.2.4.1 DNA Extraction*

All DNA extractions were performed manually with QIAamp DNA Stool kits (Qiagen) in accordance with the manufacturer's protocol except for environmental water samples. In brief, mixed cellulose ester 0.45 μm membranes were used to vacuum filter environmental water samples. In all cases, dependent on turbidity, the maximum possible volume of the collected environmental water sample was filtered. Typically, 500 mL or 1500 mL of sample was filtered for DNA extraction. The membranes were then placed in 2.0 mL microcentrifuge tubes (Qiagen) and 1.6 mL ASL buffer added. Samples were vortexed at 3000 rpm with a Vortex Mixer (VWR) and then incubated at 70$^{o}$C for 5 min in a dry heating block (VWR). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAamp DNA Stool kit protocol was followed as is outlined in the kits' manual. All extracted DNA shipped on dry ice to from Puerto Rico to Drexel University for qPCR analysis. All extracted DNA samples were stored at -20$^{o}$C at the laboratories. The concentration of the total DNA extraction yields for all samples was determined with a NanoDrop ND-000 UV spectrophotometer (Thermo Fisher Scientific) with absorbance readings at 260 nm.

#### *8.2.4.2 Standard Curves*

A plasmid DNA construct was developed to function as a plasmid DNA standard for calculation of assay calibration curves. Amplicons obtained from WWTP influent were purified with QIAquick Gel extraction kits (Qiagen); cloned and the plasmids purified with Qiagen PCR cloning plus kits. The eluate was quantified and diluted to generate samples ranging from approximately 1.0E+01 to 1.0E+09 molecules of template DNA.

#### *8.2.4.3 qPCR Protocol*

The human specific qPCR MST protocol (designated pMR) used a primer and probe set previously developed in this study. The forward primer, HF68, (5'-GGC AGC ATG GTC TTA GCT TG-3') and HF183rc, the reverse complement of the original forward primer HF183 (5′ – CGG ACA TGT GAA CTC ATG AT – 3′), were combined with a Roche UPL probe #156 (5'-GCT GAT GG-3') for human specific qPCR assays. A Roche LightCycler 480 (LC480) Real-Time PCR System was to conduct all qPCR assays. The protocol for all assays used reagents from Roche LC480 Probe Master kits in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed was as follows: pre-incubation for 5 min at 95°C; 40 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, to include no template controls and positive controls (plasmids used for the standard curves).

## 8.3 Results

### 8.3.1 Non-Molecular

In all samples except one, the results for the conventional methods showed total coliform, fecal coliform, *E. coli*, fecal streptococci and enterococci to be present in as little as 1.0 mL of sample (Table 19). This implies log densities >100/100ml of sample in most of the samples evaluated. Tests for thermotolerant coliform and *E. coli* indicated log densities >1000/100ml in samples from all four sites. Higher levels of enterococci were observed at Quebrada Arriba and Apeadero Arriba than at the other two sites.

*Salmonella* was detected at all four sites in densities of at least 1 CFU/10L of sample, and *Cryptosporidium* was detected at Mulas Jagual. *Cyclospora* was not detected at any of the sites. The turbidity levels for the four sites indicated samples relatively free of colloidal and suspended matter.

### 8.3.2 Molecular assays

The animal assays showed all the rabbit and one of the peafowl samples to be positive for the human specific qPCR marker. The concentration of the marker ranged from 2.63E+04 to 4.96E+07 gene copies per gram of rabbit sample and was 3.85E+01 gene copies per gram of the positive peafowl sample. None of the other animal samples showed positive results for this marker. Normalization of the rabbit and peafowl DNA concentrations to 2ng/μL and rerunning the qPCR assay showed only the rabbit samples to be positive. This value of the normalization concentration was chosen based on the upper limits of that of the DNA concentration of environmental water samples.

**Table 19: Table showing the results of samples collected from four different sources.**

**1) The lowest dilution (mL) at which standard indicators was detected. 2) The presence/absence of pathogenic microbes. 3) The results of physicochemical parameters.**

| **Parameters** | **5-Dec-11** | **8-Dec-11** | **12-Dec-11** | **14-Dec-11** |
|---|---|---|---|---|
| | **El Real** | **Quebrada Arriba** | **Mulas Jagual** | **Apeadero Arriba** |
| **Standard Indicators** | | | | |
| Total coliform | 0.1 | 0.1 | 0.1 | 0.1 |
| Fecal coliform | 10 | 0.1 | 1.0 | 1.0 |
| *Escherichia coli* | 0.1 | 0.1 | 0.1 | 0.1 |
| Fecal streptococci | 1.0 | 0.1 | 0.1 | 0.1 |
| Enterococci | 1.0 | 0.1 | 1.0 | 0.1 |
| **Pathogenic microbes** | | | | |
| *Salmonella* | present | present | present | present |
| *Cryptosporidium* | absent | absent | present | absent |
| *Cyclospora* | absent | absent | absent | absent |
| **Physicochemical parameters** | | | | |
| pH | 7.655 | 7.588 | 7.374 | 7.538 |
| Turbidity (NTU) | 1.76 | 2.12 | 1.73 | 2.01 |

Environmental samples collected from the Apeadero Tanque Abajo raw water tap tested positive for the human specific molecular marker (Table 20). The concentration of the marker was relatively low on all the occasions it was detected with an overall range of 13 - 264 gene copies per mL of sample. Conversely, the distribution samples collected from this system did not show the marker’s presence. None of the samples form the other systems showed positive results for the molecular marker.

**Table 20: Table showing the average qPCR results (gene copies per mL) for the environmental water samples assayed.**

| **Sampling ID** | **Date collected** | | | | | |
|---|---|---|---|---|---|---|
| | **11/22/11** | **12/1/11** | **12/5/11** | **12/8/11** | **12-Dec-11** | **12/14/11** |
| Apeadero Tanque Arriba (Raw[1]) | <10 | - | - | <10 | <10 | <10 |
| Apeadero Tanque Abajo (Raw) | 34 | 26 | - | 13 | <10 | 264 |
| Apeadero Tanque Abajo (dist[2]) | - | <10 | - | - | - | - |
| El Real (Muñoz Rivera) (Raw) | <10 | - | <10 | <10 | <10 | <10 |
| Mulas Jagual (Raw) | <10 | - | - | <10 | <10 | <10 |
| Quebrada Arriba (Raw) | <10 | | - | <10 | <10 | <10 |
| negative controls | <10 | <10 | <10 | <10 | <10 | <10 |

**Raw – source water samples   Dist – distribution water sample**

**8.4 Discussion**

Owing to their anaerobic characteristics and their abundance in the human gut microbiome, several studies have developed MST protocols using *Bacteroides* molecular markers (Ahmed et al. 2008a; Bernhard and Field 2000a; Bernhard and Field 2000b; Converse et al. 2009; Haugland et al. 2010; Kreader 1995; Shanks et al. 2009). The relatively high specificity of the most of these assays could be a result of coevolution of *Bacteroides* with its host. Its anaerobic trait could have minimized lateral transfer of the organism from one host to another and increased the chance of coevolution. Although the molecular assays show high human differentiability, the genetic targets are sometimes detected in other animal hosts. Thus, it is always recommended that specificity assays are conducted of animals that are inhabitants of watersheds investigated (Santo Domingo et al. 2007; USEPA 2005b).

The assays of animal fecal samples showed all rabbits to be positive for the human specific genetic marker. This is similar to our previous findings and supports our postulate that *Bacteroides* associated with rabbits may have a similar 16S rRNA gene region to that targeted in humans. The peafowl sample that tested positive may be due to a similar reaction shown by chickens in another study (Shanks et al. 2010b). However, when the DNA concentration of the peafowl sample was normalized to levels similar to those of environmental water samples (42 ng/μL to 2 ng/μL) no qPCR signals were observed. The normalized results indicate human specific *Bacteroides* molecular markers in environmental water samples are more likely to be associated with human or rabbit enteric sources. However, since the residents indicate only domesticated rabbits are to be found in these watersheds then it may be assumed that the presence the human specific *Bacteroides* molecular markers in environmental water samples from these watersheds are more likely to be associated with human or rabbit enteric sources.

The non-molecular assays (conventional assays) showed thermotolerant coliform, fecal coliform, *E. coli*, fecal streptococci and Enterococci to be present in all samples. These results are similar to those of historical samples of these areas. They indicate constant sources of fecal pollution of these watersheds, which is the concern, especially given the limitations of small isolated rural system like these. Given the low turbidity of the raw waters in can be assumed that surface runoff may not be the source. It is possible that these organisms are a part of the environmental microbiota of tropical waters (Anderson et al. 2005; Rivera et al. 1988). However, the detection of *Salmonella* at all these sites highlights the seriousness of the human health concerns to all the residents that are served by these systems. *Cryptosporidium* was found in the Mulas Jagual samples and was indicated to be another historical health concern of the area.

The qPCR MST assays detected none of the human specific molecular markers in any of the samples assayed except for those from Apeadero Tanque Abajo. The source/headwaters of these systems are isolated at a distance from human residence and the locals have been educated on the necessity of preserving their integrity by CECIA representatives. It was thus not expected to detect the possibility of human enteric pollution of one of the Apeadero systems, especially since these intakes are even more remotely located that the others. The Tanque Arriba intake is upstream of and closer to the headwaters of the Apeadero source than that of Tanque Abajo, thus the source of the pollution are located between. The levels of the marker are low which could indicate a diffuse source. The consistency with which sample from Tanque Abajo is shown positive indicates the results are not assay anomalies.

The molecular MST protocol complements those of the conventional and pathogenic tests thus leading to a more informed set of results. The *Cyclospora* and the qPCR MST protocols indicate the absence of human enteric pollution of these watersheds. However, the latter may be more sensitive than the former, which explains the difference with the Tanque Abajo samples. The sources of

*Cryptosporidium, Salmonella* and the other indicator organisms could be an animal enteric nature thus the persistence in their detection.

Overall, the results of the assays conducted with our qPCR MST protocol shows geographical transferability. The findings support the use of the protocol in both tropical and temperate regions. This is important to the MST community since it enable ease of standardization and region wide applicability if the protocol should be adopted for regulatory use. However, these findings were from a small sample size and the project was for a short duration. Thus, a more robust evaluation of the protocol in these and other regions is recommended. Animal specific qPCR MST assays could also be included in the evaluation especially since these preliminary findings indicate animal enteric pollution may be the main problem of these watersheds.

# CHAPTER 9: EVALUATING THE MST POTENTIAL OF ANIMAL SPECIFIC AND UNIVERSAL MOLECULAR MARKERS

## Abstract

There is the possibility that no single marker or approach may be considered an ideal MST method and the recommendation is for the application of multiple markers or approaches in a toolbox. The objective of this study was to augment the human specific assays with other molecular markers to better characterize sources of pollution. Three animal specific (Canadian goose, pig and cattle), a universal *Bacteroides* spp. (UniBaC) and a universal *Enterococcus* spp. (UniEC) NELI qPCR MST protocols were evaluate for their MST applicability for our research. The cow assay showed 100% specificity and sensitivity. However, both the Canadian goose and pig assays returned negative results even for samples associated with their respective host. The universal molecular techniques showed results more consistent with expectations in comparison to the conventional cultivation method. This study highlighted the advantages and limitations of using multiple molecular markers to characterize samples/sites.

### 9.1 Introduction

*Bacteroides* spp. have become frequently used in culture-independent, 16S based methods for microbial source tracking (MST) use (Ahmed et al. 2008b; Bae and Wuertz 2012; Bernhard and Field 2000b; Shanks et al. 2007). The Drexel MST research focused on the development and application of a non-enrichment library independent (NELI) MST protocol that differentiated human from non-human associated enteric pollution targeting *B. dorei* 16S rRNA genes. This strategy had the appeal that the assays could be completely performed in a matter of hours, were sensitive, inexpensive (excluding capital cost), quantitative, and amenable to automation in comparison to enrichment library dependent (ELD) MST method (Field and Samadpour 2007; Santo Domingo et al. 2007). The Drexel MST study developed a qPCR MST protocol that shows improvement over possibly all assays that have been used to determine human sources of enteric pollution.

The literature review indicates the possibility that no single marker or approach may be considered an ideal MST method and the recommendation is for the application of multiple markers or approaches in a toolbox approach to complement each other (Schriewer et al. 2010; USEPA 2005b). This study evaluated animal specific and universal molecular markers for their potential

MST applicability. The objective was to augment the human specific assays, in effect creating a toolbox approach, to better characterize the sources of pollution detected in samples collected in the Philadelphia watershed. Firstly, specificity and sensitivity analyses of the animal specific protocols were done to determine transferability to the Drexel MST facilities and applicability to this geographic region. Then, conventional enterococci enumeration, the universal and the host specific molecular assays were used to characterize three of the Philadelphia sites investigated throughout this study.

## 9.2 Method

The human specific qPCR MST protocol (designated pMR) of this study was used to evaluate environmental samples in a toolbox approach with conventional enumeration methods, animal specific and universal molecular markers. Enumeration using conventional mEI enterococci cultivation (USEPA 2005a) were also included for comparative analysis.

### 9.2.1 Animal and Universal Primes-Probe Sets

Three animal specific (Canadian goose, pig and cattle), a universal *Bacteroides* spp. (UniBaC) and a universal *Enterococcus* spp. (UniEC) NELI qPCR MST protocols were evaluate for their MST applicability for our research. Table 1 lists the primes, probes, and references of these protocols. All of the animal specific assays and the UniBaC assay targeted *Bacteroides* spp. 16S rRNA genes. The UniEC assay targeted the 23S rRNA genes of *Enterococcus* spp. The animal specific assays were of the most recent cited in the literature and were evaluated to be of the most promising. The UniEC protocol is that recommended by USEPA draft Method A (2010).

**Table 21: Primer and probe sequences for animal specific and universal assays.**

| Assay Target | Primer and probe sequences (5' to 3') | Reference |
|---|---|---|
| Human - pMR | HF68- GGCAGCATGGTCTTAGCTTG<br>HF183rc - CGGACATGTGAACTCATGAT<br>Roche UPL probe #156 – GCTGATGG | This study |
| Canadian goose | CG2probe--[6~FAM] AATACCTGATGCCTTTGTTTCCCTGCA [TAMRA~6~FAM]<br>CG2F--ACTCAGGGATAGCCTTTCGA<br>CG2R—ACCGATGAATCTTTCTTTGTCTCC | (Fremaux et al. 2010) |
| pig | Pig-2Bac113MGB--[6~FAM]GTCCACGGGATAGCC [TAMRA~6~FAM]<br>Pig-2-Bac41F--GCATGAATTTAGCTTGCTAAATTTGAT<br>Pig-2-Bac163Rm—ACCTCATACGGTATTAATCCGC | (Mieszkin et al. 2009) |
| Cattle (ruminants) | BacB2-626P--[6~FAM] ATGAGGTGGATGGAATTCGTGGTGT [TAMRA~6~FAM]<br>BacB2-590F--ACAGCCCGCGATTGATACTGGTAA<br>Bac708Rm—CAATCGGAGTTCTTCGTGAT | (Mieszkin et al. 2010) |
| Universal *Bacteroides* spp. - UniBaC | GenBactp2 --[6~FAM] CAATATTCCTCACTGCTGCCTCCCGTA [TAMRA~6~FAM]<br>GenBactF3--GGGGTTCTGAGAGGAAGGT<br>GenBactR4—CCGTCATCCTTCACGCTACT | (Shanks et al. 2010a; Siefring et al. 2008) |
| Universal *Enterococcus* spp. - UniEC | pECU--[6~FAM] TGGTTCTCTCCGAAATAGCTTTAGGGC TA[TAMRA~6~FAM]<br>UentF--GAGAAATTCCAAACGAACTTG<br>UentR--CAGTGCTCTACCTCCATCATT | (Ludwig and Schleifer 2000; USEPA 2010a) |

### 9.2.2 Animal Samples

The specificity and sensitivity of the animal specific protocols were evaluated by assaying known animal enteric sample collected throughout the course of the research. A set of 52 distinct animal hosts were tested (Table 2). The DNA of these samples were extracted on the day of collection with QIAamp DNA Stool kits (Qiagen) in accordance with the manufactures instruction. Extracted DNA were stored at -20$^{o}$C until analyzed by qPCR.

**Table 22: Table showing the number of samples positive/assayed for each animal type using the three animal specific protocols.**

| Animal type | Canadian goose | pig | cattle |
|---|---|---|---|
| **Cow** | 0/4 | 0/4 | 4/4 |
| **Pig** | 0/4 | 0/4 | 0/4 |
| **horse** | 0/4 | 0/4 | 0/4 |
| **goat** | 0/4 | 0/4 | 0/4 |
| **chicken** | 0/5 | 0/5 | 0/5 |
| **rabbit** | 0/1 | 0/1 | 0/1 |
| **dog** | 0/16 | 0/16 | 0/16 |
| **cat** | 0/6 | 0/6 | 0/6 |
| **goose** | 0/10 | 0/10 | 0/10 |
| **Total** | 0/52 | 0/52 | 4/52 |

**9.2.3 Environmental Sites**

The three animal specific, the UniBaC and the UniEC protocols were used to assay a set of the environmental samples collected from sites at Fox Chase Farm (FCF), the Smith's Run creek located at the Schuylkill Center for Environmental Education (SCEE), and sites along the Monoshone creek (Table 3). The FCF and SCEE samples were collected on 04/28/2011 and the Monoshone samples on 08/18/2011. The FCF sampling points were the headwaters and a pond along a tributary at FCF, before and after the confluence of this tributary and the Pennypack Creek. The Monoshone sampling points were the headwaters, before and after a stormwater outfall (SWO) along Lincoln Drive, before it enters the Wissahickon Creek, and along the Wissahickon Creek above the Monoshone Creek confluence. The SCEE sites were expected to be impacted by a small population of wildlife (deer, Canadian goose etc.) and the FCF by a larger fecal source comprised of livestock (cow, goat, sheep and horse) from the farm and to a lesser extent wild avian species (Canadian goose and duck). Duplicate samples were collected in 1L polypropylene bottles at each site and transported to the lab in igloos with ice packs (<$8^{o}$C) and processes within 6 hours of collection.

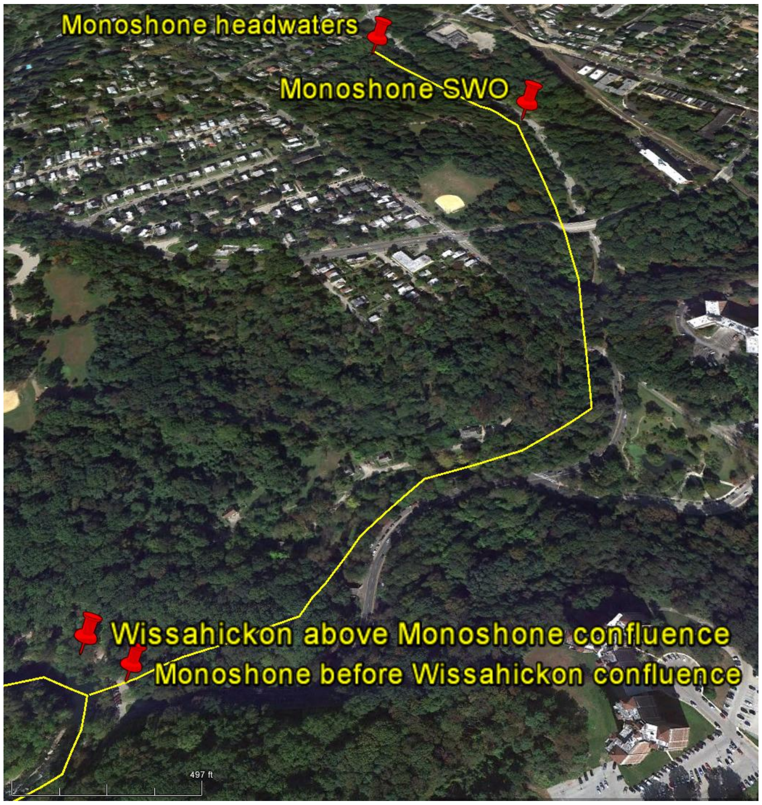



**Figure 35: Monoshone sampling sites (image obtained using Google Earth)**

All DNA extractions were performed using a QIAcube instrument (Qiagen) with QIAamp DNA Stool kits (Qiagen) with modifications to the manufacturer's. In brief, mixed cellulose esters 0.45 μm membranes were used to vacuum filter environmental water samples. In all cases, dependent on turbidity, the maximum possible volume of the collected environmental water sample that could be filtered was attempted. The membranes were then placed in 2.0 mL safe lock microcentrifuge tubes (Qiagen) and 1.8 mL ASL buffer added. Samples were vortex for 3 min @ 3000 rpm with a Vortex Genie 2 (Scientific Industries) and then incubated at 90$^{o}$C for 5 min @ 1400 rpm in a Thermomixer (Eppendorf). One InhibitEX tablet was placed into a new 2.0 mL microcentrifuge tube and 1.2 ml of the ASL incubated samples was added. The rest of the QIAcube QIAamp DNA Stool protocol was followed as outlined in the kit's manual. Extracted DNA were stored at -20$^{o}$C until analyzed by qPCR.

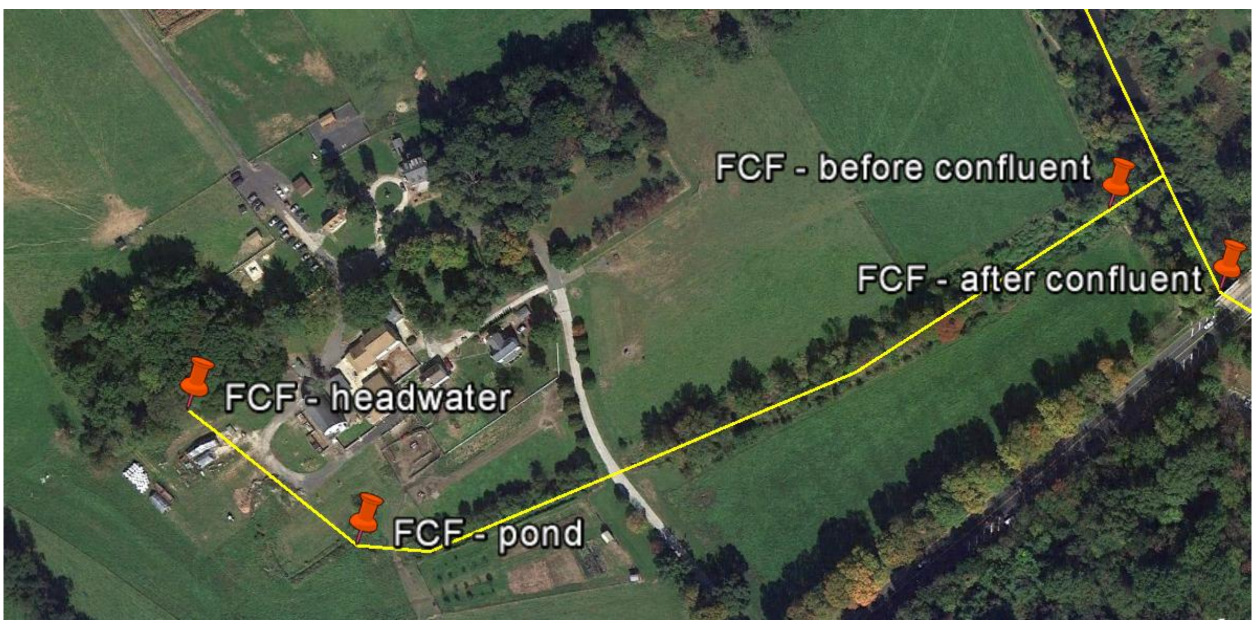



**Figure 36: Fox Chase Farm (FCF) sampling points – livestock impacted (image obtained using Google Earth)**

### 9.2.4 qPCR Assays

A Roche LightCycler 480 (LC480) Real-Time PCR System was used to conduct all qPCR assays. The protocol for all assays used the Roche LC480 Probe Master reagents in concentrations in accordance with the manufacturer's instructions for a total qPCR reaction volume of 20μL. In brief, each reaction mixture contained 5 μL of template DNA, 0.1 μM of the probe, and 0.5 μM of the forward and reverse primers. The program employed: pre-incubation for 5 min at 95°C; 40 amplification cycles of 30 s of annealing at 60°C and 10 s of melting at 95°C; and finally cooling for 10 s at 40°C. All assays were conducted in triplicate, and each included no template (negative) controls along with positive controls. Qiagen PCR Cloning Plus kits were used to produce plasmids with inserts of amplicons from assays of WWTP influent samples, which were then used as positive controls.

## 9.3 Results

### 9.3.1 Specificity and Sensitivity of Animal Specific Assays

The results of the animal specific assays showed the cattle specific assay to perform the best of the three assays. All four of the cow associated samples showing positive qPCR amplification (sensitivity – 100%) and all 48 of the non-cow associated samples showing negative qPCR signals (specificity – 100%)

(Table 3). Both the pig and Canadian goose assays returned negative results even for samples associated with their respective host (sensitivity – 0% and specificity – 100%) (Table 3). These results indicate these pig and Canadian goose protocols may not be appropriate to evaluate the environmental samples assayed.

**9.3.2 FCF and SCEE**

The results showed no qPCR amplification for the Canadian goose-, pig-, or cattle-specific assays for any of the environmental sites (Table 3). The universal protocols showed high levels of *Bacteroides* spp. and *Enterococcus* spp. at all of the FCF and SCEE sites (Table 3). These findings infer that enteric pollution is present at these sites but is not of Canadian goose, pig, or cattle origin.

The human specific pMR assays showed no human-associated pollution at the SECC or the FCF sites except for the Pennypack site located after its confluence with the FCF creek (previously reported). The UniBaC assay showed high levels of general *Bacteroides* spp. at these sites, indication a source of enteric pollution. The SCEE site showed lower levels of general *Bacteroides* spp. occurrences than the FCF sites, indication a source of enteric pollution at lower levels than the FCF sites. Overall, these findings infer that sources of enteric pollution are present at FCF and SCEE but they are not of human origin, except for the Pennypack site.

The UniEC assay showed more *Enterococcus* spp. occurrence than the conventional enterococci enumeration counts (three orders of magnitude higher). This indicate that the molecular *Enterococcus* method is more sensitive that the conventional cultivation method (Wilcoxon Rank Sum test, p-values $<0.05$). Irrespective of technique, the SCEE site showed lower levels of general *Enterococcus* spp. occurrences at the SCEE than the FCF sites. These findings support the universal *Bacteroides* assays that more enteric pollution are present at the FCF that the SCEE sites.

Both the UniBaC and the UniEC assays showed a spike in the level of the respective targets in the FCF pond sample in comparison the other FCF sites. The mEI enterococci enumeration assays showed the opposite result with the lowest count for all FCF sample collected on this occasion at the pond site. Both the UniBaC and the UniEC assays showed the lowest level of their respective targets at the FCF headwater. The enterococci enumeration technique showed lower counts at the FCF pond and the Pennypack than the other FCF sites.

**Table 23: Results for assays conducted on a set of the environmental water samples collected from Schuylkill Center for Environmental Education, the Fox Chase Farm and the Monoshone sites.**

| Sampling Point | CFU/100mL | Average pMR gene targets / mL | Average UniBaC gene targets / mL | Average UniEC gene targets / mL |
|---|---|---|---|---|
| **Smith's Run Creek – stream** | 66 | <4 | 29,100 | 72,400 |
| **Fox Chase Farm – head water** | 146 | <4 | 67,300 | 102,000 |
| **Fox Chase Farm – pond** | 55 | <4 | 5,570,000 | 187,000 |
| **Fox Chase Farm – before confluence** | 324 | <4 | 109,000 | 133,000 |
| **Pennypack after Fox Chase confluence** | 67 | 358 | 191,000 | 127,000 |
| **No template control** | - | <4 | <4 | <4 |
| | | | | |
| | | | | |
| **Monoshone headwaters along W. Johnson Street by Lincoln Drive** | 193 | <2 | 1,750 | 5,180 |
| **Monoshone before Outfall** | >20,000 | <2 | 27,000 | 20,200 |
| **Monoshone after Outfall** | >20,000 | 531000 | 5,680,000 | 44,200 |
| **WSMC001 - Monoshone Creek before it enters the Wissahickon Creek** | >20,000 | 133000 | 2,160,000 | 22,800 |
| **Wissahickon Creek above Monoshone Creek confluence** | 3,650 | 657 | 46,900 | 18,600 |
| **No template control** | - | <2 | <2 | <2 |

In all these cases, the molecular methods produced results that were more indicative of expectations, in comparison to the enumeration method. However, higher levels for enterococci enumeration counts or for the universal UniBaC and UniEC did not imply human pollution as was shown with the FCF results.

### 9.3.3 Monoshone

The human specific pMR assays showed human-associated enteric pollution at all site located downstream of the Monoshone SWO and none above. The UniBaC and the UniEC assays both showed the presence of their respective targets at all of these sites. The enterococci enumeration technique and both of the universal assays showed the lowest levels at the Monoshone headwater. The UniEC and UniBaC assay showed an increase (~ one order of magnitude) in their respective targets at the site just upstream of the outfall.

Both the UniBaC and the UniEC showed an increase in their respective targets at the site just downstream of the outfall in comparison to the location just upstream. The UniEC showed a two-fold increase while the UniBaC showed two orders of magnitude increase in the downstream site. Both universal assays showed a decrease in their respective targets as the sites progressed further downstream of the outfall. Samples collected of the Wissahickon creek above the Monoshone confluence had lower levels of *Bacteroides* spp. and *Enterococcus* spp.

than those of the Monoshone before the confluence. The enterococci enumeration results for these set of samples were relatively similar to those on the UniEC assays.

**9.4 Discussion**

The literature review indicates the possibility that no single marker or approach may be considered an ideal MST method and the recommendation is for the application of multiple markers or approaches in a toolbox approach to complement each other (Schriewer et al. 2010; USEPA 2005b). This study evaluated animal specific and universal molecular markers for their potential MST applicability. The objective was to augment the human specific assays, in effect creating a toolbox approach, to better characterize the sources of pollution detected in samples collected in the Philadelphia watershed. Firstly, specificity and sensitivity analyses of the animal specific protocols were done to determine transferability to the Drexel MST facilities and applicability to this geographic region. Then, conventional enterococci enumeration, the universal and the host specific molecular assays were used to characterize three of the Philadelphia sites investigated throughout this study.

The cow assay showed 100% specificity and sensitivity. However, both the Canadian goose and pig assays returned negative results even for samples associated with their respective host. The 0% sensitivity indicated 0% positive predictive value of the Canadian goose and pig assays. The findings of the Canadian goose and pig assays highlight limitations of applying MST protocols directly from the literature without first evaluating their applicability. The specificity, sensitivity, host distribution, temporal and geographical stability of a MST protocol are a few of the parameters that needs evaluating and/or standardizing before state or national acceptance (Santo Domingo et al. 2007).

The UniEC assay showed more *Enterococcus* spp. occurrence than the conventional enterococci enumeration counts (five orders of magnitude higher). This indicated an increased detection sensitivity of the molecular protocol over the conventional cultivation technique. Irrespective of technique, the SCEE site showed, on average, lower levels of general *Enterococcus* spp. occurrences than the FCF sites. This indicates lower levels of enteric pollution at the SCEE site than the FCF site. This finding supports the expectation since the SCEE site was more isolated and assumed to be less impacted by animals or humans enteric pollution than the FCF sites. The UniBaC assay also showed lower levels of enteric pollution at the SCEE site than the FCF site.

Both the UniBaC and the UniEC assays showed a spike in the level of the respective targets in the FCF pond sample in comparison the other FCF sites.

This was expected since this site is relatively stagnant and is in close proximity to the activities and animal shelters of the farm. The spike was much more elevated for the UniBaC (~ x100) versus that of the UniEC assay (~ x2) indicating a higher concentration of *Bacteroides* spp. associated with enteric pollution at these sites in comparison to enterococci. However, the conventional enterococci enumeration technique showed the opposite result with the lowest count for all FCF sample collected on this occasion at the pond site. In addition, both the UniBaC and the UniEC assays showed the lowest level of their respective targets at the FCF headwater, which was expected. Again, the enterococci enumeration technique showed a contrary result for this site with higher counts than the FCF pond and the Pennypack. The discrepancy in the cultivation method could be inhibition of enterococci growth due to high turbidity and algal growth in the pond samples.

While the universal assays and the enterococci cultivation methods indicated general fecal pollution at the FCF and the SCEE sites they did not indicate the source. The animal assays indicated no cattle, Canadian goose, or pig sources of pollution at FCF and SCEE. However, the Canadian goose and pig findings were inconclusive since these animal specific assays showed 0% sensitivity to their targets and require further evaluation. The cattle specific assay results is contrary to the expectation for FCF.

The human specific pMR assays showed no human-associated pollution at the SECC or the FCF sites except for the Pennypack site located after its confluence with the FCF creek. Conversely, the UniBaC assay showed high levels of general *Bacteroides* spp. which supported the posit that there were fecal pollution at these sites. However, they were not of human origin. In addition, higher enterococci enumeration levels did not imply human pollution as was shown with the FCF before and after confluence results. This finding highlighted a limitation in using the conventional enterococci test to indicate human enteric pollution

For the Monoshone sites, both the UniBaC and the UniEC showed a source of fecal pollution at all these sites. The spike observed at the site just downstream of the Monoshone SWO and the decrease in levels as the sites progressed further downstream of the outfall indicated the SWO as the potential source. The enterococci enumeration corroborates the universal molecular markers results for these set of samples. Given the caveat previously mentioned, the animal specific assays indicated no cow, Canadian goose or pig at the Monoshone sites, which was expected. However, the human specific pMR assays indicated human-associated enteric pollution as the source at the Monoshone SWO.

Overall, the universal molecular techniques showed results more consistent with expectations in comparison to the conventional cultivation method. Of the two universal assays, the UniBaC showed more vividly the introduction of fecal pollution. This was highlighted by the FCF pond and the Monoshone just after outfall sites, which were expected to have the most elevated levels of pollution at either location.

This study highlighted the advantages in using multiple molecular markers to characterize samples/sites. Each protocol (pMR, UniBaC, UniEC and cow) gave a different level of detail to the situation and by combining all we got a more comprehensive picture of what is happening. This study also showed the limitations in applying protocols directly for the literature and implied through evaluation before application. All the animal specific assays may require optimization, modification or changing before they can be used to make definitive inferences.

# CHAPTER 10: CONCLUSIONS AND FUTURE RESEARCH

## 10.1 Conclusion

The literature strongly indicates the possibility of a host association of *E. faecium* and *Bacteroides*, and the potential for their use in MST. This study conducted research to determine if these relationships existed and, if so, whether these organisms are suitable for MST application.

### 10.1.1 Objective and Hypothesis 1

The results of our bioinformatics and phylogenetic analyses provide good support for *Bacteroides dorei* human host specificity. A redesigned primer and probe set was developed for a probe based qPCR assay that specifically targets bacterial sequences from a human-specific *B. dorei* lineage (HF68-HR183rc). The results from comparative assays of HF68-HR183rc against a HF183 protocol obtained from the literature showed similar sensitivity, but an improvement in the specificity of our newly designed protocol. The qPCR assays of both the HF68-HR183rc and the HF183 protocols showed 100% sensitivity for human-associated WWTP samples with gene copies/100 mL of sample ranging from 1.0E5 to 1.0E8. The qPCR assays showed the specificity of the pMR protocol to be 93% (PPV = 77%) in comparison to 81% for the pOS (PPV = 54%). However, assays of rabbit samples showed a high number of positive assays with both the redesigned (58%) and the HF183 (67%) protocol. Six of the extracted rabbit DNA amplicons were sequenced and submitted to NCBI (GenBank Accession #'s JX889727-JX889732)

The results of our bioinformatics and phylogenetic analyses provide very little support for *E. faecium* human host specificity. A redesigned primer and probe set was developed for a probe based qPCR assay that specifically targets bacterial sequences targeting the *E. faecium* $esp_{fm}$ gene. This study showed limitations in using *E. faecium* 16S rRNA and *esp* genes in differentiating the hosts of fecal pollution. Assays conducted of *E. faecium* isolates from the MST library showed a specificity of 96.4% with both PCR-gel electrophoresis and SYBR Green assays. In addition, a very low sensitivity was obtained of assays conducted of the *E. faecium* isolates from the MST library, event for the more sensitive SYBR Green qPCR protocol (20.5%). Assays of the redesigned $esp_{fm}$ protocol were not as successful as a quantitative assay, requiring an enrichment process.

**10.1.2 Hypothesis 2**

The specificity of the HF68-HR183rc qPCR MST protocol was evaluated using fecal samples of non-human animals that inhabit the tropical island of Puerto Rico. Comparison of these results with those obtained of fecal samples of animal that inhabit the Philadelphia environs showed similar findings. On a consistent basis, rabbit associated samples showed high levels of the human specific targets irrespective of geographic location. For the other non-human animal associated samples, neither the Philadelphia nor the Puerto Rico associated inhabitants showed the human specific targets in particular when the DNA concentrations were normalized to be indicative of environmental levels. Sensitivity assays using human enteric samples from Puerto Rico were not conducted during this study due to time limitations and health permit restrictions.

#### 10.1.3 Hypothesis 3

The HF68-HR183rc qPCR MST protocol was used in assisting to characterize the sources of pollution of sub-watersheds of the temperate environs of Philadelphia/Delaware watershed; watersheds of the tropical island of Puerto Rico; and different types of green infrastructures in Philadelphia and New York. The standard *Enterococcus* enumeration protocol showed fecal pollution of these environmental sites by universal (all host groups) sources. Conversely, HF68-HR183rc qPCR MST protocols were shown to indicate fecal pollution of sites suspected to be impacted by human sources. In all the cases, neither the presence of enterococci nor the levels of the enumeration counts using the standard cultivation methods could be used to indicate the presence or levels of human enteric pollution. However, when the standard method was used in a toolbox with the HF68-HR183rc qPCR MST protocol we were able to get a better picture. The standard method indicated where there were possibly composite sources of enteric pollution and the HF68-HR183rc qPCR MST protocol was then used to differentiate sites that were possibly impacted by human sources of enteric pollution. The conclusions made needs to take into consideration the caveat that enterococci may grow in environmental waters. Other standard cultivation assays (e.g. *Salmonella*, *Cryptosporidium*, *Cyclospora*, etc.) were able to add information that enabled better characterization of sites investigated.

#### 10.1.4 Hypothesis 4

There are advantages in using multiple molecular markers to characterize samples/sites. Each protocol gave different levels of detail to the situation and by combining all we got a more comprehensive picture of what is happening. In comparison to conventional cultivation methods, the molecular methods produced results in more quickly, hours in comparison to days. In addition, the molecular methods showed results that were more indicative of expectations when environmental samples of known sources of pollution were assayed. The universal *Bacteroides* qPCR assays showed sharper increases in marker concentrations due to recent sources of enteric pollution in comparison to universal enterococci qPCR assays. However, rate of decay of the universal enterococci markers in environmental waters was lower than that of the

*Bacteroides*. Thus used together, the *Bacteroides* may suggest recent and the enterococci older or more persistent sources of enteric pollution. This study also showed the limitations in applying protocols directly for the literature and implied through evaluation before application.

### 10.1.5 Conclusion

Overall, all these assays, conventional cultivation, universal molecular and host specific molecular support the use of a toolbox approach to MST. The newly designed HF68-HR183rc assay could be of considerable use in a standardized toolbox to screen a watershed for human enteric sources of pollution if rabbit fecal sources are taken into consideration.

In this study, we achieved our research objective of developing a qPCR MST assays to identify and quantify human microbial pollution in watersheds using genetic markers targeting *Bacteroides dorei,* and have shown:

- Human specific genetic markers found in *Bacteroides* will differentiate human from non-human sources of fecal pollution (Hypothesis 1).
- Human specific genetic markers found in *Bacteroides* are geographically stable (Hypothesis 2).

- A toolbox approach, using our human specific molecular qPCR assays and current cultivable FIB tests, will provide better predictive information on the source of pollution than the use of a single protocol (Hypothesis 3).
- A toolbox approach, using multiple molecular qPCR markers targeting multiple hosts, will provide better predictive information of the source of pollution than the use of a single molecular marker or cultivable FIB tests (Hypothesis 4).

**10.2 Future Research**

The HF68-HR183rc qPCR MST protocol showed success in differentiating human from non-human enteric waste. The protocol should be used to reassess some of the site evaluated during this study to determine the temporal stability of the marker. Also, additional sites should be assayed from other diverse geographical regions to determine regional transferability of the marker. These assays would assist in evaluating the feasibility of the HF68-HR183rc marker for national and global adaptation.

Future studies should also include developing assays to differentiate individual animals groups (e.g. cow, avian, dog, rabbit, etc.). All the animal specific assays of this study showed the need for optimization, modification or changing before they can be used to make definitive inferences. However, animal specific assays would augment a standardized MST toolbox. An assay of one host target may be able to complement another assay targeting another host, one helping to reduce or eliminate the errors/false results of the other.

*E. faecium* may still prove beneficial to the MST community. Since enterococci are still used for conventional water quality assessment of recreational waters, the use of general/universal molecular assays for this group of organisms may be more advantageous. In addition, future research should be redirected to evaluating other genes of this organism for their MST potential.

## LIST OF REFERENCES

Ahmed W, Goonetilleke A, Powell D, Chauhan K, Gardner T (2009a) Comparison of molecular markers to detect fresh sewage in environmental waters. Water Res 43(19):4908-17 doi:10.1016/j.watres.2009.09.047

Ahmed W, Goonetilleke A, Powell D, Gardner T (2009b) Evaluation of multiple sewage-associated *Bacteroides* PCR markers for sewage pollution tracking. Water Res 43(19):4872-7 doi:10.1016/j.watres.2009.08.042

Ahmed W, Hodgers L, Sidhu JP, Toze S (2012a) Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland, Australia. Appl Environ Microbiol 78(1):219-26 doi:10.1128/AEM.06554-11

Ahmed W, Huygens F, Goonetilleke A, Gardner T (2008a) Real-time PCR detection of pathogenic microorganisms in roof-harvested rainwater in Southeast Queensland, Australia. Appl Environ Microbiol 74(17):5490-6 doi:10.1128/AEM.00331-08

Ahmed W, Neller R, Katouli M (2005) Host species-specific metabolic fingerprint database for enterococci and Escherichia coli and its application to identify sources of fecal contamination in surface waters. Appl Environ Microbiol 71(8):4461-8 doi:10.1128/AEM.71.8.4461-4468.2005

Ahmed W, Sidhu JPS, Toze S (2012b) An Attempt to Identify the Likely Sources of Escherichia coli Harboring Toxin Genes in Rainwater Tanks. Environ Sci Technol doi:10.1021/es300292y

Ahmed W, Stewart J, Gardner T, Powell D (2008b) A real-time polymerase chain reaction assay for quantitative detection of the human-specific enterococci surface protein marker in sewage and environmental waters. Environ Microbiol 10(12):3255-64 doi:10.1111/j.1462-2920.2008.01715.x

Ahmed W, Stewart J, Gardner T, Powell D (2008c) A real-time polymerase chain reaction assay for quantitative detection of the human-specific enterococci surface protein marker in sewage and environmental waters. Environmental Microbiology 10(12):3255-64 doi:EMI1715 [pii] 10.1111/j.1462-2920.2008.01715.x

Ahmed W, Stewart J, Gardner T, Powell D, Brooks P, Sullivan D, Tindale N (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(16):3771-9 doi:10.1016/j.watres.2007.02.051

Ahmed W, Stewart J, Powell D, Gardner T (2008d) Evaluation of the host-specificity and prevalence of enterococci surface protein (*esp*) marker in sewage and its application for sourcing human fecal pollution. Journal of Environmental Quality 37(4):1583-8 doi:37/4/1583 [pii] 10.2134/jeq2007.0474

Ahmed W, Vieritz A, Goonetilleke A, Gardner T (2010) Health Risk from Potable and Non-Potable Uses of Roof-Harvested Rainwater in Australia Using Quantitative Microbial Risk Assessment. Appl Environ Microbiol:AEM.00944-10 doi:10.1128/aem.00944-10

Allsop K, Stickler D (1985) An assessment of *Bacteroides fragilis* group organisms as indicators of human faecal pollution. J Appl Microbiol 58(1):95-99 doi:10.1111/j.1365-2672.1985.tb01433.x

Alm EW, Burke J, Hagan E Replication of the fecal indicator bacteria Escherichia coli in beach sand. In, 2004.

Altman DG, Bland JM (1994) Statistics Notes: Diagnostic tests 1: sensitivity and specificity. Br Med J 308(6943):1552 doi:10.1136/bmj.308.6943.1552

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403-10

Anderson KL, Whitlock JE, Harwood VJ (2005) Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl Environ Microbiol 71(6):3041-8 doi:10.1128/AEM.71.6.3041-3048.2005

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55(4):539-52 doi:10.1080/10635150600755453

APHA, AWWA, WEF (2005) Standard Methods for the Examination of Water and Wastewater, 21th edition edn. American Public Health Association, American Water Works Association, & Water Environment Federation, Washington, DC

Avelar KE, Moraes SR, Pinto LJ, Silva e Souza WdG, Domingues RM, Ferreira MC (1998) Influence of stress conditions on *Bacteroides fragilis* survival and protein profiles. Zentralbl Bakteriol 287(4):399-409

Bae S, Wuertz S (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(4):922-32 doi:10.1128/AEM.05157-11

Bakir MA, Sakamoto M, Kitahara M, Matsumoto M, Benno Y (2006) *Bacteroides dorei* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 56(Pt 7):1639-43 doi:56/7/1639 [pii] 10.1099/ijs.0.64257-0

Balleste E, Blanch AR (2010) Persistence of *Bacteroides* spp. populations in a river measured by molecular and culture techniques. Appl Environ Microbiol:AEM.00883-10 doi:10.1128/aem.00883-10

Benno Y, Endo K, Mizutani T, Namba Y, Komori T, Mitsuoka T (1989) Comparison of fecal microflora of elderly persons in rural and urban areas of Japan. Appl Environ Microbiol 55(5):1100-5

Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW (2012) GenBank. Nucleic Acids Res 40(Database issue):D48-53 doi:10.1093/nar/gkr1202

Berg G (1978) Indicators of Viruses in Water and Food. Ann Arbor Science, Ann Arbor, Michigan

Bergholz PW, Noar JD, Buckley DH (2011) Environmental patterns are imposed on the population structure of Escherichia coli after fecal deposition. Appl Environ Microbiol 77(1):211-9 doi:10.1128/AEM.01880-10

Bernhard AE, Field KG (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(4):1587-1594 doi:10.1128/aem.66.4.1587-1594.2000

Bernhard AE, Field KG (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(10):4571-4574

Bonde GJ (1966) Bacteriological methods for estimation of water pollution. Health Lab Sci 3(2):124-8

Bucci V, Vulić M, Ruan X, Hellweger FL (2011) Population Dynamics of *Escherichia coli* in Surface Water. Journal of the American Water Resources Association 47(3):611-619 doi:DOI 10.1111/j.1752-1688.2011.00528.x

Byappanahalli M, Fujioka R (2004) Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. Wat Sci Technol 50(1):27-32

Byappanahalli MN, Fujioka RS (1998) Evidence that tropical soil environment can support the growth of Escherichia coli. Wat Sci Tech 38:171-174

Byappanahalli MN, Przybyla-Kelly K, Shively DA, Whitman RL (2008) Environmental occurrence of the enterococcal surface protein (esp) gene is an unreliable indicator of human fecal contamination. Environ Sci Technol 42(21):8014-8020 doi:Doi 10.1021/Es800481p

Byappanahalli MN, Shively DA, Nevers MB, Sadowsky MJ, Whitman RL (2003) Growth and survival of Escherichia coli and enterocci populations in the macro-alga Cladophora (Chlorophyta). FEMS Microbiol Ecol 46:203-211

Carrillo M, Estrada E, Hazen TC (1985) Survival and enumeration of the fecal indicators Bifidobacterium adolescentis and Escherichia coli in a tropical rain forest watershed. Appl Environ Microbiol 50(2):468-76

Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 69(2):330-9 doi:10.1016/j.mimet.2007.02.005

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 33:D294-D296 doi:Doi 10.1093/Nar/Gki038

Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res 31(1):442-443 doi:Doi 10.1093/Nar/Gkg039

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37(Database issue):D141-5 doi:10.1093/nar/gkn879

Colford JM, Schiff KC, Griffith JF, Yau V, Arnold BF, Wright CC, Gruber JS, Wade TJ, Burns S, Hayes J, McGee C, Gold M, Cao YP, Noble RT, Haugland R, Weisberg SB (2012) Using rapid indicators for Enterococcus to assess the risk of illness after exposure to urban runoff contaminated marine water. Water Res 46(7):2176-2186 doi:DOI 10.1016/j.watres.2012.01.033

Converse RR, Blackwood AD, Kirs M, Griffith JF, Noble RT (2009) Rapid QPCR-based assay for fecal *Bacteroides* spp. as a tool for assessing fecal contamination in recreational waters. Water Res 43(19):4828-37 doi:10.1016/j.watres.2009.06.036

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. Science 326(5960):1694-7 doi:10.1126/science.1177486

Curriero FC, Patz JA, Rose JB, Lele S (2001) The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. Am J Public Health 91(8):1194-1199 doi:10.2105/ajph.91.8.1194

CWA § 303 (d) Impaired Waters and Total Maximum Daily Loads. United States Environmental Protection Agency. http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/index.cfm

Davies CM, Long J, Donald M, Ashbolt N (1995) Survival of fecal microorganisms in marine and freshwater sediment. Appl Environ Microbiol 61:1888-1896

Desmarais TR, Solo-Gabriele HM, Palmer CJ (2002) Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. Appl Environ Microbiol 68(3):1165-72

Dombek PE, Johnson LK, Zimmerley ST, Sadowsky MJ (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Journal of Applied and Environmental Microbiology 66(6):2572 - 2577

Ebdon JE, Taylor HD (2006) Geographical stability of enterococcal antibiotic resistance profiles in Europe and its implications for the identification of fecal sources. Environ Sci Technol 40(17):5327-5332 doi:10.1021/es0607180

Eisenhart C, Wilson PW (1943) Statistical methods and control in bacteriology. Microbiol Mol Biol Rev 7(2):57-137

Field KG, Samadpour M (2007) Fecal source tracking, the indicator paradigm, and managing water quality. Water Res 41(16):3517-38 doi:S0043-1354(07)00425-3 [pii] 10.1016/j.watres.2007.06.056

Fiksdal L, Maki J.S. Lacroix S.J. Staley J.T. (1985) Survival and detection of *Bacteroids* spp., Prospective indicator bacteria. Appl Environ Microbiol 49(1):148-150

Fiksdal L, Maki JS, LaCroix SJ, Staley JT (1985) Survival and detection of *Bacteroides* spp., prospective indicator bacteria. Appl Environ Microbiol 49(1):148-50

Fish JT, Pettibone GW (1995) Influence of freshwater sediment on the survival of Escherichia coli and Salmonella sp. as measured by three methods of enumeration. Lett Appl Microbiol 20(5):277-81

Franks AH, Harmsen HJM, Raangs GC, Jansen GJ, Schut F, Welling GW (1998) Variations of Bacterial Populations in Human Feces Measured by Fluorescent In Situ Hybridization with Group-Specific 16S rRNA-Targeted Oligonucleotide Probes. Appl Environ Microbiol 64(9):3336-3345

Fremaux B, Boa T, Yost CK (2010) Quantitative Real-Time PCR Assays for Sensitive Detection of Canada Goose-Specific Fecal Pollution in Water Sources. Appl Environ Microbiol 76(14):4886-4889 doi:10.1128/aem.00110-10

Fujioka RS, Shizumura LK (1985) CLOSTRIDIUM-PERFRINGENS, A RELIABLE INDICATOR OF STREAM WATER-QUALITY. Journal Water Pollution Control Federation 57(10):986-992

Gaffield SJG, R L; Richards, L A; Jackson, R J (2003) Public Health Effects of Inadequately Managed Stormwater Runoff. Am J Public Health 93(9):1527 - 1533

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14(7):685-95

Gilmore MS, Clewell DB, Courvalin P, Dunny GM, Murray BE, Rice LB (2002) The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance. ASM Press, Washington, D.C.

Gourmelon M, Caprais MP, Ségura R, Le Mennec C, Lozach S, Piriou JY, Rincé A (2007) Evaluation of Two Library-Independent Microbial Source Tracking Methods To Identify Sources of Fecal Contamination in French Estuaries. Appl Environ Microbiol 73(15):4857-4866 doi:10.1128/aem.03003-06

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27(2):221-224 doi:10.1093/molbev/msp259

Green HC, Dick LK, Gilpin B, Samadpour M, Field KG (2011) Genetic Markers for Rapid PCR-based Identification of Gull, Canada Goose, Duck, and Chicken Fecal Contamination in Water. Appl Environ Microbiol doi:10.1128/aem.05734-11

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 59(3):307-321 doi:10.1093/sysbio/syq010

Haas CN (1989) Estimation of microbial densities from dilution count experiments. Applied Environmental Microbiology 55(8):1934-42

Haas CN, Heller B (1986) Statistics of enumerating total coliforms in water samples by membrane filter procedures. Water Res 20(4):525-530 doi:Doi: 10.1016/0043-1354(86)90203-4

Haas CN, Heller B (1988) Test of the validity of the Poisson assumption for analysis of most-probable-number results. Appl Environ Microbiol 54(12):2996-3002

Haas CN, Rose JB (1996) Distribution of *Cryptosporidium* oocysts in a water supply. Water Res 30(10):2251-2254

Haas CN, Rose JB, Gerba CP (1999) Quantitative Microbial Risk Assessment. John Wiley & Sons, New York, NY

Hamady M, Lozupone C, Knight R (2010) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J 4(1):17-27 doi:10.1038/ismej.2009.97

Hardina CM, Fujioka RS (1991) Soil: the environmental source of Escherichia coli and Enterococci in Hawaii's streams. Environmental Toxicology \& Water Quality 6:185-195

Haugland RA, Siefring SC, Wymer LJ, Brenner KP, Dufour AP (2005) Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. Water Res 39(4):559-68 doi:10.1016/j.watres.2004.11.011

Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC (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(6):348-57 doi:S0723-2020(10)00074-3 [pii] 10.1016/j.syapm.2010.06.001

Herson D, Nichols C, Verville K, Stromberg T, Bright A, Ramírez Toro G, Martinez L, Ramirez Z, Minnigh H Occurrence of Salmonella spp. in Small Potable Water Systems in the Tropics Correlated with Microbiological Indicators of Water Quality. In: 104th General Meeting, Atlanta, GA, 2005. vol Presentation Number Q-324. American Society for Microbiology,

Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. Science 292(5519):1115-1118

Hunter P, Ramirez Toro G, Minnigh H (2010) Impact on diarrhoeal illness of a community educational intervention to improve drinking water quality in rural communities in Puerto Rico. BMC Public Health 10(1):219

Isenbarger TA, Carr CE, Johnson SS, Finney M, Church GM, Gilbert W, Zuber MT, Ruvkun G (2008) The most conserved genome segments for life detection on Earth and other planets. Orig Life Evol Biosph 38(6):517-33

Jeffers SM (2012) Characterization of the rainfall-runoff response of an urban combined sewer catchment using observed and analytical methods. Drexel University

Jiang S, Noble R, Chui WP (2001) Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. Appl Environ Microbiol 67(1):179-184 doi:10.1128/aem.67.1.179-184.2001

Johnson LK, Brown MB, Carruthers EA, Ferguson JA, Dombek PE, Sadowsky MJ (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(8):4478-4485 doi:10.1128/aem.70.8.4478-4485.2004

Khatib LA, Tsai YL, Olson BH (2002) A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic Escherichia coli. Appl Environ Microbiol 59(1):97-104 doi:10.1007/s00253-002-0959-y

Kildare BJ, Leutenegger CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S (2007) 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res 41(16):3701-15 doi:10.1016/j.watres.2007.06.037

Kim SY, Lee JE, Lee S, Lee HT, Hur HG, Ko G (2010) Characterization of *Enterococcus* spp. from human and animal feces using 16S rRNA sequences, the esp gene, and PFGE for microbial source tracking in Korea. Environ Sci Technol 44(9):3423-3428 doi:Doi 10.1021/Es903282p

Kirs M, Smith DC (2007) Multiplex quantitative real-time reverse transcriptase PCR for F+-specific RNA coliphages: a method for use in microbial source tracking. Appl Environ Microbiol 73(3):808-14 doi:10.1128/AEM.00399-06

Kreader C (1995) Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol 61(4):1171-1179

Kreader CA (1998) Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. Appl Environ Microbiol 64(10):4103-5

Lasalde C, Rodriguez R, Toranzos GA (2005) Statistical analyses: possible reasons for unreliability of source tracking efforts. Appl Environ Microbiol 71(8):4690-5 doi:10.1128/AEM.71.8.4690-4695.2005

Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G (2006) Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 72(6):4214-4224 doi:Doi 10.1128/Aem.01036-05

Lee SH, Kim SJ (2002) Detection of infectious enteroviruses and adenoviruses in tap water in urban areas in Korea. Water Res 36(1):248-256 doi:10.1016/s0043-1354(01)00199-3

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics (Oxford, England) 23(1):127-8

Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39(suppl 2):W475-W478 doi:10.1093/nar/gkr201

Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI (2005) Obesity alters gut microbial ecology. Proceedings of the National Academy of Sciences of the United States of America 102(31):11070-5

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI (2008) Evolution of mammals and their gut microbes. Science 320(5883):1647-51 doi:10.1126/science.1155725

Lozupone C, Hamady M, Knight R (2006) UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7:- doi:Artn 371

Doi 10.1186/1471-2105-7-371

Ludwig W, Schleifer KH (2000) How quantitative is quantitative PCR with respect to cell counts? Syst Appl Microbiol 23(4):556-562

Lynch PA, Gilpin BJ, Sinton LW, Savill MG (2002) The detection of Bifidobacterium adolescentis by colony hybridization as an indicator of human faecal pollution. J Appl Microbiol 92(3):526-533 doi:10.1046/j.1365-2672.2002.01557.x

Mieszkin S, Furet J-P, Corthier G, Gourmelon M (2009) Estimation of Pig Fecal Contamination in a River Catchment by Real-Time PCR Using Two Pig-Specific Bacteroidales 16S rRNA Genetic Markers. Appl Environ Microbiol 75(10):3045-3054 doi:10.1128/aem.02343-08

Mieszkin S, Yala JF, Joubrel R, Gourmelon M (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(3):974-84 doi:10.1111/j.1365-2672.2009.04499.x

Mushi D, Byamukama D, Kivaisi AK, Mach RL, Farnleitner AH (2010) Sorbitol-fermenting Bifidobacteria are indicators of very recent human faecal pollution in streams and groundwater habitats in urban tropical lowlands. J Water Health 8(3):466-478 doi:10.2166/wh.2010.116

Newton RJ, Vandewalle JL, Borchardt MA, Gorelick MH, McLellan SL (2011) Lachnospiraceae and Bacteroidales alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. Appl Environ Microbiol 77(19):6972-81

NRC (2004) Indicators for waterborne pathogens. National Research Council of the National Academies, Washington, DC

NRC (2008) Urban Stormwater Management in the United States: Committee on Reducing Stormwater Discharge Contributions to Water Pollution. The National Research Council of the National Academies Press, Washington, DC

Ochman H, Worobey M, Kuo CH, Ndjango JB, Peeters M, Hahn BH, Hugenholtz P (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8(11):e1000546 doi:10.1371/journal.pbio.1000546

Okabe S, Okayama N, Savichtcheva O, Ito T (2007) Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl Microbiol Biotechnol 74(4):890-901 doi:DOI 10.1007/s00253-006-0714-x

Olsen GJ, Overbeek R, Larsen N, Marsh TL, McCaughey MJ, Maciukenas MA, Kuan W-M, Macke TJ, Xing Y, Woese CR (1992) The Ribosomal Database Project. Nucleic Acids Res 20(suppl):2199-2200 doi:10.1093/nar/20.suppl.2199

Pope JM (2009) Tracking the sources of fecal contamination in the Wissahickon Creek Watershed using phenotypic and genotypic analytical methods. Drexel University

Puig A, Queralt N, Jofre J, Araujo R (1999) Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater. Appl Environ Microbiol 65(4):1772-1776

PWD (2007a) Poquessing Creek Watershed Comprehensive Characterization Report. Philadelphia Water Department

PWD (2007b) Wissahickon Creek Watershed Comprehensive Characterization Report. Philadelphia Water Department

PWD (2012a) Poquessing Watershed Base Map. Philadelphia Water Department. http://www.phillywatersheds.org/your_watershed/poquessing/key_documents_maps 2012

PWD (2012b) Wissahickon Watershed Base Map. Philadelphia Water Department. http://www.phillywatersheds.org/your_watershed/wissahickon/key_documents_maps 2012

Reischer GH, Haider JM, Sommer R, Stadler H, Keiblinger KM, Hornek R, Zerobin W, Mach RL, Farnleitner AH (2008) Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. Environ Microbiol 10(10):2598-608 doi:EMI1682 [pii] 10.1111/j.1462-2920.2008.01682.x

Reischer GH, Kasper DC, Steinborn R, Farnleitner AH, Mach RL (2007) A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. Lett Appl Microbiol 44(4):351-356 doi:DOI 10.1111/j.1472-765X.2006.02094.x

Ritter KJ, Carruthers E, Carson CA, Ellender RD, Harwood VJ, Kingsley K, Nakatsu C, Sadowsky M, Shear B, West B, Whitlock JE, Wiggins BA, Wilbur JD (2003) Assessment of statistical methods used in library-based approaches to microbial source tracking. J Water Health 1(4):209-23

Rivera SC, Hazen TC, Toranzos GA (1988) Isolation of fecal coliforms from pristine sites in a tropical rain forest. Appl Environ Microbiol 54(2):513-7

Russell JA, Goldman-Huertas B, Moreau CS, Baldo L, Stahlhut JK, Werren JH, Pierce NE (2009) Specialization and geographic isolation among *Wolbachia* symbionts from ants and Lycaenid butterflies. Evolution 63(3):624-640 doi:DOI 10.1111/j.1558-5646.2008.00579.x

Santo Domingo JW, Bambicb DG, Edgec TA, Wuertz S (2007) Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. Water Res 41(16):3539-52 doi:S0043-1354(07)00373-9 [pii] 10.1016/j.watres.2007.06.001

Santo Domingo JW, Siefring SC, Haugland RA (2003) Real-time PCR method to detect *Enterococcus faecalis* in water. Biotechnol Lett 25(3):261-5

Sauer EP, Vandewalle JL, Bootsma MJ, McLellan SL (2011) Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. Water Res 45(14):4081-91 doi:10.1016/j.watres.2011.04.049

Schriewer A, Miller WA, Byrne BA, Miller MA, Oates S, Conrad PA, Hardin D, Yang H-H, Chouicha N, Melli A, Jessup D, Dominik C, Wuertz S (2010) Bacteroidales as a predictor of pathogens in surface waters of the central California coast. Appl Environ Microbiol:AEM.00635-10 doi:10.1128/aem.00635-10

Scott TM, Jenkins TM, Lukasik J, Rose JB (2005) Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. Environ Sci Technol 39(1):283 - 287

Scott TM, Parveen S, Portier KM, Rose JB, Tamplin ML, Farrah SR, Koo A, Lukasik J (2003) Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl Environ Microbiol 69(2):1089 - 1092

Scott TM, Rose JB, Jenkins TM, Farrah SR, Lukasik J (2002) Microbial Source Tracking: Current Methodology and Future Directions. Appl Environ Microbiol 68(12):5796-5803 doi:10.1128/aem.68.12.5796-5803.2002

Seurinck S, Defoirdt T, Verstraete W, Siciliano SD (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(2):249-59 doi:EMI702 [pii] 10.1111/j.1462-2920.2004.00702.x

Shah HN, Collins DM (1990) *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus *Bacteroides*. Int J Syst Bacteriol 40(2):205-208 doi:10.1099/00207713-40-2-205

Shanks OC, Domingo JW, Lu J, Kelty CA, Graham JE (2007) Identification of bacterial DNA markers for the detection of human fecal pollution in water. Appl Environ Microbiol 73(8):2416-22 doi:AEM.02474-06 [pii]

10.1128/AEM.02474-06

Shanks OC, Kelty CA, Sivaganesan M, Varma M, Haugland RA (2009) Quantitative PCR for Genetic Markers of Human Fecal Pollution. Appl Environ Microbiol 75(17):5507-5513 doi:10.1128/aem.00305-09

Shanks OC, White K, Kelty CA, Hayes S, Sivaganesan M, Jenkins M, Varma M, Haugland RA (2010a) Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. Appl Environ Microbiol 76(5):1359-66 doi:AEM.02033-09 [pii]

10.1128/AEM.02033-09

Shanks OC, White K, Kelty CA, Sivaganesan M, Blannon J, Meckes M, Varma M, Haugland RA (2010b) Performance of PCR-Based Assays Targeting Bacteroidales Genetic Markers of Human Fecal Pollution in Sewage and Fecal Samples. Environ Sci Technol 44(16):6281-6288 doi:10.1021/es100311n

Sherer BM, Miner JR, Moore JA, Buckhouse JC (1992) INDICATOR BACTERIAL SURVIVAL IN STREAM SEDIMENTS. Journal of Environmental Quality 21(4):591-595

Siefring S, Varma M, Atikovic E, Wymer L, Haugland RA (2008) Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. J Water Health 6(2):225-37 doi:10.2166/wh.2008.022

Simpson JM, Santo Domingo JW, Reasoner DJ (2002) Microbial source tracking: state of the science. Environ Sci Technol 36(24):5279-5288 doi:10.1021/es026000b

Solo-Gabriele HM, Wolfert MA, Desmarais TR, Palmer CJ (2000) Sources of *Escherichia coli* in a coastal subtropical environment. Appl Environ Microbiol 66(1):230-7

Stewart-Pullaro J, Daugomah JW, Chestnut DE, Graves DA, Sobsey MD, Scott GI (2006) F(+)RNA coliphage typing for microbial source tracking in surface waters. J Appl Microbiol 101(5):1015-1026 doi:DOI 10.1111/j.1365-2672.2006.03011.x

Stewart JR, Gast RJ, Fujioka RS, Solo-Gabriele HM, Meschke JS, Amaral-Zettler LA, Del Castillo E, Polz MF, Collier TK, Strom MS, Sinigalliano CD, Moeller PD, Holland AF (2008) The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. Environ Health 7 Suppl 2:S3 doi:10.1186/1476-069X-7-S2-S3

Stoeckel DM, Harwood VJ (2007) Minireview: Performance, Design and Analysis in Microbial Source Tracking Studies. Appl Environ Microbiol 73(8):2405-15 doi:10.1128/AEM.02473-06

Tambalo DD, Fremaux B, Boa T, Yost CK (2012) Persistence of host-associated Bacteroidales gene markers and their quantitative detection in an urban and agricultural mixed prairie watershed. Water Res 46(9):2891-904 doi:10.1016/j.watres.2012.02.048

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol doi:10.1093/molbev/msr121

Tartera C, Lucena F, Jofre J (1989) Human origin of *Bacteroides fragilis* bacteriophages present in the environment. Appl Environ Microbiol 55(10):2696-2701

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] Chapter 2:Unit 2 3 doi:10.1002/0471250953.bi0203s00

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-80

U.S.CensusBureau (2010) State & County QuickFacts Philadelphia County, Pennsylvania. http://quickfacts.census.gov/qfd/states/42/42101.html 2012

Unisphere Inc., (FMCP) FMCPC, (EDD) eD, University D, (Parks) CoNYPR, (NRG) NRG (2010) Gowanus Canal and Flushing Bay Watershed Initiative 2009: Design for Treatment Wetlands and Rain Gardens for Treatment of Stormwater Entering Meadow Lake, Flushing Meadows Park, Queens.

USEPA (2000) Improved enumeration methods for the recreational water quality indicators: Enterococci and *Escherichia coli*. EPA/821/R-97/004. Office of Science and Technology. United States Environmental Protection Agency, Washington DC

USEPA (2004) Report to Congress: Impacts and Control of CSOs and SSOs. EPA 833 R-04-001. United States Environmental Protection Agency, Washington, DC

USEPA (2005a) Method 1600: Enterococci in water by membrane filtration using membrane-*Enterococcus* Indoxyl- -D-Glucoside Agar (mEI) EPA-821-R-04-023. United States Environmental Protection Agency, Washington DC

USEPA (2005b) Microbial Source Tracking Guide Document. EPA-600/R-05-064. Office of Research and Development. United States Environmental Protection Agency, Cincinnati, OH

USEPA (2009) Water Quality Scorecard: Incorporating Green Infrastructure Practices at the Municipal, Neighborhood, and Site Scale. EPA 231B09001. United States Environmental Protection Agency, Washington, DC

USEPA (2010a) Method A: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay. EPA-821-R-04-023. Water Oo. United States Environmental Protection Agency, Washington DC

USEPA (2010b) Proposed revised total coliform rule: Assessment and corrective action guidance manual (Draft). EPA 815-D-10-001. Office of Water. United States Environmental Protection Agency

USEPA (2012) Recreational Water Quality Criteria (Draft). Office of Water. United States Environmental Protection Agency

van den Bogaard AE, Willems R, London N, Top J, Stobberingh EE (2002) Antibiotic resistance of faecal enterococci in poultry, poultry farmers and poultry slaughterers. J Antimicrob Chemother 49(3):497-505 doi:10.1093/jac/49.3.497

Walters SP, Field KG (2006) Persistence and growth of fecal Bacteroidales assessed by bromodeoxyuridine immunocapture. Appl Environ Microbiol 72(7):4532-9 doi:72/7/4532 [pii] 10.1128/AEM.00038-06

Walters SP, Yamahara KM, Boehm AB (2009) Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: Implications for their use in assessing risk in recreational waters. Water Res 43(19):4929-4939 doi:10.1016/j.watres.2009.05.047

Weidhaas JL, Macbeth TW, Olsen RL, Harwood VJ (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(6):2094-102 doi:AEM.02555-10 [pii] 10.1128/AEM.02555-10

Whitman RL, Przybyla-Kelly K, Shively DA, Byappanahalli MN (2007) Incidence of the enterococcal surface protein (*esp*) gene in human and animal fecal sources. Environ Sci Technol 41(17):6090-6095 doi:Doi 10.1021/Es070817t

Whitman RL, Przybyla-Kelly K, Shively DA, Nevers MB, Byappanahalli MN (2008) Sunlight, season, snowmelt, storm, and source affect E. coli populations in an artificially ponded stream. Sci Total Environ 390(2-3):448-455 doi:10.1016/j.scitotenv.2007.10.014

Whitman RL, Shively DA, Pawlik H, Nevers MB, Byappanahalli MN (2003) Occurrence of Escherichia coli and enterococci in Cladophora (Chlorophyta) in nearshore water and beach sand of Lake Michigan. Appl Environ Microbiol 69(8):4714-9

Wiggins BA (1996) Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl Environ Microbiol 62(11):3997-4002

Wiggins BA, Cash PW, Creamer WS, Dart SE, Garcia PP, Gerecke TM, Han J, Henry BL, Hoover KB, Johnson EL, Jones KC, McCarthy JG, McDonough JA, Mercer SA, Noto MJ, Park H, Phillips MS, Purner SM, Smith BM, Stevens EN, Varner AK (2003) Use of Antibiotic Resistance Analysis for Representativeness Testing of Multiwatershed Libraries. Appl Environ Microbiol 69(6):3399-3405 doi:10.1128/aem.69.6.3399-3405.2003

Yamahara KM, Sassoubre LM, Goodwin KD, Boehm AB (2012) Occurrence and persistence of bacterial pathogens and indicator organisms in beach sand along the California coast. Appl Environ Microbiol 78(6):1733-45 doi:10.1128/AEM.06185-11

Yamahara KM, Walters SP, Boehm AB (2009) Growth of enterococci in unaltered, unseeded beach sands subjected to tidal wetting. Appl Environ Microbiol 75(6):1517-1524 doi:10.1128/aem.02278-08

Zvelebil M, Baum JO (2007) Understanding Bioinformatics. Garland Science

## APPENDIX A: GLOSSARY

**Accession number** - An identification number that used to be assigned (for cataloging purposes) to volumes of studies submitted to OPP.

**Alignment** -The process of lining up two or more sequences to achieve maximal levels of identity (and conservation, in the case of amino acid sequences) for the purpose of assessing the degree of similarity and the possibility of homology. (NCBI)

**Aerobes** - Organisms which require molecular oxygen as an electron acceptor for energy production

**Amplification** - An increase in the number of copies of a specific DNA fragment.

**Anaerobes** - A group of organisms that do not require molecular oxygen.

**Base pair (bp)** - Two nitrogenous bases (adenine and thymine or guanine and cytosine) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs. (ORNL)

**Best Management Practices (BMP)** - Procedures or controls other than effluent limitations to prevent or reduce pollution of surface water (includes runoff control, spill prevention, and operating procedures).

**Bioinformatics** - The merger of biotechnology and information technology with the goal of revealing new insights and principles in biology. (NCBI)

**Bioinformatics** - The science of managing and analyzing biological data using advanced computing techniques. (ORNL)

**BLAST** - Basic Local Alignment Search Tool. (Altschul et al.) A sequence comparison algorithm optimized for speed used to search sequence databases for optimal local alignments to a query. The initial search is done for a word of length "W" that scores at least "T" when compared to the query using a substitution matrix. Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S". The "T" parameter dictates the speed and sensitivity of the search. For additional details, see one of the BLAST tutorials (Query or BLAST) or the narrative guide to BLAST. (NCBI)

**BLAST** - A computer program that identifies homologous (similar) genes in different organisms, such as human, fruit fly, or nematode. (ORNL)

**Electrophoresis** - A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids. (ORNL)

**Genetic marker** - A gene or other identifiable portion of DNA whose inheritance can be followed. (ORNL)

**Green infrastructure** - An approach to wet weather management that is cost-effective, sustainable, and environmentally friendly. Green Infrastructure management approaches and technologies infiltrate, evapotranspire, capture and reuse stormwater to maintain or restore natural hydrologies. (USEPA)

**Homology** - Similarity in DNA or protein sequences between individuals of the same species or among different species. (ORNL)

**Hybridization** - The process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule. (ORNL)

**Library** - An unordered collection of clones (i.e., cloned DNA from a particular organism) whose relationship to each other can be established by physical mapping. (ORNL)

**Max Identity** - The highest percent identity for a set of aligned segments to the same subject sequence. (NCBI)

**Oligonucleotide** - A molecule usually composed of 25 or fewer nucleotides; used as a DNA synthesis primer. (ORNL)

**Plasmid** - Autonomously replicating extra-chromosomal circular DNA molecules, distinct from the normal bacterial genome and nonessential for cell survival under nonselective conditions. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as cloning vectors. (ORNL)

**Polymerase chain reaction (PCR)** - A method for amplifying a DNA base sequence using a heat-stable polymerase and two 20-base primers, one complementary to the (+) strand at one end of the sequence to be amplified and one complementary to the (-) strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample. (ORNL)

**Primer** - Short preexisting polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase. (ORNL)

**Probe** - Single-stranded DNA or RNA molecules of specific base sequence, labeled either radioactively or immunologically, that are used to detect the complementary base sequence by hybridization. (ORNL)

**Query coverage** - The amount of the query sequence, expressed as a percent, which overlaps the subject sequence of a BLAST query. (NCBI)

**Sequencing** - Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein. (ORNL)

**Similarity** - The extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. In BLAST similarity refers to a positive matrix score. (NCBI)

**Urban drainage systems** - The overland surface flow system, the sewer network, and the underground porous media drainage system of an urban area.

Note:

National Center for Biotechnology Information (NCBI)

Oak Ridge National Laboratory (ORNL)

# APPENDIX B: QUANTIFICATION AND DETECTION LIMIT

Error and efficiency values were used to evaluate technique execution, accuracy, and efficiency of assay. These values are important since they are a measure of the accuracy of the qPCR result given the standard curve.

The efficiency value indicates the fold increase in template material per PCR cycle. A perfect PCR reaction would replicate once every cycle resulting in an efficiency of 2.

$$C_n = 2^n C_0$$

Where $C_n$ is the amount of target amplicons at cycle $n$, $C_0$ is the initial amount of target molecules (amount of template molecule), and $n$ is the number of amplification cycles. However, actual PCR is affected by different factor that influences its efficiency E, thus:

$$C_n = E^n C_0$$

The qPCR instrument uses the slope generated from a standard curve to calculate efficiency as:

$$E = 10^{-1/slope}$$

An efficiency of 2 would thus give a slope of -3.3, i.e. $1/\log_{10}$. Slopes from -3.1 - 3.58 or efficiencies from 1.9 – 2.1 are considered good for qPCR assays.

*E. faecium* genomic DNA extracted for positive controls was used to generate standard curves for the qPCR protocol. Tenfold serial dilutions ($10^{-10}$ - $10^{0}$) were made using buffer AE (Qiagen). Replicate assays of these serial dilutions showed $10^{-7}$, the lowest concentration detected consistently, to be the detection limit of the assay. Figure B.1 shows the fluorescence history of the serial dilutions for one assay. The amplification curves show expected profiles for dilutions $10^{-7}$ - $10^{0}$. The smaller/flatter profiles observed after cycle 40 were due to dilutions $10^{-10}$ – $10^{-8}$ and were determined by MP profiles to be possible primer dimers.

The number of copies of genomic DNA is calculated based on the size of the genome of the standard organism used and its concentration is measured with the NanoDrop. The equation is:

$$gene\ copies\ = \frac{Genome\ DNA\ conc\ measured\ \ * Avogadro's\ \# * volume\ measured}{Average\ molecular\ weight\ of\ the\ genome}$$

For *E. faecium* the genome is estimated to be 2.8 Mbp (accession# ADMM00000000) and the copies of genes are equivalent to the number of organisms (1 copy of gene per genome) (Ahmed et al. 2008c). The number of copies of the *esp* gene per reaction would thus be dependent on the mass of genomic DNA per μL of standard as determined by the NanoDrop. The calculation was applied to one standard curve and showed the qPCR protocol being able to detect and estimate unit gene copies (~3-6 copies) as depicted in Table B.1.

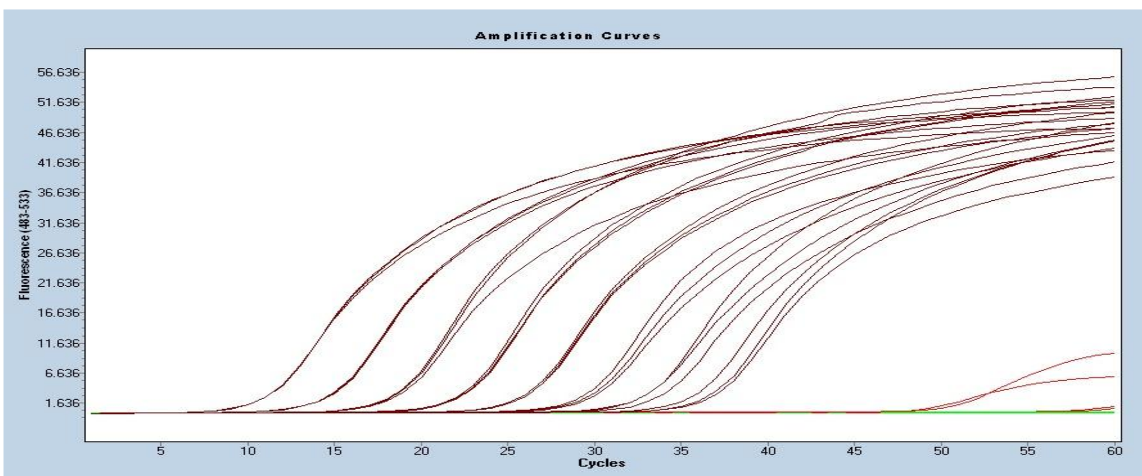



**Figure B.1: Amplification curve of the serial dilution ($10^{-10}$ - $10^{0}$) for the positive controls for one assay.**

**Table B.1: Melt temperatures and dilutions for the serial dilutions ($10^{-7}$ - $10^{0}$) for the positive control for one assay.**

| Sample Name | Melt Temperature | Serial Dilution | Real-time PCR dilution | gene copies |
|---|---|---|---|---|
| STD0 | 81.22 | 1.00E+00 | 9.81E-01 | 2.04E+07 |
| STD0 | 81.23 | 1.00E+00 | 1.05E+00 | 2.18E+07 |
| STD0 | 81.29 | 1.00E+00 | 9.71E-01 | 2.02E+07 |
| STD1 | 80.94 | 1.00E-01 | 1.07E-01 | 2.22E+06 |
| STD1 | 81.05 | 1.00E-01 | 9.94E-02 | 2.06E+06 |
| STD1 | 81.17 | 1.00E-01 | 1.04E-01 | 2.16E+06 |
| STD2 | 80.88 | 1.00E-02 | 1.01E-02 | 2.10E+05 |
| STD2 | 80.97 | 1.00E-02 | 9.32E-03 | 1.94E+05 |
| STD2 | 81.06 | 1.00E-02 | 9.82E-03 | 2.04E+05 |
| STD3 | 81.00 | 1.00E-03 | 8.77E-04 | 1.82E+04 |
| STD3 | 81.04 | 1.00E-03 | 1.03E-03 | 2.14E+04 |
| STD3 | 81.05 | 1.00E-03 | 9.81E-04 | 2.04E+04 |
| STD4 | 80.99 | 1.00E-04 | 1.03E-04 | 2.14E+03 |
| STD4 | 81.00 | 1.00E-04 | 1.08E-04 | 2.24E+03 |
| STD4 | 81.09 | 1.00E-04 | 1.04E-04 | 2.16E+03 |
| STD5 | 80.88 | 1.00E-05 | 1.05E-05 | 2.18E+02 |
| STD5 | 80.89 | 1.00E-05 | 1.12E-05 | 2.33E+02 |
| STD5 | 80.96 | 1.00E-05 | 1.32E-05 | 2.74E+02 |
| STD6 | 80.74 | 1.00E-06 | 1.63E-06 | 3.39E+01 |
| STD6 | 80.82 | 1.00E-06 | 9.86E-07 | 2.05E+01 |
| STD6 | 81.09 | 1.00E-06 | 1.46E-06 | 3.03E+01 |
| STD7 | 80.94 | 1.00E-07 | 1.43E-07 | 2.97E+00 |
| STD7 | 81.10 | 1.00E-07 | 2.75E-07 | 5.71E+00 |
| STD7 | 81.18 | 1.00E-07 | 1.55E-07 | 3.22E+00 |

Conceptually, a Poisson probability distribution describes the sampling variability in the number of organisms inoculated from a given volume of water samples (Eisenhart and Wilson 1943; Haas 1989; Haas and Rose 1996; Haas et al.

1999). Assays may not consistently generate positive results at very low concentrations (e.g. 3-6 gene copies) due to this variability. Also, studies have shown deviation from the Poisson and thus contributing to the variations in enumeration of replicate samples (Haas 1989; Haas and Heller 1986; Haas and Heller 1988). Therefore, improper technique, inaccurate/inconsistent instruments, overdispersion or underdispersion of cells/DNA during dilutions, PCR inhibition, etc., will result in variations (errors) in the qPCR results obtained beyond those accounted for by the Poisson distribution.

The relative errors in qPCR is calculated as the ratio of the difference in pipetted dilution and the qPCR dilutions computed to that of the pipetted dilutions. It should be noted that the error that results in conducting the first dilution is carried over to that of the second and, thus, error from a lower dilution propagates to the higher ditution. Figure B.2 shows a plot of the standard deviation of the computed relative error per serial dilution against the log of that pipetted serial dilutions of the Table B.1 assay. As expected, the relative error increased with dilution, with a maximum standard deviation value of 0.72 for the $10^{-7}$ serial dilution of this assay.

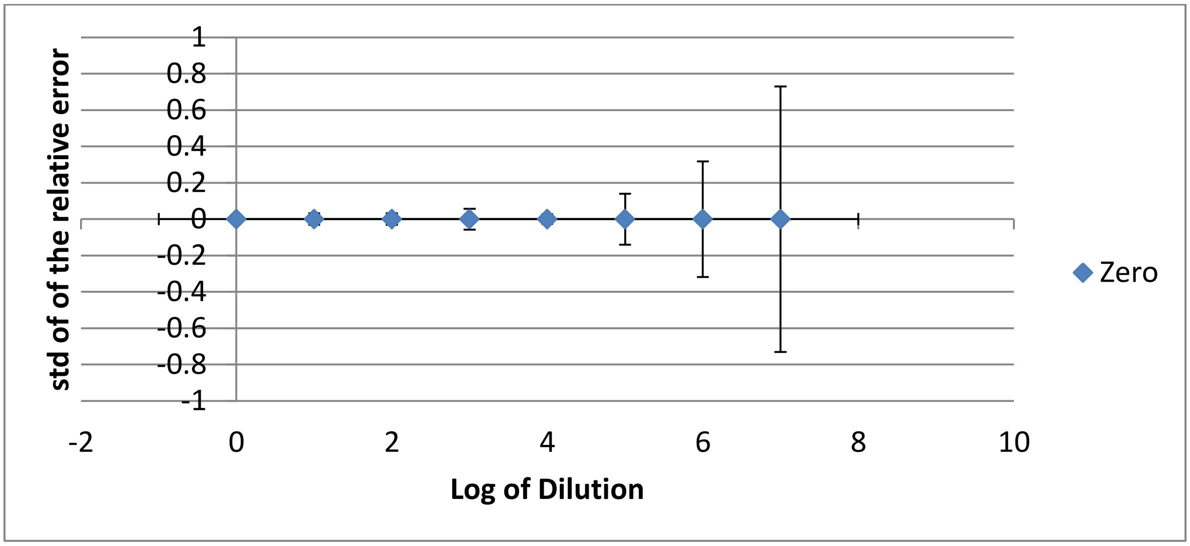



**Figure B.2: Plot showing std of the relative error (bars) for the PCR replicates at each serial dilution.**

The qPCR instrument calculates an overall error value for each standard curve as the mean squared error of the single data points fit to the regression line of the curve (Roche LightCycler 480 Operator's Manual). The standard curve data of Table B.1 computed an overall error of 0.0322 (Figure B.3) indicating good overall assay technique with precise instruments and negligible PCR inhibition. Values less than 0.2 are considered good (Roche LightCycler 480 Operator's Manual).

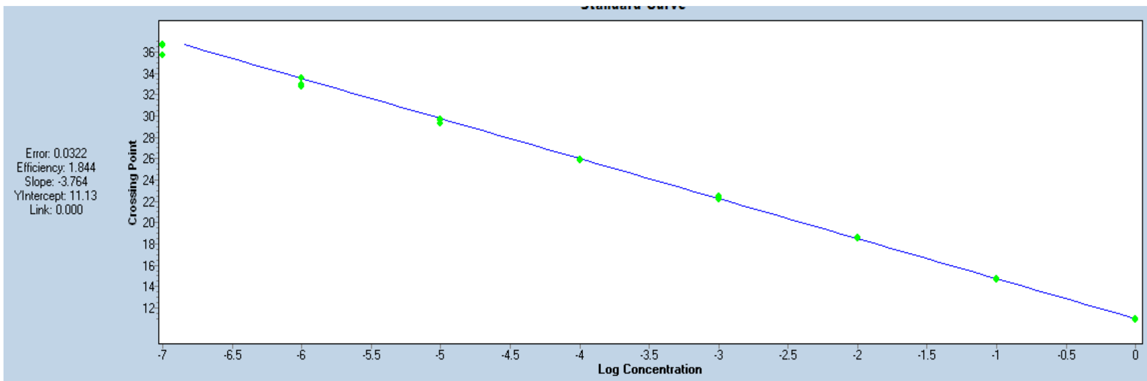



**Figure B.3: Roche LC480 generated standard curve for the data of table 2.**

Similar calculations can be used for *Bacteroides* analyses taking into consideration that there is an estimated average of five 16S rDNA operons per *Bacteroides* cell (Bernhard and Field 2000b) and that the HF183 marker is closely assosiated with the newly discovered species *Bacteroides dorei* (Bakir et al. 2006; Haugland et al. 2010), that has a genome which is estimated to be 5.5 Mbp (accession# ABWZ00000000).

# APPENDIX C: STANDARD OPERATING PROCEDURE

## *C.1 Lab Bench Clean-up for molecular work*

**Material & Equipment:**

Hand soap

10% bleach (e.g. Clorox).

Paper towel

70% ethanol (e.g. SaniHol)

DNA AWAY or DNA-EXITUSPLUS

Kimwipe

Disposable latex gloves

**Protocol:**

- Wash hands with soap and water and put on a pair of latex gloves.
- Make a solution of 10% bleach (e.g. Clorox).
- Remove pipets, pipettors, beakers of tubes, boxes of pipet tips, etc. from the bench surface.

- Using paper towel, wipe down the lab bench with 70% ethanol and allow to air dry.
- Using fresh paper towel, wipe down the entire lab bench with the bleach solution and allow to air dry.
- Using fresh paper towel, wipe down the lab bench with DNA AWAY and allow to air dry.
- Using fresh paper towel, wipe the bottom of your lab equipment with the Clorox solution before placing it on the bench.
- Using a Kimwipe, carefully wipe down all surfaces of each pipettor with DNA AWAY. Wipe off excess DNA AWAY with a fresh Kimwipe.
- Dispose of all waste in accordance with the lab's waste disposal guidelines.
- Wash your hands with soap and water.
- Put on a new pair of latex gloves before beginning molecular work.

### *C.2 Environmental water sample collection and membrane filtration (MF) prep for DNA extraction*

**Material & Equipment:**

Ice packs

Igloo

500mL or 1L sterile sample bottles

2 forceps

Alcohol

Spirit lamp/Bunsen burner

Vacuum pump (Gast)

0.45 μL filter membrane

10% bleach (e.g. Clorox)

70% ethanol (e.g. SaniHol)

Paper towel

Disposable latex gloves

50 mL beaker

Sterile DI water/phosphate buffered water (PBW)

MF equipment or disposable Microfil S filtration devices (Millipore)

**Pre &/or Post preparation**

- Wash filter funnel/s and base/s with tap water
- Soak in 10% bleach solution for ~30min
- Rinse thoroughly with DI water.
- Autoclave and store for MF protocol.

**Protocol:**

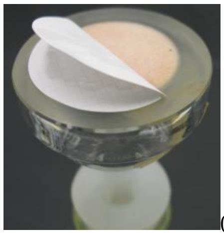
- Collected duplicate ~500 mL or 1L of environmental water samples from each site.
- Transported in igloo with frozen ice packs @ <8$^{o}$C to lab
- Using paper towel, clean lab bench with 70% alcohol followed by 10% bleach solution
- Label closed 2 mL microcentrifuge tubes in accordance to samples to be filtered

- Assemble autoclaved MF funnel and bases (or disposable Microfil S filtration devices) , MF manifold/flask, spirit lamp/Bunsen burner, and forceps in beaker with alcohol

Reusable   (Pictures from Michael Ryan)

Disposable   (Pictures from Michael Ryan)

- Aseptically filter samples with 0.45 μm filter membrane
- QA/QC - filter ~500 mL DI water for pre- (before filtration of samples) and post-filtration (after filtration of samples) controls
- If disposable Microfil S filtration devices are used then only one control is required.
- Using the two forceps for control, aseptically fold the membrane filter in half


(Picture from Michael Ryan)

- Repeat folding to ~1/8 to 1/16 size




(Pictures from Michael Ryan)

- Open appropriate labeled 2 mL microcentrifuge tubes and transfer folded filters to it.



- Close tube and store @ -20°C or -80°C until required for DNA extraction
- For reusable apparatus, thoroughly wash filter funnel/s and base/s with sterile DI water/PBW (with vacuum on) before proceeding to the next sample.

### *C.3 Animal fecal or WWTP sample collection and prep for DNA extraction*

**Material & Equipment:**

Igloo

Ice packs

Sterile Q-tips

Sterile whirl packs

2.0 mL microcentrifuge tubes

Weight Scale

**Protocol:**

- Collect 500mL of WWTP liquid sample
- For solid fecal sample, collect ~400 g of fecal samples in whirl pack
- Stored in igloo with ice packs and transported to lab@ <8$^o$C (digital thermometer)
- Distribute 180-220 mg of solid stool (200 mL WWTP) sample to 2 mL microcentrifuge tubes (12 replicates)
- Store @ -80$^o$C until required.

### *C.4 DNA extraction with QIAcube and QIAamp Stool kit*

**Material & Equipment:**

QIAcube (Qiagen)

Centrifuge (Eppendorf)

Thermomixer (Eppendorf)

Vortex Genie 2 and adaptors (Scientific Industries)

NanoDrop spectrophotometer (Thermo Scientific)

Micropipettes and tips (Rainin)

QIAamp DNA Stool Kit (Qiagen)

2.0 & 1.5 ml microcentrifuge tubes (Eppendorf)

Membrane filter (MF) with concentrated sample in 2 mL microcentrifuge tubes

**Protocol:**

- Add 1.8 mL of buffer ASL to sample in 2 mL microcentrifuge tubes and vortex for 3 min @ 3000 rpm to thoroughly mix
- For solid fecal or WWTP samples, use 1.4 mL of buffer ASL
- Incubate for 5 min @ 95 ºC in preheated Thermomixer

- Vortex for 5 sec @ 3000 rpm to homogenize samples
- Centrifuge for 1 min @ 20,000 g to settle membrane and large particles
- Add 1 InhibitEX Tablet into a new 2 mL microcentrifuge tube and add 1.2 ml of the ASL incubated samples supernatant
- Vortex for 1 min or until the tablet is completely suspended.
- Incubate suspension for 1 min at room temperature to allow inhibitors to adsorb to the InhibitEX matrix. (15 – 25 ºC)
- Centrifuge for 3 min @ 20,000 g to pellet inhibitors bound to InhibitEX matrix.
- Transfer 350 µL of the supernatant to position 2 of the rotor adaptor of the QIAcube.
- Prepare all other reagents as pre the QIAcube QIAamp DNA Stool pathogen protocol (Appendix 1 – 12 sample assay e.g. in tables below)
- On completion of the DNA extraction protocol add 1.0 µL DNA eluate to a NanoDrop to quantify the DNA concentration
- Store the remainder of the DNA eluate @ 4 ºC until required for qPCR (1 day) or store @ -20 ºC for long term storage until required

***C.5 DNA extraction (manual) with QIAamp Stool kit***

**Material & Equipment:**

Thermomixer (Eppendorf)

Vortex Genie 2 and adaptors (Scientific Industries)

NanoDrop spectrophotometer (Thermo Scientific)

Micropipettes and tips (Rainin)

QIAamp DNA Stool Kit (Qiagen)

2.0 & 1.5 ml microcentrifuge tubes (Eppendorf)

Membrane filter (MF) with concentrated sample in 2 mL microcentrifuge tubes

Centrifuge (Eppendorf)

**Protocol:**

- Add 1.6 mL of buffer ASL to sample in 2 mL microcentrifuge tubes with filter and vortex for 3 min @ 3000 rpm to thoroughly mix
- Note: For fecal samples, use 1.4 mL of buffer ASL
- Incubate for 5 min @ 70 ºC in preheated Thermomixer
- Vortex for 5 sec @ 3000 rpm to homogenize samples

- Centrifuge for 1 min @ 20,000 g to settle membrane and large particles
- Add 1 InhibitEX Tablet into a new 2 mL microcentrifuge tube and add 1.2 ml of the ASL incubated samples supernatant
- Vortex for 1 min or until the tablet is completely suspended.
- Incubate suspension for 1 min at room temperature to allow inhibitors to adsorb to the InhibitEX matrix. (15 – 25 ºC)
- Centrifuge for 3 min @ 20,000 g to pellet inhibitors bound to InhibitEX matrix.
- Pipet all the supernatant into a new 1.5 ml microcentrifuge tube (not provided) and discard pellet.
- Centrifuge for 3 min@ 20,000 g.
- Add 15 μL proteinase K into a new 1.5 ml microcentrifuge tube (not provided).
- Add 200 μL of the supernatant into the 1.5 ml microcentrifuge tube with proteinase K.
- Add 200 μL Buffer AL. Do not add proteinase K directly to Buffer AL.
- Pulse vortex to thoroughly mix.
- Incubated @ $70^{o}$C for 10 min

- Centrifuge briefly (optional)
- Add 200 μL ethanol (96-100%) to the lysate and mix by pulse vortexing.
- Transfer all of the lysate to a QIAamp spin column placed in a 2ml collection tube by pipetting (~604μL) without moistening the rim.
- Centrifuge for 1min @ 20,000 g and discard collection tube & flow-through.
- Place QIAamp spin column in new 2ml collection tube.
- Add 500 μL Buffer AW1.
- Centrifuge for 1 min @ 20,000 g and discard collection tube & flow-through
- Add 500 μL Buffer AW2.
- Centrifuge for 3 min @ 20,000 g and discard flow-through (program 3).
- Optional - place QIAamp spin column in a new collection tube and centrifuge for 3 min @ 20,000 g and discard collection tube & flow-through.
- Placed QIAamp spin column in a 1.5 ml microcentrifuge tube (not provided by Qiagen) labeled with sample ID.

- Add 200 μL Buffer AE and incubated at room temperature for 1min.
- Centrifuge for 1 min @ 20,000 g and discard QIAamp spin column
- Add 1.0 μL DNA eluate to a NanoDrop to quantify the DNA concentration
- Store the remainder of the DNA eluate @ 4 ºC until required for qPCR (1 day) or store @ -20 ºC for long term storage until required

### *C.6 Gel extraction and purification of DNA with QIAquick*

**Material & Equipment:**

Agarose gel with identified bands prepared "DATE"

QIAcube (Qiagen)

2.0 & 1.5 ml microcentrifuge tubes (Eppendorf)

QIAquick Gel extraction Kit (Cat#)

**Method: Gel extraction and purification for DNA sequencing**

- Weigh pre-labeled 2.0 mL microcentrifuge tubes
- Excise DNA fragment from gel with a clean, sharp scalpel and place in tubes
- Weigh the gel slice in its tube to ensure that it is <400mg.
- Prepare all other reagents as pre QIAcube Cleanup QIAquick Gel Extraction Standard protocol, version 3 (table below).
- Run protocol and collect eluate
- Add 1.0 µL DNA eluate to the NanoDrop to quantify the DNA concentration and store @ -20$^{O}$C until required

**Table: QIAcube QIAquick Gel Extraction protocol**

| Shaker | Rotor Adapter | | Reagent Bottle Rack | |
|---|---|---|---|---|
| Up to 400 mg agarose gel; | 1 | QIAquick spin column | 1 | - |
| ≤2% agarose | 2 | | 2 | Buffer QG |
| 2 ml safe-lock tube | 3 | 1.5 ml collection tube | 3 | Isopropanol |
| | | | 4 | |
| | | | 5 | Buffer PE |
| | | | 6 | Buffer EB |
| Microcentrifuge Tube Slots - | | | | |
| 12 samples = Shaker and centrifuge positions - | | | | |

### *C.7 Cloning of DNA fragments into plasmids with Qiagen PCR Cloning Plus*

**Material & Equipment:**

Micropipettes and tips (Rainin)

Centrifuge (Eppendorf)

Thermomixer (Eppendorf)

Vortex Genie 2 and adaptors (Scientific Industries)

Micropipettes and tips (Rainin)

Incubator (VWR model# 2005)

Qiagen PCR Cloning plus kit

1.5 and 2.0 mL microcentrifuge tubes

LB-ampicillin liquid broth

LB-ampicillin agar plates

IsoRack (white 0$^{o}$C)

Gel extracted and purified DNA (OR regular amplicons) molar ratio 5-10 times (97.5 – 195 ng)

**Method: DNA cloning with Qiagen PCR Cloning plus kit**

- Remove gel purified DNA from -20°C storage
- Thaw 2x Ligation Master Mix, pDrive Cloning Vector DNA, and distilled water (provided) @ <8°C in refrigerator
- Return 2x Ligation Master Mix to -20°C storage immediately after use
- Add 1 μL pDrive Cloning Vector DNA (50 ng/μL) to 1.5 mL microcentrifuge tube
- Add 4 μL PCR product
- Add 5 μL Ligation Master Mix and briefly mix ADDED LAST
- Incubate ligation-reaction mixture for 30 min – 2 hr.@ <8°C in refrigerator
- Store ligation-reaction mixture .@ -20°C
- OR proceed to transformation protocol
- Place SOC medium and LB-ampicillin agar plates in incubator @ 37°C to equilibrate
- Thaw Qiagen EZ competent cells in IsoRack @ <8°C in refrigerator
- Quickly add 2 μL ligation-reaction mixture to EZ cells on ice and mix gently by flicking tubes
- Incubate for 5 min in IsoRack @ <8°C in refrigerator - the last 1 min @ -20°C

- Mix gently by flicking tubes
- Heat for 30 s @ 42$^{o}$C in Thermomixer WITHOUT SHAKING
- Incubate for 2 min @ -20$^{o}$C
- Add 250 µL wormed SOC medium and directly spread plate 100 µL of mixture on wormed LB-ampicillin agar plates
- Incubate plates at room temperature to absorb mixture into agar
- Incubate plates in incubator for 15-18 hr. @ 37$^{o}$C
- Incubate plates in refrigerator for 2 hr.@ <8$^{o}$C
- Select white colonies from plates and transfer to LB-ampicillin liquid broth
- Incubate in incubator for 15-18 hr.@ 37$^{o}$C with vigorous shacking
- Store cells in LB-ampicillin liquid broth @ -80$^{o}$C until required for DNA extraction
- Store cells on LB-ampicillin agar plates @ <8$^{o}$C until required for growth in LB-ampicillin liquid broth
- Store ligation-reaction mixture @ -20$^{o}$C until required for *E. coli* transformation

### *C.8 Plasmid purification with QIAprep*

**Material & Equipment:**

QIAcube (Qiagen)

Centrifuge (Eppendorf)

NanoDrop spectrophotometer (Thermo Scientific)

Micropipettes and tips (Rainin)

QIAprep Miniprep Kit (Qiagen)

2.0 & 1.5 ml microcentrifuge tubes (Eppendorf)

Cloned and transformed cells

**Method: Plasmid purification with QIAcube**

Remove cloned and transformed cells from incubator @ 37$^{o}$C

Centrifuge for 3 min @ >8,000 rpm to harvest cells

Decant supernatant by tipping tubes

Transfer pelleted/harvested cells in tube to shaker of the QIAcube.

Prepare all other reagents as pre the QIAcube QIAamp DNA Stool pathogen MODIFIED protocol (30 µL eluate)

On completion of the DNA extraction, add 1.0 µL DNA eluate to the NanoDrop to quantify the DNA concentration

Store the remainder of the DNA eluate @ 4$^{O}$C until required for qPCR after store @ -20$^{O}$C for long-term storage

**Table: QIAcube QIAamp Stool pathogen protocol**

| **Shaker** | | **Rotor Adapter** | | **Reagent Bottle Rack** |
|---|---|---|---|---|
| Pelleted bacteria cells | 1 | QIAprep spin column | 1 | Buffer P1 |
| 2.0 ml microcentrifuge tubes | 2 | | 2 | Buffer P2 |
| | 3 | 1.5 ml collection tube (L3) | 3 | Buffer N3 |
| | | | 4 | Buffer PB |
| | | | 5 | Buffer PE |
| | | | 6 | Buffer EB |
| | | | | |
| 12 samples = Shaker and centrifuge positions | | | | |

# VITA

In 1994, Michael O'Neil Ryan earned his Bachelor of Science degree (B.Sc.) from University of the West Indies, Jamaica where he majored in Microbiology; and minored in Biochemistry and Botany. His academic career continued with the acquisition of an Executive Masters of Business Administration (M.B.A.) at the University of New Orleans in 1998. His professional development primarily entailed managing the National Water Commission's Southern Microbiology Laboratory based in Kingston, Jamaica from 1994 to 2004. He began the pursuit of a Doctor of Philosophy (Ph.D.) degree in Environmental Engineering at Drexel University, Philadelphia in 2007. His reach was focused on developing and applying microbial source tracking (MST) molecular biology and microbiology protocols to assist regulatory agencies and water utilities to make informed pollution mitigation decisions.

*Publications & Proceedings*

- Ryan, M., Duzinski, P., Gurian, P., Haas, C., Rose, J., "Acceptable Microbial Risk: Benefit-Cost Analysis of a Boil Water Order for Cryptosporidium." Journal American Water Works Association, (Accepted)
- Ryan, M., Russell J., Haas, C., "Evaluating the Microbial Source Tracking Attributes of Bacteroides Dorei" Journal Applied and Environmental Microbiology (Under Review)
- Ryan, M., Haas, C., "MST by Using qPCR to Appraise the Potential Exposure to Waterborne Pathogens." Water Environment Federation Disinfection 2011 conference, Cincinnati, Ohio, April 10-12, 2011