2011

LOCATING HOT SPOTS OF FECAL POLLUTION IN AN URBAN WATERSHED OF CENTRAL KENTUCKY USING *BACTEROIDES* 16S rRNA MARKERS

Tricia L. Coakley
University of Kentucky, tcoakley@engr.uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Tricia L. Coakley, Student

Dr. Alan E. Fryar, Major Professor

Dr. Alan E. Fryar, Director of Graduate Studies

LOCATING HOT SPOTS OF HUMAN FECAL POLLUTION IN AN URBAN WATERSHED OF CENTRAL KENTUCKY USING *BACTEROIDES* 16S rRNA MARKERS

THESIS

A Thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Earth and Environmental Sciences
at the University of Kentucky

By

Tricia L. Coakley

Lexington, Kentucky

Co-director: Dr. Alan E. Fryar, Associate Professor of Geology

Co-director: Dr. Gail M. Brion, Professor of Civil Engineering

Lexington, Kentucky

2011

ABSTRACT OF THESIS

LOCATING HOT SPOTS OF HUMAN FECAL POLLUTION IN AN URBAN WATERSHED OF CENTRAL KENTUCKY USING *BACTEROIDES* 16S rRNA MARKERS

The field of molecular fecal source tracking in the water environment has developed rapidly since the first PCR assays for general and host-specific *Bacteroides* 16s rRNA markers were published. Numerous host-specific molecular markers and PCR assays have been developed, adding greater specificity, sensitivity and quantitative methods to the array of options. The public demand for readying methods for transfer to the commercial lab, so that they may be used to generate data for public utilities, citizen action groups and regulatory agencies, has fueled the development of an entire new research community. These methods, however plentiful, have not found community agreement and there is no consensus concerning the appropriate implementation of molecular fecal source tracking in the field. Some issues plaguing the implementation include imperfect marker specificity, environmental variability, DNA extraction variability, PCR inhibition and high cost of molecular analysis. This thesis presents an approach for locating hot spots of human fecal pollution in an urban watershed by using published methodologies for the collection of molecular fecal source tracking data along with a tiered watershed screening tool for cost reduction and two data normalization techniques which ameliorate several known sources of error and strengthen the efficacy of watershed application.

KEYWORDS: fecal source tracking, *Bacteroides*, Wolf Run watershed, *E. coli*, AC/TC

Tricia L. Coakley

16 December 2011

LOCATING HOT SPOTS OF HUMAN FECAL POLLUTION IN AN URBAN WATERSHED OF CENTRAL KENTUCKY USING *BACTEROIDES* 16S rRNA MARKERS

By

Tricia L. Coakley

Dr. Alan E. Fryar
Co-director of Thesis

Dr. Gail M. Brion
Co-director of Thesis

Dr. Alan E. Fryar
Director of Graduate Studies

16 December 2011

ACKNOWLEDGEMENTS

I would like to thank my co-advisors, Dr. Alan Fryar, for his skilled mentoring and patience with this project and Dr. Gail Brion, for presenting me with the opportunity to study fecal source tracking, the example for thinking about mathematical tools for interpreting microbial data, and for allowing me the time to complete the project. I would also like to thank my third committee member, Dr. Jim Dinger, for his thoughtful comments during the thesis review period and for building my confidence and pushing me towards the finish line throughout the process.

The Environmental Research and Training Laboratory at the University of Kentucky provided the bench space and equipment, without which this project would not have been possible. I would like to specifically thank all of the students, faculty, and staff associated with the ERTL facility for troubleshooting and brainstorming assistance. John May's critical eye for data quality and clarity proved to be invaluable to this project.

I am eternally grateful to Ken Cooke and the Friends of Wolf Run watershed and to Jean Watts, Tracey Knowles and the students at the Bluegrass Community and Technical College for water sampling and logistics support. Sample sites were selected based upon the extensive knowledge of the Wolf Run watershed maintained by FOWR (a non-profit community action group). Water samples were collected by the faculty and students of BCTC so that I was able to remain blind to the geographic locations and prevent bias.

Alice Layton and Dan Williams at the University of Tennessee provided support during my early stages of learning to perform qPCR methods by allowing me to shadow them on their own bench. This kindness has undoubtedly saved me many

long hours of labor and countless supply costs for which I am very grateful. They also provided the cloned plasmid for qPCR assay calibration standards.

The conclusions derived from my research were made possible with calculations I learned from Dr. Arne Bathke and his students at the University of Kentucky Applied Statistics Lab.

Assistance with the maps for this project was provided by Matt Crawford and Jim Currens at the Kentucky Geological Survey.

Funding for the research presented in this thesis was provided by the Kentucky Water Resources Research Institute and the United States Geological Survey.

Finally, I would like to thank my family and friends for their unending support throughout the thesis process. Without their loving support in the form of food, timely distraction, and a sympathetic ear, I would not have been able to complete the task.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Literature Review and Justification of Approach

1.1. Why fecal source by PCR methods?

The EPA has regulated storm water discharge pollution with respect to *E. coli* concentrations for many years but the data are insufficient for use by researchers attempting to ameliorate the source of the fecal pollution. *E. coli* bacterial signals, even when quantitatively assayed, lack spatial and temporal stability in the watershed and are therefore not useful in the process of locating the source of the pollution (Anderson et al., 2005). *E. coli* bacteria are facultative anaerobes, meaning that they can persist in the environment regardless of oxygen levels that are greater than that of the enteric environment. There is even evidence indicating that in warm climates with sufficient nutrient loads, *E. coli* may actually replicate in the stream (Carrillo et al., 1985). Another limitation to the use of *E. coli* data is that the bacterial species is prevalent in the feces of all animals and viable enumerations cannot successfully differentiate between humans and other animals (Gordon & Cowling, 2003).

Many methods have been proposed to track the source of fecal pollution in water. They can be summarized in four categories: 1) culture-based, library-dependent; 2) culture-based, library-independent; 3) culture-independent, library-dependent; and 4) culture-independent, library-independent methods. Culture-based, library-dependent methods include antibiotic resistance and *E. coli* ribotyping methods. Both of these methods have shown promise in trials with fecal positive controls but are very labor intensive and expensive (Griffith et al., 2003; Scott et al., 2002; Stoeckel et al., 2004). The libraries developed for the implementation of these methods have been criticized for lack of temporal and geographical stability. Culture-based, library-independent methods include bacteriophage assays and bacterial ratios like fecal streptococci/fecal coliform and atypical colonies/total coliform. Bacteriophage identification was found to be capable of identifying sewage but failed when tested against single individual subjects (Noble et al., 2003).

Bacterial ratios have been criticized for their temporal and spatial variability in the watershed, which results from varying rates of persistence between the groups (Field & Samadpour, 2007). This variable persistence of coliforms and atypical colonies has subsequently been shown to have great utility for determining the age of the fecal pollution. The atypical colonies/total coliform (AC/TC) method (Brion et al., 2002) is utilized in this thesis as a measure of fecal age in watershed samples. An example of a culture-independent, library-dependent method is terminal restriction fragment length polymorphism (T-RFLP). This method is similar to *E. coli* ribotyping except that it does not require culturing of the bacteria prior to molecular analysis. This method has not advanced past the proof of concept stage due to its inability to identify between hosts even in fecal positive controls (Griffith et al., 2003).

The field of fecal source tracking has moved almost completely to the use of culture-independent, library-independent methods. This category can be divided into chemical and microbiological subcategories. Chemical methods include the quantitative measurement of caffeine, fecal sterols, and optical brighteners. Problems with detection limits, persistence, and transport differences from fecal pathogens have caused these chemical indicators of human fecal pollution to be limited in application (Chan et al., 1998; Glassmeyer et al., 2005; Peeler et al., 2006). Microbiological methods that do not require culturing or the development of a library include viral pathogen and bacteriophage monitoring by polymerase chain reaction (PCR). Although the human viral pathogens have proven to be very host-specific, the large sample volume requirement has limited their utility. F+ coliphage analysis for the detection of human fecal pollution has been found to successfully identify sewage but results in false negative determinations when challenged with individual fecal samples (Cole et al., 2003; Griffith et al., 2003; Noble et al., 2003). Host-specific DNA markers for enteric bacteria have become the most popular methods for fecal source tracking since researchers at the University of Oregon developed a primer set for the PCR assay of the *Bacteroides* genus of bacteria from the feces of humans and cattle (Bernhard & Field, 2000).

In a study of 12 different microbial source tracking methods, blind duplicate samples of mixed fecal sources were analyzed by 22 different researchers and no method identified the sources of fecal pollution perfectly (Griffith et al., 2003).

1.2. Why *Bacteroides*?

Bacteroides bacteria are ideal for fecal source tracking for several reasons. They are obligately anaerobic, gram-negative rods. Unlike *E. coli*, they are not able to persist for long periods in oxygenated environments. They are the most numerous of the intestinal bacteria with as many as 1,011 cells per gram of dry feces (Finegold et al., 1983). This bacterial genus was not used prior to the development of molecular markers due to the difficulty in culturing it. A study of the microbial composition of the human intestinal flora found that 80% of the 395 bacterial phylotypes identified represented species that have never been cultivated in the lab (Eckburg et al., 2005). *Bacteroides* are present in feces in much greater concentrations than any of the culturable bacteria (Eckburg et al., 2005), so they allow for greater sensitivity of molecular markers in the environment. *Bacteroides* are known to be involved in the function of nutrient absorption in the host and it is believed that this allows for the expression of different genes in different host species based on the variation in food sources (Dick et al., 2005; Hooper et al., 2001).

1.3. Development of *Bacteroides* markers

Molecular markers of *Bacteroides* bacteria have been developed for several host species likely to contribute to the fecal pollution of source water and recreational water, including those for cattle, geese, elk, swine and humans (Bernhard & Field, 2000; Carson et al., 2005; Layton et al., 2006; Lu et al., 2009; Okabe et al., 2007; Shanks et al., 2009).

The 16s rRNA PCR primers (Bac32f and Bac708r) published by Bernhard and Field (2000) identify a non-host-specific *Bacteroides* marker of general fecal pollution.

They also developed corresponding primers for the identification of human (HF183 and HF134) and bovine (CF128 and CF193) fecal pollution (Bernhard & Field, 2000). These primers were designed for use with conventional PCR, which relies upon the detection of a band of DNA by gel electrophoresis. Therefore, they yield a simple presence or absence signal and do not provide quantitative information about the fecal pollution.

Seurinck et al. (2005) presented a real-time PCR assay for the HF183 marker using SYBR green intercalating dye. This method allows for the quantification of the marker against a calibration curve generated from a serial dilution set of a plasmid clone of the marker. This was a significant advance in the field of fecal source tracking, but the intercalating dye fluorescence is emitted when any double-stranded DNA product is generated during PCR, creating the potential for non-specific binding and false positive determinations. Dick and Field (2004) presented a primer/probe design for the quantification of a *Bacteroides* marker that provides proof of concept for the use of TaqMan-based probe assays with fecal source tracking. This assay is not affected by non-specific binding associated with intercalating dye methods because the probe, which binds to the target between the forward and reverse primers, provides an added specificity requirement. The assay tested well but was not used further in the field because it was not designed for a host-specific marker.

Researchers at the University of Tennessee–Knoxville developed three *Bacteroides* fecal markers which utilize the TaqMan probe technology with quantitative real-time PCR (qPCR) (Layton et al., 2006). Of these markers, one is a non-host-specific general marker of fecal pollution (Allbac), one is a bovine-associated fecal marker (Bobac), and one is a human-associated fecal marker (Hubac). Allbac and Hubac markers are used in this thesis to quantitatively analyze the fecal contributions in an urban watershed of central Kentucky. The Hubac marker was found in a study of fecal positive and negative controls to be "selective rather than specific" (Layton et al., 2006). The marker yielded a 100% true positive rate but also a 32% false

positive rate, indicating that the potential to isolate the marker from animals other than humans is quite high.

1.4. Field application case studies

Most of the host-specific *Bacteroides* fecal genetic markers developed have been tested in fecal samples and sewage, but some have also been studied at the watershed scale. In a study of Tillamook Bay in Oregon, Bernhard et al. (2003) found that their previously published markers for human and cattle feces were detected in both river and estuary samples. Twenty-two samples were taken during high tide and at least one of the host-specific markers was found in 17 of the samples. All but one of the sites classified as urban or near a sewage treatment plant were positive for the human-specific marker, whereas only three of the 14 rural sites were positive for the human marker. The study found that both human and ruminant markers were absent at sites that were not likely to be impacted by human or cattle inputs. The success in detection and correct classification of the *Bacteroides* genetic markers was very encouraging for the field of fecal source tracking, but quantification of the markers was still desired.

Shanks et al. (2006) conducted a more in-depth study of the Tillamook Bay watershed. This study was also limited by the use of conventional PCR markers that yield a simple presence/absence signal, but the authors were able to determine a frequency of occurrence result by implementing a large-scale sampling plan. A total of 2,912 samples were collected from 30 sites bimonthly and analyzed for *E. coli* concentration, physical parameters, and genetic markers for human and ruminant hosts. This study successfully identified a point source of human fecal pollution from a wastewater treatment plant. The authors found that the probability of detecting ruminant markers was double that of detecting human markers in the watershed, but the probability for a human marker increased significantly when sampled near a known point source of human fecal inputs. Another important finding derived from this work was the relationship between rainfall and the distribution of the most

predominant fecal source marker in the watershed (ruminant). This was indicated by the increase in the probability of detection following rain events.

The Allbac, Bobac and Hubac fecal markers were developed with the objective of absolute quantification (Layton et al., 2006). They utilize a fluorescently labeled TaqMan probe in a real-time qPCR assay, which compares threshold cycle (Ct) values of samples of unknown concentration against a calibration curve generated from Ct values of known concentrations of the marker plasmid clone or diluted feces. These markers were tested in three watersheds of differing land use patterns. The authors found that the Allbac marker concentrations were linearly correlated ($r^2 = 0.85$) with the *E. coli* enumerations in the study. The Hubac and Bobac markers were detected in each of the mixed land-use samples, indicating contributions from both humans and cattle. Allbac, Bobac and Hubac concentrations were elevated during high flow as compared to low flow samples. The urban and resort land use samples yielded significantly greater Hubac marker concentrations than Bobac, indicating that the fecal contamination in these watersheds is predominantly from humans. The authors also presented the results as a percentage of total fecal contamination by dividing the Hubac or Bobac concentration by the Allbac concentration. They found that in all samples except one resort location, the sum of the contributing portions did not equal 100% of the Allbac signal. They concluded from this that there may be other animal contributors in the watershed which provide the Allbac signal, but do not provide the Hubac or Bobac signal. These results indicate that the Hubac marker, while not perfectly specific, is useful for defining abundance of human fecal contributions in watershed-scale studies.

1.5. Tiered and multiparameter approaches

It is becoming widely accepted in the field of fecal source tracking that the measurement of several indicators of water quality, combined in a toolbox approach, leads to a greater ability to define the dynamics of fecal pollution in the watershed than is found by focusing on one fecal indicator alone. Several studies

have found that the implementation of multiple indicators used together in a model, or used sequentially in a tiered approach, identifies the hot spots of specific fecal sources with a greater degree of accuracy than was found by any one of the indicators viewed in isolation (Boehm et al., 2003; Black et al., 2007; Blanch et al., 2006; Noble et al., 2006; Vogel et al., 2007).

Black et al. (2007) found that the presence of human enteric virus could be predicted at the intake of a water treatment plant with a model using multiple indicators including those for fecal load (fecal coliform bacteria), source (epicoprostanol), and age (AC/TC). In a study of known fecal sources from wastewaters and slurries, Blanch et al. (2006) found that none of the 38 variables measured were able to correctly identify the source, but when at least 2 variables were combined in models to include those for both fecal load and source indicators, 100% correct classification was achieved. Boehm et al. (2003) found that a tiered approach using fecal indicator bacterial assays across a wide spatial and temporal range in tiers one and two allowed for the prediction of potential hot spots, which were then resampled and analyzed for molecular markers of fecal source and for enteric virus in the third tier of the investigation. A study of a watershed draining to Santa Monica Bay in 2006 found that the fecal contamination levels appeared consistently high across the entire watershed when measuring only one parameter, but the authors were able to define one tributary as the primary contributor when they used a tiered approach that incorporated several variables (Noble et al., 2006).

This thesis proposes a tiered approach for identifying the locations of human fecal hot spots while reducing analytical costs by screening the watershed with traditional culture-based microbial methods prior to the analysis of more costly molecular fecal markers.

1.6. Data normalization

Molecular fecal source tracking, although considered more effective than previously available methods, is confounded in watershed application due to several known sources of error including temporal and spatial variability of molecular markers, DNA extraction efficiency variability, PCR inhibition and imperfect marker source specificity. The call to normalize quantitative PCR data to overcome these sources of error is found in the literature but there is no consensus on how to do this (Santo Domingo et al., 2007; Soule et al., 2006; Stoeckel & Harwood, 2007).

A study of two different spike and recovery controls that were added to water samples prior to extraction found that the variability between duplicate samples was decreased when the fecal source marker quantitative result was normalized by the recovery of the added controls (Stoeckel et al., 2009). The authors further determined that between 72% and 89% of the variability in fecal marker results could be attributed to extraction and recovery efficiencies. The authors of a 2007 review paper called for the development of "novel computational methods" for use with molecular fecal source tracking data (Santo Domingo et al., 2007). Soule et al. (2006) suggested using a relative ratio of real-time PCR results to provide greater utility of the fecal source markers by limiting the effects of differing quantities of PCR inhibitors in various samples. The authors were careful to point out, however, that the relative ratio would not be useful to apportion fecal sources within a sample due to a lack of information about the relative abundances of the different markers in feces from any one host.

This thesis proposes two normalization techniques that both employ a relative ratio of a human associated fecal marker and a general fecal marker. One method uses the quantitative results from qPCR and the other uses the raw Ct values.

2. Problem Statement and Hypothesis

Examination of fecal source marker specificity, variability, and watershed-scale utility is needed before proceeding with the widespread application of molecular fecal source tracking data collection efforts in central Kentucky watersheds. I hypothesize that Allbac and Hubac fecal source DNA markers may be used for identifying hot spots of human fecal pollution in an urban watershed of central Kentucky.

i. Objective 1: Define the specificity of the human marker. Task: Analyze fecal samples from animals and humans in central Kentucky for the presence of general and human-specific *Bacteroides* markers.
ii. Objective 2: Determine the fate and persistence of a DNA marker in the environment. Task: Measure the concentration of a *Bacteroides* DNA marker at multiple time points in a sewage positive control incubated at various environmental conditions.
iii. Objective 3: Determine the utility of fecal DNA marker data at the watershed scale.
    a. Task 1: Investigate the potential for data normalization of the human-associated fecal marker.
    b. Task 2: Collect samples from an urban watershed in central Kentucky known to be contaminated with human fecal pollution.
    c. Task 3: Analyze all samples for indicators of fecal age and load.
    d. Task 4: Measure human-associated and general *Bacteroides* marker concentrations in all samples.
    e. Task 5: Define hot spots of human fecal pollution
    f. Task 6: Determine if the use of a fecal age and load screening tool could reliably eliminate non-hot spots.

3. Approach

The interpretation of molecular fecal source marker data has been problematic due to several known sources of error, including PCR inhibition, extraction variability, and imperfect marker specificity, as well as spatial and temporal variability at the watershed scale. Several host-specific genetic markers have been developed and applied to determine the sources of fecal pollution in water. Some of these markers have been subsequently shown to have cross reactivity with fecal samples from hosts other than those that the markers were designed to identify. This thesis addresses the question of marker specificity, fate and persistence and watershed utility for the Allbac (general) and Hubac (human-associated) DNA markers developed by Alice Layton at the University of Tennessee (Layton et al., 2006).

3.1. Marker selection and specificity

The utility of any host-specific DNA marker in water-quality fecal source tracking research and monitoring is limited by the specificity of the marker to only one host. If the marker is found in the feces of animals other than the target species, the use of the marker may lead to false positive determinations. The specificity of a fecal marker must be determined before generating environmental water-quality data. If a fecal source marker is found to lack specificity, it is commonly thought to render the marker useless as a fecal source tracking tool because it is impossible to determine if the marker, even if found in high concentrations, is actually from the related host. The first study in this thesis examines the specificity of a human-associated *Bacteroides* 16S rRNA genetic marker (Hubac). By ensuring first that the Hubac marker is found only in human hosts, one can presume that water samples that test positive for the Hubac marker were indeed contaminated with human feces. Fecal samples from various animals likely to contribute to the fecal bacterial load of streams in Kentucky were collected and analyzed to answer this question: is the Hubac marker specific to human fecal contamination?

The Allbac marker is a non-host-specific fecal marker, which is expected to be found in significant quantities in fecal samples from all animals. Allbac assays were used in this study to determine the integrity of each fecal sample. The Hubac marker, in contrast to Allbac, should only be found in feces from human hosts.

The Allbac and Hubac markers were chosen for this study for several reasons. The region for which the markers were developed (around Knoxville, Tennessee) is ~300 km south of the study area in this thesis (around Lexington, Kentucky). The geology, climate, and land uses of the two areas are broadly similar, which should eliminate any major geographical differences in marker specificity. Also, these markers were developed to be analyzed quantitatively using a TaqMan fluorescently labeled probe, while other *Bacteroides* genetic fecal source markers were useful only for qualitative analysis by conventional PCR or quantitative analysis with an intercalating dye qPCR assay at the inception of this study.

3.2. Marker fate and persistence in the environment

The widespread application of bacterial DNA markers as environmental indicators of fecal pollution requires that the markers are source-specific as well as spatially and temporally conserved. DNA markers found in the genome of the *Bacteroides* genus of bacteria are being investigated for this use because these bacteria are strict anaerobes. Survival and regrowth are not likely in the oxygen-rich stream environment, but it is not known if the DNA markers from the bacterial cells persist in the environment after cell death and lysis. A bench-scale study of the fate of a general *Bacteroides* marker over time and in various environmental conditions was performed to answer this question.

3.3. Data normalization

The identification and absolute quantification of a host-specific fecal marker in the environment do not provide enough information to determine the degree of the

contamination, because of possible confounding effects of environmental variables such as temperature and dilution as well as analytical errors including extraction variability and PCR inhibition. The host-specific marker concentration must be put into perspective first before conclusions about the watershed can be made. This thesis examines an approach for the normalization of the source-specific marker with a general fecal marker from the same sample extract.

3.4. Watershed application and cost reduction

Additional barriers to the broad application of fecal source tracking by quantitative PCR include cost and level of technology. The technical expertise required is not readily available to community action groups and local governments with limited budgets seeking to pinpoint human fecal hot spots within their watersheds. Previous studies have demonstrated the utility of a multi-indicator approach for human health risk assessment that relies upon signals of fecal load and fecal age to model the presence of culturable human enteric virus (Black et al., 2007). The less expensive analyses for AC/TC ratio and *E. coli* enumerations provide additional information on the average fecal age and load, respectively. This information may be used as a screening tool to eliminate sites with less potential to be hot spots of human fecal contamination. If these indicators are used prior to molecular fecal source tracking methods, screening and removal of low-risk sites may allow sampling across great temporal and spatial variability while keeping the cost of the DNA marker analysis to a minimum.

4. Methods and Materials

4.1. Fecal sample collection

To investigate the specificity and lack of cross-reactivity of the human-associated marker, fecal samples were collected from various hosts. A total of 22 fecal samples were collected representing 11 species. Six human subjects were utilized, including one breast-fed infant. The other five were omnivorous adults between the ages of 25 and 55. All human fecal samples were collected in clean zipper-lock-type bags and stored at 4°C until delivery to the lab. Three individual cattle fecal controls were sampled from a beef farm in Woodford County. Each was identified as a cattle manure isolate separated from any other manure. The freshest manure was selected from those available within two pastures. Two fecal samples from deer were collected from a wooded area of Scott County. One sample of goose feces was collected from the bank of an urban pond in Lexington that retains a large flock of geese year-round. One river otter sample was collected from a creek bank in Scott County. The otter was identified visually at the time of sampling. One horse and one llama fecal sample were collected from a pasture in Jessamine County. The pasture contained one horse and six llamas. The fecal samples from this field were source-identified by visual differentiation and diet. One rabbit fecal sample was collected from a suburban yard in Scott County where rabbits were observed. One composited, chicken fecal sample was collected from the University of Kentucky chicken farm. Three domestic feline samples were collected, two from indoor cats that live with a primary caregiver and one from the Lexington Humane Society. One composited canine fecal sample was collected from the kennel waste at the Lexington Humane Society and one canine sample was collected from a domestic dog residing with human companions. All samples were collected from the freshest source reasonably obtainable in clean plastic bags. They were stored in the laboratory at -80°C until DNA extraction.

4.2. Fecal marker fate and persistence experiment

Primary effluent from the Blue Sky wastewater treatment plant in Fayette County, Kentucky, was collected and diluted by half with sterile deionized water. The diluted sewage effluent was then strained through three layers of sterile cotton gauze to remove large debris and 10 mL was pipetted into each of 139 sterile, 25-mL cell culture flasks. Forty-five of the flasks were placed in a dark cabinet with the caps loose to simulate ambient temperature (20-25°C), aerobic environment and the absence of light. Forty-five more flasks were placed on a window sill with the caps loose to simulate ambient temperature, aerobic environment and the potential for photosynthesis and UV decay of genetic material. Another 45 flasks were evacuated with nitrogen, capped securely and placed in a dark cabinet to simulate ambient temperature, anaerobic environment and the absence of light. The anaerobic state of these flasks is assumed as the dissolved oxygen level in each flask was not measured. Two flasks stored at 4°C to limit microbial predation and degradation of genetic material by UV light were used as positive controls; one was evacuated with nitrogen, capped securely and stored in the refrigerator; the other was stored in the refrigerator with a loose cap. Another portion of the diluted primary effluent was transferred to a 1-mL microcentrifuge tube and placed in the freezer so that the starting concentration of the marker could be determined. Ten milliliters of deionized water was added to one sterile cell culture flask and stored in the refrigerator with the cap loose as a negative control. One flask from each condition was removed on the second, third, and fifth day and agitated, after which 1 mL of the contents was transferred to a microcentrifuge tube, which was placed in the -20°C freezer for subsequent PCR analysis. The positive controls and the negative control stored at 4°C were removed upon completion of the experiment. Samples were removed from the freezer, thawed quickly at room temperature, vortexed to mix and used directly in the PCR reaction without extraction.

4.3. Environmental water sample collection

Environmental water samples were collected for examination by microbiological and molecular methods to determine the utility of the fecal source tracking and data normalization methods at the watershed scale in central Kentucky. A total of 20 samples were analyzed from the Wolf Run watershed in Lexington, Kentucky (Figure 4.1, Table 4.1). The Wolf Run watershed is known to receive inputs from human sewage in storm water overflows because of cross-connections between the sanitary and storm sewers. All samples were collected within 2 hours of each other and following a large storm event on April 3, 2008. All samples were collected in sterile 1-L polypropylene bottles and transported to the lab on ice. Aliquots of 250 mL were filtered through 0.45-µm cellulose membrane filters, which were stored at -80°C until DNA extraction. Culturable assays for *E. coli* and AC/TC were performed with aliquots of the remaining sample and were completed within 8 hours of sampling.

ID	Latitude (°N)	Longitude (°W)
D01	38.04890	-84.55360
D02	38.04225	-84.52547
D03	38.05435	-84.53133
D04	38.05737	-84.54246
D05	38.05949	-84.54815
D06	38.03695	-84.52271
D07	38.02240	-84.51240
D08	38.03220	-84.52430
D09	38.03255	-84.52652
D10	38.04480	-84.53600
D11	38.04936	-84.54265
D12	38.05153	-84.54563
D13	38.05480	-84.54970
D14	38.02300	-84.52860
D15	38.03010	-84.53730
D16	38.03318	-84.54210
D17	38.04290	84.54917
D18	38.05350	-84.55090
D19	38.06685	-84.55435
D20	38.02940	-84.53740

Table 4.1. GPS coordinates of sample locations.

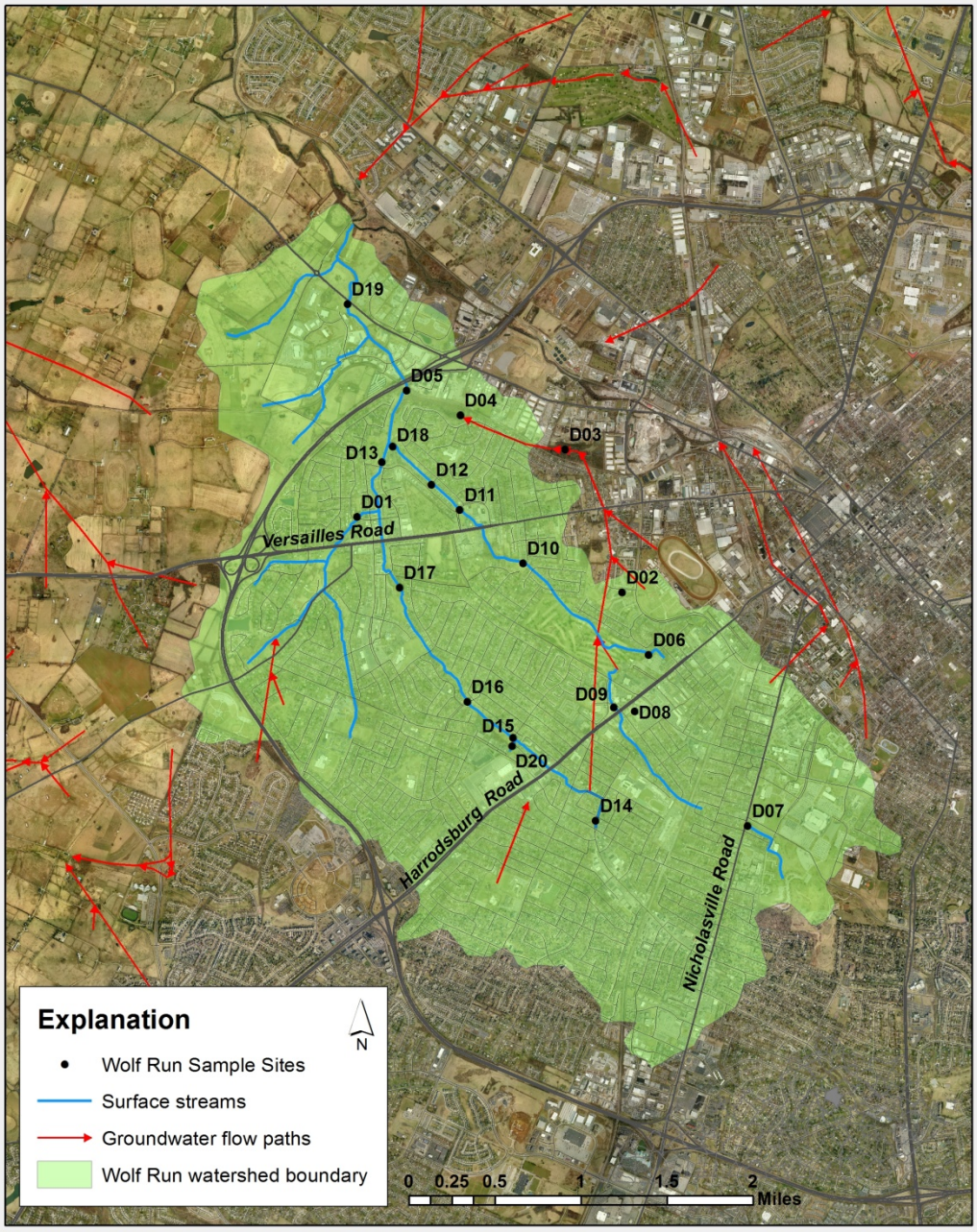

Figure 4.1. Map of sample locations

Map provided by Matt Crawford at the Kentucky Geological Survey.

4.4. Microbiological examination of environmental water samples

E. coli concentrations were enumerated by Idexx Colilert™ media in Quantitrays 2000™. Samples were diluted 1:10 with sterile, phosphate-buffered saline solution prior to the addition of media. They were then shaken, poured into Quantitrays, sealed with a Quantitray sealer and incubated at 35°C for 24 hours. Wells which fluoresced under UV light were counted and used to calculate *E. coli* concentrations as most probable number per 100-mL sample (MPN/100 mL) utilizing the table provided by Idexx for use with Quantitray 2000™.

AC/TC ratios were determined from the total coliform (sheen) and atypical (non-sheen) colony counts on m-Endo media at multiple dilutions as described by Brion & Mao (2000). Water samples were diluted 1:100 with sterile, phosphate-buffered saline and shaken prior to filtration. 1-mL and 10-mL volumes of the diluted sample (representing 0.01- and 0.1-mL volumes of the original water sample) and a 1-mL portion of the original sample were filtered through a 0.45-µm cellulose membrane filter. Duplicate filters were processed for each filter volume from each sample. The filters were placed on 50-mm petri dishes containing m-Endo broth soaked nutrient pads and were then incubated at 35°C for 24 hours. Colonies exhibiting a metallic sheen were counted as total coliforms (TC) and those with a red or pink color and no metallic sheen were counted as atypical colonies (AC). Each bacterial group (AC or TC) was counted on the filter volume plates exhibiting between 10 and 100 colonies of the appropriate morphology. The counts from duplicate plates were averaged for a final count and used to calculate each bacterial group (AC or TC) in units of CFU/100 mL. The calculated CFU/100mL for AC was then divided by that found for TC from the same sample to determine the ratio.

4.5. DNA extraction

DNA extraction for all fecal samples was performed with MoBio™ fecal DNA extraction kits. Fecal samples were removed from the freezer and thawed at room temperature prior to extraction. 0.25 g of each fecal sample was added to the supplied bead-beating tube and extracted following the manufacturer's directions, which included a 10-minute bead beating on a vortexer, followed by several filtration and centrifugation steps to yield a final extract volume of 50 µL. All extracts were stored at -20°C until analysis. A blank extraction was also completed to serve as a negative control.

Filters stored in the freezer from environmental water samples were extracted immediately upon removal from the freezer in groups of twelve. Extractions were completed using UltraClean Water DNA isolation kits (MoBio™) following the manufacturer's directions. The kit utilized a 10-minute bead beating with garnet shards on a vortexer followed by a series of filtration and centrifugation steps, yielding a 3-mL final extract volume. Extracts were stored at -20°C until DNA analysis by qPCR.

4.6. DNA analysis

Bacteroides genetic markers were analyzed by qPCR using the Allbac and Hubac primers and probes developed by Alice Layton at the University of Tennessee Center for Environmental Biotechnology (Layton et al., 2006). Real-time PCR was performed using a BioRad iCycler IQ™. Each 25 µL PCR reaction consisted of 12.5 µL IQ supermix (BioRad™), 10 pmol forward primer (Allbac or Hubac), 10 pmol of the corresponding reverse primer, 5 pmol of the corresponding FAM fluorescently-labeled molecular probe, and either 1 µL of the fecal extract or 2 µL of the filtered water extract. PCR protocols consisted of 50°C for 2 minutes and a 10-minute

activation at 95°C, followed by 50 cycles of a 95°C denaturation for 30 seconds and a 60°C annealing for 45 seconds.

Absolute quantification of the marker concentrations in all environmental water samples and marker fate experimental samples was achieved with the analysis of calibration standards consisting of 0 to 1×10^7 DNA copies/µL plasmid clones. Cloned plasmid DNA from human feces (TN hu) was obtained from Alice Layton at the University of Tennessee. The plasmid was then subcloned for use in this study using One Shot® chemically competent *E. coli* cells from Invitrogen and quantified by absorbance at 260 nm with a nanodrop® spectrophotometer. All PCR reactions, including calibration standards, were run in triplicate. Five duplicate environmental water samples and two filtration blanks were carried through the entire method to ensure precision and absence of contamination at each step. Extracted diluted wastewater treatment plant effluent was used as a positive control for general and human signals (Allbac and Hubac, respectively) for the environmental water sample study.

Fecal samples were analyzed for the human and general fecal marker relative concentrations as threshold cycle (Ct) values rather than by absolute quantification with a calibration curve. The threshold cycle is the PCR cycle at which the signal from a given reaction well crosses the defined threshold of detection. The initial intention of the fecal analysis was simply to determine the marker specificity by presence or absence of the Hubac marker in each fecal sample, so absolute quantification would not be required.

4.7. Data analysis

Threshold cycle (Ct) values were determined for each PCR reaction as the PCR cycle at which fluorescence from the fluorescently labeled probe is visible above baseline. This value is calculated for each PCR run by the BioRad iCycler IQ instrument

software. The Ct value for each reaction is inversely proportional to the starting concentration of DNA target. For example, a starting concentration of 1×10^7 copies/µL may yield a Ct value of 18 whereas a starting concentration of 1×10^1 copies/µL may yield a Ct value of 37. For absolute quantitation by qPCR, Hubac and Allbac marker concentrations were reported in units of DNA copies/µL after comparison to a calibration curve generated by plotting Ct values against the known starting concentration of the DNA target derived from a serial dilution from the plasmid clone. These concentrations were then converted to copies/mL of original sample for the environmental water samples.

Hubac results were normalized by two methods. Concentration values were normalized by dividing the Hubac concentration by the Allbac concentration and then multiplying by 100. This normalized Hubac value is further referred to as relative Hubac copy abundance:

Relative Hubac Copy Abundance = [(Hubac copies/mL / Allbac copies/mL) × 100] **(1)**

Threshold cycle (Ct) values were normalized by dividing the exponential function of the Allbac Ct by the exponential function of the Hubac Ct and then multiplying by 100. This normalized Hubac value is further referred to as relative Hubac Ct:

Relative Hubac Ct = [($e^{\text{Allbac Ct}}/e^{\text{Hubac Ct}}$) × 100] **(2)**

4.8. Statistics

Statistically significant differences between fecal sample results were determined by grouping the host animals into host categories of human, cow, domestic pets and other animals including wildlife. Sigma Plot® software was used to perform a one-way ANOVA test of the normalized and non-normalized real-time PCR-determined Ct values.

Significant differences in the Allbac marker fate and persistence data were also determined by a one-way ANOVA test using Sigma Plot software. Allbac concentration values as copies/µL from qPCR were compared between incubation conditions as well as between days of incubation.

Outliers among the water sample data were determined by calculating the median, 25^{th}, and 75^{th} percentiles with respect to all samples. An outlier is defined as any value that is greater than the sum of the 75^{th} percentile value (Q3) and 1.5 times the interquartile range (IQR). An extreme outlier is defined as any value that is greater than the sum of the 75^{th} percentile value and 3 times the IQR (Hoaglin et al., 1983).

5. Results and Discussion

5.1. Marker specificity

The Allbac marker analysis was performed first to determine the integrity of the fecal samples as the Allbac marker is expected to be found in the feces of every animal. All fecal samples were positive for the Allbac marker with the exception of the otter and chicken feces (Table 5.1). Both of these failed samples are presumed to have been compromised by age and desiccation. The chicken litter was dry when obtained from the barn and there was no access to fresher or wetter litter. The otter fecal sample was collected from the stream bank but had been exposed to the sun for an unknown period of time and was dry upon collection. The absence of detectable Allbac fecal DNA marker in these samples indicates that the signal may be decreased in the environment over time with exposure to light and air. Concentrations of the Allbac marker in the remaining fecal samples were variable; Ct values ranged from 10.9 to 23.1, with the minimum and maximum both observed in human samples.

All of the human fecal samples were positive for the Hubac marker and the concentrations were variable based on the range of Ct values obtained, with a minimum of 11.6 and a maximum of 23.4. One cow, both deer, llama, horse, and chicken samples were all negative for the Hubac marker (Ct >30). Two of the three beef cattle fecal samples were Hubac positive with Ct values of 18.3 and 23.2, which are within the Ct range shown for human fecal samples. Other non-human fecal samples that tested positive for the Hubac marker included the goose, rabbit, all three cats, and the single and composite dogs. The corresponding Ct values are given in Table 5.1.

Fecal ID	Host	Allbac Ct	Hubac Ct	Relative Hubac Ct
1"	Human	23.1	23.4	77.9
2"	Human	18.5	17.5	271.8
3"	Human	10.9	11.6	49.7
4"	Human	13.1	12.4	201.4
5"	Human	13.6	13.0	191.6
6"	Human	21.2	21.1	110.5
7"	Cow	12.1	>30	<0.0
8"	Cow	13.3	18.3	0.6
9"	Cow	13.5	23.2	<0.0
10"	Deer	14.8	>30	<0.0
11"	Deer	13.7	>30	<0.0
12"	Goose	13.6	18.5	0.8
13"	Otter	>30	>30	N/A
14"	Horse	19.2	>30	<0.0
15"	Llama	13.7	>30	<0.0
16"	Rabbit	14.4	18.1	2.5
17"	Cat	15.7	25.4	<0.0
18"	Cat	21.0	21.4	63.8
19"	Cat	17.6	20.6	5.0
20"	Dog	14.5	15.8	25.9
21"	Dog	22.8	25.1	10.0
22"	Chicken	>30	>30	N/A

Table 5.1. Real-time PCR threshold cycle results for Allbac and Hubac markers in fecal controls.

The Hubac marker is not entirely specific to humans in this study when viewed qualitatively as a presence/absence signal. However, the marker is found in relatively high concentrations based on Ct values obtained from human samples as compared to those from other host animals. The average Ct value across six human fecal samples was 16.5. All non-human fecal samples, with the exception of one dog, had Ct values greater than the human average, with an average of 21.3, a minimum of 18.1 and a maximum of 25.4. These quantitative differences alone are not sufficient to consider the Hubac marker specific enough for analytical determinations of host contributions in the environment. A tool for the normalization of the Hubac marker with respect to cross-reactivity among hosts is necessary. This thesis addresses the concern by using the signal from the non-host-specific Allbac marker to normalize the Hubac signal as relative Hubac Ct.

5.2. Normalization of Hubac Ct values in fecal samples

Fecal sample data were utilized to assess the proposed Hubac data normalization approach. All human fecal samples were positive for the Hubac marker and had Ct values within a range of ± 1 Ct of their corresponding Allbac marker. Many of the non-human fecal samples were positive for the Hubac marker as well, but with higher Ct values than their corresponding Allbac marker and therefore with lower concentrations of Hubac than Allbac (Table 5.1). The difference in Ct value (Allbac – Hubac) varied from a minimum of 0.4 for one feline fecal sample to a maximum of 17.9 for one bovine fecal sample.

Fecal sample results were grouped by host categories for comparisons. The categories created were human, cow, domestic pets, and other animals including wildlife. When the Hubac Ct values of human fecal samples are compared to the fecal samples from other host groups, no statistically significant differences are found ($P>0.05$) except when compared to the wildlife group ($P=0.034$)(Figures 5.1 and 5.2). When the relative Hubac Ct is calculated (eq. 2), the values for human fecal samples are statistically significantly greater than for wildlife and cattle ($P<0.05$),

but not for domestic pets (Figures 5.1 - 5.3). Normalization of the Hubac signal improves the ability to differentiate between feces from human and non-human hosts by highlighting true positive signals from humans while the false positive signals from other animal hosts drop into insignificance.

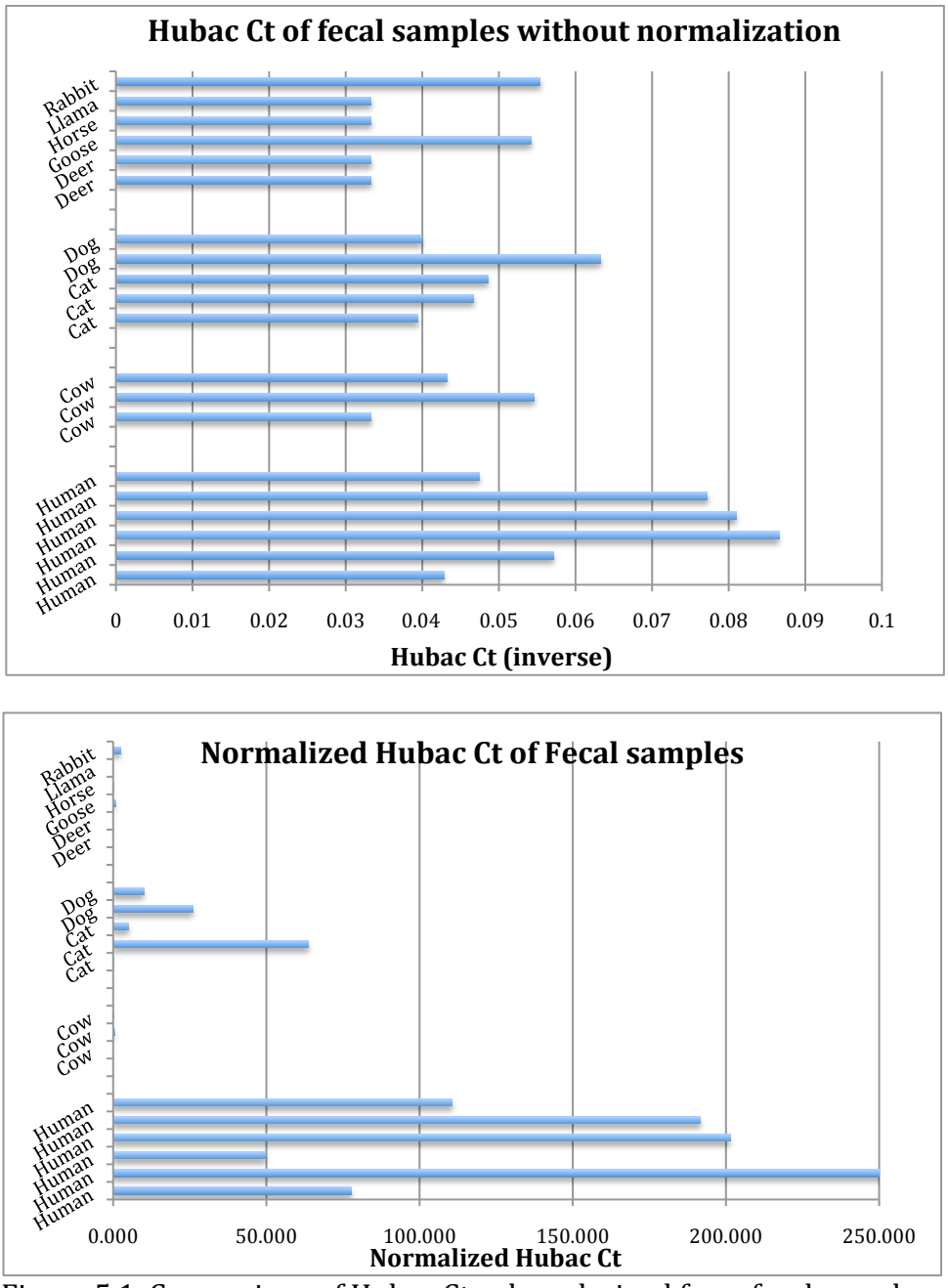

Figure 5.1. Comparison of Hubac Ct values derived from fecal samples plotted with and without mathematical normalization by the Allbac Ct value from the same sample.

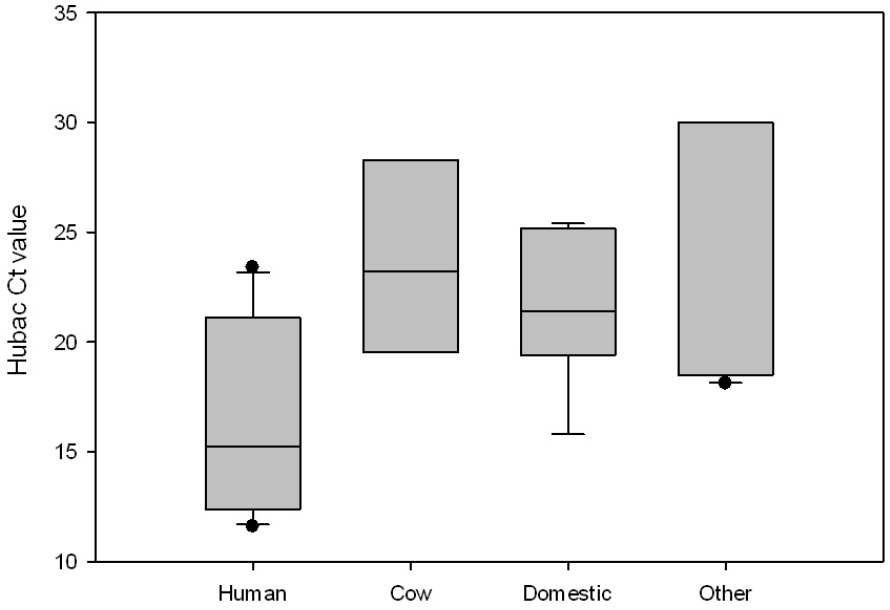

Figure 5.2 Hubac concentration by host category without normalization.

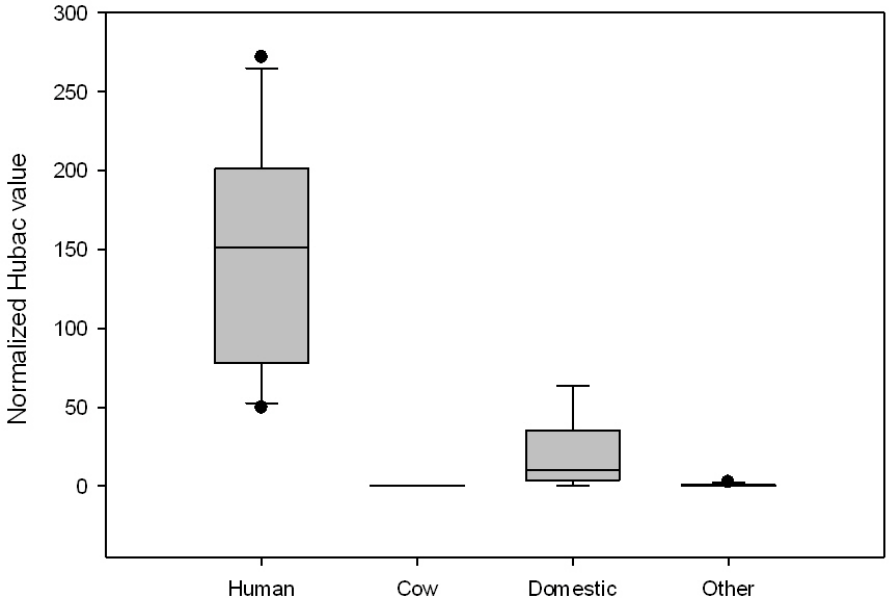

Figure 5.3 Normalized relative Hubac Ct by host category.

5.3. Marker fate and persistence

The starting qPCR concentration of Allbac DNA marker in the diluted sewage effluent from Blue Sky wastewater treatment plant was 4.4×10^4 copies/µL, yielding a log-transformed value of 4.65. The qPCR-measured concentrations of Allbac decreased by at least two orders of magnitude within 5 days of incubation for all marker fate experimental conditions (Table 5.2, Figure 5.4).

day	log Allbac light	log Allbac dark	log Allbac anaerobic dark
1	4.65	4.65	4.65
2	3.50	3.29	3.48
3	2.53	2.41	2.85
5	2.31	2.18	2.06

Table 5.2. Log Allbac copies/µL concentrations over time in light aerobic, dark aerobic and dark anaerobic conditions.

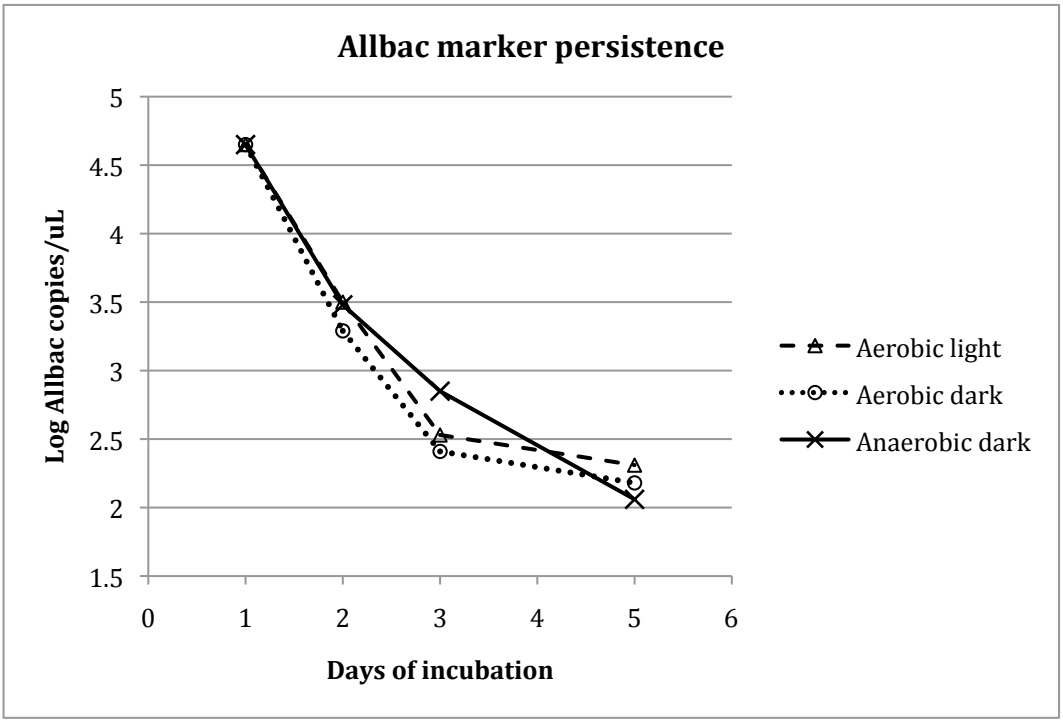

Figure 5.4. Allbac marker concentration reduction over time for three environmental conditions

This marker fate study assumes that the *Bacteroides* bacteria, originally present in the sewage sample, have been killed and possibly lysed prior to the start of the marker fate experiment. This assumption is based on the strict anaerobic nature of *Bacteroides*. The bacterial cells will not remain viable upon being exposed to oxygen after leaving the intestines of their hosts and through the transport to and processing at the wastewater treatment plant. The Allbac general fecal DNA marker was found to be unstable in the environment, as shown by significant degradation ($P<0.001$) for three different simulated environmental conditions. Because the marker did not degrade to undetectable levels for any of the tested conditions after 5 days, its presence in an environmental sample may not indicate that the sample was taken near the pollution source. However, if the marker is found at a high concentration, it is likely fresh and near the source. These results highlight the importance of quantification of the DNA markers and development of an environmentally significant limit of detection.

The Allbac marker may be more or less stable than other fecal DNA markers. Although all DNA is subject to predation and the effects of light and temperature, some sequences may be degraded more rapidly than others. Studies of fate and persistence are needed for all proposed markers for fecal source tracking so that comparisons can be made among markers like the comparison used for normalization of the Hubac signal in this thesis. Because the Allbac marker was the only fecal DNA marker studied in this fate and persistence experiment, the results are limited by the lack of knowledge of the fate and persistence of the Hubac marker in comparison to that of the Allbac marker.

5.4. Watershed application

Surface water samples from the Wolf Run watershed were analyzed for concentrations of the Hubac and Allbac genetic markers, *E. coli*, and AC/TC to determine the utility of screening urban watersheds with less expensive indicators prior to qPCR analysis of fecal source markers. Fecal loadings, as indicated by *E. coli*

concentrations, were variable, ranging from 10 to 17,329 MPN/100 mL. The Allbac genetic marker (non-host-specific, general *Bacteroides*) was present in all samples analyzed and its log-transformed concentrations were weakly proportional to log-transformed *E. coli* concentrations (R^2=0.65) (Figure 5.5). The AC/TC ratios varied from a minimum of 1.1 to a maximum of 66.5, denoting a range of relative fecal ages from very fresh to very aged.

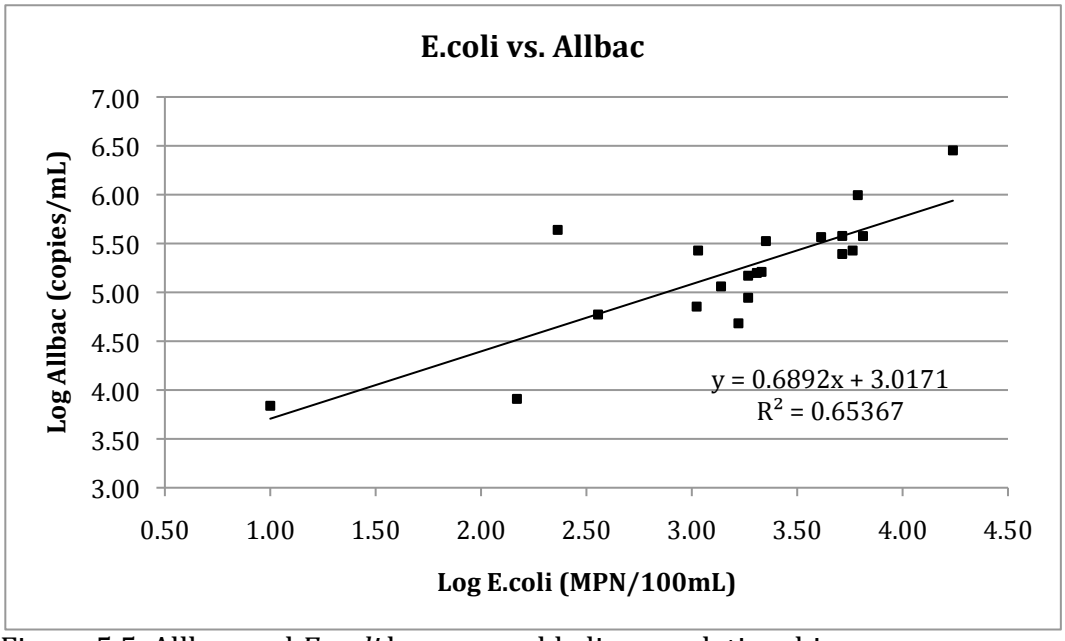

Figure 5.5. Allbac and *E. coli* have a weakly linear relationship.

The Hubac marker (human-associated *Bacteroides*) was detected in all samples, with concentrations ranging across two and one-half orders of magnitude (~8×10^2 – ~1×10^5 copies/mL water sample). No genetic marker concentrations were detected in the method blanks. The average relative percent difference between sample duplicates for log-transformed marker concentrations was 9%. The instrumental minimum detection limit for this study is 1×10^1 copies/µL DNA extract (which equals 1.2×10^2 copies/mL water sample). It should be noted, however, that the precision of values below 1×10^2 copies/µL (which equals 1.2×10^3 copies/mL water sample) may be poor.

Although the Hubac marker does not have perfect specificity, its utility for determining the human contribution among a set of samples from various locations in an urban watershed and among fecal samples from various animals is greatly improved by normalizing data relative to the nonspecific Allbac marker.

5.4.1 Hubac signal normalization in water samples

5.4.1.1 Relative Hubac copy abundance

Hubac marker concentrations alone were insufficient to define valid statistically significant differences between samples within the urban watershed. When the real-time PCR-quantified Hubac signal was divided by the Allbac signal at each site, a greater ability to differentiate sites was realized. Relative Hubac copy abundances ranged from <10 to 60 (Table 5.3).

Identifier	Hubac Copy/mL	Ct	Allbac Copy/mL	Ct	Relative Hubac copy abundance	Relative Hubac Ct
D01	4.67E+03	30.54	5.93E+04	26.02	7.87	1.09
D01duplicate	8.40E+02	29.60	3.36E+04	25.11	2.50	1.12
D02	1.64E+05	25.13	2.84E+06	20.8	5.78	1.32
D03	1.36E+03	32.38	6.90E+03	28.93	19.65	3.17
D04	1.74E+03	32.04	8.12E+03	28.7	21.42	3.54
D05	6.13E+04	26.61	2.47E+05	24.1	24.81	8.13
D06	8.83E+04	26.06	3.68E+05	23.56	23.97	8.21
D07	2.66E+03	31.35	2.68E+05	23.99	1.00	0.06
D08	2.18E+04	28.18	4.36E+05	23.33	5.01	0.78
D09	1.92E+03	31.87	8.80E+04	25.49	2.18	0.17
D10	7.82E+04	26.25	9.89E+05	22.23	7.91	1.80
D11	8.30E+04	26.16	2.68E+05	23.99	31.03	11.42
D12	5.39E+04	26.81	3.78E+05	23.52	14.25	3.73
D13	1.69E+05	25.08	3.77E+05	23.53	44.90	21.22
D13duplicate	9.78E+04	25.91	3.10E+05	23.79	31.59	12.00
D14	6.92E+04	26.43	1.15E+05	25.13	60.29	27.25
D15	3.84E+04	27.32	1.58E+05	24.7	24.24	7.28
D16	1.86E+04	28.42	1.48E+05	24.79	12.60	2.65
D17	2.20E+04	28.17	3.35E+05	23.69	6.56	1.13
D18	1.84E+04	28.44	1.62E+05	24.68	11.33	2.33
D19	6.65E+03	29.97	7.15E+04	25.77	9.30	1.50
D20	8.42E+02	33.14	4.82E+04	26.3	1.75	0.11

Table 5.3. Hubac marker concentrations and threshold cycle values before and after normalization by two methods.

Sample D14 had a Hubac concentration of 7×10^4 copies/mL, which is close to the mean Hubac concentration for the entire watershed on the date of sampling (4.6×10^4 copies/mL) and is not an outlier with respect to results from other locations on the same date (Figures 5.6 and 5.7). However, the relative Hubac copy abundance for this sample (60.17) is an outlier of the data set, with a value greater than the sum of the 75th percentile value and 1.5 times the difference between the 25th and 75th percentiles (5.77 and 23.99, respectively) (Figure 5.7). The normalized value indicates that D14 is much more contaminated than most other sites in the study. It was later confirmed that site D14 is the location of a cross-connection between the sanitary- and storm-water sewer systems, which would have been missed if the data had not been normalized.

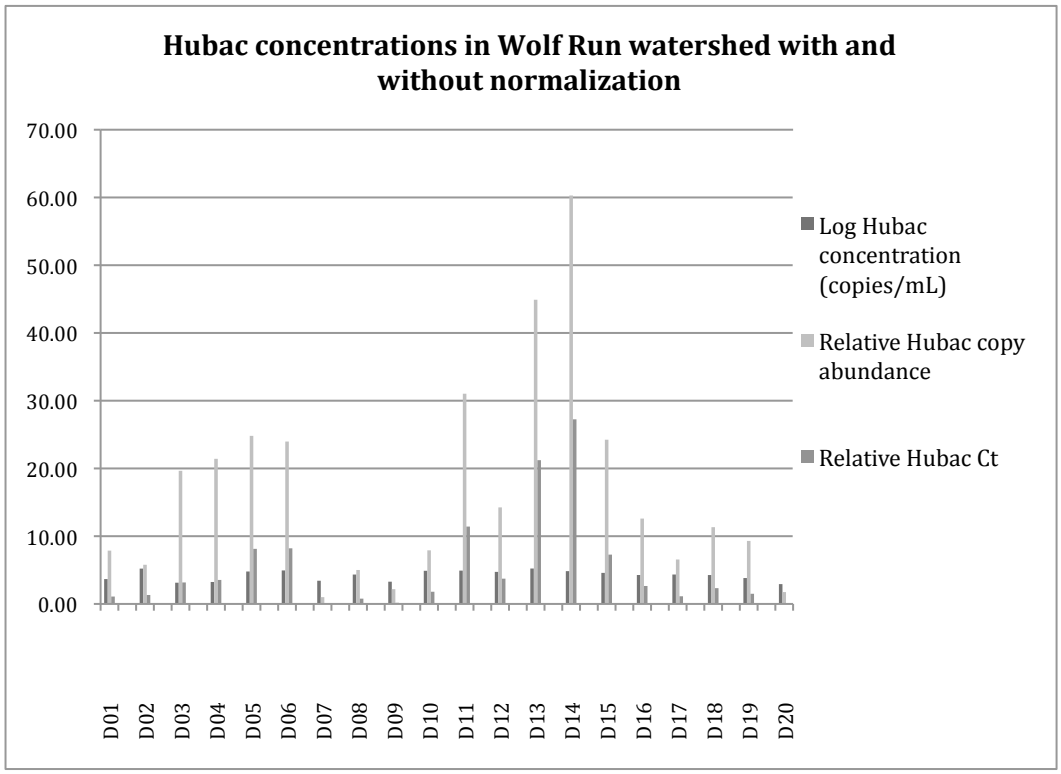

Figure 5.6. Hubac concentrations in water samples from the Wolf Run watershed compared with and without normalization by two methods.

Conversely, sample D02 had the second highest Hubac marker concentration among all samples when not normalized (1.64×10^5 copies/mL). This concentration result is nearly an outlier of the data set and would be considered a human hot spot. However, the relative Hubac copy abundance for this sample is only 5.77, which lies at the 25th percentile of the data set and is not an outlier (Figure 5.7). D02 was later determined to be the location of a drainage tile from the Red Mile horse racing track and stables. It is possible that the high Hubac concentration result was a false positive from the horse population nearby and that the normalization technique used to calculate the relative Hubac copy abundance removes this source of error as shown in section 5.1.

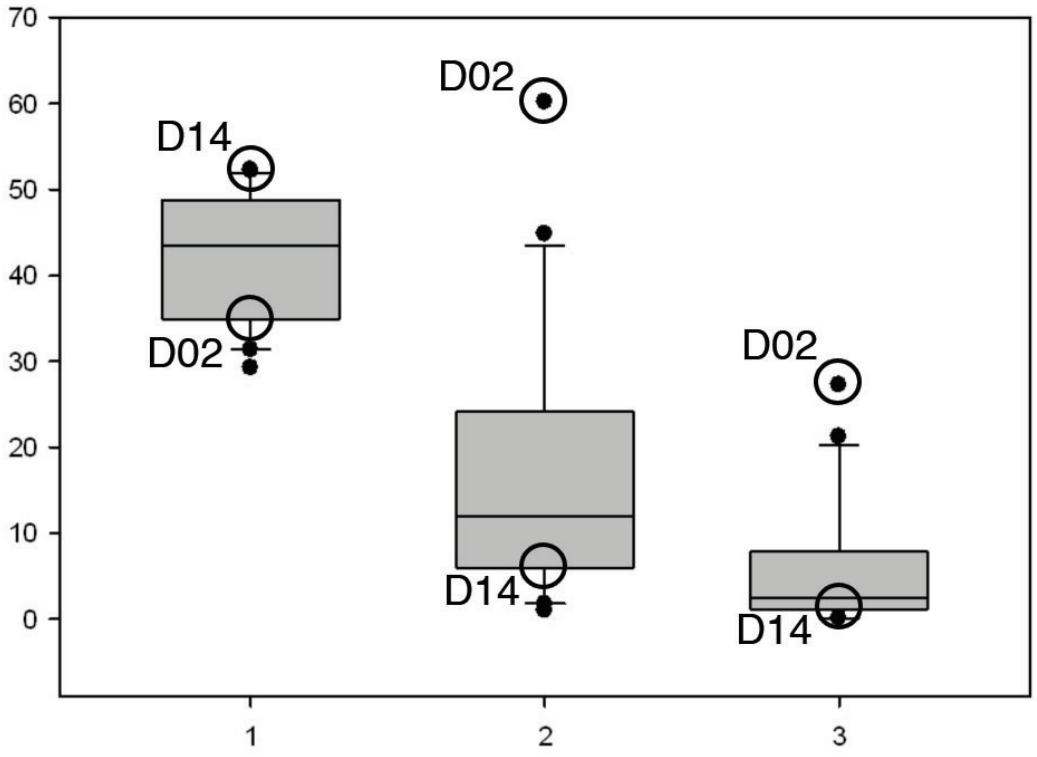

Figure 5.7. Hubac results from water samples 1) without normalization (log-transformed and multiplied by 10 for scale), 2) normalized as Relative Hubac copy abundance and 3) normalized as Relative Hubac Ct.

In terms of Hubac concentration, sample D13 is an outlier of the data set, and it is nearly an outlier when normalized as relative Hubac copy abundance. D13 is not likely to be a cross-connection between the sanitary and storm-water sewers, but is downstream of a neighborhood where the sanitary sewer lines are undersized for the population served. Broken or collapsed sections of the pipe have been found in the area since the completion of this study.

5.4.1.2 Relative Hubac Ct

Normalizing the Hubac signal by the Allbac signal (eq. 2) gave a similar trend to that found when normalizing with respect to concentration (eq. 1). The range of normalized Hubac Ct values found when using the exponential function (0.1 to 27.2) is greater than can be achieved by dividing the Ct values directly (0.79-0.95) and provides a more visible distribution of results.

Samples D14 and D02 again serve as examples of the utility of Hubac marker normalization. Although sample D14 is not an outlier of the data set when Hubac marker concentrations are used, it is an extreme outlier, with a value greater than the sum of the 75th percentile and 3 times the difference of the 25th and 75th percentiles (1.09 and 7.28, respectively), when normalized as relative Hubac Ct (27.25) (Figure 5.7). As noted above, subsequent to this study, D14 was confirmed to be the location of a cross-connection between the sanitary and storm-water sewers. Likewise, sample D02, the drainage from the horse track, would be considered a human hot spot when Hubac concentration results are used directly, but it is not an outlier in terms of its relative Hubac Ct value (1.32). Sample D13 is an outlier considering its relative Hubac Ct value. This is the only location in the study that would be labeled as a hot spot of human fecal pollution when the Hubac signal is normalized as well as when using non-normalized concentration data.

5.4.2 Watershed screening tool

The AC/TC ratio supported the relative Hubac copy abundance findings, with values that were suppressed from those expected based on previous studies of non-

sewage-impacted, urban creeks in the region. Higher Hubac signals were associated with AC/TC ratios of less than 20, while ratios greater than 20 were associated with lower Hubac signal strength. The threshold value of 20 was chosen for differentiating AC/TC results based upon the findings from work of previous investigators (Nieman & Brion, 2003; Reed et al., 2011). Of the 20 sample locations tested, 14 sites showed AC/TC values of less than 20 and *E. coli* concentrations of greater than 500 MPN/100 mL. These sites would have been considered positive for potential hot spot identification by screening with AC/TC and *E. coli*. Of these 14 sites, three were confirmed to be positive hot spots of human fecal contamination (potentially from broken or leaking sewer lines or cross-connections) by human-specific marker concentrations, as relative Hubac copy abundance, greater than 28.42 (Table 5.4). This threshold value represents an outlier (i.e., greater than the sum of the 75th percentile and 1.5 times the inter-quartile range).

Site ID	Relative Hubac copy abundance	AC/TC	*E.coli*
D01	7.87	1.13	359
D02	5.78	N/A	17329
D03	19.65	8.93	10
D04	21.42	4.62	148
D05	24.81	9.94	5172
D06	23.97	18.24	4106
D07	1.00	65.64	1071
D08	5.01	66.52	231
D09	2.18	12.97	1850
D10	7.91	3.49	6131
D11	31.03	3.87	5794
D12	14.25	2.69	5172
D13	44.90	4.38	6488
D14	60.29	12.57	1376
D15	24.24	12.33	2035
D16	12.60	8.68	1850
D17	6.56	13.10	2247
D18	11.33	8.99	2142
D19	9.30	6.76	1054
D20	1.75	12.00	1664

Table 5.4. Six of the 14 potential hot spots flagged by the screening tool were confirmed by Hubac/Allbac ratios greater than 20%.

Screening the samples with less expensive viable bacterial indicators of load and age would have reduced the molecular analysis costs by one-third for this study by decreasing the number of samples for molecular analysis from 20 to 14. All three sites found to be true positive human hot spots by molecular analysis would have been selected as potential hot spots by the screening tool. Although the positive predictive value of the screening tool is only 21% and the specificity is also poor at only 31%, the negative predictive value and sensitivity of the screening tool were both found to be 100% in this study (Table 5.5). It is not proposed that the screening tool be used alone to locate the hot spots of human fecal pollution in a watershed. The tool shows great utility, however, to eliminate sites that do not need to be analyzed further because they are not likely to be true hot spots.

	<27 relative Hubac abundance	>27 relative Hubac abundance	
E. coli<500 or AC/TC>20	6	0	Negative Predictive Value= 100% =TN/(FN+TN)
E. coli>500 & AC/TC<20	11	3	Positive Predictive Value= 21% =TP/(TP+FP)
	Specificity= 31% =TN/(FP+TN)	Sensitivity=100% =TP/(TP+FN)	

Table 5.5. Screening tool has a strong negative predictive value (100%) and high sensitivity (100%).

6. Conclusions and Future Work

6.1. Reduction of error by Hubac signal normalization

Several sources of error may be ameliorated with the data normalization techniques presented in this thesis. Although the Hubac marker is found in the feces of animals other than humans, non-human and human fecal samples can be differentiated by normalization relative to the corresponding Allbac signal (with the exception of some domestic pets). Whereas contamination from farm animals may cause a false positive "human" signal, when normalized by the non-specific Allbac marker, such contamination would not be above normal baseline, as shown with sample D02 in this thesis. Layton et al. (2006) presented a relative ratio technique utilizing the concentration data derived from the Allbac and Hubac marker assays. The authors suggested that dividing the Allbac result by the Hubac result from the same sample would yield a single value representing the percent of the human contribution to the total fecal pollution present in the sample. Use of the ratio technique for calculation of percent contribution has since been discarded because the marker concentrations vary from host to host even within one species. The normalization approaches presented in this these, therefore, should not be used to define a standard value for use in a regulatory sense, but rather as a tool to help define trends across a watershed and to compare watershed sites to each other.

Another source of error in quantitative fecal source tracking is environmental variability. A sample site that is contaminated with sewage from a broken or leaking pipe may show a very high concentration of Hubac marker initially but then drop below the level of concern after a rain event due to dilution. By normalizing the marker concentration with that of another marker which fluctuates equally according to environmental variables, this temporal variation may be stabilized, as shown with sample D14 of this thesis. The Allbac marker persistence results presented in this thesis verify the results presented by Bell (2007), who found the Allbac marker to degrade by 2 log units within 2 days in unfiltered stream water. Both studies, however, analyzed the Allbac marker alone and the decay rate of the

Hubac marker is still unknown. If the markers persist in the environment similarly, the use of this technique should be valid regardless of environmental variation and stream conditions. If, however, the Hubac marker degrades more rapidly under some environmental conditions than the Allbac marker, a false negative bias in the normalized Hubac signal will occur over time after fecal input to the environment. Likewise, a more rapid degradation of the Allbac marker with respect to Hubac would lead to a false positive bias. More robust sampling and data collection are needed to determine the extent to which this source of error is limited by the normalization technique presented. Specifically, in light of the potential for differential decay of the markers in the environment (Walters et al., 2009), the decay rates of any markers utilized for normalization against each other should be determined. A study analyzing the concentrations of both human and non-host-specific fecal markers at multiple locations in a watershed across multiple seasons, as well as a bench-scale study of the fate and persistence of each marker, is needed to determine the stability of the markers at one location over time.

Stoeckel et al. (2009) indicated that the greatest source of variability in molecular fecal source tracking data comes from the DNA extraction step. This source of variability may be controlled with the data normalization approaches presented in this thesis. Because the individual marker concentrations are determined from the same DNA extract, the ratio between the markers should be consistent for a given sample even if the precision between replicate extractions is imperfect. In this way, the Allbac marker is serving as an internal standard with respect to the Hubac marker. It is not known, however, if the markers are subject to differential PCR inhibition (Huggett et al., 2008). The assumption that the markers are inhibited similarly needs to be validated to allow greater confidence in the normalization approaches presented in this thesis.

The relative Hubac Ct normalization approach presented in this thesis indicates that it may be possible to define the locations of human hot spots in a watershed using relative concentrations (as Ct values) of two fecal markers rather than absolute

quantification by calibration. If this technique is validated for further application, the cost of analysis for each sample would be decreased. It would also be cost-effective to analyze both markers simultaneously in the same real-time PCR reaction with a multiplexed assay. This would require that the TaqMan probes for each marker be labeled with a different dye and that the two reactions do not experience competitive forces for reaction components that prevent the efficient generation of each PCR product.

6.2. Watershed application and cost reduction

Similar results are found by normalizing the data using marker concentrations derived by qPCR (relative Hubac copy abundance) and with Ct values derived directly from real-time PCR without quantification (relative Hubac Ct). If it is determined at the beginning of analysis that all data will be normalized for application, it is possible to reduce analytical costs and turnaround time significantly by eliminating the generation of a calibration curve and using the Ct values alone. These two normalization approaches remove some sources of error related to environmental variability and extraction efficiency and allow for greater comparison of the signal between sample sites and across time. It is important to note, however, that the ratio of two separate DNA markers does not relate directly to the proportion of the total fecal pollution that is human.

As with the conclusions from prior studies utilizing multi-parameter and tiered approaches to fecal source tracking, this thesis finds greater utility and cost savings in the use of a tiered, screening tool method than with the measurement of any one fecal indicator alone (Black et al., 2007; Blanch et al., 2006; Boehm et al., 2003; Noble et al., 2006). Significant cost reduction of watershed analysis is achieved with the screening tool presented in this thesis. Many of the entities seeking access to molecular fecal source tracking techniques are local municipalities and utilities with budgets that do not support large research efforts. The initial reaction to this dilemma is to limit the number of samples to be analyzed by decreasing the spatial

and temporal range. This decision leads to decreased understanding of the watershed and can cause true hot spots of human fecal pollution to be missed. Screening the samples of a watershed study by viable indicators of fecal load and age prior to further fecal source analysis will save time and reduce the overall cost of the study while maintaining the ability to cover a large temporal and spatial range in the sampling plan. This is a critical development for the application of molecular fecal source tracking, as it provides a path for parties with lower budgets to utilize the more costly methods after eliminating sites of low risk, using a screening tool with a strong negative predictive value.

APPENDICES

Appendix A: Fecal sample statistics

Descriptive Statistics: Wednesday, July 27, 2011, 3:54:45 PM

Data source: Wolf Run watershed in Thesis data playsheet

Column	Size	Missing	Mean	Std Dev	Std. Error	C.I. of Mean
Human	6	0	16.500	4.957	2.024	5.202
Cow	3	0	23.833	5.876	3.392	14.596
Domestic	5	0	21.660	3.916	1.751	4.863
Other	6	0	26.100	6.043	2.467	6.342
Human norm	6	0	150.469	84.895	34.658	89.091
Cow norm	3	0	0.216	0.368	0.213	0.915
Domestic norm	5	0	20.940	25.840	11.556	32.084
Other norm	6	0	0.543	0.996	0.406	1.045
Human Allbac	6	0	16.733	4.912	2.005	5.155
Cow Allbac	3	0	12.967	0.757	0.437	1.881
Domestic Allbac	5	0	18.320	3.508	1.569	4.356
Other Allbac	6	0	14.900	2.160	0.882	2.266
Domestic Allbac	5	0	18.320	3.508	1.569	4.356
Other Allbac	6	0	14.900	2.160	0.882	2.266

Column	Range	Max	Min	Median	25%	75%
Human	11.800	23.400	11.600	15.250	12.200	21.675
Cow	11.700	30.000	18.300	23.200	18.300	30.000
Domestic	9.600	25.400	15.800	21.400	18.200	25.250
Other	11.900	30.000	18.100	30.000	18.400	30.000
Human norm	222.169	271.828	49.659	151.036	70.825	218.988
Cow norm	0.641	0.641	0.000	0.00600	0.000	0.641
Domestic norm	63.757	63.763	0.00600	10.026	2.493	44.843
Other norm	2.472	2.472	0.000	0.001000	0.000	1.205
Human Allbac	12.200	23.100	10.900	16.050	12.550	21.675
Cow Allbac	1.400	13.500	12.100	13.300	12.100	13.500
Domestic Allbac	8.300	22.800	14.500	17.600	15.100	21.900
Other Allbac	5.600	19.200	13.600	14.050	13.675	15.900
Domestic Allbac	8.300	22.800	14.500	17.600	15.100	21.900
Other Allbac	5.600	19.200	13.600	14.050	13.675	15.900

Column	Skewness	Kurtosis	K-S Dist.	K-S Prob.	SWilk W	SWilk Prob
Human	0.486	-1.930	0.260	0.226	0.885	0.293
Cow	0.479	--	0.210	0.612	0.991	0.821
Domestic	-0.759	0.0560	0.210	0.550	0.909	0.460
Other	-0.970	-1.864	0.407	0.002	0.648	0.002
Human norm	0.254	-1.397	0.186	0.617	0.946	0.711
Cow norm	1.732	--	0.382	0.094	0.757	0.016
Domestic norm	1.548	2.191	0.264	0.286	0.842	0.171
Other norm	1.993	3.868	0.373	0.009	0.661	0.002
Human Allbac	0.191	-2.067	0.238	0.330	0.923	0.524
Cow Allbac	-1.597	--	0.337	0.197	0.855	0.253
Domestic Allbac	0.338	-2.101	0.181	0.672	0.938	0.653
Other Allbac	2.196	4.979	0.352	0.019	0.676	0.003

Domestic Allbac	0.338	-2.101	0.181	0.672	0.938	0.653
Other Allbac	2.196	4.979	0.352	0.019	0.676	0.003

Column	**Sum**	**Sum of Squares**
Human	99.000	1756.340
Cow	71.500	1773.130
Domestic	108.300	2407.130
Other	156.600	4269.860
Human norm	902.813	171880.605
Cow norm	0.647	0.411
Domestic norm	104.698	4863.085
Other norm	3.257	6.724
Human Allbac	100.400	1800.680
Cow Allbac	38.900	505.550
Domestic Allbac	91.600	1727.340
Other Allbac	89.400	1355.380
Domestic Allbac	91.600	1727.340
Other Allbac	89.400	1355.380

One Way Analysis of Variance Wednesday, July 27, 2011, 4:02:46 PM

Data source: fecal hubac in Thesis data playsheet

Normality Test (Shapiro-Wilk) Passed (P = 0.076)

Equal Variance Test: Passed (P = 0.921)

Group	**Name**	**N**	**Missing**	**Mean**	**Std Dev**	**SEM**
Human	6	0	16.500	4.957	2.024	
Cow	3	0	23.833	5.876	3.392	
Domestic	5	0	21.660	3.916	1.751	
Other	6	0	26.100	6.043	2.467	

Source of Variation	**DF**	**SS**	**MS**	**F**	**P**
Between Groups	3	291.963	97.321	3.573	0.038
Residual	16	435.839	27.240		
Total	19	727.802			

The differences in the mean values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = 0.038).

Power of performed test with alpha = 0.050: 0.518

All Pairwise Multiple Comparison Procedures (Holm-Sidak method):
Overall significance level = 0.05

Comparisons for factor:

Comparison	**Diff of Means**	**t**	**P**	**P<0.050**
Other vs. Human	9.600	3.186	0.034	Yes
Cow vs. Human	7.333	1.987	0.283	No
Domestic vs. Human	5.160	1.633	0.406	No

Other vs. Domestic		4.440	1.405	0.447	No
Other vs. Cow	2.267	0.614	0.795	No	
Cow vs. Domestic		2.173	0.570	0.576	No

One Way Analysis of Variance Wednesday, July 27, 2011, 4:04:02 PM

Data source: fecal hubac in Thesis data playsheet

Normality Test (Shapiro-Wilk) Failed (P < 0.050)

Test execution ended by user request, ANOVA on Ranks begun

Kruskal-Wallis One Way Analysis of Variance on Ranks Wednesday, July 27, 2011, 4:04:02 PM

Data source: fecal hubac in Thesis data playsheet

Group	**N**	**Missing**	**Median**	**25%**	**75%**	
Human norm	6	0	151.036	70.825	218.988	
Cow norm	3	0	0.00600	0.000	0.641	
Domestic norm	5	0	10.026	2.493	44.843	
Other norm	6	0	0.001000		0.000	1.205

H = 14.999 with 3 degrees of freedom. (P = 0.002)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = 0.002)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Dunn's Method) :

Comparison	**Diff of Ranks**	**Q**	**P<0.05**	
Human norm vs Other norm		12.083	3.538	Yes
Human norm vs Cow norm		11.667	2.789	Yes
Human norm vs Domestic norm		5.833	1.628	No
Domestic norm vs Other norm		6.250	1.745	No
Domestic norm vs Cow norm		5.833	1.350	Do Not Test
Cow norm vs Other norm	0.417	0.0996	Do Not Test	

Appendix B: Marker fate and persistence statistics

One Way Analysis of Variance Wednesday, July 27, 2011, 4:08:23 PM

Data source: marker fate in Thesis data playsheet

Normality Test (Shapiro-Wilk) Passed (P = 0.083)

Equal Variance Test: Passed (P = 0.997)

Group Name	N	Missing	Mean	Std Dev	SEM
day 1 light	4	0	3.248	1.068	0.534
day 1 dark	4	0	3.133	1.119	0.560
day 1 anaerobic	4	0	3.260	1.094	0.547

Source of Variation	DF	SS	MS	F	P
Between Groups	2	0.0395	0.0198	0.0165	0.984
Residual	9	10.770	1.197		
Total	11	10.809			

The differences in the mean values among the treatment groups are not great enough to exclude the possibility that the difference is due to random sampling variability; there is not a statistically significant difference (P = 0.984).

Power of performed test with alpha = 0.050: 0.050

The power of the performed test (0.050) is below the desired power of 0.800.
Less than desired power indicates you are less likely to detect a difference when one actually exists.
Negative results should be interpreted cautiously.

One Way Analysis of Variance Wednesday, July 27, 2011, 4:15:27 PM

Data source: marker fate in Thesis data playsheet

Normality Test (Shapiro-Wilk) Passed (P = 0.803)

Equal Variance Test: Passed (P = 0.580)

Group Name	N	Missing	Mean	Std Dev	SEM
day 1	3	0	4.650	0.000	0.000
day 2	3	0	3.423	0.116	0.0669
day 3	3	0	2.597	0.227	0.131
day 5	3	0	2.183	0.125	0.0722

Source of Variation	DF	SS	MS	F	P
Between Groups	3	10.648	3.549	175.707	<0.001
Residual	8	0.162	0.0202		
Total	11	10.809			

The differences in the mean values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = <0.001).

Power of performed test with alpha = 0.050: 1.000

All Pairwise Multiple Comparison Procedures (Holm-Sidak method):
Overall significance level = 0.05

Comparisons for factor:

Comparison	**Diff of Means**		**t**	**P**	**P<0.050**
day 1 vs. day 5	2.467	21.256	<0.001	Yes	
day 1 vs. day 3	2.053	17.694	<0.001	Yes	
day 2 vs. day 5	1.240	10.685	<0.001	Yes	
day 1 vs. day 2	1.227	10.571	<0.001	Yes	
day 2 vs. day 3	0.827	7.124	<0.001	Yes	
day 3 vs. day 5	0.413	3.562	0.007	Yes	

Appendix C: Analytical cost comparison with and without screening tool

	Without Screening by E.coli and AC/TC	Screening by E.coli and AC/TC
Molecular analysis cost per sample (including labor and supplies)	$250.00	$250.00
# samples analyzed to find hot spots	20	14
Total molecular analysis cost	$5,000.00	$3,500.00
Microbial screening cost per sample (including labor and supplies)	$30.00	$30.00
# samples analyzed to find hotspots	0	20
Total microbial screening cost	$0.00	$600.00
Total analytical cost	$5,000.00	$4,100.00

REFERENCES

Anderson, K. L., Whitlock, J. E., & Harwood, V. J. (2005). Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied and Environmental Microbiology*, *71*(6), 3041-3048.

Bell, A.K., (2007). Factors influencing persistence of fecal *Bacteroides* in stream water. Thesis presented for Master of Science, University of Tennessee, Knoxville.

Bernhard, A. E., & Field, K. G. (2000). A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Applied and Environmental Microbiology*, *66*(10), 4571-4574.

Bernhard, A. E., Goyard, T., Simonich, M. T., & Field, K. G. (2003). Application of a rapid method for identifying fecal pollution sources in a multi-use estuary. *Water Research*, *37*(4), 909-913.

Black, L. E., Brion, G. M., & Freitas, S. J. (2007). Multivariate logistic regression for predicting total culturable virus presence at the intake of a potable-water treatment plant: Novel application of the atypical coliform/total coliform ratio. *Applied and Environmental Microbiology*, *73*(12), 3965-3974.

Blanch, A. R., Belanche-Munoz, L., Bonjoch, X., Ebdon, J., Gantzer, C., Lucena, F., Ottoson, J., Kourtis, C., Iversen, A., Kuhn, I., Moce, L., Muniesa, M., Schwartzbrod, J., Skraber, S., Papageorgiou, G.T., Taylor, H., Wallis, J., & Jofre, J. (2006). Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Applied and Environmental Microbiology*, *72*(9), 5915-5926.

Boehm, A. B., Fuhrman, J. A., Mrše, R. D., & Grant, S. B. (2003). Tiered approach for identification of a human fecal pollution source at a recreational beach: case study at Avalon Bay, Catalina Island, California. *Environmental Science & Technology*, *37*(4), 673-680.

Brion, G. M., & Mao, H. H. (2000). Use of total coliform test for watershed monitoring with respect to atypicals. *Journal of Environmental Engineering*, *126*, 175-181.

Brion, G. M., Neelakantan, T. R., & Lingireddy, S. (2002). A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water. *Water Research*, *36*(15), 3765-3774.

Carrillo, M., Estrada, E., & Hazen, T. C. (1985). Survival and enumeration of the fecal indicators Bifidobacterium adolescentis and Escherichia coli in a tropical rain forest watershed. *Applied and Environmental Microbiology*, *50*(2), 468-476.

Carson, C. A., Christiansen, J. M., Yampara-Iquise, H., Benson, V. W., Baffaut, C., Davis, J. V., Broz, R.R., Kurtz, W.B., Rogers, W.M., & Fales, W.H. (2005). Specificity of a Bacteroides thetaiotaomicron marker for human feces. *Applied and Environmental Microbiology 71*(8), 4945-4949.

Chan, K. H., Lam, M. H. W., Poon, K. F., Yeung, H. Y., & Chiu, T. K. T. (1998). Application of sedimentary fecal stanols and sterols in tracing sewage pollution in coastal waters. *Water Research*, *32*(1), 225-235.

Cole, D., Long, S. C., & Sobsey, M. D. (2003). Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. *Applied and Environmental Microbiology*, *69*(11), 6507-6514.

Dick, L. K., & Field, K. G. (2004). Rapid estimation of numbers of fecal Bacteroidetes by use of a quantitative PCR assay for 16S rRNA genes. *Applied and Environmental Microbiology 70*(9), 5695-5697.

Dick, L. K., Bernhard, A. E., Brodeur, T. J., Santo Domingo, J. W., Simpson, J. M., Walters, S. P., & Field, K.G. (2005). Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. *Applied and Environmental Microbiology 71*(6), 3184-3191.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., & Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science 308*(5728), 1635-1638.

Field, K. G., & Samadpour, M. (2007). Fecal source tracking, the indicator paradigm, and managing water quality. *Water Research 41*(16), 3517-3538.

Finegold, S. M., Sutter, V. L., & Mathisen, G. E. (1983). Normal indigenous intestinal flora. In Hentges, D. J., Human Intestinal Microflora in Health and Disease, Academic Press, p. 3-31.

Glassmeyer, S. T., Furlong, E. T., Kolpin, D. W., Cahill, J. D., Zaugg, S. D., Werner, S. L., Meyer, M. T., & Kryak, D.D. (2005). Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. *Environmental Science & Technology*, *39*(14), 5157-5169.

Gordon, D. M., & Cowling, A. (2003). The distribution and genetic structure of Escherichia coli in Australian vertebrates: host and geographic effects. *Microbiology*, *149*(12), 3575-3586.

Griffith, J. F., Weisberg, S. B., & McGee, C. D. (2003). Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *Journal of Water and Health*, *1*(4), 141-151.

Hoaglin, D.C., Mosteller, F., & Tukey, J.W. (eds) (1983). Understanding Robust and Exploratory Data Analysis. Wiley.

Hooper, L. V., Wong, M. H., Thelin, A., Hansson, L., Falk, P. G., & Gordon, J. I. (2001). Molecular analysis of commensal host-microbial relationships in the intestine. *Science*, *291*(5505), 881-884.

Huggett, J.F., Novak, T., Garson, J.A., Green, C., Morris-Jones, S.D., Miller, R.F., & Zumla, A. (2008). Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. *BMC Research Notes*, 1(70), doi:10.1186/1756-0500-1-70.

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. *Applied and Environmental Microbiology 72*(6), 4214-4224.

Lu, J., Santo Domingo, J. W., Hill, S., & Edge, T. A. (2009). Microbial diversity and host-specific sequences of Canadian goose feces. *Applied and Environmental Microbiology, 75*(18), 5919-5926.

Nieman, J., & Brion, G.M. (2003). Novel bacterial ratio for predicting fecal age. *Water Science and Technology,* 47(3), 45-49.

Noble, R. T., Allen, S. M., Blackwood, A. D., Chu, W., Jiang, S. C., Lovelace, G. L., Sobsey, M.D., Stewart, J.R., & Wait, D.A. (2003). Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study. *Journal of Water and Health*, *1*(4), 195-207.

Noble, R. T., Griffith, J. F., Blackwood, A. D., Fuhrman, J. A., Gregory, J. B., Hernandez, X., Liang, X., Bera, A.A., & Schiff, K. (2006). Multitiered approach using quantitative PCR to track sources of fecal pollution affecting Santa Monica Bay, California. *Applied and Environmental Microbiology 72*(2), 1604-1612.

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. *Applied Microbiology and Biotechnology 74*(4), 890-901.

Peeler, K. A., Opsahl, S. P., & Chanton, J. P. (2006). Tracking anthropogenic inputs using caffeine, indicator bacteria, and nutrients in rural freshwater and urban marine systems. *Environmental Science & Technology*, *40*(24), 7616-7622.

Reed, T. M., Fryar, A. E., Brion, G. M., & Ward, J. W. (2011). Differences in pathogen indicators between proximal urban and rural karst springs, Central Kentucky, USA. *Environmental Earth Sciences*, *64*, 47-55.

Santo Domingo, J. W., Bambic, D. G., Edge, T. A., & Wuertz, S. (2007). Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Research 41*(16), 3539-3552.

Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R., & Lukasik, J. (2002). Microbial source tracking: current methodology and future directions. *Applied and Environmental Microbiology*, *68*(12), 5796-5803.

Seurinck, S., Defoirdt, T., Verstraete, W., & Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environmental Microbiology 7*(2), 249-259.

Shanks, O. C., Nietch, C., Simonich, M., Younger, M., Reynolds, D., & Field, K. G. (2006). Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Applied and Environmental Microbiology 72*(8), 5537-5546.

Shanks, O. C., Kelty, C. A., Sivaganesan, M., Varma, M., & Haugland, R. A. (2009). Quantitative PCR for genetic markers of human fecal pollution. *Applied and Environmental Microbiology*, *75*(17), 5507-5513.

Soule, M., Kuhn, E., Loge, F., Gay, J., & Call, D.R., (2006). Using DNA microarrays to identify library-independent markers for bacterial source tracking. *Applied and Environmental Microbiology* 72(3), 1843-1851.

Stoeckel, D. M., Mathes, M. V., Hyer, K. E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T.L., Fenger, T.W., Samadpour, M., Strickler, K.M., & Wiggins, B.A. (2004). Comparison of seven protocols to identify fecal contamination sources using Escherichia coli. *Environmental Science & Technology*, *38*(22), 6109-6117.

Stoeckel, D.M., & Harwood, V.J. (2007). Performance, design, and analysis in microbial source tracking studies. *Applied and Environmental Microbiology*, 73(8), 2405-2415.

Stoeckel, D.M., Stelzer, E.A., & Dick, L.K. (2009). Evaluation of two spike-and-recovery controls for assessment of extraction efficiency in microbial source tracking studies. *Water Research*, 43, 4820-4827.

Vogel, J. R., Stoeckel, D. M., Lamendella, R., Zelt, R. B., Santo Domingo, J. W., Walker, S. R., & Oerther, D.B. (2007). Identifying fecal sources in a selected catchment reach using multiple source-tracking tools. *Journal of Environmental Quality 36*(3), 718-729.

Walters, S.P., Yamahara, K.M., & Boehm, A.B. (2009). Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: Implications for their use in assessing risk in recreational waters. *Water Research, 43*(19), 4929-4939.

NAME **POSITION TITLE**
Tricia L. Coakley Research Facility Manager

EDUCATION/TRAINING

INSTITUTION	DEGREE	YEAR	FIELD OF STUDY
Western KY University	BS	1998	Biology/Chemistry (double major)

Positions and Employment
1995-1998 Student Laboratory Technician, Ogden Environmental Laboratory, Western KY University
1998-1999 Substitute Teacher, Metcalfe County Schools, Edmonton, KY
1999-2002 Laboratory Technician, Ogden Environmental Laboratory, Western KY University
2002-present Research Facility Manager, Environmental Research and Training Laboratory, University of KY

Research Skills and Interests
My research interests include microbial and chemical water quality and hydrogeology. I am specifically interested in the study of molecular fecal source tracking and in method development and technological advancement of this field. My skills include differential bacteriology, epifluorescence microscopy, cell culture, conventional and real-time PCR, metals analysis by ICP-OES and AA, and environmental sampling.

Selected publications and presentations
Coakley, T., G.M. Brion, and A.E. Fryar, 2009, "Important Considerations for the Interpretation of Molecular Fecal Source Tracking Data: A Case Study of Central Kentucky Urban and Mixed Use Watersheds." *IWA WaterMicro: 15th Int. Symposium on Health-related Water Microbiology*, Naxos Island, Greece.

Coakley, T., G.M. Brion, and A.E. Fryar, 2009, "Relationships between Indicators of Fecal Load, Source, and Age: Developing a Multi-Indicator Approach for Risk Characterization" *IWA WaterMicro: 15th Int. Symposium on Healthrelated Water Microbiology*, Naxos Island, Greece.

Reischer, G.H., J.M. Haider, N. Pokorny, A.M. de Roda Husman, S. Wuertz, B. Pradhan, C. Ferguson, D. Byamukama, D.W. Mushi, G. Ko, J. Ebdon, J. Aström, T. Coakley, R. Poma, W. Ahmed, V. Vrajmasu, E.M. Toth, M.A. Schade, A.R. Blanch, A.H. Farnleitner, 2009, "Global scale investigation of source- specificity and -sensitivity of microbial source tracking qPCR assays targeting intestinal *Bacteroidetes". IWA WaterMicro: 15th Int. Symposium on Health-related Water Microbiology*, Naxos Island, Greece.

Brion, G.M., L.E. Black, S.J. Freitas, and T. Coakley, 2007, "Modeling Enteric Virus

Presence in the Kentucky River". International IWA WaterMicro 2007 Conference, Tokyo, Japan.

Coakley, T., C. Groves, L. Elliott, and S. Wright, 2001, "Microbial Ecology of Conduit Stream Sediment Interstitial Fluids of the South Central Kentucky Karst Aquifer". National Meeting of the Geological Society of America, Boston, Massachusetts.

Coakley, T., C. Groves, L. Elliott, and S. Wright, 2001, "Reasons Why We Should Be Mindful of Microbes When We Consider Karst Systems: Impacts on Karst Development". 13th National Cave and Karst Management Symposium, Tucson, Arizona.

Coakley, T., C. Groves, L. Elliott, and S. Wright, 2002, "Microbial Ecology of Conduit Stream Sediment Interstitial Fluids of the South Central Kentucky Karst Aquifer: Impacts on Aquifer Development". Institute of Karst Geology, Chinese Academy of Geological Sciences, Guilin, China.

Elliott, L., S. Wright, C. Groves, and T. Coakley, 2000, Microbial Ecology of Conduit Stream Interstitial Fluids of the South Central Kentucky Karst Aquifer and Impacts on Aquifer Development, *Proceedings of the Eighth Mammoth Cave Science Conference, Mammoth Cave, Kentucky*, pp. 57-60.