FATTY ACID METHYL ESTER PROFILING OF ***Enterococcus*** and ***Esherichia coli*** FOR MICROBIAL SOURCE TRACKING

By

Deniz YURTSEVER

Thesis

Submitted to Department of Civil and Environmental Engineering

College of Engineering

Villanova University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

In

Water Resources and Environmental Engineering

September, 2007

Villanova, Pennsylvania

FATTY ACID METHYL ESTER PROFILING OF ***Enterococcus*** and ***E. coli*** FOR MICROBIAL SOURCE TRACKING

By

Deniz YURTSEVER

Approved: _____
Metin Duran, Ph.D.
Associate professor, Department of Civil and Environmental Engineering

Approved: _____
Andrea Welker, Ph.D.
Associate professor, Department of Civil and Environmental Engineering

Approved: _____
Ronald A. Chadderton, Ph.D., P.E.
Chairman of Department of Civil and Environmental Engineering

Approved: _____
Gary Gabriele, Ph.D.
Dean of College of Engineering

ACKNOWLEDGEMENTS

This M.Sc. thesis marks the end of my education at Villanova University as a graduate student which was the result of a long process and during which I have gratefully received the help from many people. First and foremost, I owe a heartfelt thanks and appreciation to my supervisor, Dr. Metin Duran for his great guidance and endless support as I pursued my degree. Besides leading me into the field, he has become my example in many aspects with his essential nature and character. I would also like to thank Jamie R. Lefkowitz and Timur Dunaev for their assistance in collecting and analyzing the samples.

This research is sponsored in parts by grants from Pennsylvania Department of Environmental Protection (Grant MU050144) and US Geological Survey through Pennsylvania Water Resources Research Center (Grant 2885-VU-USDI-0099). The author also acknowledges the input of Anne Murphy and Craig E. Marleton.

Last but not least, my deepest love to MY FAMILY for their infinite support. They provided me the encouragement throughout the entire of my life. Thank you for all your continuous tolerance and understanding.

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

AFLP: Amplified Fragment Length Polymorphism

APHA: American Public Health Association

APEC: Alkylophenol Polyethoxy Carboxylates

ARA: Antibiotic Resistance Analysis

ARCC: Average Rate of Correct Classification

CWA: Clean Water Act

CUP: Carbon Source Utilization Profiling

DA: Discriminant Analysis

DGGE: Denaturing Gradient Gel Electrophoresis

DNA: Deoxyribonucleic Acid

ECL: Equivalent Chain Lenghts

EDTA: Ethylene Diaminetetraacetic Acid

EIA: Esculin Iron Agar

FAME: Fatty acid methyl ester

FC: Fecal coliform

FS: Fecal streptococci

GC: Gas Chromatography

HAAs: Haloacetic Acids

LH-PCR: Length Heterogeneity Polymerase Chain Reaction

MAR: Multiple Antibiotic Resistance

MS and AS: Maximum or Average Similarity

MST: Microbial source tracking

NDC: Naphthalene Dicarboxylate

NN: Nearest Neighbor

NPDES: National Pollutant Discharge Elimination System

NTA: Nitrilotriacetic Acid

PCA: Principal Component Analysis

PFGE: Pulsed Field Gel Electrophoresis

RCC: Rate of Correct Classification

REP-PCR: Repetitive-element Polymerase Chain Reaction

RNA: Ribosomal Deoxyribonucleic Acid

TC: Total coliform

T-RFLP: Terminal Restriction Fragment Length Polymorphism

TMDL: Total Maximum Daily Load

USEPA: United States Environmental Protection Agency

WHO: World Health Organization

CHAPTER I

INTRODUCTION

Fecal contamination is a serious problem that affects the coastal waters, freshwater bodies, wetlands, and drinking water supplies worldwide. In 2000, the latest National Water Quality Inventory reported that approximately 40% of streams, 45% of lakes, and 51% of estuaries were not clean enough to support recreational uses such as fishing and swimming (USEPA, 2002a). The microbial maintenance of waters that have primary and secondary contact with humans is imperative, as contamination can create serious health risks. Presence of pathogens not only poses potential high risks to human health, but also it causes significant economic losses in industries that depend on water quality (Rabinovici et al., 2004).

Currently, total coliforms (TC), fecal coliforms (FC), *Esherichia coli*, and enterococci are bacterial indicators used in water quality and health risk assessments (USEPA 1986, USEPA 2002b, WHO 2004). Fecal coliforms are a subset of the total coliforms and include *Klebsiella, Citrobacter, Enterobacter,* and *Escherichia*. These organisms are gram-negative aerobic or facultatively anaerobic rods and are distinguished by the ability to produce gas upon lactose fermentation. Enterococci are gram-positive facultatively aerobic cocci. *Enterococci* are taxonomically separate from the genus streptococci, and include at least eighteen species (Holt, 1994). Enterococci are good indicators in marine environments due to the ability to survive in 6.5% salt (Holt, 1994). Each group of bacteria is normally prevalent in the intestines of warm-blooded mammals, including

wildlife, livestock, and humans (USEPA, 2005). The indicator bacteria themselves are usually not pathogenic. Indicator bacteria such as fecal coliforms, enterococci and *E. coli* are used because they are much easier and less costly to detect and enumerate than the pathogens themselves.

Indicator bacteria are enumerated using either the membrane filter technique or the multiple-tube fermentation test (APHA, 1998). However, these techniques fail to identify the source of the pollution. In the early 1990's, technologies collectively referred to as microbial source tracking (MST) were introduced which were meant to enable researchers to pinpoint sources of fecal pollution and even calculate loads from the sources. The application of MST technologies is essential for successful Total Maximum Daily Load (TMDL) determination for fecally contaminated waters. There are several methods currently available for MST related projects. These include molecular (genotypic) methods, biochemical (phenotypic) and chemical methods. Genetic fingerprinting is currently restricted by practical limitations of cost, speed, and ease of analysis. Similarly, the methods based on antibiotic resistance patterns, a biochemical method, are time consuming, and elaborate. Thus, there is a need to improve the existing MST techniques and to develop new MST methods that are reliable and easily applicable (Griffith et al., 2003). In arecent research, it has been shown that fatty acid methyl ester (FAME) profiling, an economical, relatively simple and easily applicable method, can differentiate indicator organisms according to their primary hosts (Haznedaroglu, 2005). Haznedaroglu (2005) showed that the host-specific differences of whole-cell FAME profiles of indicator organisms, *i.e.* TC, FC, and *Escherichia coli* may be used to predict the sources of fecal contamination.

The first objective of this study was to build a library of approximately 500 known source isolates of whole-cell FAME profiles of *Enterococcus* isolates from 6 possible sources of fecal pollution, namely: sewage; feces of livestock including bovine (dairy cattle), poultry, and swine; and feces of wildlife, including waterfowl and deer. Secondly, as this is a continuing study enlarging the existing library for *E. coli* up to 600 isolates. In addition, the host-specific differences of whole-cell FAME profiles of indicator organisms, for TC, FC, and *E. coli,* and *Enterococcus* were compared. Lastly, a field study has been conducted to determine the sources of fecal contamination in Goose Creek which is located in Chester County, PA. FAME profiles of isolates from creek samples were developed and compared against the known source library with the intention of predicting sources of fecal pollution in Goose Creek.

\

CHAPTER II

LITERATURE REVIEW

2.1. Water Quality

In recent decades, the recognition that national waters are impaired with respect to health or conservation issues have led to the establishment of numerous governmental regulations. The Clean Water Act (CWA) of 1972 sets guidelines for TMDLs in national waters. The Act calls for the accountability of states, territories, and tribes in identifying the presence and level of pollution in their waters. The main goals outlined were to establish all waterways as "fishable and swimmable" by 1983, and to remove all contaminant discharge by 1985 (Cassady, 2004). These goals have not yet been met in all waterways, and current research and practices are working to achieve a national standard. Protection of these waters includes the assessment of several pollutants, including temperature, sediment, nutrients, organics, metals, and pathogens (Simpson et al., 2002). The Safe Drinking Water Act of 1974 established legal limits based on human health issues caused by specific water contaminants (USEPA, 2006). The World Health Organization (WHO) has established guidelines for irrigation practices to limit the concentration of microorganisms allowed in reused wastewater (Blumenthal et al., 2000). The CWA of 1972 also led to the development of the National Pollutant Discharge Elimination System (NPDES), which requires permits for discharge of point sources,

such as sewage, industrial outlets, and landfills (Cassady, 2004). The overall goal of these agencies is to provide clean water for all citizens.

TMDL is formally defined as a calculation of the maximum amount of a pollutant that a water body can receive and still meet water quality standards, and allocation of that amount to the pollutant sources (USEPA, 2003a). Section 303 of the CWA establishes TMDL programs and requires each state to create their own water quality standards. Through this legislation, states also identify and list impaired waters every two years (USEPA, 2003b), and prioritize and perform TMDL studies for those waters listed. Loadings are based on evaluations of point source estimates, non-point source estimates, natural background loading, a margin of safety, and seasonal variations (USEPA, 2003c). According to the National Section 303(d) List, pathogens are the primary cause of water quality impairment nationwide, with 5084 TMDL developed due to presence of pathogens between 1996 and 2007 (USEPA, 2007a). Seventy-seven of these TMDL programs are in the state of Pennsylvania; one of the most significantly affected states (USEPA, 2007b).

The EPA has the responsibility of recommending water quality criteria to states; however, they are free to adapt the standards. Recent changes have been made to bacterial standard recommendations that now advise no more than 126 *E. coli* per 100 mL or 33 *Enterococcus* per 100 mL in freshwater based on a geometric mean of at least five samples taken within a 30-day period. Marine water recommendations call for no more than 35 *Enterococcus* per 100 mL as a geometric mean over a 30-day period (USEPA, 2003a).

MST has been the leading source of progress being made to develop more meaningful and useful TMDL information for remediation efforts and for addressing human health risks associated with microbial pollution.

2.1.1. *E. coli* and *Enterococcus* as indicators of pollution

While numerous studies have shown that the presence of thermotolerant coliforms generally indicates that fecal contamination has occurred, their presence in water does not always imply a health hazard. Also, the absence of fecal coliforms does not necessarily indicate that the water is safe for consumption, in particular from enteric viruses and pathogenic protozoans (Gleeson and Gray, 1997). As early as 1972, the EPA recognized the limitations of the fecal coliform standards and initiated a 10-year project (Cabelli et al., 1982) to develop more reliable recreational water quality standards. The introductory phase of this investigation was to develop more specific tests for different groups of bacteria such as *E. coli*, enterococci, and *Clostridium perfringens*. These organisms were presumed to be superior to coliform bacteria as indicators of fecal contamination. The development of these test methods was a prerequisite for completing a comprehensive water monitoring phase of a research study. The research study was implemented to evaluate which of nine different microbial water quality indicators in marine waters would best predict incidences of diarrheal diseases among swimmers. The results showed that of all the microorganisms measured, only the concentrations of enterococci in marine waters correlated positively with incidences of diarrheal disease among swimmers (Cabelli, 1983).

Fecal streptococci (FS) and enterococci, which are gram positive bacteria, have received widespread acceptance as useful indictors of microbiological water quality

because; i) they show a high and close relationship with health hazards, mainly gastrointestinal symptoms, associated with bathing in aquatic environments; ii) they are not as ubiquitous as coliforms; iii) they are always present in feces of warm-blooded animals; iv) they are unable to multiply in sewage-contaminated waters; and v) their die off is slower than that of coliforms in water, and persistence patterns are similar to those of potential waterborne pathogenic bacteria (Richardson et al., 1991). FS comprise species of different sanitary significance and survival characteristics. In addition, the numbers of species of this group are not the same in animal and human feces (Godfree et al., 1997; Pourcher et al., 1991; Wheeler et al., 2002). Therefore a clearer definition of FS is necessary to establish a specific standard methodology of enumeration. The taxonomy of this group has been subject to extensive revision (Holt et al., 1993). According to Borrego and Figueras (1999) the following species of the *Enterococcus* and *Streptococcus* genera may be included in the FS group: *Enterococcus faecalis, Enterococcus faecium, Enterococcus durans, Enterococcus hirae, Enterococcus avium, Enteroccoccus gallinarum, Enterococcus cecorum, Streptococcus bovis, Streptococcus equines, Streptococcus alactoyticus, Streptococcus intestinalis, Streptococcus hyointestinalis,* and *Streptococcus acidominimus.*

Although FS may not prove to be an ideal indicator in all circumstances, the use of streptococci may be advised in several situations. For instance, it is beneficial when assessing the quality of reservoirs, drinking water, and other waters in which viral contamination is particularly undesirable (Clausen et. al, 1977).

2.2. Microbial Source Tracking: Current Methods

There are three basic groups of MST methods: molecular, biochemical, and chemical. Molecular (genotypic methods) are commonly referred to as DNA fingerprinting and are based on the unique genetic makeup of different strains, or subspecies, of fecal bacteria. Biochemical (phenotypic) methods are based on an effect of an organism's genes that actively produce a biochemical substance. The type and quantity of these substances produced is what is actually measured. Chemical methods are based on finding chemical compounds such as detergents and optical brighteners that are associated with human wastewaters (Hagedorn, 2000). MST studies may use one of several methods to differentiate between potential sources of fecal contamination, all of which follow a common sequence of analysis. First, a differentiable characteristic, or fingerprint (such as antibiotic resistance patterns or DNA), must be selected to identify various strains of bacteria. A representative library of bacterial strains and their fingerprints must then be generated from the human and animal sources that may impact the water body. Indicator bacteria fingerprints from the polluted water body are then compared to those in the library and assigned to the appropriate source category based on fingerprint similarity.

Several statistical methods, including discriminant analysis (DA) (Carson et al., 2001; Harwood et al., 2000; Wiggins 1996; Whitlock et al., 2002), maximum or average similarity (MS and AS) (Carson et al., 2003; Dombek et al., 2000) and principal component analysis (PCA) have been used to classify sources. These statistical approaches differ with respect to distributional assumptions, measures of distance or similarity, and strategies for prediction. Many approaches, such as DA and AS, take into

account the central tendency and variability of each source group as a whole. Other methods, such as nearest neighbor (NN) and MS predict source membership based on similarity to an individual isolate within each source. Consequently, the predictive abilities of the various statistical methods for each of the library methods are assessed by calculating the percentage of correctly classified isolates, called Average Rate of Correct Classification (ARCC) and the obtained ARCC values may differ depending on the method used.

2.2.1. Molecular methods

2.2.1.1 Ribotyping

Ribotyping generates a molecular fingerprint based on genomic 16S rRNA restriction fragment length polymorphisms (Stoeckel, 2004). This method involves matching the "fingerprint" patterns from known sources of feces in the library to patterns of water isolates. In this case, the size of the library is very important as is the statistical analysis used to develop the fingerprint. Samadpour et al. (2002) demonstrated that choosing restriction enzymes to develop a fingerprint is critical; in fact, double enzyme analyses should always be performed, as they are more accurate than a single enzyme digestion. The ARCC of this method has been reported as 67% for human source and 97% for nonhuman source in one study and 73% overall in another (Carson et al., 2003; Parveen et al., 1999).

2.2.1.2. Pulse-field gel electrophoresis

Pulse-field gel electrophoresis (PFGE) is similar to ribotyping in that banding patterns are analyzed after restriction digestion, except that instead of just 16 rRNA, the whole DNA genome is restricted (Kariuki et al., 1999). After digestion, DNA is imbedded into specialized electrophoresis gels and electrophoresed for an extended period of time with alternating currents from different directions using specialized equipment (Kariuki et al., 1999). Tynkkynen et al. (1999) compared ribotyping and PFGE for typing two strains of *Lactobacillus* and found that PFGE was the most discriminatory method. In a comparative study of MST methods only PFGE correctly classified all isolates into the correct species-level category (Meays et al., 2004). As with ribotyping adding more enzymes increases accuracy of classification. However, PFGE is time-consuming and the number of isolates that can be analyzed simultaneously is limited.

2.2.1.3. Denaturing-gradient gel electrophoresis

Denaturing gradient gel electrophoresis (DGGE) is another electrophoretic technique that separates polymerase chain reaction (PCR) products of similar size that differ in base sequence based on changes in electrophoretic mobility, influenced by the melting properties of the DNA fragments (Meays et al., 2004). Farnleitner et al. (2000) showed that DGGE could detect and differentiate *E. coli* populations from freshwater samples polluted with fecal matter, but did not use DGGE to differentiate between sources. However, Buchan et al. (2001) used DGGE to differentiate environmental *E. coli* isolates from three host sources (bovine, human, and poultry), but were unable to identify the

source of contamination in the watershed studied. High levels of genetic diversity in environmental isolates make DGGE unreliable for use in the field, although it is useful for differentiating the strain level of bacteria (Buchan et al., 2001).

2.2.1.4. Repetitive polymerase chain reaction

Repetitive extragenic palindromic polymerase chain reaction (REP-PCR) uses conserved sequences in bacterial repetitive elements as PCR primers to distinguish among different strains of the same bacterial species (Simpson et al., 2002). REP-PCR has been used to examine fecal bacteria strains isolated from different sources of fecal pollution (Field, 2004). REP-PCR has not performed as well as ribotyping and PFGE in microbial source tracking studies and has similar limitations since it is library-dependent (Field, 2004).

2.2.1.5. Amplified fragment length polymorphism

Amplified fragment length polymorphism (AFLP) uses PCR amplification to generate between 50 and 100 DNA fragments that are commonly analyzed by DNA sequencers containing fluorescence-based detectors (Simpson et al., 2002). Using DNA sequencers allows for automation, so that over one hundred strains can be analyzed per day. Based on ARCC, Guan et al. (2002) found AFLP to be the most effective of three methods tested-ARA, host-specific *E. coli* 16S rRNA, and ALFP- when discriminating among *E. coli* isolates from animal and human sources. However, the ARCC statistic does not predict a method's ability to classify isolates from outside the library, but instead is inflated for small libraries. This method requires a large isolate library to rely

on the accuracy of the results. Otherwise, the high accuracy of discriminating host sources could be due to a small library that is not diverse.

2.2.1.6. Host-specific 16S rRNA genetic markers

The host-specific 16S rRNA genetic markers technique distinguishes members of mixtures of bacterial gene sequences by detecting differences in the number of base pairs in a particular gene fragment (Bernhard and Field 2000a, Bernard and Field 2000b). Length heterogeneity PCR (LH-PCR) separates PCR products for host specific genetic markers based on length of amplicons (Bernhard and Field 2000a). LH-PCR can quickly provide a profile of amplicon diversity in complex mixtures of PCR products (Suzuki et al., 1998). Terminal restriction fragment length polymorphism analysis (T-RFLP) uses restriction enzymes on PCR amplicons to determine unique size fragments among fluorescently labeled terminal end fragments (Bernhard and Field 2000a). LH-PCR and T-RFLP analyze differences in the lengths of gene fragments due to insertions and deletions to estimate the relative abundance of each fragment (Bernhard and Field 2000a). This method helps to decrease some of the problems associated with the under sampling of diversity in a microbial community and the uncertainty of bias due to the reannealing kinetics in the cloning process by PCR (Suzuki et al. 1998). Bernhard and Field (2000a) developed 16S rRNA markers that were based on fecal anaerobes, *Bacteroides* and *Bifidobacterium*, to distinguish human and cow fecal pollution. Strict anaerobes were chosen because they are restricted to warm-blooded animals, make up a large portion of the fecal bacteria, and do not survive long once deposited in waters. *Bacteroides* and *Bifidobacterium* have had limited use as indicators of fecal pollution

because they are difficult to grow in culture media. The use of molecular methods versus culture-based methods improved the ability for their use in water quality monitoring (Bernhard and Field 2000a). Bernhard and Field (2000a) found that the *Bacteroides-Prevotella* group was a better indicator than the *Bifidobacterium* species due to the ease of detection and longer survival in water. The authors also tested their approach on feces from human, sewage and cattle sources and found their method was successful in being able to distinguish sources. Recently, another study developed and tested primers specific for pigs, geese, and also developing other groups including seals, gulls and chicken (Shanks et al., 2006).

2.2.2. Biochemical methods

The biochemical, or phenotypic, methods have been in use for many years due to developing technology and an understanding of physical interactions. These methods examine the physical attributes, biochemical products, and chemical requirements of microorganisms (Leung et al., 2004; USEPA, 2005). These techniques generally require less training for laboratory personnel, cost less per isolate, and typically can be performed on hundreds of isolates per week. These methods under investigation for use in MST include the fecal coliform/fecal streptocci ratio, F+ RNA coliphage analysis, Carbon Utilization Profiling (CUP), ARA, and FAME Profiling.

2.2.2.1. The Fecal coliform/fecal streptocci ratio

The fecal coliform/fecal streptococci ratio was proposed as a source tracking method as far back as 1969. Prior to more modern MST techniques it was the most widely

accepted and used as a means of differentiating between human and non-human sources. It is based on the premise that human feces have higher levels of fecal coliform counts while animals have higher levels of FS. A ratio of greater than 4 would thus indicate human pollution while a ratio of less than 0.7 would indicate non-human pollution (Scott et al., 2002). The method was, however, proven unreliable due to variable survival rates of fecal streptococci and differences in fecal *Enterococcus* densities found in individuals with different diets (Simpson et al., 2002).

2.2.2.2. F+ RNA coliphage analysis

Like the fecal coliform/fecal streptococci ratio, the F+RNA coliphage method only allows for the differentiation of human and non-human pollution. The F-specific RNA coliphage has four main subgroups (I-IV). Groups II and III have been shown to be highly associated with human fecal contamination and/or domestic sewage, group IV are associated with animal and livestock waste, and group I has been found in all types of wastes (Hager, 2001). As a virus, its numbers are much lower in the environment than its bacterial counterpart *E. coli* so detection methods are very sensitive and efforts to isolate the F+ RNA coliphage has revealed that only a small percentage of human fecal samples contain these phages (Havelaar et al., 1990).

2.2.2. 3. Carbon utilization profiling

CUP is used to differentiate sources of bacteria based on bacteria's use of a wide range of carbon and nitrogen sources for energy and growth. This method has worked well in the laboratory environment, but many environmental factors impact bacterial

nutrient requirements, which may make this method impractical for field determination. BIOLOG is a commonly used system for performing CUP studies. It rapidly scores and tabulates a 96-carbon source utilization test per isolate, which generates a nutritional profile that can be used to build a known source library (Hagedorn, 2003). A study performed by Hagedorn et al. (2003) involved the identification of 375 *E. coli* isolates down to species and the creation of nutritional patterns based on 30 of the 96 nutrient wells provided by the BIOLOG system. A two way human and non-human split gave and ARCC of 92.7%, a three way source split gave an ARCC of 85.7%, and a four way source split gave an ARCC of 81.9%. The results were based on a modest *Enterococcus* library and a preliminary field validation test, and the authors did suggest that this method has potential as a phenotypic MST method.

2.2.2.4. Antibiotic resistance analysis

ARA was developed as a method for source tracking based on the assumption that bacteria from different hosts exposed to antibiotics will develop resistance to those antibiotics, and that this selective pressure would be a mechanism for discriminating among fecal bacteria from various hosts. Antibiotics are used to prevent and treat infections in humans and domestic animals and to increase growth rates in animal production. Bacteria resistant to antibiotics used in animal feed (Bryan et al., 2004) have been found in poultry litter (Kelley et al., 1998), cattle feces (Dargatz et al., 2003), and in swine manure (Smalla et al., 2000). Throughout the literature, different combinations of antibiotics and concentrations (range in μg/ml) have been used for antibiotic resistance tests. There is currently no standard suite of antibiotics or concentrations used for

antibiotic resistance testing. Antibiotics are best chosen after determining potential animal fecal sources and antibiotics used in their treatment. Furthermore, the antibiotics chosen must be appropriate to the source identifier utilized, i.e., *E. coli* and other fecal coliforms are intrinsically resistant to vancomycin; therefore, its use with this class of source identifier is not informative. This method has been used extensively because it is rapid, relatively simple, and relatively inexpensive. Furthermore, it requires less technical expertise than molecular methods and no specialized equipment.

There are three approaches that have been used in MST studies; ARA, multiple antibiotic resistance (MAR), and Kirby-Bauer antibiotic susceptibility. In MAR studies, bacteria are tested for resistance to different antibiotics (Parveen et al., 1999). ARA differs slightly by including different concentrations of each antibiotic being tested (Wiggins, 1996; Wiggins et al., 1999). The Kirby-Bauer antibiotic susceptibility test has been a standard method for use in clinical studies and uses small filter disks that have been impregnated with antibiotics. The zone of growth inhibition around the disks is used to quantify resistance. Some MST researchers believe that ARA provides the most information of the three antibiotic-based approaches. A potential problem when using antibiotic resistance as a phenotypic source tracking method is the transfer of resistance genes between bacteria. Genes conferring antibiotic resistance have been found on a variety of mobile genetic elements including plasmids, transposons, and conjugative transposons that provide a means for lateral transfer of the genes (Bass et al., 1999; Kruse et al., 1994; Ohlsen et al., 2003; Salyers et al., 1995; Smalla et al., 2000;). Although indigenous bacteria have the potential to transfer antibiotic resistance genes to fecal bacteria after bacteria from fecal sources enter the environment, this would have to occur

at very high frequency to affect the overall proportion of resistant cells in the fecal host population. Even if gene transfer frequencies were as high as 1%, which is much higher than has been reported (Smalla et al., 2000) their detection will be unlikely with current antibiotic resistance protocols unless there is extensive regrowth of the recipients in the environment.

Among the different antibiotic resistance approaches available, ARA is the most common methods in MST studies (Booth et al., 2003; Choi et al., 2003; Graves et al., 2002; Harwood et al., 2000; Harwood et al., 2003; Wiggins, 1996; Wiggins et al., 1999; Wiggins et al., 2003; Whitlock et al., 2002), and has been utilized in many TMDL studies (Harwood et al., 2000). Regardless of the specific method, they all first require cultivation of the target organism and *E. coli*, enterococci, and FS have been tested with this method (Harwood et al., 2000; Parveen et al., 1997; Wiggins, 1996). However, the large library requirement is a disadvantage for ARA. The magnitude of the minimum size for a "representative" library of AR profiles to encompass the diversity has been estimated to be at least 2,300 microbial isolates (Wiggins et al., 2003).

2.2.3. Chemical methods

Chemical targets studied for their applicability to source tracking projects are limited at this time. These targets can only be indicative of the presence or absence of contamination sources, with most of the research focusing on chemicals associated with human wastes. Chemical targets include caffeine, fragrance agents, fluorescent whitening agents, and fecal sterols. These compounds are associated with human wastewaters so their presence would indicate human pollution.

Detecting the presence of caffeine has been used to indicate human fecal pollution. Several methods have been used to detect the presence of caffeine such as liquid and gas chromatography (Bendriss et al.,2000; Weigel et al., 2001) or capillary electrophoresis (Horie et al., 2000). While caffeine detection has been successfully used to identify human fecal pollution, the method only detects human fecal pollution. Nonpoint sources, like agricultural runoff or wildlife, cannot be determined by the monitoring caffeine.

Fecal steroids, most notably coprostanol (5-cholestan-3-ol), have been also used to detect fecal contamination (Hatcher et al., 1979; Noblet et al., 2004). Often the detection of fecal steroids is used to complement fecal coliform counting methods (Noblet et al, 2004). Coprostanol makes a good indicator because it can be quantitatively related to the amount of sewage-derived organic matter (Hatcher et al., 1979). Due to their poor water solubility, fecal steroids are often deposited in sediments, which can limit the use of this method in deep bodies of water such as lakes and oceans. Fecal sterols are found in the feces of higher animals, making the determination of the host source impossible from this method alone. Additionally, pathogens would not necessarily be correlated with these chemicals.

Another method used in the past to detect traces of human fecal contamination is by identifying wastewater indicator compounds such as ethylene diaminetetraacetic acid (EDTA), nitrilotriacetic acid (NTA), a naphthalene dicarboxylate (NDC) isomer, alkylophenol polyethoxy carboxylates (APECs), and haloacetic acids (HAAs). These compounds found in sewage effluent can be identified and quantified by gas chromatography/mass spectrometry (Standley et al., 2000). While this method has proven

effective in identifying human fecal contamination from wastewater, it does not detect nonpoint sources of contamination, making its application of limited value.

CHAPTER III

FATTY ACID METHYL ESTER (FAME) PROFILING

3.1. Fatty Acid Nomenclature and Main Structural Features

Fatty acids consist of a carboxylic group connected to a carbon chain (Figure 1). The carbon chain may be saturated or unsaturated, and may contain carbon branches as well as other functional groups. The majority of fatty acids in nature have unbranched carbon chains with 4–24 carbons, 0–6 double bonds, and no other functional groups in most organisms (Kakela et al., 1995; Rezanka et al., 2002).

Several types of fatty acid nomenclature are common, and naming of fatty acids in the literature may vary with what is convenient. Unbranched fatty acids are described by the number of carbons followed by the number of double bonds. Thus, the saturated fatty acid may be shown either ‘C16:0’ or ‘16:0’. Double bond positions may be described from either end of the molecule. Double bond positions given from the methyl end of the carbon chain are commonly referred to by ‘ω’ or by ‘n–’. The monounsaturated fatty acid may be shown as ‘20:1 (ω9)’ or ‘20:1 n–9’, the latter is usually preferred in chemical literature (Anon, 1978). Alternatively, the double bond position may be specified by the distance from the carbonyl group as ‘Δ11-20:1’ or as ‘11-20:1’. The terms ‘cis’ and ‘trans’ are commonly applied to describe the geometries of double bonds in fatty acids instead of ‘Z’ (zusammen) and ‘E’ (entgegen), which is more

common in general organic chemistry. The geometries are often described by a single letter ‘c’ or‘t’, which is combined with the double bond position.

Even though the structures described above cover the majority of common fatty acids, there are a large number of less common fatty acids with various structures. The carbon chain may contain triple bonds, branches, as well as saturated and unsaturated carbon rings. Oxygen may be introduced in the carbon chain in the form of hydroxy groups, oxo groups, furan rings or additional carboxyl groups (Brondz, 2002). Other heteroatoms, e.g. halogens may also be present (Mu et al., 1997). Due to variations in number of carbon atoms and number, location, and position of double bonds, there is an enormous variety of fatty acids that lipids contain. This results in many bacteria having unique fatty acid profiles.

The majority of fatty acids are esterified to larger lipid molecules and therefore only small amounts are present in free form in living organisms as well as food matrices. Lipid molecules are traditionally classified into neutral and polar lipids. Common neutral lipids are triacylglycerols, used as energy reserve in most organisms; wax esters, the energy reserve in certain marine species; cholesteryl esters and free fatty acids. In most tissues, the majority of polar lipids are phospholipids from cell membranes.

Fatty acids are mainly analysed by gas chromatography as their corresponding fatty acid methyl esters (FAME). The preparation of FAME involves extraction of the lipid molecules from the sample matrix, breaking of the ester bonds, and formation of methyl esters. The two last steps may be combined by trans-esterifying the lipids directly with acid or base in methanolic solution. The details of the extraction process are described in Chapter 4.

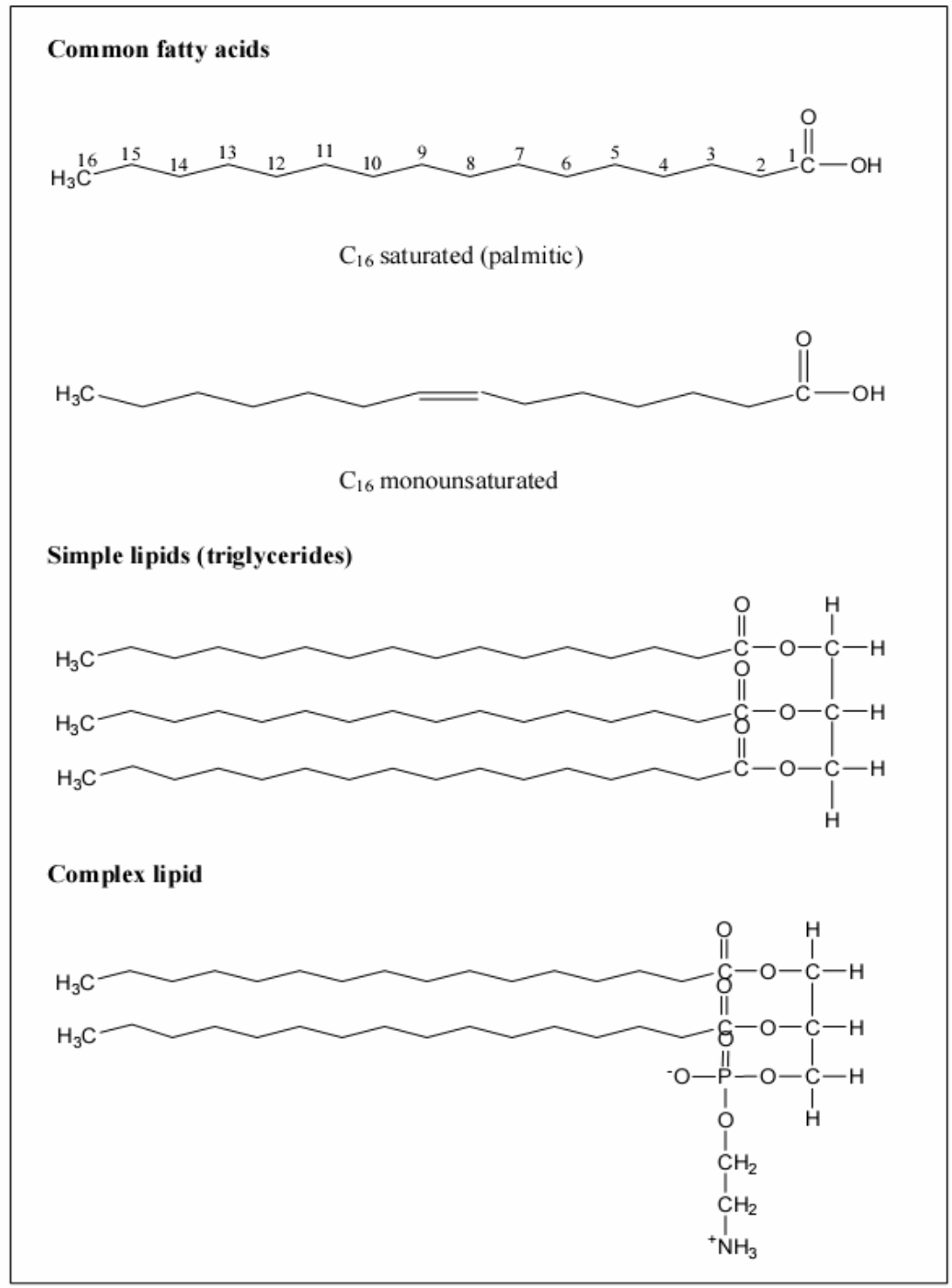

Figure 1. Chemical structures of fatty acids, simple and complex lipids (Adopted from Madigan et al, 2002).

3.2. Identification of FAMEs by Gas Chromatography

3.2.1. Principles of separation

Modern gas chromatography (GC) is typically based on open tubular capillary columns varying in length from 10–100 m and with internal diameters from 0.1– 0.5 mm. The principle of separation in open tubular gas chromatography is explained in Figure 2 (Etre, 1993). The retention of a compound is determined by its distribution between the stationary phase and the mobile phase. The distribution can be expressed by the retention factor, k :

$$k = \frac{\text{amount of solute in stationary phase}}{\text{amount of solute in mobile phase}}$$

k depends on the solubility of the solute in the stationary phase (k increases with increased solubility), the thickness of the stationary phase (k increases with increased thickness), column diameter (k decreases with increased diameter), and the temperature (k decreases with increased temperature).

Retention as described in Figure 2 and by the equation above is an idealized model, where it is assumed that solutes in the stationary phase behave as ideal solutions. Deviations from ideal conditions may be caused by surface effects between the stationary phase and the carrier gas, uneven distribution and composition of the stationary phase, adsorption of the analytes, displacement effects and interactions between analytes. Deviations from ideal conditions may be especially large for analytes with low volatility and low solubility in the stationary phase or when the capacity of the stationary phase is overloaded (Etre, 1993).

It should also be emphasized that the proportions shown in Figure 2 do not correspond with dimensions in real capillary columns. In modern columns the internal diameter is typically 1000 times larger than the thickness of the stationary phase. The solutes also elute in broad bands.

The term 'carrier gas velocity' in GC normally refers to the average gas velocity in the column, which is the length of the column divided by the elution time for an unretained component. Because of the high compressibility of the carrier gas and large pressure drop in the column the actual carrier gas velocity is higher in the end than at the head of the column (Etre, 1993).

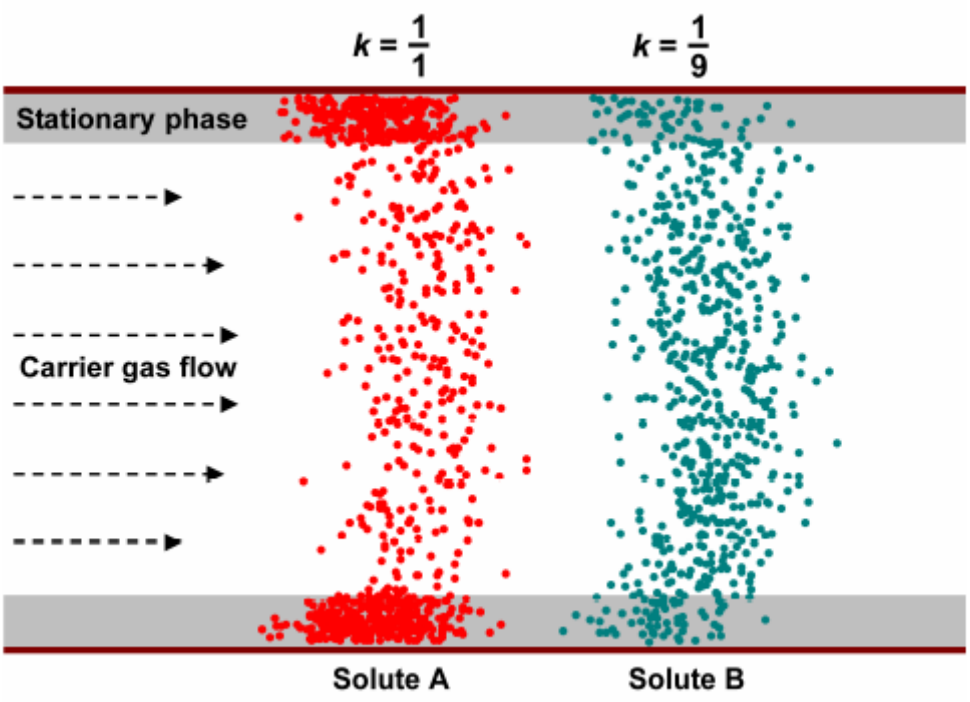

Figure 2. Principle of separation in gas chromatography (The solutes A and B are separated because they differ in retention factor.) (Adopted from Etre, 1993).

3.2.2. Identification of fatty acids from retention times and ECL-values

The most successful approach may be the use of equivalent chain lengths (ECL) for fatty acid analysis. The ECL system is based on saturated unbranched FAMEs as reference compounds and ECL values for the references are by definition equal to the number of carbons in the alkyl chain. ECL values for all of the peaks in the calibration mix permits the software to automatically calculate a "nominal retention time" for each peak. The calibration mixture is composed of compounds having the same general chromatographic properties. The saturated fatty acids are assigned an ECL value corresponding to their length (e.g. 11:0 = ECL 11.000) (Miwa et al., 1960).

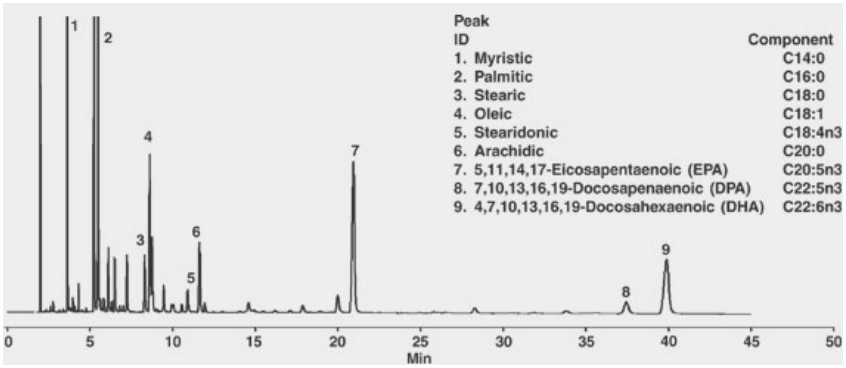

Figure 3. Elution pattern of a FAME mixture (Adopted from http://www.sigmaaldrich.com/Graphics/Supelco/objects/8100/8046.pdf)

3. 3. Microbial Source Tracking Applications

Extraction and quantification of fatty acids from cellular lipids reveal information regarding identity and physiology of microbial species. Species-specific FAME profiles are routinely used for in environmental and clinical microbiology. Since a FAME profile is 'fingerprint' of each species, it can be used for microbial source tracking, *i.e.*, determining the source of microbial contamination in surface and groundwater. Due to different environmental and physiological conditions they are exposed to, indicator microorganisms from different sources may have quantifiably different FAME profiles

allowing source tracking. Application of FAME technology to MST has been cited in various studies (Booth et al., 2003; Briganti and Wacker, 1995; Glucksman et al., 2000; Parveen et al., 2001; Stewart and Stoeckel 2004; Seurinck et al., 2006; USEPA 2004b). However, there are only few published works on the ability and accuracy of FAME profiling to predict sources of fecal pollution, one by Parveen et al. (2001) and several recent studies by Haznedaroglu et al. (2005), Duran et al. (2006) and Seurinck et al. (2006).

Parveen et al. (2001) reported that 104 *E. coli* isolates from human and non-human sources contained neither a signature FAME nor statistically significant differences in the FAME profiles (Parveen et al., 2001). In opposition, Seurinck et al. (2006) and Haznedoroglu et al. (2005) showed that FAME profiling can be useful tool for microbial source tracking. Although Seurinck et al. (2006) used a relatively low number of *E. coli* isolates (only 238) for a known source library, the study showed 61% overall correct classification for five sources (raw sewage, cow, dog, gull, horse) classification tree and 75% correct classification for two sources (raw sewage, animal) classification tree.

Another comprehensive study was carried out by Haznedaroglu et al. (2005) with different indicator organisms. The ARCC for total coliforms under 6-way classification analysis was 74%, whereas the results under 3-way and 2-way classification analyses were 81% and 92%. Also, Duran et al. (2006) showed that when fecal coliform was used for FAME profiling the ARCC under 2-way classification scenario was 100%. Three-way and 6- way discriminant analysis of fecal coliform resulted in 95% ARCC and 86% ARCC, respectively (Duran et al., 2006).

CHAPTER IV

MATERIALS AND METHODS

4.1. Sample Collection and Preparation

For human host category, multiple samples were collected from five different domestic wastewater treatment plants, Phoenixville, Bridgeport, Oaks and Ocean County Wastewater Treatment Plants and Valley Forge Sewer Authority Treatment Plant, located in Pennsylvania and New Jersey. Raw influent samples were collected with sterile media bottles and transferred to the laboratory for processing. For the livestock host category, bovine, poultry, and swine stool samples were collected from animals randomly selected at several private dairy, poultry and swine farms within a 40 miles radius around Villanova University. The wildlife category included waterfowl and deer. Fecal samples from geese and ducks were collected around a well-protected natural pond at a public park where there was a large waterfowl population. Deer stool samples were picked from a national park 20 miles northwest of Villanova University's campus. For the sewage samples, several 10-fold dilutions were prepared using sterile Milli-Q® water. For the fecal samples, 10 grams of fecal material were suspended with 90 ml of sterile phosphate buffer, pH 7.35. The samples were shaken vigorously at room temperature, 23° C, for 30-45 minutes. Then, several 10-fold dilutions were prepared with sterile Milli-Q® water.

4.2. Culturing Indicator Organisms

Enterococci were cultured according to a slightly modified form of EPA Method 1106.1 (USEPA, 2002). All samples were filtered onto 0.45 μm membrane filters (Millipore) and transferred to 47 mm Petri dishes containing selective and differential media, m-Enterococcus Agar (Becton, Dickinson and Company, Sparks, MD). This method uses Esculin Iron Agar (EIA) as the selective media. However higher recovery was observed on m-Enterococcus Agar. The plates were incubated at 37 ºC for 48 hours. The filters were then transferred to EIA plates and incubated at 37 ºC for 20-30 minutes. Colonies that developed a black shadow were considered enterococci and were streak plated on 100 mm Petri dishes with m-Enterococcus agar for further analysis.

E. coli bacteria were isolated according to the EPA Method 1103.1 (USEPA, 2002). Membranes containing filtered samples were transferred onto 47 mm Petri dishes with selective and differential media, m-TEC® agar (Becton, Dickinson and Company, Sparks, MD). After incubation for 2 hours at 35±0.5 ºC and then 22 hours at 44.5±0.2 ºC, membranes were placed onto an absorbent pad saturated with Urea Substrate Medium and allowed to sit for 15-20 min. Then isolates with yellow, yellow-green, or yellow-brown color were streak plated on 100 mm Petri dishes with m-TEC agar and incubated for harvesting.

4.3. Fatty Acid Extraction and Quantification

Approximately 40 mg (wet-weight) of cells were harvested from the streaked plates and transferred into sterile culture tubes for whole-cell fatty acid extraction. The cells were saponified with 1 mL of 3.50 N NaOH in 50% methanol in water (v/v). Then culture tubes were capped, vortexed for 10 s, incubated at 100 ºC in a water bath for 5 min, vortexed for another 10 s, incubated again at 100 ºC for an additional 25 min, and cooled to room temperature. The saponification step lysed the cells and converted the released fatty acids into their sodium salts. Fatty acids were then esterified with 2 mL of 3.25 M HCl in 46% (v/v) methanol. The tubes were vortexed for 5 s, placed in a water bath at 80 ºC for 10 min and cooled to room temperature immediately to minimize degradation of the FAMEs. Then the FAMEs were extracted using 1.25 mL 1:1 (v/v) solution of methyl-tert-butyl ether and hexane. After capping, the culture tubes were rotated end-over-end continuously for 10 min on a laboratory rotator and the aqueous (lower) phase was removed using a sterile glass Pasteur pipette. Three mL of 0.27 N NaOH in saturated NaCl solution was added to the tubes to wash the free fatty acids and remaining cellular debris and then the tubes were rotated again continuously for 5 min. Two-thirds of the organic (upper) phase was transferred with a glass Pasteur pipette into a 2-ml glass gas chromatography vial, which was sealed with a Teflon®-lined cap. FAMEs were quantified using an Agilent 6890 gas chromatograph (Agilent Technologies, Wilmington, DE) equipped with flame ionization detector and Ultra 2 capillary column. The stationary phase of this column was cross linked to the silica tube, which provides low noise and drift during temperature programmed runs. The temperature program was a

ramp from 170 ºC to 270 ºC at 5 ºC per minute, then a sudden increase to 300 ºC for 2 minutes to clear the column of extraneous material, and decrease the possibility of carryover. The inlet temperature was 250 ºC and it was kept at constant pressure, 9.0 psig. This setting gave an initial hydrogen carrier flow rate of approximately 0.4 mL per minute. The flame ionization detector temperature was 280 ºC. The make up gas was ultra-high-purity nitrogen set at a constant flow of 30 mL per minute. Ultra-high-purity hydrogen and air flows to the detector were 40 and 350 mL per minute respectively.

The individual FAMEs were identified by Sherlock Version 4.5® pattern recognition software (Microbial ID, Newark, DE), based on their ECL values. Calibration standards were analyzed prior to each batch of analysis and between every 11 samples.

4.4. Statistical Analysis

FAME profiles of the isolates were constructed from the relative masses of individual FAMEs extracted from each isolate. The relative masses were calculated by dividing the area response of a particular FAME by total area for all of the FAMEs identified from that particular isolate. Using relative masses eliminated the possible errors that might have been introduced from variations in harvested cell masses.

Classification of each isolate into appropriate host categories was tested using discriminant analysis (DA). Discriminant analysis was performed using prior probabilities, linear discriminant function, and cross validation. Cross validation is a technique used to compensate for an optimistic apparent error rate defined as the percent of misclassified observations. The cross-validation routine works by omitting each

observation one at a time, recalculating the classification function using the remaining data, and then classifying the omitted observation (Haznedaroglu, 2005). The statistical tests were conducted using the software package MINITAB® (Minitab Inc., State College, PA).

4.5. Quality Assurance/Quality Control

4.5.1. Glassware/laboratory apparatus

All glassware and supplementary apparatus were cleaned and dried prior to every use to lessen or eliminate any chance of cross contamination. Cleaning procedures included hand washing in hot water/Alconox® (Alconox Inc., New York, NY) solution, rinsing with tap water, rinsing three times with distilled water, and a final rinse with Milli-Q® water. Glassware that washed in the mechanical dishwasher was rinsed with tap and distilled water automatically. After the end of the cycle, glassware was rinsed a final time with de-ionized water. All glassware was dried in a 250 ºC oven for a minimum of two hours. Single use glassware, such as Pasteur pipettes were sterilized using the same oven conditions, 250 ºC for a minimum of two hours. The glassware coming into contact with fatty acids was dried in 450 ºC oven for a minimum of four hours.

Inoculation loops used to transfer bacterial cultures were flame sterilized between each use. In addition, these loops were washed as described above for glassware, on a weekly or bi-weekly basis to insure cleanliness, sterility, and proper mechanical function. Glass bacterial culture spreading rods were handled in a manner similar to that used for

the sample loops, with the additional step of submersing them in isopropyl alcohol prior to flame sterilization.

Sample containers were single use, disposable, pre-sterilized, 8 oz. polyethylene cups with lids purchased from Fisher Scientific (Fisher Scientific International Inc., Pittsburg, PA). Petri dishes were sterile, disposable, 100x150 millimeter plastic dishes, purchased from Spectrum Chemical Inc. (Spectrum Chemicals & Laboratory Products, (Gardena, CA), 47 millimeter Petri dishes purchased from Fisher Scientific.

4.5.2. Solution and agar preparation

All solutions and the selective and differential growth media were prepared according to manufacturers' directions. Phosphate buffer was prepared according to the following formula: potassium phosphate monobasic 8.5 g, potassium phosphate dibasic 21.7 g, sodium phosphate dibasic 17.7 g, ammonium chloride 1.7 g, and de-ionized water 950 mL. pH was between 7.2 and 7.4, checked for each batch with a calibrated pH meter. Buffer and media that require autoclaving were processed through a Genetige, Castle 122 gravity steam sterilizer that received preventative maintenance and calibration checks at regular, three month intervals.

4.5.3. Sample collection

Fecal samples were collected in the field, immediately after the host animal defecated. Sample material was scooped directly into the sample container from the uppermost part of the stool; only material that did not touch the ground or some part of the external environmental was used. Sampling in this manner removed the possibility of

environmental contamination during the sample collection. Gloves worn by the person taking the sample were changed to a new pair between two consecutive samplings.

Multiple hosts were sampled from each category to investigate the possibility of variation in bacterial populations due to environmental conditions or differences in host age or gender. All samples were transported to the laboratory within two hours of collection and immediately processed to minimize the possibility of changes in the bacterial population, or chemical composition, due to factors external to the host.

4.5.4. Fatty acid extraction and quantification

Cellular fatty acids were extracted and converted to their methyl esters according to recent literature and a procedure developed by MIDI Inc., Newark DE. The main analytical instrument for this study was a Agilent 6890 gas chromatograph (Agilent Technologies, Wilmington, DE), connected to a personal computer running Agilent ChemStation Ver. D.00.01 (Agilent Technologies, Wilmington, DE), and MIDI Sherlock® 4.5 Microbial Identification software packages (MIDI, Inc., Newark, DE).

4.5.5. Data collection, analysis and storage

Quality control for collecting raw data was excellent, due to safeguards built in to the Sherlock® 4.5 Microbial Identification software. The software configured the instrument to run two replicates of a sixteen component calibration solution, which must meet or exceed set values for several quality specific parameters, before the instrument was allowed to run experimental samples. The calibration solution was then run again after every eleven samples, as a continuing check on the quality of the instrumental analysis.

In addition to these software controls, a hexane/MTBE blank was run every six samples to provide further assurance the instrument was providing valid raw data, and there was no carryover between samples.

Data was collected and stored on personal computers, in two hard drive locations, raw data and back up raw data. The Villanova University mainframe computer was used as a separate location for back up files. Additionally, files were copied to 3.5” diskettes after every batch to provide an extra layer of security.

CHAPTER V

RESULTS AND DISCUSSION

5.1. *Escherichia coli*

The known source library for E. coli was constructed with 605 isolates cultured from 78 samples of six possible host categories; 180 isolates from 15 sewage samples, 85 from 17 dairy cow samples, 98 from 17 chicken samples, 76 from 15 swine samples, 94 from 7 deer samples, and 72 from 7 waterfowl samples. A total of 12 FAMEs identified from these isolates are shown in Table 1 with their systematic names and structural properties.

Table 1. Structural properties of fatty acids identified in *E. coli* isolates

Symbol	**Common name**	**Systematic name**	**Structure**
C10:0	Capric	Decanoic	
C12:0	Lauric	Dodecanoic	
C14:0	Myristic	Tetradecanoic	Saturated
C16:0	Palmitic	Hexadecanoic	
C18:0	Stearic	Octadecanoic	
C16:1 ω5c		*cis*-11-Hexadecenoic	
C16:1 ω7c	Palmitoleic	*cis*-9-Hexadecenoic	Unsaturated
C18:1 ω7c	Vaccenic	*cis*-11- Octadecenoic	
C16:0 ISO I		14-methlypentadecenoic	Branched
C19:ISO		17-methlyoctadecanoic	chain
C14:0 3OH	β-Hydroxymyristic	3-Hydroxytetradecanoic	Hydroxy
C17:0 cyc		Cyclo-heptadecanoic	Cyclopropane
C19:0 cyc ω8c		9-(2-eptylcyclopropyl) Nonanoic	

Table 2 represents the mean relative masses and the standard deviations of the 12 FAMEs along with their relative occurrences. The relative occurrence of each FAME is defined as the ratio of the isolates containing that FAME to the total number of isolates within each host category. Because current chromatographic techniques are not able to draw apart 16:1 ISO I from 14:0 3OH, their combination is designated as "Sum in Feature 2" in Sherlock nomenclature. As shown in Table 2, excluding 17-methyl octadecanoic acid, 19:0 ISO, none of the FAMEs identified was exclusively associated with a particular host group. 19:0 ISO was distinguished only in isolates from dairy cow, poultry, and swine fecal samples. However; it was seen in very low concentrations, less than 1%, and was not observed in the majority of the livestock fecal samples. Consequently, considering it as a signature FAME might be misleading, although it was only seen in livestock samples.

The average relative masses and standard deviations are plotted in Figure 4. Among 12 of the identified FAMEs; 19:0 cyclo w8c, 18:1 w7c, 17:0 cyclo, 16:0, 16:1 w7c, 16:1 ISO/14:0 3OH, 14:0, 12:0 had significantly higher relative masses when compared to remaining FAMEs and also they were detected in all isolates. Whereas, 10:0, 19:0 ISO, 18:0, 16:1 ω5c were found in quite low percentage of the isolates with less relative mass ratios. None of the FAMEs identified has shown significant differences in the mean relative masses for each host category.

Table 2. Mean relative masses and relative occurrences of FAMEs within each host category

	Host					
FAME	**Human**	**Bovine**	**Poultry**	**Swine**	**Waterfowl**	**Deer**
10:0	0.02±0.04 (15%)[a]	0.02±0.04 (16%)	0.01± 0.02 (6%)	0±0	0±0.01 (1%)	0.01±0.03 (11%)
12:0	4.88±1.42[b]	6.01 ±4.85	4.79±1.02	6.09±1.63	5.34±0.72	5.17±5.31
14:0	10.24±2.10	9.07±2.10	9.92±1.42	11.16±2.00	10.68±1.41	9.04±1.76
16:1 ISO/14:0 3OH	12.29±4.22	13.61±6.54	11.81±3.73	15.34±5.25	13.03±2.04	11.19±2.72
16:1 ω7c	4.30±1.85	3.47±2.76	2.85±1.45	3.12±1.08	4.33±1.68	3.40±1.31
16:1 ω5c	0.02±0.05 (20%)	0.01±0.06 (8%)	0.02±0.04 (18%)	0.01±0.03 (6%)	0±0	0.01±0.03 (3%)
16:0	41.11±4.23	39.56±4.72	41.26±2.46	37.76±4.33	40.98±2.22	42.35±2.83
17:0 CYCLO	18.70±2.25	20.35±3.53	19.95±2.23	18.77±2.73	18.44±2.30	20.65±1.83
18:1 ω7c	3.59±1.43	2.34±0.95 (99%)	2.47±0.91	2.49±1.02 (99%)	2.96±1.27	2.80±1.11
18:0	0.48±0.29 (79%)	0.25±0.24 (56%)	0.45±0.21 (88%)	0.18 ± 0.25(38%)	0.37 ± 0.29 (67%)	0.46 ± 0.24 (83%)
19:0 CYCLO ω8c	4.42±1.21	5.74±3.38	6.59±2.71	4.94±2.54	3.80±1.09	5.46±1.41
19:0 ISO	0±0	0.01±0.04 (12%)	0.02±0.05 (12%)	0.01±0.04 (4%)	0±0	0±0

a: Numbers in parenthesis indicate relative occurrence, *i.e.* per cent of isolates within a particular host group that carries the particular FAME. Lack of parenthesis means that all of the isolates in that host category carried that particular FAME

b: Mean relative mass ± standard deviation

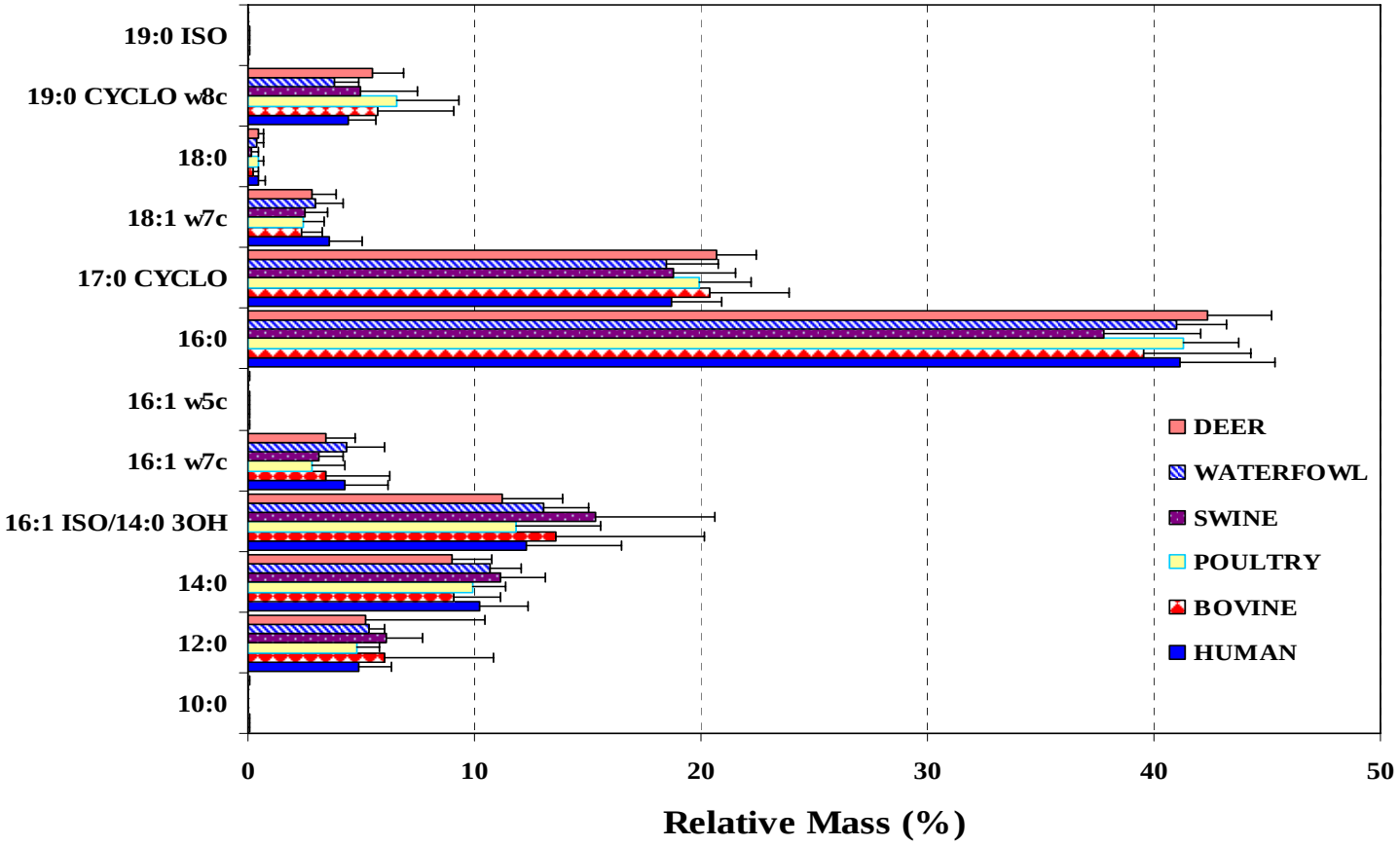

Figure 4. Average relative masses of FAMEs in *E. coli* isolates (Stems represent standard deviation)

The discrimination through the FAME profiles of isolates was investigated using discriminant analysis. For this purpose, different discriminant analysis scenarios were performed and results are given in Table 3 through Table 10. Initially, all the isolates were classified into six different host categories, a six-way classification scenario. Table 3 shows the percentage of the isolates correctly placed in the corresponding host group in a six-way discrimination. The rate of correct classification (RCC) ranged from 42% for deer isolates to 58% for waterfowl isolates. The RCC for human category was 47% as a result of classifying 84 isolates as human out of 180 sewage isolates. It should be mentioned that RCC results are based on false negatives and do not include false positives. Overall, out of 605 isolates 288 were correctly assigned to their respective hosts giving 48% ARCC. Although it seems a low classification accuracy, it is comparable to the other phenotypic MST studies in current literature. For example, Guan et al. (2002) studied the MAR patterns of 319 *E. coli* and reported 33.9% ARRC for multiple antibiotic resistance (MAR) patterns of 319 *E. coli* isolates under nine-way classification (human, beef, dairy, chicken, pig, turkey, deer, geese, and moose). In another study, Stoeckel et al. (2004) reported 27% accuracy of antibiotic resistance analysis (ARA) under eight-way classification (humans, cattle, chickens, dogs, horses, swine, Canada geese, and white-tailed deer) when a set of 150 isolates from known-sources were classified using a known-source library of 900 *E. coli* isolates. The authors reported 13% ARCC for carbon utilization profiling (CUP), under the same classification scheme, when 150 isolates collected from the known sources were classified by a discriminant function based on CUP of 630 *E. coli* isolates from known sources (Stoeckel et al., 2004).

Table 3. 6-way DA analysis of *E. coli* isolates into respective host categories

Assigned Host	True Host					
	Human	Bovine	Poultry	Swine	Waterfowl	Deer
Human	**84**	6	13	4	13	15
Bovine	7	**42**	8	10	0	12
Poultry	13	7	**46**	11	3	13
Swine	26	7	9	**35**	3	11
Waterfowl	25	4	12	8	**42**	4
Deer	25	19	10	8	11	**39**
Total N	180	85	98	76	72	94
N correct	84	42	16	35	42	39
RCC	**0.47**	**0.49**	**0.47**	**0.46**	**0.58**	**0.42**
N= 605	**N Correct= 288**					**ARCC= 0.48**

Another discrimination scenario given in Table 4 is a 4-way classification where sewage isolates are discriminated against livestock isolates. The ARCC was 59%, slightly higher than that in a 6-way classification.

Table 4. Discrimination of *E. coli* isolates from sewage against bovine, poultry and swine

Assigned Host	True Host			
	Human	Bovine	Poultry	Swine
Human	**110**	9	24	6
Bovine	14	**54**	7	11
Poultry	24	10	**58**	22
Swine	32	12	9	**6**
Total N	180	85	98	76
N correct	110	54	58	37
RCC	**0.61**	**0.64**	**0.59**	**0.49**
N= **439**	N Correct= **259**			ARCC= **0.59**

The third scenario was performed for discrimination of sewage isolates against waterfowl and deer. This 3-way scenario resulted in 66% ARCC as indicated in the Table 5.

Table 5. Discrimination of *E. coli* isolates from sewage against waterfowl and deer

Assigned Host	True Host		
	Human	Waterfowl	Deer
Human	**106**	11	18
Waterfowl	42	**55**	9
Deer	32	6	**67**
Total N	180	72	94
N correct	106	55	67
RCC	**0.71**	**0.76**	**0.71**
N= **346**	N Correct= **228**		ARCC= **0.66**

Also with the purpose of evaluating the effects of grouping similar host categories into larger groups; bovine, poultry and swine samples were pooled as livestock whereas waterfowl and deer samples were pooled as wildlife and they were discriminated against sewage samples. Table 6 shows the 2-way DA scenario assessing sewage samples against livestock. This scenario demonstrated 73% ARCC which is more accurate than the other scenarios mentioned above as a result of grouping similar host categories into a larger group.

Table 6. Discrimination of *E. coli* isolates from sewage against livestock

Assigned Host	True Host	
	Human	Livestock
Human	**132**	72
Livestock	48	**187**
Total N	180	259
N correct	132	187
RCC	**0.73**	**0.72**
N= **439**	N Correct= **319**	ARCC= **0.73**

Likewise, DA of sewage samples against wildlife was tested which is displayed by Table 7. The ARCC value for this scenario was 70%.

Table 7. Discrimination of *E. coli* isolates from sewage against wildlife

Assigned Host	True Host	
	Human	Wildlife
Human	**118**	41
Wildlife	62	**125**
Total N	180	166
N correct	118	125
RCC	**0.66**	**0.75**
N= **346**	N Correct= **243**	ARCC= **0.70**

Furthermore, a 3-way classification was performed by sewage isolates against livestock and wildlife pooled categories. For this scenario, 61% of ARCC was observed as shown in the Table 8.

Table 8. Discrimination of *E. coli* isolates from sewage against livestock and wildlife

Assigned Host	True Host		
	Human	Livestock	Wildlife
Human	**105**	30	36
Livestock	28	**158**	22
Wildlife	47	71	**108**
Total N	180	259	166
N correct	105	158	108
RCC	**0.58**	**0.61**	**0.65**
N= **605**	N Correct= **371**		ARCC= **0.61**

Lastly, all the host categories except human were pooled together as non-human group and distinguished from human host. This 2-way DA scenario was resulted in 77% ARRC and shown by Table 9.

Table 9. Discrimination of *E. coli* isolates as human versus non-human

Assigned Host	True Host	
	Human	Non-human
Human	**132**	91
Non-human	48	**334**
Total N	180	425
N correct	132	334
RCC	**0.73**	**0.79**
N= **605**	N Correct= **466**	ARCC= **0.77**

A summary of all classification scenarios are shown in Table 10 with RCC and ARCC values.

Table 10. Summary of RCC values for seven classification scenarios tested for *E. coli* isolates

Classification scenario	RCC									ARCC
	H	B	P	S	WF	D	LS	WL	NH	(%)
H/B/P/S/WF/D (6-way)	47	49	47	46	58	42				48
H/B/P/S (4-way)	61	64	59	49						59
H/WF/D (3-way)	71				76	71				66
H/LS (2-way; pooled B+P+S)	73						72			73
H/WL (2-way; pooled WF+D)	66				75					70
H/LS/WL (3 way; pooled B+P+S and WF+D)	58						61	65		61
H/NH (2-way; pooled B+P+S+WF+D)	73								79	77

H: Human, B: Bovine, P: Poultry, S: Swine, WF: Waterfowl, D: Deer, LS: Livestock, WL: Wildlife, NH: Non-human,

The lowest ARCC value was 48% which belongs to the 6-way classification scenario. When the isolates are pooled into three larger categories, 58% of the sewage isolates classified correctly as human origin whereas the RCCs were 61% for livestock and 65% for wildlife categories for a 3-way scenario. ARCC was 61% for this classification scenario. On the other hand, discrimination of human isolates from non-human resulted in ARCC as 77%. 132 of the 180 sewage isolates were classified correctly as human with 73% RCC while 334 of the 445 isolates belonging to the remaining hosts were classified as non-human with 79% RCC. Overall ARCC was 77% for the 2-way classification scenario.

Adequate size and representativeness of the known source library was critical for establishing confidence in the results. A library is representative when the sources are

randomly assigned to the isolate profiles and RCCs are approximately equal to the probability that an isolate would be assigned to a source category by chance. Random grouping, also known as artificial clustering, is one approach to test the sufficiency of number of isolates in the known source library (Booth et al., 2003; Duran et al., 2006; Hagedorn et al., 2003; Harwood et al., 2000; Whitlock et al., Wiggins, 1996). For this purpose, 605 *E. coli* isolates were assigned to random hosts and results are shown by Table 11.

Table 11. *E. coli* library ARCCs from artificial clustering to source categories

Classification scenario	**The probability of an isolate being assigned to a source by chance**	**ARCC (%)**
6- way classification	16.7	15
3-way classification	33.3	33.1
2-way classification	50	47.6

The randomly-assigned ARCCs were lower than the probability of the isolates being assigned to a host category by chance. Therefore, the known source library of *E. coli* is large enough to prevent random grouping. These findings suggest that required library size is related to the method and it is not necessarily true that all the library-based methods require isolate numbers in thousands, as in the case of ARA.

5. 2. *Enterococcus*

The known source library for *Enterococcus* was built with 511 isolates from 76 samples. The names and the structural properties of the 18 FAMEs that were identified from these isolates are summarized in the Table 12.

Table 12. Structural properties of fatty acids identified in *Enterococcus* isolates

Symbol	**Common name**	**Systematic name**	**Structure**
C11:0		Undecanoic	
C12:0	Lauric	Dodecanoic	
C14:0	Myristic	Tetradecanoic	
C15:0	Pentadeyclic	Pentadecanoic	Saturated
C16:0	Palmitic	Hexadecanoic	
C18:0	Stearic	Octadecanoic	
C16:1 ω7c	Palmitoleic	*cis*-9-Hexadecenoic	
C18:1 ω9c	Oleic	cis-Methyl oleate	Unsaturated
C18:1 ω7c	Vaccenic	*cis*-11- Octadecenoic	
C14:1 ISO E	Myristoleic	Iso-tetradecanoic	
C16:0 ISO I		14-Methlypentadecenoic	Branched chain
C15:0 ISO		13-Methyltetradecanoic	
C18:ISO H		16-Methylheptadecanoic	
C19:ISO I		17-Methlyoctadecanoic	
C12:1 3OH	3-Hydroxylauric acid	3-Hydroxydodecenoic acid	
C14:0 2OH		2-Hydroxytetradecanoic	Hydroxy
C14:0 3OH	β-Hydroxymyristic	3-Hydroxytetradecanoic	
C17:0 cyc		Cyclo-heptadecanoic	
C19:0 cyc ω8c		9-(2-eptylcyclopropyl) Nonanoic	Cyclopropane

Table 13 demonstrates the mean relative masses and abundance of the 18 FAMEs in each host category. The results showed that none of the FAMEs identified was solely associated with a particular host group, except 11 methyl 18:1 w7c and 14:1 ISO E. 11 methyl 18:1 w7c and 14:1 ISO E were identified only in isolates from geese and swine samples. However; they were found in very low relative masses, less than 1%, and were not observed in majority of the fecal samples. Therefore, considering them as a signature FAME might be inaccurate although they were seen in particular host category. Eight of these FAMEs were also found in *E. coli* isolates. The host specific differences among *E. coli* and *Enterococcus* isolates were solely quantitative differences. That is, no signature FAME was identified and variations in relative masses of FAMEs common among different host groups were the only host-specific differences.

The mean relative masses and standard deviations of the 18 FAMEs identified in *Enterococcus* isolates within each host category are shown in Figure 5. One noticeable differences in mean relative masses of FAMEs was that saturated FAME 11:0 was significantly higher in human isolates than in non-human isolates. 18:1 ISO H was another FAME that showed noticeable difference in that it was found in fewer isolates from human and swine (68% and 84%), whereas it was found in all of the bovine, poultry, geese and deer isolates. Its mean relative mass was also lower for human (12.30±10.02) and swine samples (15.13±8.27) compared to the bovine (24.37±7.66), chicken (21.16±3.61), waterfowl (22.37±5.47) and deer (24.48±5.02) isolates. 19:0 CYCLO ω8c was seen significantly low in bovine and deer isolates when it is compared to the other host categories.

Table 13. Mean relative masses and relative occurrences of FAMEs within each host category

	Host					
FAME	**Human**	**Bovine**	**Poultry**	**Swine**	**Geese**	**Deer**
11 methyl 18:1 w7c	0±0	0±0	0±0	0±0	0.78±1.38 (27%)	0±0
11:0	2.21±2.89 (58%)	0.44±1.06 (22%)	0.13±0.43 (9%)	0.65±0.83 (41%)	1.47±2.35 (37%)	0.58±2.29 (6%)
12:0	0.28±0.34 (55.83%)	0.55±0.55 (65%)	0.08±0.13 (31%)	0.14±0.25 (37%)	0.32±0.90 (60%)	0.39±0.25 (82%)
12:1 3OH	0.26±0.44 (31%)	0±0	0.02±0.11 (5%)	0±0	0.31±1.25 (31%)	0.07±0.29 (6%)
14:1 ISO E	0±0	0±0	0±0	0.20±0.29 (36%)	0±0	0±0
14:0	16.25±7.39 (100%)	17.84±8.63 (100%)	10.66±3.31 (100%)	13.56±7.95 (100%)	11.63±3.51 (99%)	18.14±4.20 (100%)
15:0	0.11±0.32 (13%)	0.03±0.12 (7%)	0.12±0.32 (20%)	0.07±0.19 (12%)	0.01±0.06 (6%)	0.22±0.32 (35%)
14:0 2OH	0.03±0.12 (5%)	0±0	0±0	0±0	0.05±0.13 (12%)	0±0
16:1 ISO/14:0 3OH	0.06±0.22 (10%)	0±0	0.01±0.9 (3%)	0±0	0.01±0.04 (3%)	0.03±0.13 (6%)
16:1 ω7c/ 15 ISO 2 OH	16.92±5.53 (100%)	13.51±5.11 (100%)	16.01±2.99 (100%)	17.36±4.66 (100%)	12.93±1.60 (100%)	12.02±3.35 (100%)
16:0	26.38±6.06 (100%)	23.81±5.10 (100%)	21.16±3.61 (100%)	23.45±6.43 (100%)	24.16±2.58 (100%)	27.83±3.49 (100%)
17:0 CYCLO	0.71±0.74 (57%)	0.20±0.36 (26%)	0.49±0.42 (66%)	0.72±0.62 (72%)	0.45±0.52 (55%)	0.21±0.43 (24%)
18:1 ISO H	12.30±10.02 (68%)	24.37±7.66 (99%)	21.16±3.61 (100%)	15.13±8.27 (84%)	22.37±5.47 (100%)	24.48±5.02 (100%)
18:1 ω9c	0.24±0.57 (26%)	0.30±0.45 (36%)	0.49±0.42 (45%)	0.32±1.50 (9%)	0.26±0.65 (22%)	0.21±0.43 (31%)
18:1 ω7c	13.93±5.19 (100%)	13.30±5.73 (100%)	22.59±6.63 (100%)	15.98±4.53 (100%)	14.23±2.23 (100%)	11.13±3.09 (100%)
18:0	0.34±0.39 (48%)	0.12±0.24 (20%)	0.34±0.32 (57%)	0.34±0.51 (45%)	0.72±0.35 (91%)	0.47±0.13 (67%)
19:0 ISO I	1.47±1.47 (58%)	1.20±1.30 (57%)	1.91±1.50 (78%)	2.40±1.39 (86%)	0.75±1.20 (32%)	0.17±0.44 (14%)
19:0 CYCLO ω8c	8.49±.2.72 (98%)	3.80±4.01 (64%)	8.96±1.89 (100%)	9.35±4.17 (99%)	9.59±2.03 (100%)	3.96±2.26 (91%)

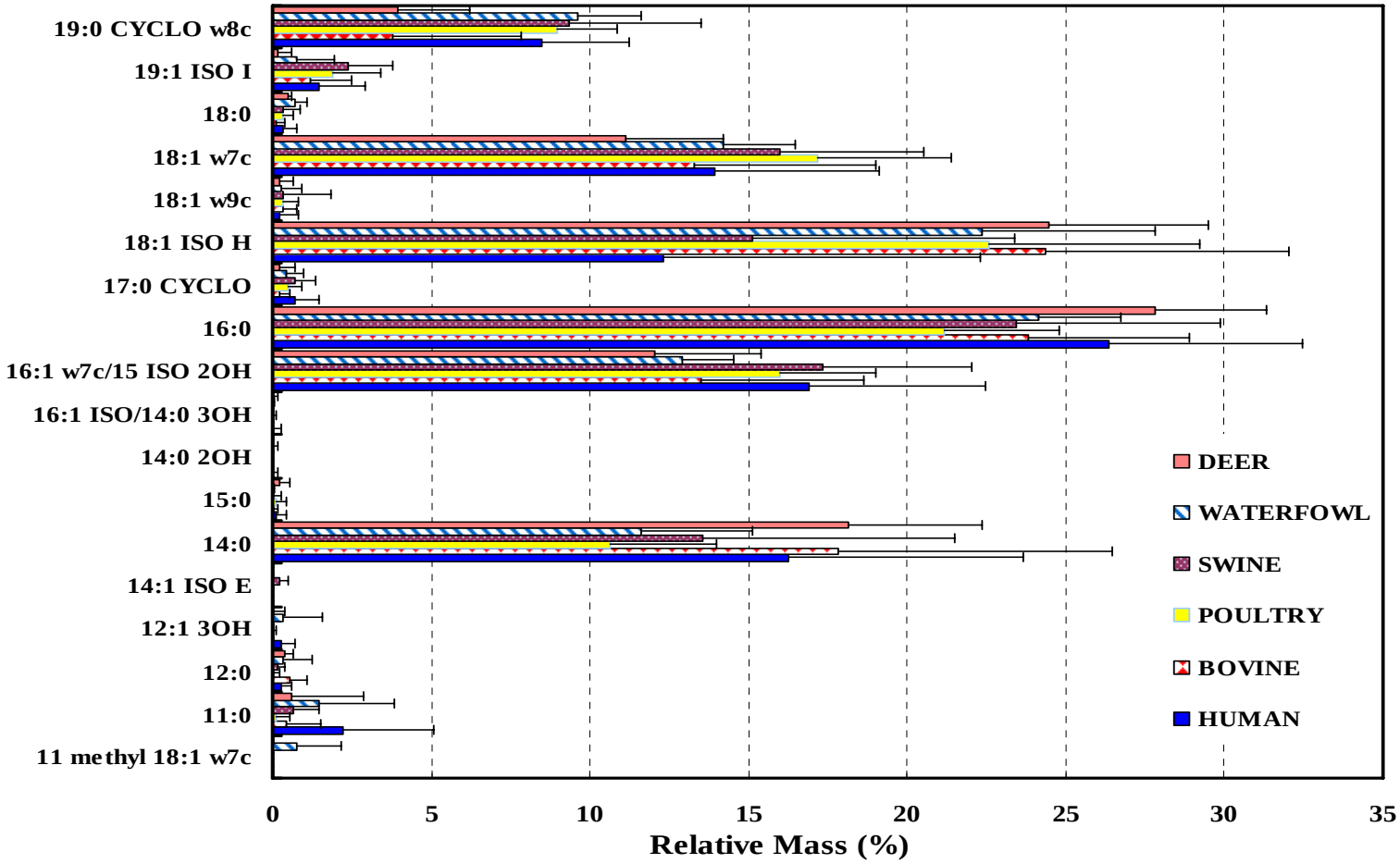

Figure 5. Average relative masses of FAMEs found in *Enterococcus* isolates (Stems represent standard deviation)

The classification analysis of *Enterococcus* isolates into their respective host categories was tested again under seven different classification scenarios similar to those for *E. coli* isolates. Table 14 shows the first scenario under 6-way discriminant analysis. The RCCs for this classification scenario varied from 43% for swine to 75% for deer isolates. Out of 120 human isolates, 78 of them were classified correctly as human, which resulted in a 65% RCC. The ARCC under this classification scenario was 66%. Again both RCC and ARCC values are based on only false negatives do not include the false positive results.

Table 14. 6-way DA analysis of *Enterococcus* isolates into respective host categories

Assigned Host	True Host					
	Human	Bovine	Poultry	Swine	Waterfowl	Deer
Human	**78**	2	4	12	9	7
Bovine	3	**40**	7	5	0	15
Poultry	25	22	**61**	17	6	0
Swine	5	0	2	**33**	0	1
Waterfowl	6	0	0	8	**56**	1
Deer	3	5	0	1	7	**70**
Total N	120	69	74	76	78	94
N correct	78	40	61	33	56	70
RCC	**0.65**	**0.58**	**0.82**	**0.43**	**0.72**	**0.75**
N= **511**	N Correct= **338**					ARCC= **0.66**

The next classification scenario demonstrates the 4-way classification where human isolates were compared to non-pooled livestock isolates which resulted in 70% ARCC. Out of 339 isolates 239 of them were correctly classified into their respective host categories. The highest RCC value was 82% for poultry. The ARCC value was slightly higher than that for the 6-way classification scenario.

Table 15. Discrimination of *Enterococcus* isolates from sewage against bovine, poultry and swine

Assigned Host	True Host			
	Human	Bovine	Poultry	Swine
Human	**88**	0	4	5
Bovine	4	**43**	7	6
Poultry	20	24	**61**	18
Swine	8	2	2	**47**
Total N	120	69	74	76
N correct	88	43	61	47
RCC	**0.73**	**0.62**	**0.82**	**0.62**
N= **339**	N Correct= **239**			ARCC= **0.70**

The third scenario was performed for discrimination of sewage isolates against waterfowl and deer isolates. This 3-way scenario resulted in 90% ARCC (Table 16). Out of 292 isolates, 262 were correctly classified into their host category.

Table 16. Discrimination of *Enteroccus* isolates from sewage against waterfowl and deer

Assigned Host	True Host		
	Human	Waterfowl	Deer
Human	**103**	8	5
Waterfowl	10	**70**	4
Deer	7	0	**85**
Total N	120	78	94
N correct	103	70	85
RCC	**0.85**	**0.90**	**0.90**
N= **292**	N Correct= **262**		ARCC= **0.90**

In the next classification scheme, the bovine, poultry and swine isolates were pooled together as livestock and compared with sewage isolates. This 2- way classification scenario resulted in 86%.

Table 17. Discrimination of *Enterococcus* isolates from sewage against livestock

Assigned Host	True Host	
	Human	Livestock
Human	**89**	18
Livestock	31	**201**
Total N	120	219
N correct	89	201
RCC	**0.74**	**0.91**
N= **339**	N Correct= **290**	ARCC= **0.86**

When sewage samples against wildlife were tested, the ARCC value for this scenario was 90% (Table 18).

Table 18. Discrimination of *Enterococcus* isolates from human against wildlife

Assigned Host	True Host	
	Human	Wildlife
Human	**107**	17
Wildlife	13	**155**
Total N	120	172
N correct	107	155
RCC	**0.89**	**0.90**
N= **292**	N Correct= **262**	ARCC= **0.90**

Another 3-way classification was tested by pooling the categories of wildlife and livestock and discriminating these two categories against human category. This scenario resulted in ARCC value of 75% (Table 19).

Table 19. Discrimination of *Enterococcus* isolates from human against livestock and wildlife

Assigned Host	True Host		
	Human	Livestock	Wildlife
Human	**80**	24	20
Livestock	30	**167**	17
Wildlife	10	28	**135**
Total N	120	219	172
N correct	80	167	135
RCC	**0.67**	**0.76**	**0.79**
N= **511**	N Correct= **382**		ARCC= **0.75**

For the last scenario, all non-human isolates from wildlife and livestock categories were pooled into one non-human category and compared to human category. Two-way classification has resulted in an exceptionally high ARCC, i.e. 80%, as shown in Table 20.

Table 20. Discrimination of *Enterococcus* isolates as human versus non-human

Assigned Host	True Host	
	Human	Non-human
Human	**82**	65
Non-human	38	**326**
Total N	120	391
N correct	82	326
RCC	**0.68**	**0.83**
N= **511**	N Correct= **382**	ARCC= **0.80**

Table 21 summarizes the RCC values for each classification scenario for individual host categories, along with ARCC values for each classification scenario tested. Highest RCC values were obtained by eliminating the livestock category from discriminant analysis. When only human and wildlife isolates were investigated in a 3-way classification scenario, the RCC varied from 85 % to 90% with a 90% ARCC. This indicates that there is a high degree of discrimination between the waterfowl and deer isolates. The significantly higher RCCs in this scenario imply that majority of misclassifications were in livestock category, particularly in discrimination of bovine and swine isolates. The fact that there was significant misclassification of bovine and swine isolates into poultry category and vice versa suggested that pooling individual hosts into larger categories may increase the accuracy of classification. Thus, when bovine, poultry, and swine isolates were pooled into a single livestock category, the accuracy of

discrimination increased significantly. The ARCC was 85% for this classification scenario and 91% of the livestock isolates were classified correctly.

Table 21. Summary of RCC and ARCC values for each classification scenario tested

Classification scenario	RCC									ARCC
	H	B	P	S	WF	D	LS	WL	NH	(%)
H/B/P/S/WF/D (6-way)	65	58	82	43	72	75				66
H/B/P/S (4-way)	73	62	82	62						70
H/WF/D (3-way)	85				90	90				90
H/LS (2-way; pooled B+P+S)	74						91			85
H/WL (2-way; pooled WF+D)	89							90		90
H/LS/WL (3 way; pooled B+P+S and WF+D)	67						76	79		75
H/NH (2-way; pooled B+P+S+WF+D)	68								83	80

H: Human, B: Bovine, P: Poultry, S: Swine, WF: Waterfowl, D: Deer, LS: Livestock, WL: Wildlife, NH: Non-human

Similar results for both ARA and CUP have been reported in the literature when the phenotypic characteristics of *Enterococcus* were used instead of that of *E. coli*. For example, Harwood et al. (2000) reported 62% accuracy of ARA in a six-way discrimination of 1,466 *Enterococcus* isolates from known sources. In another study, 1,435 *Enterococcus* isolates from six possible sources of fecal pollution were classified into their respective host categories at 74% accuracy using ARA (Wiggins, 1996). Wiggins et al. (1999) used ARA of four large sets of *Enterococcus* isolates, between 2,635 and 5,990 isolates per set, and reported ARCCs ranging from 64 to 78% in four-way discrimination scenarios.

Artificial clustering analysis was also performed for *Enterococcus* library, as already discussed in Section 5.1., to determine whether library size was adequate for satisfactory

discrimination. The results are shown by Table 22. The randomly-assigned ARCCs were lower than the probability of the isolates being assigned to a host category by chance, therefore, the known source library of *Enterococcus* is large enough to prevent random grouping.

Table 22. *Enterococcus* library ARCCs from artificial clustering to source categories

Classification scenario	The probability of an isolate being assigned to a source by chance	ARCC (%)
6- way classification	16.7	14.1
3-way classification	33.3	31.7
2-way classification	50	43.6

5.3. Comparison of FAME Profiling for Different Indicator Organisms

The effects of indicator organism type on the host specificity of FAME profiles were studied by comparing the FAME profiles of *Enterococcus* isolates to those of total coliform (TC), fecal coliform (FC), and *Escherichia coli* isolates. Four separate known-source libraries, one known-source library for each of the indicator organisms studied, included a total of 1,742 isolates cultured from six possible sources of fecal matter; sewage, cow (dairy and cattle), chicken, swine, deer, and waterfowl. The known-source library of *Enterococcus* included 511 isolates from 76 samples, while that of *E. coli* known-source library was comprised of 605 isolates from 78 samples. The TC library included 303 isolates from 48 samples and the FC known-source library size was 314

isolates from 48 samples (Duran et al., 2006; Haznedaroglu et al., 2005) and the details of each known-source library are represented in Table 23.

Table 23. Known-source library data for each indicator organism

Host / Indicator	Sewage	Bovine	Poultry	Swine	Geese	Deer	Total
Total coliform [a]	104/11[c]	30/10	29/10	53/8	55/6	32/3	303/48
Fecal coliform [b]	99/11	29/10	29/10	50/8	46/5	61/4	314/48
E. coli	180/15	85/17	98/17	76/15	72/7	94/7	605/78
Enterococcus	120/22	69/10	74/8	76/15	78/9	94/12	511/76

[a] Results from Haznedaroglu et al. (2005).
[b] Number isolates/ Number of samples.
[c] Results from Duran et al. (2006).

Table 24 shows the results of DA where differentiation of isolates based on their primary hosts was tested using a linear discriminant function. The most accurate classification of known-source isolates was achieved when FC and TC were used as indicator organisms. The 95% accuracy of discriminating FC isolates of sewage origin against those of livestock and wildlife is particularly significant. The accuracy increased to 100% when FC isolates were pooled into human versus non-human categories.

Table 24. Comparison of ARCC values for four different indicator organisms

Classification scenario	ARCC (%)			
	Total coliform [a] (n=303)	Fecal coliform [b] (n=314)	*E. coli* (n=605)	*Enterococcus* (n=511)
Six-way (H/C/P/S/WF/D)	74	86	48	66
Three-way (H/LS/WL)	81	95	61	75
Two-way (H/NH)	92	100	76	80

n: number of isolates; H: Human; C: Cow; P: Poultry; S: Swine; WF: Waterfowl; D: Deer; LS: Livestock; WL: Wildlife; NH: Non-human.
[a] Results from Haznedaroglu et al. (2005).
[b] Results from Duran et al. (2006).

5.4. Determining Sources of Fecal Contamination in the Goose Creek Using FAME Profiles of *Enterococcus*

Applying MST methods to TMDL programs or human health risk assessments is the ultimate goal and direction of source tracking efforts. Though MST is not an official part of the TMDL program, some localities that are facing major water quality issues have begun to employ source tracking.

The sources of fecal coliforms to waterways are ubiquitous. Any means by which fecal matter can be transported to the receiving waters is a potential source. These sources include combined sewer overflows, separate sanitary sewer overflows, which can result from leaky or undersized sanitary sewer pipes, direct deposits from wildlife and livestock, and stormwater runoff, which include overland flow and flow conveyed through storm sewer pipes.

5.4.1. Site Description and Sample Collection

One of the objectives of this study was to apply FAME as a tool for determining the sources of fecal pollution in the Goose Creek which is located approximately 20 miles Southwest of Villanova University in the Chester Creek Watershed area. The 67.2 square-mile Chester Creek watershed (the area that contributes runoff to Chester Creek and all its tributary streams) begins in a number of headwater streams in eastern Chester County and western and central Delaware County. The watershed is within the jurisdiction of fourteen townships, five boroughs and one city. The watershed is a mixed urban, suburban and rural land. The East branch's major tributary, Goose Creek, begins in West Chester which is shown in Figure 6 as the circled area. Goose Creek, flowing out of downtown West Chester has only fair water quality. The reason is that Goose Creek receives discharges from two large wastewater treatment plants (WWTP); The West Chester Borough (WCWWTP) and West Goshen Sewer Authority (WGWWTP). The non-point sources of pollution are the run off from roads, parking lots, lawns, farms, and rooftops in rainwater. Pennsylvania Department of Environment Protection (DEP) has a TMDL program for Chester Creek Watershed as shown by Figure 7. Due to the all these reasons, Goose Creek was chosen as the study area for assessing the efficiency of FAME profiling. Environmental samples were collected monthly from six different sites which are highlighted on a map (see Figure 8). Sample locations were chosen according to the location of treatment plants and densely populated residential areas and shown by Table 25.

Table 25. Site numbers and locations

Site Number	**Location**
Site 1	Close to Highway 202
Site 4	Below WGWWTP
Site 6	Below WCWWTP
Site 7	Close to West Bourne Middle School
Site 9	Close to major roads and residency area
Site 10	Close to Ridley State Park

Sample collection began in October 2006. The data represented here contains 101 cultured *Enterococcus* isolates from total 4 different sampling events. The FAME profiles of the isolates from the environmental samples were constructed as explained in the previous sections and then compared to FAME profiles in the known source library of *Enterococcus.*

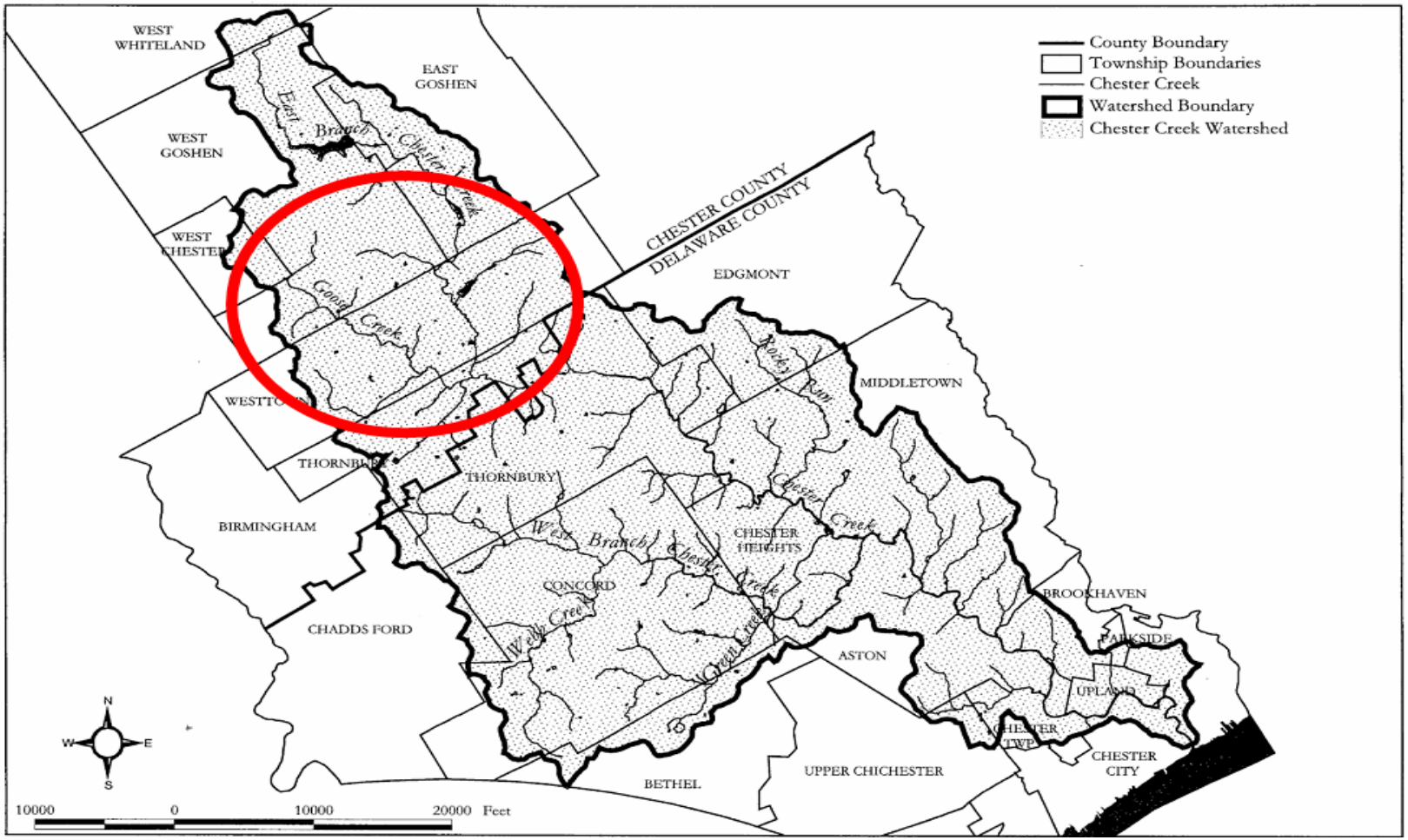

Figure 6. Chester Creek Watershed Map (Adopted from http://www.crcwatersheds.org/img/chester_map.pdf)

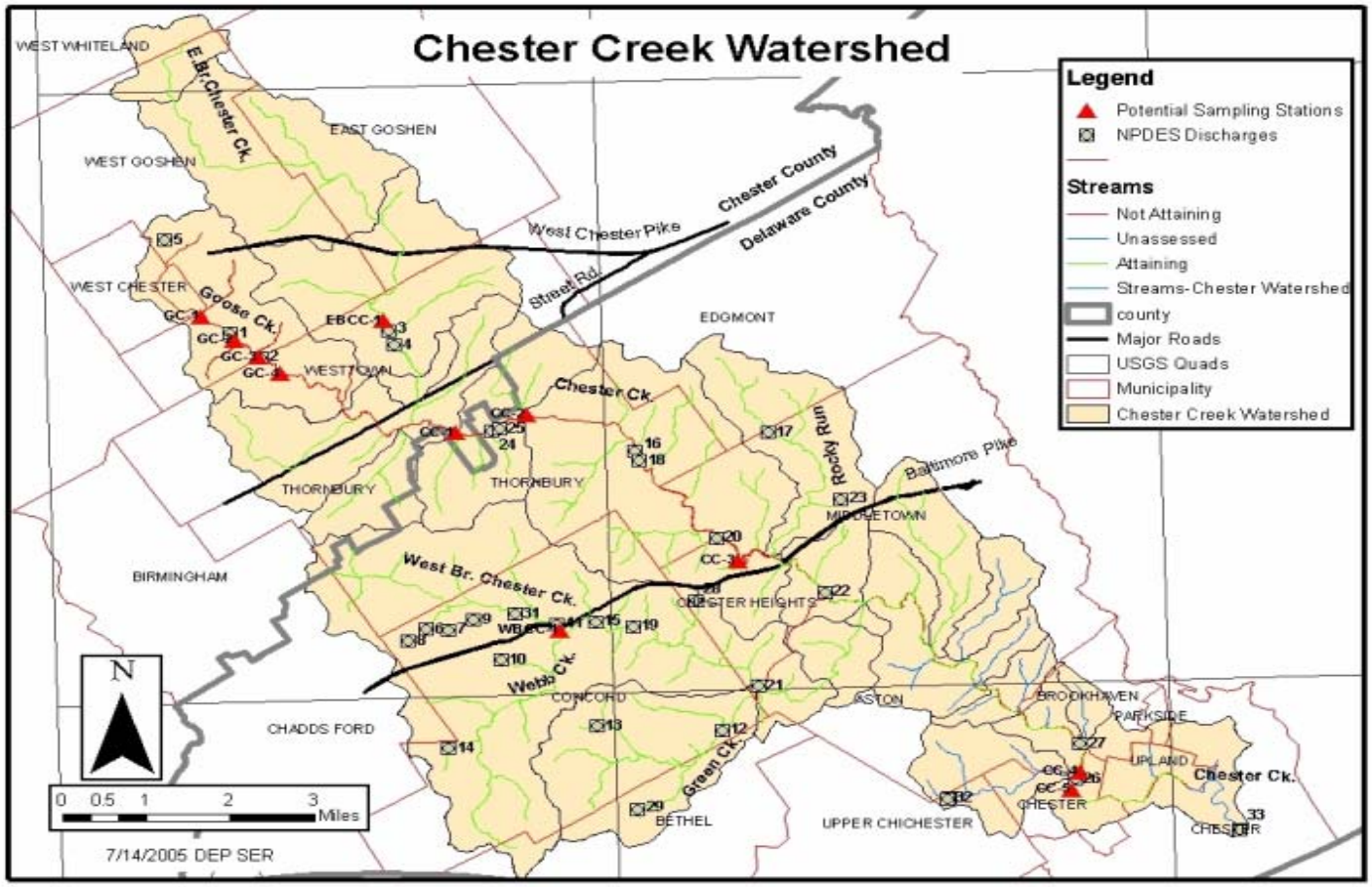

Figure 7. PA Department of Environment Protection TMDL study outline for Chester Creek Watershed.

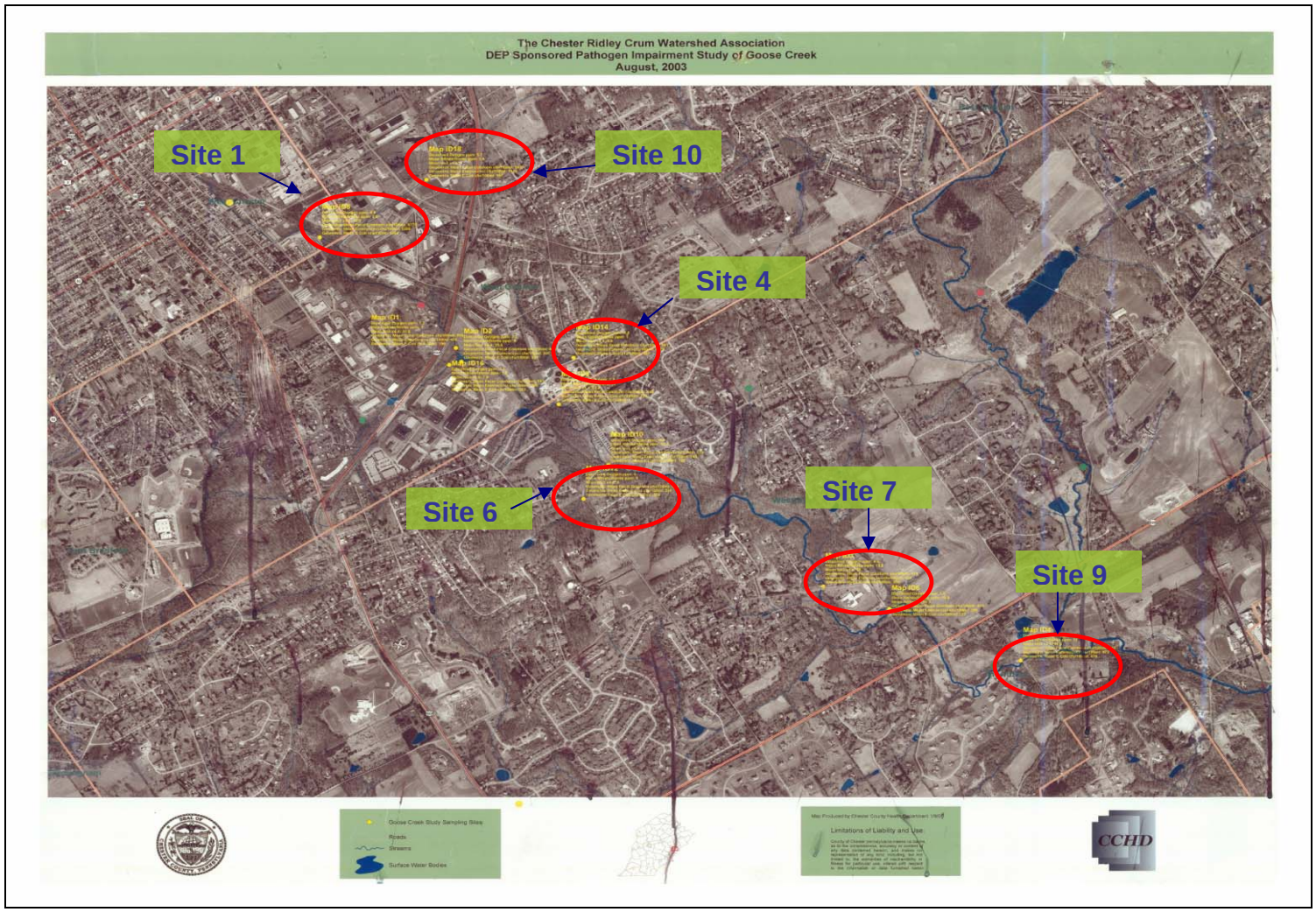

Figure 8. Goose Creek sampling sites (Adopted from http://www.crcwatersheds.org/img/chester_map.pdf).

5.4.2. Summary of Project Results Based on Source Tracking with *Enterococcus*

The relative occurrence of each host group for a particular site is given by pie charts in Figure 9 to Figure 14. The relative occurrence of the human host category was higher than other hosts in all sites except for Site 10. The reason for that is Site 10 is located close to the Ridley State Park and thus predictable pollution sources are farm animals and/or wildlife. The classification resulted in 50% bovine, 11% poultry, 11% deer and 11% swine origin, whereas 17% of the total isolates were classified as human host for Site 10. For Site 1, Site 4 and Site 6, the human host category classification was relatively higher from the other sites because sampling locations are close to discharge points from wastewater treatment plants. The other two dominated source categories were bovine and poultry for all sampling sites.

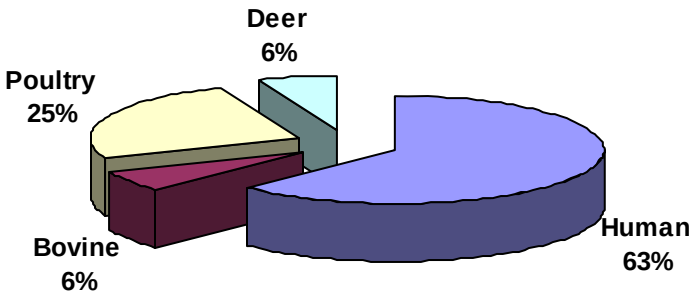

Figure 9. Relative percentage of host groups for Site 1

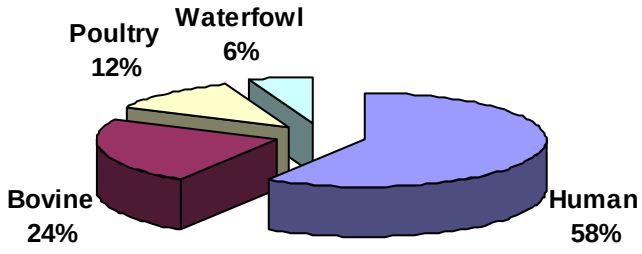

Figure 10. Relative percentage of host groups for Site 4

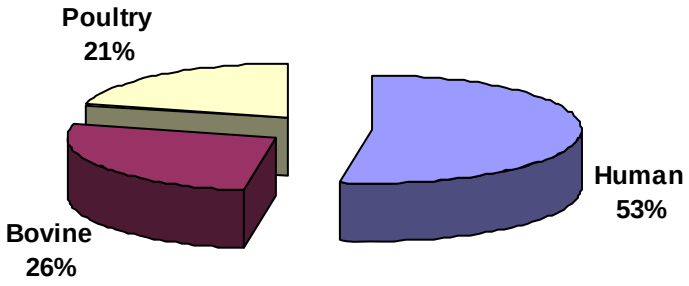

Figure 11. Relative percentage of host groups for Site 6

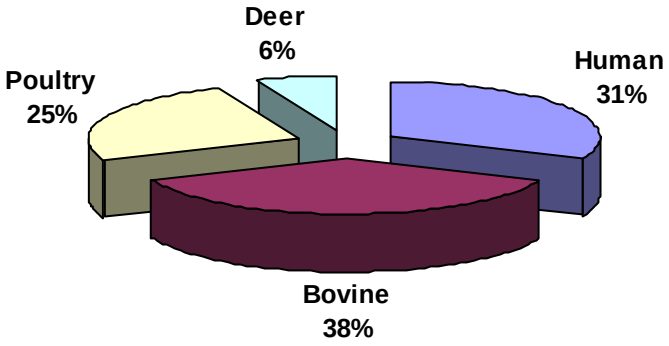

Figure 12. Relative percentage of host groups for Site 7

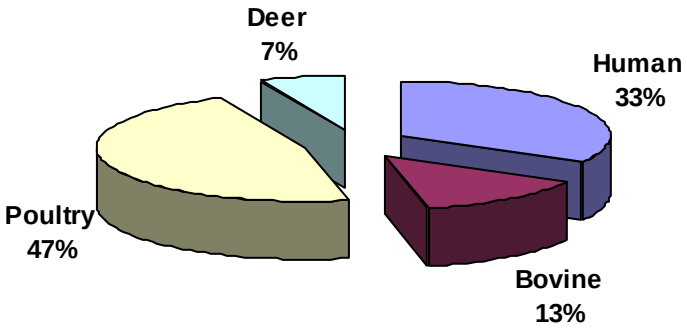

Figure 13. Relative percentage of host groups for Site 9

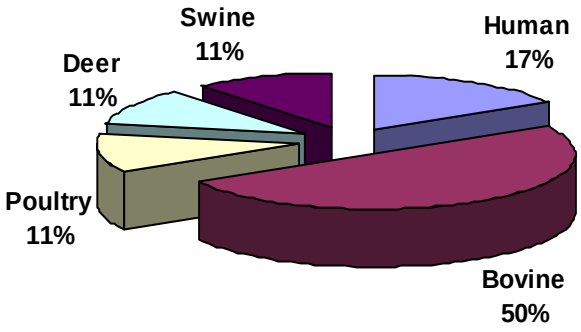

Figure 14. Relative percentage of host groups for Site 10

In addition, to see the overall results for 101 cultured *Enterococcus* isolates from Goose Creek, one single pie chart was created (Figure 15). The main sources of *Enterococcus* for Goose Creek, in decreasing order, were human (42%), bovine (27%), poultry (27%), deer (5%), swine (2%) and waterfowl (1%). These findings suggest that human and livestock hosts are the main pollution sources for Goose Creek.

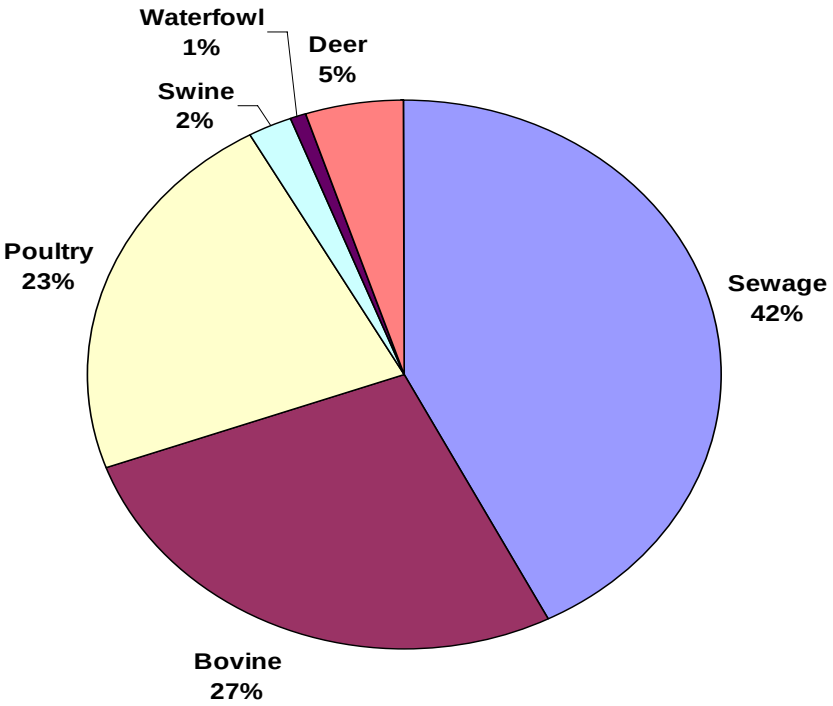

Figure 15. Relative percentage of host groups for overall isolates.

CHAPTER VI

CONCLUSIONS

The primary objectives of this study were to build a library of FAME profiles of *E. coli* and *Enterococcus* from known sources that was of appropriate size and representativeness and to use those libraries to determine sources and source distributions of unknown fecal environmental isolates deposited in the Goose Creek.

The results reported herein show that accuracy of FAME profiling is comparable to, and in some cases higher than, those reported for alternative phenotypic MST methods such as ARA and CUP. A library of 605 *E. coli* isolates resulted in 48% ARCC under a 6-way classification scenario, whereas a library of 511 *Enterococcus* isolates resulted in 66% ARCC. The host specific differences among *E. coli* and *Enterococcus* isolates were solely quantitative differences, that is, no signature FAME was identified and variations in relative masses of FAMEs common among different host groups were the only host specific differences. In addition, FAME profiles of 101 unknown *Enterococcus* isolates from monthly Goose Creek samples were compared to the known-source library. Three dominant source of fecal pollution in Goose Creek were; human (42%), bovine (27%), and poultry (27%). Site to site variations were explained in many instances by determining the sources nearby to each site and evaluating their potential impact.

The further significance of the findings reported in this study is that it compares two indicator organisms involving multiple genera to two other indicator organisms, one of which involves four species of one genus (*Enterococcus*), while the other one is the

only species level indicator studied, *E. coli*. The finding that the overall accuracy of classification is lower in the *E. coli* library than in the *Enterococcus* library implies that the accuracy of differentiation increases as the diversity of species included in the indicator organism increases. This suggests that the majority of the host-specific differences identified by FAME profiling may be due to different species being the predominant colonizers of different host groups. This may be a deficiency when the method is applied in the aquatic environments due to possible differential survival of different species.

Nevertheless FAME profiling relies on the differences in FAME profiles to identify the primary hosts of the indicator organisms and developing large known-source libraries may not be necessary to predict sources of fecal pollution, as is the case for the other phenotypic MST methods. Besides its accuracy, FAME profiling as an MST tool offers advantages such as being relatively rapid and highly economical.

It should also be noted that for further research other possible sources of fecal pollution such as domestic pets (dogs, cats, and horses) should be included in the known-source libraries. In addition, the unknown environmental sampling should continue over a period of time to see possible temporal difference in the sources of pollution.

REFERENCES

American Public Health Association (APHA). (1998). *Standard Methods for the Examination of Water and Wastewater*, 20th Edition, Washington, DC.

Anon. (1978). "The nomenclature of lipids". *J. Lipid Res*.,19, 114–128.

Bass, L., Liebert, C. A., Lee, M. D., Summers, A. O., White, D. G., Thayer, S. G., Maurer, J. (1999). "Incidence and characterization of integrons, genetic elements mediating multiple-drug resistance, in avian *Escherichia coli"*. *Antimicrob. Agents Chemother*., 43, 2925-2929.

Bendriss, E., Markoglou, N., Wainer, I. W. (2000). "Liquid chromatographic method for the simultaneous determination of caffeine and fourteen caffeine metabolites in urine". *J. Chromatog. B.*, 746, 331-338.

Bernhard, A. E., Field, K. G. (2000a). "Identification of nonpoint sources of fecal pollution in Coastal waters by using Host-Specific 16S Ribosomal DNA Genetic markers from fecal anaerobes". *Appl. Environ. Microbiol*., 66, 1587-1594.

Bernhard, A. E.,Field, K. G. (2000b). "A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA". *Appl. Environ. Microbiol.,* 66, 4571-4574.

Blumenthal, U.J., Mara, D. D., Peasey, A., Ruiz-Palacios, G., Stott, R. (2000). "Guidelines for the Microbiological Quality of Treated Wastewater Used in Agriculture: Recommendations for Revising WHO Guidelines".World Health Organization, 78,9.

Booth, A., Hagedorn, M. C., Graves, A. K., Hagedorn, S., Mentz, K. H. (2003). "Sources of fecal pollution in Virginia's Blackwater river". *J. Environ. Eng*., 129, 547-552.

Brondz, I. (2002) "Development of fatty acid analysis by high-performance liquid chromatography, gas chromatography, and related techniques". *Anal. Chim. Acta*., 465, 1–37.

Bryan, A., Shapir, N., Sadowsky, M. J. (2004). "Frequency and distribution of tetracycline resistance genes in genetically diverse, nonselected, and nonclinical *Escherichia coli* strains isolated from diverse human and animal sources". *Appl. Environ. Microbiol*., 70, 2503-2507.

Buchan, A., Alber, M., Hodson, R. E. (2001). "Strain-specific differentiation of environmental *Eschericia coli* isolates via denaturing 57 gradient gel electrophoresis (DGGE) analysis of the 16S-23S intergenic spacer region". *FEMS Microbiol. Ecol.*, 35, 313-321.

Cabelli, V.J. (1983). *Health Effects Criteria for Marine Recreational Water*. Publication EPA- 600/1-80-031. Environmental Protection Agency, Washington, D.C.

Cabelli, V. J., Dufour, A. P., McCabe, L. J.,Levin, M. A. (1982). "Swimming-associated gastroenteritis and water quality". *Am. J. Epidemiol*., 115, 606-616.

Clausen, E. M., Green, B. L., Litsky, W. (1977). *Fecal Streptococci: Bacterial Indicators of Pollution. Health Hazards associated with Water ASTM STP635.* American Society for Testing and Materials, pp. 247-264.

Carson, C. A., Shear, B. L., Ellersieck, M. R., Schnell, J. D. (2003). "Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal Escherichia coli from humans and animals". *Appl. Environ. Microbiol.,* 69,1836-1839.

Cassady, A. (2004). *Troubled Waters: An Analysis of Clean Water Act Compliance*, January 2002-June 2003. U.S. Public Interest Research Group (PIRG) Education Fund. Washington D.C. www.uspirg.org.

Choi, S., Chu, W., Brown, J., Becker, S. J., Harwood, V. J., Jiang, S. C. (2003). "Application of enterococci antibiotic resistance patterns for contamination source identification at Huntington Beach, California". *Mar. Poll. Bull*., 46, 748-755

Dargatz, D. A., Fedorka-Cray, P. J., Ladely, S. R., Kopral, C. A., Ferris, K. E., Headrick, M. L. (2003). "Prevalence and antimicrobial susceptibility of *Salmonella* spp. isolates from US cattle in feedlots in 1999 and 2000". *J. Appl. Microbiol*., 95, 753-761.

Duran, M., Haznedaroglu, B. Z., Zitomer, D. H. (2006). "Microbial source tracking using host specific FAME profiles of fecal coliforms". *Water Res.,* 40, 67–74.

Ettre, L.S. (1993). "Nomenclature for chromatography". *Pure Appl. Chem.,* **65**, 819–872.

Farnleitner, A. H., N. Kreuzinger, G. G. Kavka, S. Grillenberger, J. Rath, and R. L. Mach. (2000). "Simultaneous detection and differentiation of *Escherichia coli* populations from environmental freshwaters by means of sequence variations in a fragment of the beta-D-glucuronidase gene". *Appl. Environ. Microbiol.,* 66, 1340-1346.

Field, K. G. (2004). "Fecal Source Identification". *In* Cotruvo, A. Dufour, G., Rees, J., Bartram, R., Carr, D. O., Cliver, G. F., Craun, R., Fayer, V. P. G. (ed.), *Waterborne Zoonoses: Identification, Causes and Control*, World Health Organization, IWA Publishing Alliance House, London.

Gleeson, C., Gray, N. (1997). *The Coliform Index and Waterborne Disease: Problems of microbial drinking water assessment*. E&FNSPON, London.

Godfree, A. F., Kay, D., and Wyer, M. D. (1997). " Faecal streptococci as indicators of fecal contamination in water". *J Appl. Microbiol*., 83, 110-119.

Graves, A. K., C. Hagedorn, A. Teetor, M. Mahal, A. M. Booth, R. B. Reneau, Jr. (2002). "Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed". *J. Environ. Qual*., 31, 1300-1308.

Griffith, J. F., Weisberg, S. B., McGee, C. D. (2003). "Evaluation of Microbial Source Tracking Methods Using Mixed Fecal Sources in Aqueous Test Samples". *J. of Water and Health.,* 1(4), 141-151.

Guan, S., R. Xu, S. Chen, J. Odumeru, Gyles, C. (2002). "Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources". *Appl. Environ. Microbiol*., 68, 2690-2698.

Gunstone, F. D. (1996). *Fatty Acid and Lipid Chemistry*, Chapman&Hall, London, UK.

Hager, M.C. (2001). *Detecting Bacteria in Coastal Waters 1 and 2*. Stormwater Magazine. 2(3), 16-25. http://www.forester.net/sw_0105_detecting.html, and http://www.forester.net/sw_0106_detecting.html[Online] Last accessed: January 15, 2007.

Hagedorn, C. (2000). *Bacterial Source Tracking*. http://filebox.vt.edu/users/chagedor/biol_4684/BST/BSTmeth.html [online] Last accessed: April 23, 2006.

Hagedorn, C.,. Crozier, J.B., Mentz, K.A, Booth, A.M., Graves, A.K., Nelson, N.J., . Reneau Jr., R.B. (2003). "Carbon Source Utilization Profiles as a Method to Identify Sources of Faecal Pollution in Water". *J. Appl. Microbiol.,* 94, 792-799.

Havelaar, A.H., Pot-Hogeboom, W.M., Furuse, K., Pot, R., Hormann M.P. (1990). "Fspecific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin". *J. Appl. Bacteriol.*, 69,30-37.

Harwood, V. J., Whitlock, J., Withington, V. (2000). "Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters". *Appl. Environ. Microbiol*., 66, 3698-704.

Harwood, V. J., Wiggins, B., Hagedorn, C., Ellender, R. D., Gooch, J., Kern, J., Samadpour, M., Chapman, A. C. H, Robinson, B. J. (2003). "Phenotypic library-based microbial source tracking methods: Efficacy in the California collaborative study". *J. Water Health,* 1, 153-166.

Hatcher, P. G., McGillivary, P. A. (1979). "Sewage contamination in the New York Bight. Coprostanol as an indicator". *Environ. Sci. Technol*.,13, 1225-1229.

Haznedaroglu, B.Z., Zitomer, D.H., Hughes-Strange, G.B., Duran, M. (2005). "Whole-cell Fatty Acid Composition of Total Coliforms to Predict Sources of Fecal Contamination". *J. Environ. Eng.*, 131(10), 1426-1432.

Haznedaroglu, B. Z. (2005). *Fatty Acid Methyl Ester Profiling Of Indicator Organisms for Microbial Source Tracking*. MS Thesis, Villanova University, PA.

Holt, J.G., Krieg, N.R., Sneath, P., Staley, J.T., Williams, S.T. (1994). *Bergey's Manual of Determinative Bacterialogy*, 9th Edition, Williams & Wilkins, Baltimore. pp. 538- 539**.**

Horie, H., Kohata, K. (2000). "Analysis of tea components by high performance liquid chromatography and high-performance capillary electrophoresis". *J. Chromatog. A.,* 881, 425-438.

Käkelä, R., Ackman, R.G., Hyvärinen, H. (1995). "Very long chain polyunsaturated fatty acids in the blubber of ringed seals (Phoca hispida sp.) from Laike Saimaa, Lake Ladoga, The Baltic Sea, and Spitsbergen". *Lipids,* 30, 725–731.

Kariuki, S., C., Gilks, J., Kimari, A., Obanda, J., Muyodi, P., Hart, C. A. (1999). "Genotype analysis of *Escherichia coli* strains isolated from children and chickens living in close contact". *Appl. Environ, Microbiol.,* 65, 472-476.

Kelley, T. R., Pancorba, O. C., Merka, W. C. ,Barnharts, H. M. (1998). "Antibiotic resistance of bacterial litter isolates". *Poultry Sci.,* 77, 243-247.

Kruse, H., Sorum, H. (1994). "Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments". *Appl. Environ. Microbiol*., 60, 4015-4021.

Leung, K.T., Mackereth, R., Tien, Y. C., Topp, E. (2004). "A Comparison of AFLP and ERIC-PCR analyses for discriminating *Escherichia coli* from cattle, pig, and human sources". *FEMS Microbiol. Ecol*., 47, 111-119.

Madigan, M. T., Martinko, J.M., Parker, J. (2002). *Brock Biology of Microorganisms,* 10th Ed., Macromolecules, pp. 38-55., Prentice Hall, Upper Saddle River, NJ.

Meays, C. L., Broersma, K., Nordin, R., Mazumder, A. (2004). "Source tracking fecal bacteria in water: a critical review of current methods". *J. Environ. Manage.*, 73,71-79.

Miwa, T.K., Mikolajczak, K.L., Earle, F.R. and Wolff, I.A. (1960). "Gas chromatographic characterization of fatty acids". *Anal. Chem.,* **32**, 1739–1742.

Mu, H., Wesén, C., Sundin, P. (1997). "Halogenated fatty acids. 1. Formation and occurrence in lipids". *Trends Anal. Chem.* 16, 266–274.

Noblet, J. A., Young, D. L., Zeng, E. Y., Ensari, S. (2004). "Use of fecal sterols to infer the sources of fecal indicator bacteria in the lower Santa Ana river watershed, California: Sewage is unlikely a significant source". *Environ. Sci. Technol.*, 38, 6002-6008.

Ohlsen, K., Ternes, T., Werner, G., Wallner, U., Loffler, D., Ziebuhr, W., Witte, W., Hacker, J. (2003). "Impact of antibiotics on conjugational resistance gene transfer in *Staphylococcus aureus* in sewage". *Environ. Microbiol.,* 5, 711-716.

Parveen, S., Hodge, N. C., Stall, R. E., Farrah, S. R., Tamplin, M. L. (2001). "Phenotypic and Genotypic Characterization of Human and Nonhuman *Escherichia col"i. Water Res.,* 35(2), 379-386.

Parveen, S., Portier, K. M., Robinson, K., Edmiston, L., Tamplin, M. L. (1999). 'Discriminant Analysis of Ribotype Profiles of *Escherichia Coli* for Differentiating Human and Nonhuman Sources of Fecal Pollution". *Appl. Environ. Microbiol.,* 65(7), 3142-3147.

Parveen, S., Murphee, R. L., Edmiston, L., Kaspar, C. W., Portier, K. M., Tamplin, M. L. (1997). "Association of Multiple-Antibiotic-Resistance Profiles with Point and Nonpoint Sources of *E. coli* in Apalachicola Bay. *Appl. Environ. Microbiol.,* 63(7), 2607-2612.

Pourcher, A. M., Devriese, L.A., Hernandez, J. F., Delattre, J. M. (1991). "Enumeration by miniaturized method of E. coli, S. bovis and enterococci as indicators of the origin of fecal pollution of waters". *J. Appl. Bacteriol.*, 70, 525-530.

Rabinovici, S. J. M., R. L. Berhknopf, A. M. Wein, D. L. Coursey, and R. L. Whitman. 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan Beach. *Environ. Sci. Technol.,* 38, 2737-2745.

Rezanka, T., Votruba, J. (2002). "Chromatography of very long-chain fatty acids from animal and plant kingdoms". *Anal. Chim. Acta,* 465, 273–297.

Richardson, K. J., Stewart, M. H., Wolfe, R. L. (1991). "Application of gene probe technology to the water Industry". *J. Am. Water Works Assoc.,* 83, 71-81.

Samadpour, M., Stewart, J., Steingart, K., Addy, C., Louderback, J., McGinn, M., Ellington, J., Newman, T. (2002). "Laboratory investigation of an *Escherichia coli* 0157:H7 outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington". *J. Environ. Health*, 64,16-20.

Salyers, A. A. (1995). *Antibiotic resistance gene transfer in the Mammalian intestinal tract,* 1st edition, Springer-Verlag Berlin and Heidelberg , GmbH & Co. KG.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., Lukasik, J. (2002). "Microbial Source Tracking: Current Methodology and Future Directions". *Appl. Environ. Microbiol.,* 68(12), 5796-5803.

Shanks, O. C., Santo Domingo, J. W., Lu. J, Kelty, C. A., Graham, J. E. (2006). "Identification of bacterial DNA markers for the detection of human fecal pollution in water'. *Appl. Environ. Microbiol.,* 73(8), 2416-2422.

Simpson, J.M., Santo Domingo, J.W., Reasoner, D. J. (2002). "Microbial Source Tracking: State of the Science". *Environ. Sci. Technol*., 36, 5279-5288.

Smalla, K., Heuer, H., Gotz, A., Niermeyer, D., Krogerrecklenfort, E., Tietze, E. (2000). "Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity in IncQ-like plasmids". *Appl. Environ. Microbiol.,* 66, 4854-4862.

Standley, L. J., Kaplan, L. A., Smith, D. (2000). "Molecular tracers of organic matter sources to surface water resources". *Environ. Sci. Technol.,* 34, 3124-3130.

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"*. *Environ. Sci. Technol.,* 38, 6109-6117.

Suzuki, M., Rappe, M. S., Giovannon, S. J. (1998). "Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity". *Appl. Environ. Microbiol*., 64(11), 4522-4529.

Tynkkynen, S., R. Satokari, M. Saarela, T. Mattila-Sandholm, and M. Saxelin. (1999). "Comparison of ribotyping, randomly amplified polymorphic DNA analysis, and pulsed-field gel electrophoresis in typing of *Lactobacillus rhamnosus* and *L. casei* strains". *Appl. Environ. Microbiol.,* 65, 3908-3914.

U.S. Environmental Protection Agency. (2007a). *National Section 303(d) List Fact Sheet.* Online Document at http://oaspub.epa.gov/waters/national_rept.control (Accessed June 11, 2007).

U.S. Environmental Protection Agency. (2007b). *National Section 303(d) List Fact Sheet.* Online Document at http://oaspub.epa.gov/waters/state_rept.control?p_state=PA (Accessed June 11, 2007).

U.S. Environmental Protection Agency. (2005). *Microbial Source Tracking Guide Document. Office of Research and Development*. EPA/600-R-05-064. Cincinnati, OH. pp-150.

U.S. Environmental Protection Agency. (2003a). *Introduction to TMDLs*. http://www.epa.gov/owow/tmdl/intro.html#definition [Online] Last accessed: January 9, 2007.

U.S. Environmental Protection Agency. (2003b). *Overview of TMDLs*. http://www.epa.gov/owow/tmdl/tptmdl/slide09.html [Online] Last accessed: January 9, 2007.

U.S. Environmental Protection Agency. (2003c). *Overview of TMDLs*. http://www.epa.gov/owow/tmdl/tptmdl/slide11.html [Online] Last accessed: January 9, 2007.

U.S. Environmental Protection Agency (2002). *Method 1103.1: Escherichia coli in Water by Membrane Filtration using membrane-Thermotolerant Escherichia coli Agar (mTEC).* EPA-821-R-02-020.

U.S. Environmental Protection Agency (2002). *Method 1106.1: Enterococci in Water by Membrane Filtration using membrane- Enteroccous-Esculin Iron Agar (mE-EIA*). EPA-821-R-02-021.

U. S. Environmental Protection Agency. (2002a). *Water Quality Conditions in the United States, A Profile from the 2000 National Water Quality Inventory*.

U.S. Environmental Protection Agency. (2002b). *Implementation Guidance for Ambient Water Quality Criteria for Bacteria.* [Online] http://www.epa.gov/ost/standards/bacteria/bacteria.pdf. Accessed 19 July 2006.

U.S. Environmental Protection Agency. (1986). *Ambient Water Quality Criteria for Bacteria. Office of Water Regulations and Standards*. EPA440/5-84-002. Washington, DC. pp 18.

Weigel, S., Bester, K., Huhnerfuss, K. (2001). "New method for rapid solid-phase extraction of large-volume water samples and its application to non-target screening of North Sea water for organic contaminants by gas chromatography mass spectrometry". *J. Chromatog*. A., 912, 151-161.

Whitlock, J. E., Jones, D.T., Harwood, V.J. (2002). "Identification of the Sources of Fecal Coliforms in an Urban Watershed Using Antibiotic Resistance Analysis." *Water Res*., 36(17), 4273-4282.

Wiggins, B. A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M., Han, J., Henry, B.L., Hoover, K.B., Johnson, E.L., Jones, K.C., McCarthy, J.G., McDonough, J.A., Mercer, S.A., Noto, M.J., Park, H., Phillips, M.S., Purner,

S.M., and B. M. Smith, Stevens, E.N., Varner, A.K. (2003). "Use of Antibiotic Resistance Analysis for Representativeness Testing of Multiwatershed Libraries." *Appl. Environ. Microbiol*., 69(6), 3399-3405.

Wiggins, B. A., Andrews, R. W.,Conway, R. A., Corr, C. L., Dobratz, E. J., Dougherty, D. P., Eppard, J. R., Knupp, S. R., Limjoco, M. C., Mettenburg, J. M., Rinehardt, J. M., Sonsino, J., Torrijos, R. L., Zimmerman, M.E. (1999). "Use of Antibiotic Resistance Analysis to Identify Nonpoint Sources of Fecal Pollution." *Appl. Environ. Microbiol.*, 65(8), 3483-3486.

Wiggins, B. A. (1996). "Discriminant Analysis of Antibiotic Resistance Patterns in Fecal Streptococci, A Method to Differentiate Human and Animal Sources of Fecal Pollution in Natural Waters." *Appl. Environ. Microbiol.*, 62(11), 3997-4002.

World Health Organization (WHO). 2004. Guidelines for Drinking-water Quality. 3rd Edition. 494 p. [Online] http://www.who.int/water_sanitation_health/dwq/gdwq3/en/index.html. Accessed 27 September 2006.