A MOLECULAR APPROACH TO DETERMINE THE ORIGIN OF FECAL BACTERIA IN THE CATOMA CREEK WATERSHED

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Rasanthi Udenika Wijesinghe

Certificate of Approval:

C. Wesley Wood
Professor
Agronomy and Soils

Yucheng Feng, Chair
Associate Professor
Agronomy and Soils

Joey N. Shaw
Associate Professor
Agronomy and Soils

Donald M. Stoeckel
U.S Geological Survey
Columbus, Ohio

Stephen L. McFarland
Dean
Graduate School

A MOLECULAR APPROACH TO DETERMINE THE ORIGIN OF FECAL BACTERIA IN THE CATOMA CREEK WATERSHED

Rasanthi Udenika Wijesinghe

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirement for the

Degree of

Master of Science

Auburn, Alabama
December 16, 2005

A MOLECULAR APPROACH TO DETERMINE THE ORIGIN OF FECAL

BACTERIA IN THE CATOMA CREEK WATERSHED

Rasanthi Udenika Wijesinghe

Permission is granted to Auburn University to make copies of this thesis at its discretion, upon the request of individuals or institutions and at their expense. The author reserves all publication rights.

____

Signature of Author

____

Date of Graduation

THESIS ABSTRACT

A MOLECULAR APPROACH TO DETERMINE THE ORIGIN OF FECAL BACTERIA IN THE CATOMA CREEK WATERSHED

Rasanthi Udenika Wijesinghe

Masters of Science, December 16, 2005
(B.S., University of Peradeniya, Sri Lanka, 1995)

125 Typed Pages

Directed by Yucheng Feng

High concentrations of fecal indicator bacteria are the most common cause of surface-water impairment in Alabama. A 37 km segment of Catoma Creek in Montgomery Country has been included on Alabama 303(d) List of impaired water bodies due to elevated concentrations of fecal coliform bacteria and organic enrichment. Fecal contamination can originate from both human and non-human sources, including surface runoff from land application of animal wastes or farm animal feedlots, inadequate septic or sewer systems, improper waste disposal, and wildlife impact. The objectives of this study were to monitor the fecal contamination and identify sources of contamination in the Catoma Creek watershed.

Water samples were collected monthly at eight locations in the watershed for a period of one year. *E. coli* was enumerated using the modified m-TEC media. Data showed that *E. coli* concentrations varied from 18 to 12,650 CFU/100 ml, with 70% of

the samples exceeding the EPA criterion for swimming water. There was a positive correlation between flow rates and *E. coli* concentrations. Chemical analyses of the water samples showed that the concentration of total phosphorus in all the samples was above the proposed Ecoregion IX nutrient criterion, 78% of samples were above the NO_3-N criterion, and 50% of samples were above the total nitrogen criterion, suggesting that there is a serious risk of eutrophication in this watershed.

The rep-PCR DNA fingerprint technique was used to identify the sources of fecal contamination in the Catoma Creek watershed. A known source library of DNA fingerprints was developed using 582 *E. coli* isolates obtained from humans, dogs, cattle, chickens, horses, wild turkeys, waterfowl and deer. DNA fingerprints generated using the BOX A1R primer demonstrated great genetic diversity of *E. coli*. Cluster analyses of DNA fingerprint patterns were performed with BioNumerics software using a densitometric curve based matching function (Cosine) and unweighted pair group method with arithmetic average. Jackknife analysis was used to determine cluster/group validity, revealing that the average rate of correct classification for the entire library was 88% and that of the decloned library was 74%. The DNA fingerprints (502) obtained from *E. coli* isolated from the water samples of the Catoma Creek watershed were compared against those in the known source library. Results showed that 18% of the *E. coli* isolates were from humans, 14% each from dogs and waterfowl, 4% each from deer and wild turkeys, 2% each from cattle and chickens, 0% from horses, and the remaining 41% unidentified. Further research is needed to improve the representativeness of the library by including more source groups and *E. coli* isolates.

ACKNOWLEDGMENTS

The author would like to express her sincere thanks to Dr. Yucheng Feng for her guidance and encouragement during the course of study. Thanks are also extended to the members of her committee: Dr. Wesley Wood, Dr. Joey Shaw and Dr. Donald Stoeckel for their valuable suggestions and assistance. She is also grateful to Dr. Ludovic Capo-chichi, Brenda Wood, John Owen and Michele Owsely in the Department of Agronomy and Soils, April Jones at NRCS, Wil Smith of Montgomery County Soil & Water Conservation District, Buddy Morgan and Jerald Conway of MWWSSB, Malene Watson and Sabra Sutton of CH2M Hill, and Brian Atkins and Amy Gill at USGS for their generous help during the course of study. The author would like to thank the staff at the Baptist Medical Center South in Montgomery for providing *E. coli* isolates. Financial support of the research was provided by the Alabama Water Resources Research Institute, the Auburn University Environmental Institute, the Auburn University Competitive Research Grant Program, the Alabama Agricultural Experiment Station, and the Montgomery County Soil and Water Conservation District. Finally, she would like to express deep appreciation to her husband, daughter and her family for their encouragement, support and love throughout this study.

Style manual or journal format used: Journal of Environmental Quality

Computer software used: Microsoft office 2003, BioNumerics v. 4.0, Minitab 13.32, KaleidaGraph v. 3.6

TABLE OF CONTENTS

LIST OF TABLES………………………………………………………………... ix

LIST OF FIGURES ……………………………………………………………… x

I. LITERATURE REVIEW..…………………………………………….……. 1

II. MONITORING FECAL CONTAMINATION AND NUTRIENT

ENRICHMENT IN THE CATOMA CREEK WATERSHED ……………….. 35

Abstract……………………………………………………………………… 35

Introduction………………………………………………………………….. 34

Materials and Methods………………………………………………………. 38

Results and Discussion………………………………………………………. 41

Summary…………………………………………………………………….. 48

III. IDENTIFICATION OF SOURCES OF FECAL CONTAMINATION

IN THE CATOMA CREEK WATERSHED………………………………… 63

Abstract……………………………………………………………………… 63

Introduction………………………………………………………………….. 64

Materials and Methods………………………………………………………. 66

Results and Discussion……………………………………………………… 71

Summary…………………………………………………………………….. 88

IV. REFERENCES ……………………………………………………………... 107

V. APPENDIX……………………………………………………….…………. 115

LIST OF TABLES

1.1 EPA criteria for *E. coli* densities ………………………………………….. 9

2.1 Concentrations of *E. coli* at the Catoma Creek watershed
from May 2003 to April 2004……………………………………………… 49

2.2 Daily rainfall values for the Catoma Creek watershed
from May 2003 to April 2004……………………………………………... 50

2.3 pH values of the Catoma Creek watershed May 2003 to April 2004……… 52

2.4 Water temperatures at the Catoma Creek watershed sampling sites ……..... 53

2.5 Ranges of chemical water quality parameters at the
Catoma Creek watershed ……………………………………………………. 54

3.1 Host groups and DNA fingerprint patterns in the known source
library………………………………………………………………………. 90

3.2 Comparison of different coefficients using Jackknife analysis
in the decloned library………………..…………………………………. 91

3.3 Reproducibility of DNA fingerprints using test isolates…………………… 92

3.4 Rate of correct classification for three source groups …………………..... 93

3.5 Rate of correct classification for eight source groups …………………..... 94

3.6 ID Bootstrap analysis results…...………………………………………….. 96

3.7 Number of *E. coli* isolates identified at the eight sampling sites ………….. 98

3.8 Number of *E. coli* isolates identified at the eight sampling sites by month... 99

3.9 Source identification based on flow rate…………………………………… 100

3.10 Source distribution with different flow condition………………………... 101

LIST OF FIGURES

1.1 Land used pattern in the Catoma Creek watershed ………………………….. 34

2.1 Sampling sites in the Catoma Creek watershed ……………………………... 55

2.2 *E. coli* concentrations at the Catoma Creek sampling sites

from May 2003 to April 2004……………………………………………….. 56

2.3 Average monthly rainfall at the Catoma Creek watershed …………………... 57

2.4 Percentages of water samples having *E. coli* concentrations above the

30-day geometric mean criterion…..…………………………………………. 58

2.5 Correlation between *E. coli* concentration and flow rate at the selected

sites in the Catoma Creek watershed…………………………………………. 59

2.6 Average water temperature across the Catoma Creek sampling sites ……….. 60

2.7 Concentrations of TN, NO_3-N and TP at each sampling site ……….………. 61

2.8 Yearly average NH_4-N, NO_3-N, TN and TP concentrations at sampling site... 62

3.1 rep-PCR fingerprints generated using the BOX A1R primers ………………. 102

3.2 MANOVA plots of rep-PCR DNA fingerprint patterns of *E. coli* from

different source groups ……………………………………………………… 103

3.3 MANOVA plots of rep-PCR DNA fingerprints of *E. coli* isolated from

Catoma Creek watershed …………………………………………………….. 104

3.4 Sources of fecal contamination in the Catoma Creek watershed…………….. 105

3.5 Source distribution at different sampling sites in the Catoma Creek

Watershed……………………………………………………………………. 106

I. LITERATURE REVIEW

A. Introduction

Fecal pollution impairs the quality of streams and rivers for recreational use and adversely affects fish and other aquatic life. Globally, 1.5 billion people suffer from a lack of safe drinking water and hundreds of thousands of people die each year due to water borne diseases (WHO, 2001). According to the United States Environmental Protection Agency's Clean Water Action Plan, in 1998 40% of the waterways in the USA were unsafe for fishing and swimming due to fecal contamination. Fecal coliforms normally inhabit the intestinal tract of warm-blooded animals and their presence in soil or water is a good indicator that the soil or water has been contaminated by fecal material. *Escherichia coli* is one common type of fecal coliform bacteria that is used as the indicator bacteria for fresh water testing. The presence of these fecal indicator bacteria suggests the presence of potential human pathogens that may pose health risks to humans, and threaten the integrity of ecosystems. Fecal contamination can originate from both human and non-human sources including surface runoff from land application of animal wastes or farm animal feedlots, inadequate septic or sewer systems, improper waste disposal, and wildlife impact. Determining the source of fecal contamination is necessary to develop effective pollution control strategies.

Microbial source tracking or bacterial source tracking is a new technology that is being developed to identify the source of fecal contamination in surface waters. Both

phenotypic and genotypic methods are being used to determine the host origin of fecal bacteria. The repetitive element polymerase chain reaction (rep-PCR) genomic fingerprinting technique was selected for this study because it has been found to be reliable, reproducible, rapid, and highly discriminatory (Rademaker and de Bruijn, 1997). Rep-PCR genomic fingerprinting makes use of DNA primers that are complementary to naturally occurring, highly conserved, non-coding, repetitive DNA sequences present in multiple copies in the genomes of most bacteria. The rep-PCR genomic fingerprints generated from bacterial isolates permit differentiation to the species, subspecies, and strain level (Rademaker and de Bruijn, 1997).

This study focuses on the Catoma Creek watershed in Montgomery County, Alabama. Catoma Creek is a tributary of the Alabama River and drains 932 km^2 of both agricultural and urban land. Forest, agriculture (pasture and row crop), and urban land uses represent 54.5%, 36.2%, and 9.3%, respectively, of the watershed (ADEM, 2002). Urban runoff and pasture grazing are suspected to be the main sources of fecal contamination of Catoma Creek. The segment (37 km) of Catoma Creek from Alabama River to Ramer Creek is on the Section 303(d) List due to fecal contamination and organic enrichment (ADEM, 2002).

B. Objectives

The objectives of this study were to

1. Monitor the level of fecal contamination in the Catoma Creek watershed and isolate *E. coli* from water samples;

2. Construct a library of rep-PCR DNA fingerprints from *E. coli* strains isolated from a wide range of human and animal feces in the Catoma Creek watershed;
3. Identify the origin of fecal contamination in the Catoma Creek watershed using the constructed rep-PCR DNA fingerprint library.

C. Water quality regulations

Section 303(d) of the Clean Water Act requires that each state, territory and tribe in the United States establish a list of impaired water bodies that do not currently support their designated uses. A value for the Total Maximum Daily Load (TMDL) needs to be developed for each water body that does not meet its designated classification. This TMDL consists of the sum of point source load and non-point source load along with a margin of safety. According to a report that was issued in the 1990s, more than 40% of the watersheds in the country failed to meet the EPA standards. Each state needs to identify the present quality of their waters and the pollutant sources of these waters. Impairments may be caused by multiple pollutants such as sediments, pathogens, nutrients, metals, dissolved oxygen, pH, pesticides, temperature (thermal pollution) and other organic chemicals. The EPA enforces the regulations and laws mandated by the Clean Water Act, but day to day enforcement is the responsibility of the states, which are also required to implement one or more pollution control remedies as best management practices (BMP) (Clean Water Act, 1972; EPA, 1999).

Initially the Clean Water Act only targeted point sources; however, after 1982 it was also adopted for non-point sources. Certain construction activities and Municipal Separate Storm Sewer Systems (MS4s) for large populated metropolitan areas with populations exceeding 100,000 require a permit and are currently regulated by each state's National Pollutant Discharge Elimination System (NPDES) programs. Pollutant loadings from MS4s enter surface waters in response to storm events. Along with the storm water, pollutants in urban runoff, accumulated street dust, and litter from impervious roadway surfaces may enter surface water bodies. Metropolitan areas with populations exceeding 100,000 need to obtain NPDES storm water permits. The purpose of this permit is to eliminate or minimize the extent of pollutant discharge (ADEM, 2002).

In 1996, EPA imposed new TDML rules for calculating the load allocation for waters impaired solely or primarily by nonpoint sources (Stiles, 2003). It is not difficult to identify whether water is contaminated with fecal materials; however, identification of sources of contamination is not trivial. Determining the sources of fecal contamination is important for developing effective pollution control strategies and best management practices (BMPs).

Under the Clean Water Act, all waterways in the United State are classified to maximize the utilization of that water as public drinking water supply, propagation of fish and wildlife, recreational activities, industrial and agricultural usage, navigation and others. Each category has different water quality standards; for example, a stream classified as agricultural and industrial will have lower standards than one that is designated for swimming. One objective of stream classification is to attain the Clean

Water Act's goals of fishable and swimmable water wherever possible and to prevent further stream degradation. Stream classification is mainly assigned by applying both qualitative and quantitative criteria. According to quantitative stream classification system, Alabama streams can be classified as outstanding national resource waters where no discharge is permitted into these waters, outstanding Alabama water, swimming, shellfish harvesting, public water supply, fish and wildlife, agricultural and industrial water supply, and industrial operations (Boyd, 2000). Some parameters such as pH, temperature, wastewater effluent, dissolved oxygen, bacteria, and turbidity limits have been established to maintain the assigned stream classification system or upgrade it. According to the qualitative narrative criteria, qualitative parameters such as toxicity, taste, odor, and color are used to maintain or upgrade assigned stream classification. In Alabama and other states, more streams are now classified for fish and wildlife after being upgraded to a higher use category than in the past (Boyd, 2000).

Current water quality issues mainly involve chemical, physical and microbiological aspects. The chemical quality of a body of water is often related to the type of industries and agricultural activities that take place in its watershed. For example, agricultural wastes such as animal wastes, organic and inorganic fertilizers (especially nitrogen and phosphorus) and pesticides may enter the streams with runoff water and effluents from industries also contribute to the pollution loadings. The physical aspect of water quality is related to natural or man-made factors such as the amount of suspended solids, which may increase the turbidity of the water and reduce the light penetration. Microbial pollution is an especially important issue for water used

for recreation activities, public water supplies, aquifer protection and propagation of fish, shellfish, and wildlife.

D. Fecal indicator bacteria

Since water borne pathogenic organisms can be fatal to humans, fecal contamination is considered one of the most important factors affecting water quality. Globally, 1.5 billion people suffer from a lack of safe drinking water and 3.4 million people die each year due to these water borne diseases (WHO, 2001). Thus, monitoring the biological water quality of surface water plays an important role in environment related research. To detect contamination, it is important to use indicator bacteria because enumeration of each pathogen is unrealistic. Since the early 20th century, the United States has used fecal coliforms as convenient indicator bacteria to identify the fecal pollution in surface waters. Fecal coliform bacteria are a subgroup of total coliforms; they include the genera of *Escherichia, Klebsiella,* and *Citrobactor.* Fecal coliforms are gram-negative, non-spore forming rods, which can produce both gas and acid by fermenting lactose when incubated at 44.5°C for 48 hours (Reynolds, 2003). They are associated with fecal materials of warm-blooded animals and their presence indicates the possible of presence of pathogenic bacteria such as *shigella, Cholera, Salmonella*, and viruses such as hepatitis A and Norwalk group viruses in the water (Reynolds et al., 2003; Simpson et al., 2002).

a. *E. coli* as indicator bacteria

Between 1972 and the early 1980s, the EPA conducted a series of studies to establish the relationship between indicator organisms and the incidence of intestinal illness or gastroenteritis (EPA, 1986). For marine waters, the highest correlation was found for *Enterococcus*, and in fresh waters, the highest correlation was found for both *E. coli* and *Enterococcus*. However, the correlation between fecal coliforms and intestinal illness was found to be relatively poor. As a result of this extensive epidemiological study, the EPA has recommended the use of both *E. coli* and *Enterococcus* as indicator bacteria for surface waters and *Enterococcus* in estuarine waters (EPA, 1986).

E. coli is a rod shape, facultative anaerobic, gram negative bacterium belonging to the large bacterial family, *Enterobacteriaceae* that lives in the lower intestinal track of warm blooded animals. According to a study done in Australia with 16 families and 79 species of mammals, *E. coli* is the dominant member (in this study relative abundance was 46%) of the family *Enteribacteriaceae* (Gordon and FitzGibbon, 1999). Some members of this family are human intestinal pathogens (e.g., *Salmonella*, *Shigella*, *Yersinia*) and several others are normal colonists of the human gastrointestinal tract (e.g., *Escherichia, Enterobacter, Klebsiella*). Most strains of *E. coli* are not pathogenic to humans except for a few that cause urinary tract infections, neonatal meningitis and some intestinal diseases that are sometimes fatal due to acute kidney failure (e.g., Serotype O157:H7) (Todar, 2002).

E. coli has a rapid growth rate, with a generation time of 20 minutes under optimum conditions. It is easy to culture and has been studied extensively. There is no

evidence of its duplication in fresh water. Furthermore, *E. coli* tends to occur in the same environment as pathogenic organisms, but with greater densities. Therefore, *E. coli* is recognized as a good indicator bacterium (Scott et al., 2002).

The decay of *E. coli* and other fecal coliform bacteria is influenced by several environmental factors. The rate of die off increases with increasing temperature, elevated pH, higher dissolved oxygen, solar radiation, predacious microorganisms such as protozoa, lack of nutrients, and salinity (An et al., 2002). These bacteria can enter streams and lakes through run off water and some are removed by adsorption onto particles and subsequent sedimentation. As a result, higher densities of fecal bacteria usually accumulate in the sediment than in the water (Crabill et al., 1999; Irvine and Pettibone, 1993). A positive relationship between *E. coli* density and rainfall has been reported; during the summer the density was lower mainly due to dry weather. During the same period, high temperatures, solar radiation, high levels of dissolved oxygen and elevated pH due to algal growth also reduced the *E. coli* density in the water. A wet season after dry weather is often associated with a higher density of *E. coli* (An et al., 2002).

b. Water quality criteria for *E. coli*

Table 1.2 shows the threshold levels used to enforce recreational water quality standards in the United States (EPA, 1986).

Table 1.1 EPA criteria for *E. coli* densities (EPA, 1986)

Steady state Geometric mean* indicator densities	
Contact recreation	126 CFU/100ml
Non contact recreation	605 CFU/100ml
Single sample maximum allowable density	
Designated beach area	235CFU/100ml
Moderate full body contact recreation	298 CFU/100ml
Lightly used full body contact recreation	410 CFU/100ml
Infrequently used full body contact recreation	576 CFU/100ml
Drinking water	free from *E. coli*

*Geometric mean – At least 5 samples, collected during a one-month period

$y = n^{th}$ root of $y_1*y_2*y_3$ ----- y_n.

c. Host specificity of *E. coli*

Host specificity is considered an important factor influencing the genetic diversity of *E. coli*. Until recently, few studies reported how populations of *E. coli* varied in different hosts. Gorden and Lee (1999) isolated enteric bacteria from four orders and ten families of mammals in Australia. Based on multi-locus enzyme electrophoresis (MLEE) characterization, the taxonomic family of the host was found to explain a small but significant amount of the genetic variation among *E. coli* isolates (6%). The same multi-locus genotype was recovered from individuals of different taxonomic orders. Nucleic acid based methodology seems to offer a better approach to differentiating between the fecal indicator bacteria colonizing different animal hosts. Dombek et al. (2000) showed that the DNA fingerprints obtained using rep-PCR with the BOX A1R primer were effective for grouping *E. coli* isolates by hosts (humans, geese, ducks, cows, pigs, chickens, and sheep). In general, the band patterns of *E. coli* isolates from different animal sources are very similar, indicating that the isolates are closely related. Fingerprint patterns are similar, but not always identical. In Dombek et al.'s study approximately one quarter of the bands were shared by more than 80% of isolates and a few bands were shared by more than 90% of the isolates. Parveen et al. (1999) reported that ribotype profiles of 238 *E. coli* isolates from both human and nonhuman sources resulted in 82% of average rate of correct classifications (ARCC) and formed four clusters, with the majority of human and nonhuman samples isolated in two clusters. According to the study by Dombek et al. (2000), 100% of *E. coli* isolates derived from cows and chickens and between 78% - 90% of the *E. coli* isolates from humans, geese, ducks, pigs and sheep were correctly assigned to the correct host group.

Whether these observations reflect true host specificity for *E. coli* or merely differentiate their distribution needs further research.

d. Geographic variation of *E. coli*

Geographical locality is another factor that can be used to account for the genetic diversity. According to Gordon (1997), the spatial structure accounted for 2% of the genetic variation of *E. coli* isolates from two populations of feral house mice located 15 km apart. Gordon and Lee (1999) reported that *E. coli* isolated from mammals throughout Australia showed that 5% of the allelic diversity was due to a locality effect. *E. coli* isolated from rodents living in Australia and Mexico showed 10% of the diversity thus revealing the intercontinent differences (Souza et al., 1999). Hartel et al. (2002) used ribotyping to determine the geographic variability of *E. coli* isolates from cattle, horses, swine, and chickens. They found that decreased distance among locations increased ribotype sharing for cattle and horses, but not for swine and chicken. These studies showed that the limitations of a database designed to identify environmental isolates when the host origin isolates are from another geographic location varies considerably. It was suggested that researchers should be cautious about the universal use of a host origin database developed for a limited geographic region (Hartel et al., 2002). Isolates from non domestic animals seem to show more geological variation than those of humans. In a collection of human *E. coli* from four continents, spatial effects accounted for only 2% of the observed allelic diversity (Whittam et al., 1983). The greater mobility of the human population may account for the lack of variation among human *E. coli*.

e. Primary versus secondary habitats

E. coli found in the environment may spend part of its time in the host and the rest of the time in the external environment (Savageau, 1983). Little information is available concerning the fate of clones moving from a host to the external environment. Whittam (1989) examined *E. coli* isolated from domestic birds and their litters, but found only 10% of the 113 clones was recovered from both habitats. This suggests that the clonal composition of *E. coli* communities changes during the transition from a host to the external environment.

f. Temporal variation of *E.coli*

Temporal variability is another factor influencing the genetic variability among *E. coli*. According to Whittam (1989), about 25% to 50% of the *E. coli* isolates identified in an individual host arose due to temporal variation. There were no obvious environmental factors (e.g., pH, temperature) that could account for this variation. *E. coli* isolated from the inflow to sewage treatment plant over a four month period showed significant variations in the composition (Pupo and Richardson, 1995). Aslam et al. (2003) obtained 1403 *E. coli* isolates from feces of beef cattle, and from hides, carcasses, and ground beef after slaughter. They showed that majority of *E. coli* isolates within individual animals shared close genetic relatedness within each sampling time. However unique genetic subtypes were observed at each sampling time. Some genetic subtypes of *E. coli* were present at high frequency in feedlots, and these may contribute a resident population. Some subtypes appeared at certain sampling periods, which indicate that cattle could also harbor transient subtypes (Aslam et al., 2003).

E. Identification of sources of fecal contamination

Various approaches have been used to identify the origin of fecal contamination in surface waters. There are two basic types of source tracking approaches: chemical and biological.

a. Chemical methods

Various organic wastewater contaminants (OWCs) such as pharmaceuticals, hormones, caffeine, and fecal steroids have been used as indicators for fecal contamination of human or non-human origin. The U.S. Geological Survey conducted a nation wide study to measure the concentrations of 95 OWCs in a 139-stream network extending over 30 states during 1999 and 2000. According to this study, one or more OWCs were found in 80% of the streams tested. The most frequently detected compounds were caprostanol (fecal steroids), cholesterol (plant and animal steroids), insect repellants, caffeine, antibacterial disinfectants, fire retardants and nonionic detergent metabolites, all of which indicate the influence of human activities on these streams (Koplin et al., 2002).

Caffeine is one of the most common chemicals used to identify human activities in watersheds. It is abundant in a range of beverages, including coffee, tea, soft drinks, and many pharmaceutical products. It enters the environment with human wastes, passes through wastewater treatment processes and then enters the water-bodies. It has been suggested that the presence of caffeine in the environment indicates the extent of human activities (Scott et al., 2002).

Coprostanol is another common organic chemical that can be used to identify fecal contamination in watersheds. Coprostanol is a fecal sterol formed during the catabolism of cholesterol by indigenous bacteria present in the gut of humans and higher animals. Human feces contain ten times more coprostanol on a dry weight basis than that of either cats or pigs. Herbivores such as cows, sheep and horses also contain coprostanol, but their sterol profiles are dominated by 24-ethyl coprostanol. Thus, coprostanol can be used as a biomarker to identify the sources of fecal pollution (Leeming et al., 1996; Scott et al., 2002).

b. Biological methods – Microbial/Bacterial source tracking

Biological methods are employed to differentiate the sources of microbes, in particular whether they are human, livestock, or wildlife in origin. This is a relatively new technology that was first utilized in 1996 by Professor Charles Hagedorn in the Page Brook watershed in Clark County, Virginia (Stiles, 2003). Various genotypic and phenotypic methods have been used for source tracking purposes. Most involve the construction of a known source library, although a few adopt a library-independent approach. These methods can be divided into four categories (Bush et al., 2003) and are briefly discussed below:

- **Library-dependent genotypic methods**: Pulsed-field gel electrophoresis (PFGE), random amplified polymorphic DNA (RAPD) analysis, amplified fragment length polymorphism (AFLP) analysis, ribotyping, and rep-PCR.

- **Library-dependent phenotypic methods:** Antibiotic resistant analysis, carbon source profiling.
- **Library-independent genotypic methods**: Host specific molecular markers, tRFLP
- **Library-independent phenotypic methods**: Fecal bacteria ratio, host specific indicator organisms, viruses, F^+ coliphage serotyping, enterotoxin biomarkers.

i. Library-dependent genotypic methods

This group of methods is based on the unique genetic makeup of different strains or sub species of fecal bacteria. The three most commonly used methods are pulse field gel electrophoresis, ribotyping, and repetitive element PCR.

Pulse Field Gel Electrophoresis (PFGE)

This method involves direct analysis of the microbial genome without performing PCR. Restriction enzymes are used to cut the genomic DNA infrequently, resulting in about 10 to 30 large fragments. These fragments are too large to resolve using standard gel electrophoresis technique because the gel pore size limits their migration. Special gel apparatus is used to overcome this problem, where an electric current is passed through the gel in different directions at a low voltage for 10 to 12 hours (EPA 2005; Stiles, 2003). McLellan et al. (2001) compared rep-PCR with PFGE and found that the PFGE method has higher discriminatory power than the rep-PCR method. PFGE has rapidly become a very useful technique in determining bacterial

relatedness and in epidemiological studies. However, this method is both time consuming and expensive. In addition, the number of isolates that can be processed is limited (EPA, 2005).

Ribotyping

This method is based on the differentiation of genetic differences in the genomic sequences of 16S or 23S ribosomal RNA genes, which are highly conserved among bacteria (EPA, 2005). Total genomic bacterial DNA is cut with different restriction enzymes, followed by the gel electrophoresis. Then Southern blotting is performed to blot the DNA bands onto a nylon membrane from the gel. Southern blot hybridization analysis is then performed using rDNA probes, which results in a pattern composed of four to twelve bands. The different sizes and locations of the bands on the filter can be used to identify the sources that the fecal bacteria came from. This method has been effective in tracking human and nonhuman sources of pollution. Ribotyping, however, is an expensive and labor-intensive method (Stiles, 2003; Scott et al., 2002; Carson et al., 2001).

Hyer et al. (2003) used ribotyping with restriction enzyme EcoRI and PbuII to identify the sources of fecal contamination in three Virginia streams. According to their results, about 65% of the water isolates were assigned to host groups and 35% remained unidentified. In addition, Carson et al. (2003) and Hartel et al. (2002) also successfully used this method for microbial source tracking studies.

Repetitive Element PCR (rep-PCR)

Rep-PCR genomic fingerprinting takes advantage of DNA primers that are complementary to the naturally occurring, highly conserved, non-coding, repetitive DNA sequences, that one present in multiple copies in the genome of most Gram negative and Gram positive bacteria. No previous knowledge of the genomic structure or the nature of the indigenous repeated sequences is necessary (Bacterial Barcodes, 2003; Gresshoff, 1997). This method has been used extensively because it is rapid, simple, and less expensive compared to other genomic methods (EPA, 2005).

Three families of repetitive sequences have been identified, namely the 35-40 bp repetitive extragenic palindromic sequence (REP), the 124-127 bp enterobacterial repetitive intergenic consensus sequence (ERIC), and the 154 bp BOX element. These sequences are located in distinct, intergenic positions around the genome. Repetitive elements are present in both orientations. During polymerase chain reactions, specific primers bind to these specific repetitive sequences and amplify multiple DNA fragments with various lengths (Gresshoff, 1997).

The genetic fingerprints generated using rep-PCR contain multiple bands, which can be subsequently analyzed, categorized by host sources, and used to construct a database to identify the source of an unknown isolate. rep-PCR with the REP primers generates fewer products, but still yields reproducible differentiating fingerprints**.** The ERIC primer is more sensitive to suboptimal PCR conditions, such as the presence of contaminants in the DNA preparation, but generates highly discriminatory patterns. Generally the BOX primer is used in cases where a detailed characterization is needed since it generates highly complex fragment patterns. This method was originally used to

differentiate between closely related strains of bacteria (Gessshoff, 1997). Dombek et al. (2000) reported that the BOX A1R primer is more useful in separating *E. coli* from different sources than the REP primers. They found that 25% fewer PCR products were usually present in the fingerprints generated with REP primers than in the fingerprints obtained with the BOX primer. Classification of human and sheep samples was similar using both primers, where the rate of correct classification (RCC) was 87% and 95%, respectively. However, REP derived fingerprints did not group chicken, cows, ducks, geese and pigs as effectively as BOX derived fingerprints did (Dombek et al., 2000).

McLellan et al. (2003) used both REP- and ERIC-PCR methods to obtain amplified fragments from *E. coli* isolates from humans, gulls and dairy cattle. They reported that ERIC-PCR generated fewer amplified fragments (7-13 product bands) than REP-PCR (13-22 product bands). For all host groups, REP-PCR fingerprints produced higher similarity scores than the ERIC- PCR method. Dombek et al. (2000) used the BOX-PCR method to differentiate *E. coli* isolated from humans and animal sources. Using Jackknife analysis, 100% of chicken and cow and between 78 and 90% of human, goose, duck, pig and sheep isolates were assigned to the correct source group. McLellan et al. (2004) also used the REP-PCR method to identify the relatedness of *E. coli* isolated from environmental samples (storm water, river water, beach water), and host sources (sanitary sewage and gull feces). Results suggested that a large number of strains is needed in order to represent the contributing host sources.

Random Amplified Polymorphic DNA (RAPD) Analysis

Non-selective primers at high stringency have been used to produce a series of species or strain specific PCR products that depend on both the primer and template used. This method is relatively inexpensive compared to ribotyping and PFGE. However, poor reproducibility and lab-to-lab variation have limited its use for MST work (EPA, 2005).

Amplified Fragment Length Polymorphism (AFLP) Analysis

This method was originally developed for plant genome mapping, but its use was later extended to fingerprinting bacterial species. The majority of the AFLP analyses published so far have focused on epidemiological studies (EPA, 2005). Restriction enzymes have been used to digest the genomic DNA of bacteria, and specific primers with PCR reactions the used to amplify the digested fragments. Additional primers are used to run a second round of PCR, which increases the specificity and decreases the number of resultant PCR products. According to a study by Guan et al. (2002), AFLP was compared with multiple antibiotic analysis (MAR) and 16S rRNA gene sequences of *E. coli* isolated from wildlife, human and livestock. Discriminant analysis revealed that AFLP showed better isolate separation into host groups than MAR and 16S rRNA analysis. In another study by Leung et al. (2004), AFLP produced a better rate of correct classification for *E. coli* isolated from cattle, human and pigs than ERIC-PCR. However this method is both time consuming and more expensive compared with other MST methods.

ii. Library-dependent phenotypic methods

Antibiotic Resistant Analysis (ARA)

This method is used for fecal streptococci or *E. coli* to identify the fecal sources by screening isolates against commonly used antibiotics. Fecal bacteria are plated on agar media containing different concentrations of antibiotics. After incubation, each isolate is scored for growth or no growth and resistant patterns that emerge can be used in source differentiation. Since fecal bacteria from humans show greater resistance to antibiotics than those from animal sources, this method is particularly useful for differentiation between human and non-human sources. Three main approaches, antibiotic resistant analysis (ARA), multiple antibiotic resistance (MAR) and Kirby-Bauer antibiotic susceptibility, have been used for MST studies. Most of the researchers prefer to use the ARA method, because this provides more information than either of the other two methods. Hagerdorn et al. (1999) isolated fecal streptococci for ARA. A total of 7058 isolates from human, livestock and wildlife were used to develop an ARA library. The ARCC of the library was 88%, while a comparison of a stream sample against this library showed cattle to be the main contributor of water pollution in that watershed. Booth et al. (2003) and Parveen et al. (1997) also used this method for source identification in watersheds.

The ARA method is extensively used among MST researchers because it is rapid, relatively simple, and inexpensive. However, problems arise due to the tendency of indigenous bacteria to transfer antibiotic resistance genes to fecal bacteria, which enter the environment through fecal matter (Smalla et al., 2000; EPA, 2005).

Carbon Source Profiling (CUP)

This method is based on differences in the nutritional requirements for different bacterial groups. Carbon and nitrogen sources are mainly used for this analysis. The BIOLOG system allows users to rapidly perform and score 96-carbon source utilization tests. However environmental factors can affect bacterial nutrient requirements, this is not a good method for watershed sample analysis (Bush et al., 2003; Stiles, 2003).

iii. Library-independent genotypic methods

Host Specific Molecular Markers

This is a rapid test that detects human fecal pollution by analyzing the members of *Bacteroides-Prevotella* group and genus *Bifidobacterium*. These bacteria have a low survival rate and thus serve as indication of recent fecal pollution. The raw water samples are directly used to characterize microbial population based on the presence/absence of PCR products. In addition, assaying for specific toxic genes or additional host specific genes is used to differentiate bacteria based on their pathogenic properties and the hosts they target. Several *E. coli* strains secrete enterotoxins, which are biochemical substances poisonous to the host in large quantities. The enterotoxin producing *E. coli* strains are ideal for source tracking because their enterotoxins are described in both phenotypic and genotypic ways, and are relatively easy to isolate from a suspected source. However, host specific molecular markers currently only differentiate between human and nonhuman sources and do not provide quantitative information (Bush et al., 2003; EPA, 2005; Scott et al., 2002).

Terminal Restriction Fragment Length Polymorphism (tRFLP)

tRFLP is a method used to determine the diversity of an entire bacterial community by examining differences in the 16S rRNA gene. This method is considered a library independent method because it does not require the isolation of environmental strains but rather depends on the extent to which the DNA sequences from the environmental strains are represented in the available molecular database (Bush et al., 2003).

iv. Library-independent phenotypic methods

Fecal Coliform/Streptococci (FC/FS) Ratio

A comparison of the amount of fecal streptococci and fecal coliforms were used to differentiate the sources of fecal contamination. A ratio of FC/FS greater than or equal to 4.0 indicates human fecal contamination, whereas a ratio below 0.7 is associated with nonhuman contamination (Geldreich, 1970). However, this is not a reliable method, primarily due to differences in fecal enterococci densities found in individuals with different diets and different effects of the environmental factors on survival rates. Additionally, a recent study has showed that FC/FS ratios cannot discriminate between human and domestic animal fecal samples (Bush et al., 2003; Simpson et al. 2002).

F^+ Coliphage Serotyping

F-specific RNA coliphages are viruses that infect *E. coli*. Human and animal feces contain four different serotypes of RNA coliphages. Groups I and IV are generally

associated with animal feces while groups II, and III are more sewage specific, suggesting that phages can be used to determine sources of pollution. However, little is known about the survival rates of those viruses (Bush et al., 2003; Scott et al., 2002; EPA 2005).

Host Specific Indicator Organisms

There are a few bacteria that are more specific to human and/or certain animal species and can therefore be used as indicators of the presence of microbial contamination from particular host species. For instance, if water samples show positive results for ruminant specific *Bacteroides, Rhodococcus coprophilus* and *Streptococus* bovis this indicates that cattle are the source of fecal contamination. Similarly identification of human specific *Bacteroide* (*B. fragilis*) and *Bifidobacterium* species indicate that the source of fecal contamination is human (Boehm et al., 2003; Bush et al., 2003; Scott et al., 2002; EPA 2005).

F. Comparison of genotypic methods

Stoeckel et al. (2004) evaluated seven MST approaches by using *E. coli* isolated from humans, cattle, dogs, horses, swine, Canada geese, chicken and white-tailed deer. Evaluation was mainly based on the reproducibility, accuracy and robustness of these phenotypic (ARA, CUP) and genotypic (RT-EcoRI, RT-Hindlll, PFGE, REP-PCR, Box-PCR) protocols. Challenge isolates (replicates of arbitrarily selected isolates from a known source library that have been re-cultivated) were used to evaluate the reproducibility and accuracy (the ability to correctly identify the sources of isolates that

were collected independently from the known source library). Robustness was evaluated by using Ringer isolates (*E. coli* isolated from warm-blooded animals which did not represent the library), which were used to identify isolates, that came from sources not represented in the library. The results of the study showed that PFGE correctly classified all of the replicates into their correct host groups with a 100% reproducibility. Box-PCR, REP-PCR, RT-EcoRI, RT-Hindlll, ARA and CUP showed 62%, 48%, 54%, 13%, 23% and 20% correct classification rates, respectively. The accuracy test showed that RT-EcoRI obtained a 90% rate of correct classification, although only 5% of the isolates were classified. Other protocols obtained rates of correct classification ranging from 13% to 41%. The RT-EcoRI protocol did not classify any of the 24 Ringer isolates, while PFGE failed to classify 67% and REP-PCR did not classify 33% of the Ringer isolates.

As part of a method comparison study organized by the Southern California Coastal Water Research Project, Myoda et al. (2003) compared PFGE, rep-PCR and ribotyping techniques by using blind samples spiked with feces of 5 known sources. The results obtained by the six investigators were evaluated based on the following criteria: 1) their ability to predict certain values (positive predictive values, negative predictive values, sensitivity, specificity, test efficiency, and false positive rate); 2) their ability to accurately identify human and sewage influent sources; 3) their ability to identify the dominant sources of fecal material contaminating a sample; and 4) their ability to identify all the sources of fecal material contained in a sample. The results showed that the positive predictive rate was higher in ribotyping, followed by rep-PCR and PFGE. Positive predictive values were higher in each method when identifying

samples that only contained sewer and/or human fecal matter. Negative predictive value rate was also highest in ribotying. The specificity of ribotyping ranged from 6% to 67%, while that of rep-PCR was between 7% and 33% and PFGE was 63%. The test efficiency was also higher in ribotyping. All methods correctly identified the dominant source in majority of the samples, however, none of these methods correctly identified the dominant source in all samples. One rep-PCR method correctly identified 83% of the dominant source in samples, while one ribotyping and PFGE correctly identified 75% each. PFGE and one ribotyping technique identified all the sources in 42% of the samples, while ribotyping with enterococci identified 33% and one rep-PCR method identified 17%. Two rep-PCR and one ribotyping methods were not able to identify any of the samples. The false positive rates ranged from 19 to 57%.

Carson et al. (2003) compared ribotyping and rep-PCR for identification of fecal *E. coli* sources. Rates of correct classification (RCC) were used to determine the number of isolates correctly assigned to the proper host class. Using eight host groups, the study found that the RCC for rep-PCR with the BOX A1Rprimer was 88.14% and that for ribotyping was 72.78%. rep-PCR typically generated 18-30 bands during the gel-electrophoresis, while ribotypes gave only 6-12 DNA bands. The rep-PCR method was highly reproducible and produced a high quality pattern about 95% of the time compared to the 85% for ribotyping. In addition, ribotyping is a more rigorous process that requires more skilled technicians and time, and has more individual steps in the procedure. The cost for ribotyping is also higher, and the likelihood of universal application is lower than rep-PCR (Carson et al., 2003).

G. Data analysis

Library-based genotypic bacterial source tracking approaches rely strongly on image and statistical analyses. Several statistical methods have been used to analyze fingerprints and identify the sources.

a. Fingerprint analysis

DNA fingerprints can be analyzed using either band-based or curve-based methods. Band based methods can be used to characterize well-defined fingerprints of low complexity. For example, a collection of fingerprints can be translated into a matrix of binary variables: band present (1), band absent (0). Curve based methods are suitable for highly complex fingerprints with bands of varying intensities. Since these methods are less sensitive to background differences or variations in the amount of PCR products, curve based analyses may provide less biased analyses of DNA fingerprints (Verseveld and Roling, 2004; Albert et al., 2003).

Various classification methods have been used to describe the observations within libraries including cluster analysis, dimensional techniques (e.g., principal component analysis, multi dimensional scaling) and discriminant analysis.

b. Cluster analysis

Cluster analysis is used to find data groups. The similarity between all pairs of samples is calculated and these similarities are expressed in a matrix. These matrixes are visualized a through dendrogram that illustrates the relationships among the isolates (Verseveld and Roling, 2004). Several algorithms are used to develop the dendrogram.

1. Single linkage clustering: Calculates the similarity between two groups based the proximity between the most similar pair (nearest neighbor) of profiles between the two groups. A weakness of the single linkage approach is chaining, which results in thin, poorly separated clusters. It is often used to detect clonal complexes.
2. Complete linkage clustering: Uses the least similar pair (furthest neighbor) profiles between the two groups. It is the exact opposite of the single linkage approach and yields spatially compact clusters.
3. Unweighted pair-group method using arithmatic averages (UPGMA): Has properties that are intermediate between single and complete linkage clustering. This is a frequently used method for rep-PCR DNA fingerprints.

Although cluster analysis is a valuable technique for use in identifying the relationships among isolates, the resulting clusters do not always accurately represent the host groups (Applied Maths, 2002; Verseveld and Roling, 2004; Rademaker and Bruijn, 2004).

c. Dimensional techniques

Principal component analysis (PCA) and multi dimensional scaling (MDS) are two grouping techniques that do not produce hierarchical structures such as those used in cluster analyses. Instead, they produce two-dimensional or three-dimensional plots where entries are distributed according to their degree of relatedness. PCA is a statistical technique based on assumptions about variable interdependence that is used to reduce the dimensions and identify the dependence patterns among the variables. The interdependences between the original set of variables are measured as a correlation or

covariance. The data set is the reduced to a small set of variables, called principle components, which reproduce patterns in the variables that are easy to visualize (EPA, 2005). This method is applicable to all kinds of character data, but not directly to fingerprint data. As a result, fingerprint data must be converted to a band matching table before analysis. MDS is another algorithm that may be used to measure the inter-isolate distances numerically. Patterns of inter-isolate variation can be represented in two or three-dimensional plots. MDS analyzes the matrix of similarities obtained using similarity coefficients, but does not analyze the original character set. This method is based on several assumptions about inter-isolate similarity (Applied Maths, 2002; EPA, 2005; Redemaker and Bruijn, 2004; Verseveld and Roling., 2004).

d. Discriminant analysis (DA) and multivariate analysis of variance (MANOVA)

Discriminant analysis is very similar to PCA. The major difference is that PCA calculates the best discriminating components for the character table as a whole, without knowledge about groups, while DA calculates the best discriminating components for groups that are defined by users. DA works only on complete character data and cannot be used for sequence type data. The information provided by DA can then be used as the basis for MANOVA. MANOVA is another type of statistical technique that allows the significance of user-delineated groups to be calculated. This technique also allows the characters that are responsible for the separation of the delineated groups to be determined. However, this technique can only be applied to composite data sets (Applied Maths, 2002).

H. Evaluation of the performance of a known source library

The goal of developing a known source library is to use that library to identify the sources of fecal contamination in the watershed. This information is vital for water resource managers seeking to determine the total maximum daily load (TDML) levels and implement best management practices (BMPs). Therefore, a good library is essential. EPA's MST Guide Document recommends several universal quality measure parameters: specificity, precision (reproducibility), control samples, quality assurance documentation, and the minimum number of controls. These factors are crucial for the evaluation of the performance of the known source library (EPA, 2005). Three of the above mentioned parameters are discussed here.

a. Specificity

According to EPA's MST guide document (2005), specificity is defined as "the ability of a particular MST method to discriminate between different animal fecal sources." Specificity can be calculated using the following formula:

$$TN / (TN + FP) \times 100\,\%$$

TN represents the test negatives and TN + FP (false positive) represents the total number of negative samples (EPA, 2005).

i. Cross validation test

The use of cross validation test is popular among MST workers in order to discriminate the host groups. Jackknife analysis is used to determine the accuracy of isolates assigned to the host group based on their maximum similarity coefficient. First,

isolates are manually assigned to the host groups and then each isolate is matched against all other isolates in the database. The percentage of isolates that are identified as being in the correct group where they were originally assigned is then calculated. This is known as the rate of correct classification (RCC) and based on the individual rates, the average rate of correct classification (ARCC) for the database is calculated (McLellan et al., 2003). RCC will be low if the library size is small, but increases as the number of isolates increases. In Hagedon's (2004) study, RCC was 27% for a library containing only 100 cattle isolates but that percentage increased to 79% when the number of isolates in the library was raised to 400. This information can also be used to determine how many isolates per source are needed to achieve an acceptable correct classification rate. In practice, 100% classification rate is almost never observed, particularly in large libraries. However, RCC percentages between 50% and 80% are useful for watershed analysis (Hagedon, 2004).

ii. ID Bootstrap analysis

ID Bootstrapping is a script file designed to calculate the correct classification rate. The script applies a bootstrap algorithm in order to estimate the likelihood of obtaining an observed similarity score by chance. Similarity matrixes are first calculated for all known samples, which are assigned to particular groups, and then each unknown is compared with each group of known samples. This provides an average (or maximum) similarity for each unknown for each group, as well as probability that each isolate has been correctly classified. Each bootstrap iteration involves 30 or more group members and a single non-member (Ritter et al., 2003; Verseveld and Roling, 2004).

b. Precision or reproducibility

Reproducibility is important to all fingerprint based methods and can be measured by using replicates. There are two main categories of replicates: identical replicates and experimental replicates. Identical replicates are the DNA banding patterns obtained from the same isolates, run as two different sets of PCR under the same conditions. It is recommended that 10% of isolates in the known source library should be replicated (EPA, 2005). Experimental replicates are the DNA banding patterns obtained under the same experimental conditions from different isolates of the same host group (EPA, 2005).

c. Control samples

Control samples (both positive control and negative control) are used to measure the performance of MST methods and to screen for the presence or absence of contaminants. Negative controls are used to monitor the introduction of contaminants into the experiment, indicating the poor aseptic conditions. Positive control measures whether the MST method is performing adequately (EPA, 2005) and can be used as an indicator for reproducibility. One or more *E. coli* known isolates is included in each PCR reaction as a positive control to quantify the densitometric curve variation due to PCR reactions, gel normalization, DNA loading and thermocyler differences. Cluster analysis is used to assess the similarity coefficients generated from the same isolates (positive control) (Albert et al., 2003). McLellan (2004) reported that in their study, *E. coli* strain K-12 was included in every PCR setup as a positive control. They found no differences in banding patterns from the same template and reported that similarity

scores ranged from 87.9-99.5%, which indicated on adequate performance by the rep-PCR reactions and the reproducibility of the *E. coli* data.

I. Source identification using known source library

For source identification using a library, BioNumerics software offers 4 options: mean similarity, maximum similarity, K-nearest neighbor and neutral network. Under the mean similarity option, the software finds the similarity between the unknown entry and each entry in the library unit. It then calculates the average similarity for the entire library unit for use in the identification process. For the maximum similarity option, all similarities between unknown entries and the library units are calculated but only the highest similarity value found is used in the identification process. For each mean or maximum similarity coefficient, a quality quotient is calculated. This quality quotient indicates the confidence of the identification by taking into account the internal spread of the group (Applied Maths, 2002).

J. Study area – Catoma Creek of Montgomery County, Alabama

Montgomery County is located in the south-central part of Alabama and in the northern part of the Coastal Plain. It is about 225 km from the Gulf of Mexico and covers an area of 2045 km^2 (Soil survey, 1960). The Catoma Creek watershed consists largely of two major formations: Alluvial Coastal and Low Terrace Deposits within the immediate stream drainage areas and Mooreville Chalk located outside those immediate areas. All streams within the Catoma Creek watershed are classified as part of the

Southeastern Plains Sub Ecoregion, which consists of irregular plains with broad inter-stream areas comprising croplands, pasture, woodlands, and forest. Blackland prairies cover the greatest portion of the watershed, and the parent material is Cretaceous-age chalk, marl, and calcareous clay (ADEM, 2002).

The county has a subtropical climate with an average annual temperature of 17.4°C. The average January temperature is 7.6° C while that of July is 26.7°C. Annual rainfall is 1356 mm (onlinemontgomery.com). The total population of the county is 223,000 according to the year 2000 census (onlinemontgomery.com).

Catoma Creek is a tributary of the Alabama River and therefore a part of the Alabama River basin. The linear length of the Creek is 68 km and it drains an area of 932 km^2. Predominant land uses within the watershed consist of forest and wetlands (54.5%), pasture and hay (21.6%), and row crop activities (14.6%) (Fig.1.1) (ADEM, 2002). Presently this creek is classified for *Fish and Wildlife*. A 37-km section of Catoma Creek from Ramer Creek to the Alabama River has been included on the Section 303(d) List due to organic enrichment and high fecal coliform density (ADEM, 2002). Dissolved oxygen impairment typically occurs during the summer months (May through November). When dissolved oxygen falls below 5 mg/l, a stream is no longer suitable for fish and wildlife. This may be the result of point and non-point discharges, the decay of organic matter, algal bloom due to eutrophication, and/or sediment oxygen demand. Approximately 17% of the water samples tested for fecal coliforms in 2000 and 2001 exceeded the single sample maximum criterion of 2000 CFU/100 ml (ADEM, 2002), and 43% of the water samples collected from 1991 to 2001 exceeded the same criterion (M.A Watson, personal communication).

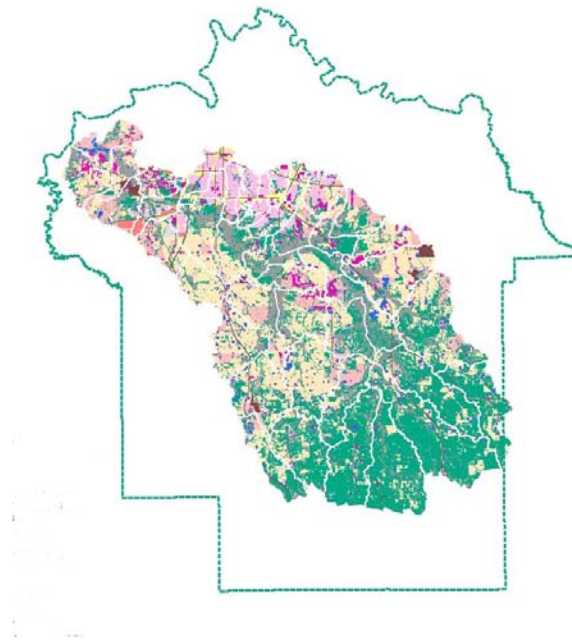


Fig. 1.1 Land use pattern in the Catoma Creek watershed of Montgomery County, Alabama.

II. MONITORING FECAL CONTAMINATION AND NUTRIENT ENRICHMENT IN THE CATOMA CREEK WATERSHED

ABSTRACT

High concentrations of fecal indicator bacteria are the most common cause of surface-water impairment in Alabama. Catoma Creek in Montgomery County has been included on the state Section 303(d) List due to elevated concentrations of fecal coliform bacteria and organic enrichment. Forest, agriculture, and urban land uses represent 54.5%, 36.2%, and 9.3%, respectively, of the watershed. Objectives of this study were to monitor the fecal contamination and the nutrient enrichment in the Catoma Creek watershed. Water samples were collected monthly at eight locations in the watershed over a one-year period. *E. coli* was enumerated using the modified m-TEC media. Data showed that *E. coli* concentrations varied from 18 to 12,650 CFU/100 ml, with 70% of the samples exceeding the EPA criterion for swimming water. A good correlation between flow rates and *E. coli* concentrations was found. Chemical analyses of the water samples showed that concentration of total phosphorus in all samples was above the proposed Ecoregion IX nutrient criterion, 78% of samples were above the NO_3-N criterion and 50% of samples were above the total nitrogen criterion. These results suggest that there is a serious risk of eutrophication in this watershed.

INTRODUCTION

Fecal contamination of surface water threatens both ecosystem and human health. This is not a problem limited to one country or region; globally hundreds of thousands of people die each year due to water borne diseases (WHO, 2001). In the United States, the Clean Water Act was enacted in 1972 to achieve the goal of "all water to be swimmable and fishable" (Clean Water Act, 1972). However, fecal contamination of surface water is still a major issue in the United States, where 40% of the waterways cannot be used for swimming (EPA, 1999). In Alabama, 45% of the state's watersheds do not meet EPA standards due to fecal contamination (Southern Forest Resource Assessment, 2001). Under the Clean Water Act each state should identify the water-bodies which do not meet their designated use, monitor the severity of this impairment, and develop an appropriate Total Maximum Daily Load (TMDL) for these waters (EPA, 1986).

Fecal contamination of surface water can be detected using indicator bacteria such as *E. coli,* enterococci or fecal coliforms. *E. coli* is the most widely known fecal coliform bacteria indigenous to the intestinal track of humans and other warm-blooded animals. It enters the environment with feces (An et al., 2002). Thus the presence of *E. coli* in any water source indicates that the water has been contaminated by waste from humans and other warm-blooded animals.

A high correlation between the *E. coli* density in fresh water and intestinal illness was identified by EPA after conducting a series of epidemiological studies. Based on these results, *E. coli* was recommended as one of the indicator bacteria for

fresh water. According to EPA recommendations, the 30-day geometric mean for swimming water is 126 CFU/100 ml while drinking water should be completely free from *E. coli* (EPA, 1986). *E. coli* is considered to be a good indicator bacterium because most of the strains are non pathogenic to humans, there is no evidence of its replication in water, it is easy to culture, and it has a short generation time (Scott et al., 2002). Potential sources of fecal contamination include municipal wastewater discharge, septic tank leachate, pasture and agricultural land runoff, urban run off, and wildlife.

A positive relationship has been found between *E. coli* concentrations in surface water and rainfall. Increase of *E. coli* concentrations in water can be found during wet periods, especially after a storm, possibly due to transportation of fecal material from the surrounding area and resuspension of bacteria into the water from sediments that have been disturbed by increased water currents (Crabill et al., 1999; An et al., 2002). *E. coli* has been found to be associated with sediments. According to a study in Oak Creek, Arizona, fecal coliform concentration in sediment was 2200 times higher than that in a water column (Crabill et al., 1999). The survival rate of fecal bacteria may be higher in sediment due to the presence of organics, a decreased level of UV radiation, and the presence of anaerobic regions in sediments (An et al. 2002; Enzinger and Cooper, 1976).

This study was conducted on the Catoma Creek watershed of the Alabama River basin. Catoma Creek was added to Alabama’s 303(d) List in 1996 due to organic enrichment, and low dissolved oxygen, and since 2002 has also been found to be contaminated by for fecal pollution (ADEM, 2002). It is therefore necessary to

determine the severity of fecal pollution and nutrient enrichment in the creek and to develop appropriate TDML levels.

The objectives of this research were: 1) to monitor the fecal contamination of the Catoma Creek watershed using *E. coli* as an indicator bacteria, and 2) to monitor the nutrient enrichment in the watershed.

MATERIALS AND METHODS

Study site

The study site is located in the Catoma Creek watershed in Montgomery County, Alabama. Catoma Creek is a tributary of the Alabama River and the watershed covers more than 50% of land in the county. Its drainage area is 932 km^2 and the linear length of the stream is 68 km. Forest, agriculture (pasture and row crop), and urban land uses represent 54.5%, 36.2%, and 9.3%, respectively, of the watershed (ADEM, 2002). The average annual rainfall in the Catoma Creek watershed is 1356 mm, while the average temperature during the winter is 7.6º C and during the summer is 26.7º C. According to the 2000 census, the total population of the county is 223,000 (onlinemontgomery. com). The present classification of this stream is “fish and wild life”. A 37-km segment from Ramer Creek to the Alabama River is listed on the 303(d) List of impaired water bodies in Alabama due to nutrient enrichment and elevated fecal coliform concentrations (ADEM, 2002).

The Catoma Creek watershed is located in the Nutrient Ecoregion IX: Southeastern Temperate Forest Plains and Hills. Nutrient Ecoregions are defined as “broad areas that have general similarities in the quantity and types of ecosystems as

well as natural and anthropogenic characteristics of nutrients" (EPA, 2000). EPA has documented nutrient reference condition criteria for each Ecoregion and provided guidelines for each state to establish its own guidelines (EPA, 2000).

Collection of water samples

Water samples were collected monthly from 8 locations in the Catoma Creek watershed, 3 of which are on the main stem of the Creek (Catoma at Old Selma Road (CO), Catoma at Court Street (CC), and Catoma at Woodley Road (CW), and 5 on its tributaries (Baldwin Slough (BS), White Slough (WS), Ramer Creek at Sprague Junction Road (RSP), Ramer Creek at Snowdown Chamber Road (RSC), and Little Catoma Creek (LT) (Fig. 2.1). Water samples were collected from May 2003 to April 2004. The hand dip method was used to collect water samples from Baldwin Slough, White Slough and Ramer Creek at Sprague Junction Road, where the water was shallow, and a sampling rod was used at the other five locations, where the water was too deep to wade. Two water samples were collected from each sampling site using sterile polyethelene bottles, kept on ice, and transported to the laboratory. Samples were processed within 6 hours. A blank sample was taken randomly at one sampling site each month. Stream water temperature and pH were measured on site. The flow velocities at Baldwin Slough, White Slough and Ramer Creek at Sprague Junction Road were measured across the streams at a 3 feet distance using a Flo-Mate portable flow-meter (Marsh-Mc Birney Inc., Frederick, MD) each month. The height of the water level with respect to each velocity and the stream width were recorded and used to calculate the mean flow rate (Mean flow rate = average velocity * stream width * average depth of

the flow). USGS flow data (USGS Water Resources of Alabama) were used for Catoma Creek at Court Street.

Enumeration of *E. coli* in water samples

From each sample, 3 dilutions (1 ml, 10 ml, 20 ml or adjusted according to the water level of the streams) were taken and filtered through 0.45 µm membrane filters using vacuum filtration. Membranes were placed on the modified membrane-thermotolerant *Escherichia coli* agar (m-TEC) media (Difco, Detroit, MI) and incubated at 37ºC for 2 hours and then incubated at 44.5ºC for 24 hours. Colonies which have given magenta color was identified as *E. coli* and only colonies between 20 and 80 were counted and reported as colony forming units per 100 ml of water (EPA, 2002).

Chemical analysis

Every three months, two additional water samples from each site were collected using sterile polyethylene bottles, kept on ice and transported to the laboratory. Total nitrogen, NH_4-N, NO_3-N, total phosphorus and micronutrient levels in the water samples were measured. NH_4-N and NO_3-N were determined using a modified indophenol method adapted to microplate format (Sims et al., 1995). Total nitrogen was measured using the Kjeldahl method described by Bremner (1965) and total phosphorus and micronutrients (Ca, K, Mg, Al, Zn, Cu, Mn, and Fe) were determined using ICAP (SPECTRO CIROS, Germany).

RESULTS AND DISCUSSION

E. coli concentration in the Catoma Creek watershed

The *E. coli* concentrations found in the Catoma Creek watershed from May 2003 to April 2004 are shown in Fig. 2.2. Since *E. coli* was found in all the water samples, this indicates that the Catoma Creek watershed was contaminated by fecal materials from humans and/or other warm blooded animals. Table 2.1 shows the concentrations of *E. coli* found at each site by month. Average concentrations of *E. coli* varied from 18 CFU/100ml in March at the CW site to 12,650 CFU/100 ml in November at the same site. The highest *E. coli* concentrations recorded at the CO, CC, BS, WS, RSC, RSP and LT sites were 10,050, 9,850, 8,517, 2,533, 12,200, 6,100 and 6,500 CFU/100 ml, respectively. Of the 96 water samples collected during the one-year period, 70% exceeded the EPA's 30-day geometric mean criterion while 42% exceeded the single sample maximum value. Similar levels of fecal pollution have been reported in other studies. According to a study in Michigan, 5 out of 9 sampling sites showed *E. coli* concentrations that were above allowable state limits for recreational water (Tam et al., 2005). Davis-Colley et al. (2004) reported *E. coli* concentrations of 50,000 CFU/100 ml at a cow-crossing stream in New Zealand, while Howell et al. (1995) found that fecal coliforms in Western Kentucky streams exceeded water quality standards between 87 and 100% of the time.

E. coli concentrations in all the water samples collected from the RSP site were above the 30-day geometric mean, indicating high levels of fecal contamination at this site. CO, CC, BS, WS, CW, and RSC had 75%, 75%, 58%, 83%, 58%, and 75% of their samples above the 30-day geometric mean. The LT site had the lowest percentage

of water samples above the 30-day geometric mean (42%), as well as the lowest percentage above, single sample maximum value (33%). However, the water at this site was turbid and stagnant during the entire study period. It is possible that *E. coli* settled to the bottom of the stream and resulted in lower concentrations in the water. Another possibility is that the Little Catoma site had lower fecal material loading.

E. coli occurrence with rainfall

Fig. 2.3 and Table 2.2 show the average monthly rainfall and the daily rainfall for the Catoma Creek watershed, respectively. The highest rainfall was recorded during the summer months. It seems reasonable to assume that when rainfall is high, fecal loading into the stream increases. However, in this watershed most of the sites showed lower *E. coli* concentrations during the summer months. This may be due to the higher grazing of *E. coli* by protozoa in the summer, and an increased die-off rate caused by greater exposure to UV radiation (An et al., 2002).

Fig. 2.4 shows the percentage of samples above the 30-day geometric mean criterion. All the sites except White Slough recorded their highest number of *E. coli* in November. Referring to the rainfall data for the Catoma Creek watershed (Table 2.2), the day before the water samples were collected this watershed had a heavy rainfall event of 48 mm, after 22 days of dry spell. Thus, bacteria associated with fecal matter that had collected in the watershed were washed into the stream with the runoff water, resulting in high numbers of *E. coli*. In September, only the RSP site and in October only the WS and RSP sites showed *E. coli* concentrations that were above the 30-day geometric mean. Both months experienced dry spells prior to collecting water samples,

therefore there was little fecal matter loading into the streams and correspondingly low concentrations of *E. coli*.

Data for the Catoma Creek watershed did not show there to be a good correlation between rainfall amount and the *E. coli* concentration in the stream. The highest correlation coefficient (r) was 0.28 at Little Catoma, while the lowest was 0.08 reported at Baldwin Slough. However, there were some exceptions. During October this watershed experienced mostly dry weather and in November 48 mm of rainfall was received a day before the water samples were collected. *E. coli* concentration increased over the sampling sites ranged from the lowest of approximately 3-fold to the highest of approximately 360-fold of that of the previous month. This suggests that rainfall is a significant factor, which controls surface runoff of fecal material into streams. It is also important to take into account the length of dry spell prior to collecting water samples and if there was any rainfall, and its intensity. Rodgers et al. (2003) studied the effect of hydrological events on fecal coliform concentrations in stream water. After the first significant rainfall event, fecal coliform concentrations in the watershed increased by 4 to 100 fold over the previous samples. They reported that fecal concentrations decreased with the increasing number of rainfall events, suggesting the depletion of the fecal matter on the land. Another study was performed in New Zealand to examine the correlation between over-land flow and the concentration of fecal bacteria in streams. The results suggested that overland flow was responsible for delivering fecal bacteria directly into the streams (Collins et al., 2005).

***E. coli* occurrence with flow rate**

E. coli concentrations at four sampling sites (CC, WS, BS and RSP) were plotted against stream flow (Fig. 2.5). The correlation coefficients (r) were 0.84, 0.81, 0.80 and 0.91, respectively, with P values of 0.001. These data demonstrate that there was a positive correlation between flow rates and *E. coli* concentrations. However, the correlation was not apparent when the data from all four sites were pooled.

Limited information is available regarding the effect of stream flow on *E. coli* concentration. Jagal (1997) showed a positive relationship between fecal indicator bacteria and stream flow. Donnison et al. (2004) monitored *E. coli* concentrations at 14 sampling sites over a 2-year period in a New Zealand rural watershed. They did not find an obvious correlation between *E. coli* concentration and stream flow or rainfall.
Crabill et al. (1999) reported a drastic increase of fecal coliforms in water after a storm event. They suggested that runoff water transports the fecal material that has accumulated in the watershed. In addition, increased water currents disturb the bottom sediments, which can be a reservoir of fecal coliforms. Sheerer et al. (1992) also found that benthic sediments harbor fecal coliforms and fecal streptococci organisms and their survival rates were higher than in the water. Muirhead et al. (2004) used artificial flood events (without overland flow from a watershed) to quantify *E. coli* associated with stream sediments. According to this study, the *E. coli* concentration in the streambed was 10^8 cfu/m^2 and there was a strong correlation between *E. coli* concentration and turbidity. This also suggested that high flow rates may disturb the sediments and suspend the *E. coli* that are normally associated with the sediments.

Water pH and temperature

A study carried out by Grandjean et al. (2005) found that *E. coli* culturability is optimum around pH 8.2 or higher. In the Catoma Creek watershed, the average monthly pH varied from 5.70 in May to 7.42 in March. Table 2.3 shows the pH values by month at each site. Low pH values found in May might adversely affect *E. coli* survival. The average rainfall in month of May was 136 mm, which was relatively high, but *E. coli* concentrations were relatively low. In contrast, in March this study area recorded the highest average pH, but except for BS (5200 CFU/100 ml) and RSP (2200 CFU/100 ml) site, all other sampling sites showed very low *E. coli* concentrations. There was thus no obvious correlation between *E. coli* concentration and water pH.

Fig. 2.6 and Table 2.4 show the average temperature and the temperature of each site for each month, respectively. There were no significant temperature differences among the sampling sites. The average temperature across the sampling months varied from 9.9ºC to 27.6ºC. Since *E. coli* is thermotolerant, this temperature range is unlikely to affect their survival. As expected *E. coli* concentrations showed no correlation with water temperature.

Nutrient enrichment in the Catoma Creek watershed

The ranges of the chemical water quality parameters at the Catoma Creek watershed are shown in Table 2.5. Fig. 2.7 shows the concentrations of NO_3-N, total N and total P at each site, along within the Ecoregion IX nutrient criteria. The level of NO_3-N ranged from 0 to 1.18 mg/l, which was lower than the EPA drinking water MCL of 10 mg/l (EPA, 2002), while 78% of the samples were above the Ecoregion IX

reference level for NO_2 + NO_3 –N of 0.125 mg/l (EPA, 2000). Spring water samples showed the lowest NO_3-N concentrations while winter samples showed the highest concentrations at most sampling sites (Fig. 2.7). Ahearn et al. (2004) reported a strong seasonal cycle of each chemical (total suspended solids, NO_3-N, total nitrogen, PO_4-P total phosphorus, specific conductivity and flow) during their four-year study, with the highest NO_3-N nitrogen concentration during the winter period. NH_4-N concentrations varied from 0 mg/l to 0.78 mg/l. Winter samples showed the lowest NH_4-N concentration while fall samples showed the highest NH_4-N concentration. None of the samples exceeded the EPA criterion for fresh water aquatic life at corresponding pH and temperature (EPA, 1999).

Total N concentrations varied between 0 mg/l in the summer at all sampling sites and 4.71 mg/l in the fall at the CO site. The Ecoregion IX criterion for total nitrogen is 0.692 mg/l (EPA, 2000). Of the total 32 samples, 50% showed total nitrogen concentrations above the Ecoregion IX criterion. All the samples collected during the summer showed 0 mg/l total nitrogen concentrations, while the fall and winter samples showed relatively high concentrations of total nitrogen. Aquatic plants such as algae and plankton grow during the spring and summer and these plants will absorb nitrogen in the water during those periods, resulting in low nitrogen levels. During the fall and winter these plants decay and release N into the water, which thus produces higher N levels in the water.

Phosphorous is the most limiting nutrient for algae growth (Matlock et al., 1998). The Ecoregion IX nutrient criterion for total phosphorus is 0.036 mg/l (EPA, 2000). Thus, any stream in this Ecoregion having phosphorus concentrations above

0.036 mg/l may run a risk of eutrophication. The Catoma Creek data show that total phosphorus concentration of each water sample was consistently higher than this criterion. Another often cited total P critical concentration for stream eutrophication is 0.1 mg/l (EPA, 1986). Again, all the water samples from Catoma Creek exceed this level. Therefore, there is a high risk of eutrophication in this watershed. Unlike N, the P level was higher in the spring than in any other season at 5 out of 8 sites. Agricultural activities such as land preparation and manure application mainly occur in the early spring, and organic and inorganic P may be washed to the streams through runoff. Similar to nitrogen, P also showed lower concentrations during the summer, possibly due to aquatic plant uptake.

Fig. 2.8 shows the yearly average concentrations of NH_4-N, NO_3-N, total N and total P at each sampling site. RSP is the furthest sampling site and CO is the nearest sampling site to the Alabama River. All the sampling sites showed high concentrations of total N. BS showed the highest total N concentration while RSC showed the lowest concentration among the sampling sites. It was noticeable that RSC had the lowest chemical concentrations in most cases. Although the total P concentration was above the Ecorgion IX criteria at all sampling sites, the LT site showed the highest average concentration, 9-fold above the EPA Ecoregion IX criterion for total P, suggesting that there was a high risk of eutrophication at this site. Overall we did not find any correlation between *E. coli* and nutrient concentrations.

SUMMARY

During the one-year study period, 70% of the water samples collected from the Catoma Creek watershed had *E. coli* concentrations above the 30-day geometric mean criterion and 42% of the samples exceeded the single sample maximum criterion. Since all the water samples were positive for *E. coli*, this indicates that the Catoma Creek watershed was contaminated with feces of human and /or other animal origin. This elevated level of *E. coli* concentration suggests the presence of pathogenic organisms in these waters. Local communities should therefore avoid using these waters for body contact recreational activities.

Overall, there was no good correlation between *E. coli* concentration and rainfall. However, after a long dry spell, the *E. coli* concentration after a significant rainfall event was 360-fold higher than the previous dry month. Therefore, the rainfall is a contributor to fecal loading into the streams. A positive correlation between *E. coli* concentration and flow rate was found in this watershed. Elevated total phosphorus, total nitrogen and NO_3-N levels at the sampling sites indicate a high risk of eutrophication in this watershed. Implementation of best management practices (BMP) in this watershed is therefore essential to reduce bacteria and nutrient loads to stream water. However, before implementing BMPs, it is important to identify the sources of the fecal contamination.

Table 2.1 Concentrations of *E. coli* at the Catoma Creek watershed from May 2003 to April 2004

Location	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
CO	423 (109)	560 (122)	183 (32)	730 (57)	115 (29)	28 (7)	10050 (1626)	138 (18)	3350 (212)	7250 (212)	68 (13)	359 (236)
CC	265 (21)	623 (378)	208 (75)	176 (30)	115 (7)	100 (14)	9850 (71)	170 (7)	3300 (424)	6800 (849)	196 (45)	70 (5)
BS	150 (21)	8517 (1296)	3750 (71)	44 (7)	92 (7)	40 (26)	3600 (283)	20 (8)	242 (13)	2250 (71)	5200 (424)	43 (11)
WS	897 (84)	2533 (589)	133 (50)	82 (28)	108 (4)	336 (58)	1086 (71)	130 (14)	194 (42)	2500 (0)	175 (22)	211 (19)
CW	138 (53)	591 (139)	138 (32)	94 (22)	97 (33)	52 (12)	12650 (636)	150 (0)	3350 (495)	3800 (424)	18 (8)	80 (18)
RSC	210 (28)	253 (25)	315 (24)	138 (4)	103 (22)	51 (18)	12200 (141)	218 (12)	3150 (71)	9000 (283)	123 (18)	134 (42)
RSP	404 (43)	585 (201)	198 (39)	530 (85)	250 (13)	142 (14)	6100 (566)	242 (26)	465 (40)	1127 (286)	2200 (0)	234 (40)
LT	103 (38)	385 (59)	175 (21)	41 (17)	20 (3)	20 (11)	6500 (141)	77 (9)	3450 (212)	6200 (424)	55 (18)	20 (5)

Standard deviations are shown in parentheses for each data point.

Table 2.2 Daily rainfall values (mm) for the Catoma Creek watershed from May 2003 to April 2004, measured near the Catoma Creek at Court Street, Montgomery (Source: www.waterdata.usgs.gov))

DATE	May 2003	Jun 2003	Jul 2003	Aug 2003	Sep 2003	Oct 2003	Nov 2003	Dec 2003	Jan 2004	Feb 2004	Mar 2004	Apr 2004
1	0	0	81.5	0	0	0	0	0	0	0	0	0
2	0.25	0	0.75	0.50	0	0	0	0	0	0	0	0
3	14.0	14.0	0	0	0	0	0	0	0	0	0	0
4	0	1.00	28.3	0	7.75	0	0.25	2.25	0	0	0	0
5	0	0	9.5	22.3	0.28	0	0	0	6.0	0	0	0
6	0	22.2	0	33.0	66.8	1.00	0	0	0	22	0	0
7	0	6.00	0	4.25	15.0	0.75	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	41.8
9	0	0	0	0	0	---	0	*10.8*	0.75	0	0	0
10	0	0	3.75	0	0	22.5	0	17.3	0	0	0	0
11	38.8	2.25	0.25	5.5	0	0	0	0	0	5.25	0	0
12	0	24.3	0	0	0	0	0	0	0	3.75	0	13.5
13	0	8.75	6.5	26.5	0	0	0	1.75	0	0.50	0	2.0
14	4.75	0.75	0.5	0	0	0.25	0	0.25	0	4.25	0	0
15	11.0	3.25	0	2.75	0	0	0	0	0	2.5	0.50	0
16	0	6.5	0	56.3	0	0	0	4.75	0	0	*7.0*	0

Table 2.2 (cont.)

17	0	*6.75*	0.25	0.25	0	3.25	0	0	0	0	0	0
18	14.5	8.00	0	0	0	0	48	0	0	0	0	0
19	0.75	0	6.50	0.75	*0*	0	*0*	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	*0*
21	28.0	0	0.50	*0*	37.3	0	0	0	0	0	0	0
22	8.25	0	*20.0*	0	17.3	0	0	0	0	0	0	0
23	0	0	7.75	0	0	0	0	17.5	0	11.3	0	0
24	0	0	0	0	0	0	9.0	0	2.0	*0*	0	0
25	14.8	0	0	7.25	0	0	0	0	6.75	40	0	0
26	0.75	0	0	0	0	9.25	0	0	2.0	5.0	0	11.0
27	0	4.25	0	0	0	0	14.5	0	*0*	0	0	0
28	*0*	0.25	2.75	0	0	0	0.75	0	0	0	0	0
29	0	0	0	0	0	0	0	8.50	0	0	0	29.8
30	0	25.0	0	0	0	*0*	0	0	0		14.3	9.0
31	0		7.75	0		0		0	0		0.25	
Total	136	133	196.8	159	147	37.0	72.5	63.0	17.5	94.5	22.0	107

Numbers in italics indicate the sample collection date.

Table 2.3 pH values of the Catoma Creek watershed from May 2003 to April 2004

Location	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	**Annual Avg.**
CO	**5.49**	7.26	7.26	7.28	7.12	6.96	6.82	6.66	6.79	6.57	7.58	6.5	**6.86**
CC	5.5	7.15	7.16	7.12	7.05	6.81	6.9	6.86	6.75	6.54	7.75	6.52	**6.84**
BS	5.61	7.68	7.73	7.42	7.51	**7.91**	7.19	7.37	7.32	6.83	7.68	6.28	**7.21**
WS	5.72	7.7	7.59	7.22	7.25	6.99	7.25	7.01	7.32	6.84	7.32	7.2	**7.12**
CW	5.76	6.88	7.1	7.18	6.85	6.75	6.47	6.9	6.78	6.62	7.59	6.33	**6.77**
RSC	5.95	7.39	7.32	7.33	7.14	7.2	6.61	6.94	6.88	6.54	6.72	6.66	**6.89**
RSP	5.77	6.99	7.15	7.61	7.04	7.19	6.8	6.97	7.05	6.51	7.14	7.09	**6.94**
LT	5.8	6.55	6.84	6.8	7.02	6.98	6.65	6.44	6.74	6.55	7.56	6.67	**6.72**
Monthly Avg.	**5.70**	**7.20**	**7.27**	**7.25**	**7.12**	**7.10**	**6.84**	**6.89**	**6.95**	**6.63**	**7.42**	**6.66**	**6.92**

Table 2.4 Water temperature of the Catoma Creek watershed sampling sites from May 2003 to April 2004

Location	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	**Annual Avg.**
CO	22	24.5	27	25	23	15.5	16	10	11	10	18.5	20	**18.5**
CC	22	24.4	27	27	22	17	15.5	10	10	10	18	21	**18.7**
BS	28	29	30	**31**	25	20	16	11	10	11	20	21	**21.0**
WS	23	25.5	28	27	22	15.5	14	**7**	10	10.5	18	19	**18.3**
CW	22	24.5	27	28	23	16.5	16	9	11	10.5	18.5	20	**18.8**
RSC	22	24.5	27	27	23	17.5	16	9.5	11	10	19	19.5	**18.8**
RSP	23.5	25	27	27	21.5	18	15	12	11	11	19	19	**19.1**
LT	25	26	27	29	25	20.5	17	10.5	12	11.5	19.5	22	**20.4**
Monthly Avg.	**23.4**	**25.4**	**27.5**	**27.6**	**23.1**	**17.6**	**15.7**	**9.9**	**10.8**	**10.6**	**18.8**	**20.2**	**19.2**

Table 2.5 Ranges of chemical water quality parameters at the Catoma Creek watershed in summer, fall, winter and spring of 2003-2004.

Location	EC	Total N	NO_3-N	NH_4-N	P	Ca	Al	Cu	Fe	K	Mg	Mn	Zn
	mmhos/cm	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l
CO	0.10-0.20	0.00-4.71	0.00-0.28	0.00-0.40	0.20-0.29	18.8-44.4	0.00-0.43	0.00-0.03	0.00-0.68	1.58-2.94	1.05-2.24	0.00-0.09	0.00-0.02
CC	0.11-0.20	0.00-2.27	0.10-0.27	0.00-0.39	0.16-0.23	19.7-41.6	0.00-0.34	0.00-0.03	0.00-0.99	1.71-4.52	1.05-1.75	0.00-0.04	0.00-0.02
BS	0.12-0.49	0.00-3.77	0.20-0.53	0.00-0.42	0.14-0.28	25.5-104	0.00-0.23	0.00-0.03	0.00-0.51	2.06-3.16	0.52-2.03	0.00-0.05	0.00-0.02
WS	0.14-0.36	0.00-1.57	0.02-1.18	0.03-0.34	0.15-0.27	32.4-79.5	0.00-0.41	0.00-0.03	0.00-0.51	1.69-2.63	0.62-1.54	0.00-0.03	0.00-0.01
CW	0.06-0.15	0.00-2.35	0.00-0.32	0.01-0.33	0.15-0.25	11.5-36.0	0.00-0.41	0.00-0.03	0.00-1.00	1.00-2.45	0.87-2.15	0.00-0.05	0.00-0.01
RSC	0.13-0.24	0.00-1.43	0.00-0.32	0.00-0.35	0.13-0.34	26.1-50.4	0.00-0.28	0.00-0.03	0.00-0.81	1.59-3.91	1.44-2.57	0.00-0.03	0.00-0.02
RSP	0.12-0.18	0.00-2.04	0.00-0.54	0.00-0.42	0.14-0.52	23.5-35.6	0.00-0.21	0.00-0.03	0.00-0.80	0.97-3.09	1.55-2.41	0.00-0.09	0.00-0.02
LT	0.07-0.2	0.00-3.41	0.21-0.65	0.00-0.78	0.22-0.66	13.5-45.3	0.00-0.23	0.00-0.03	0.00-0.46	1.32-4.74	0.82-1.86	0.00-0.09	0.00-0.04

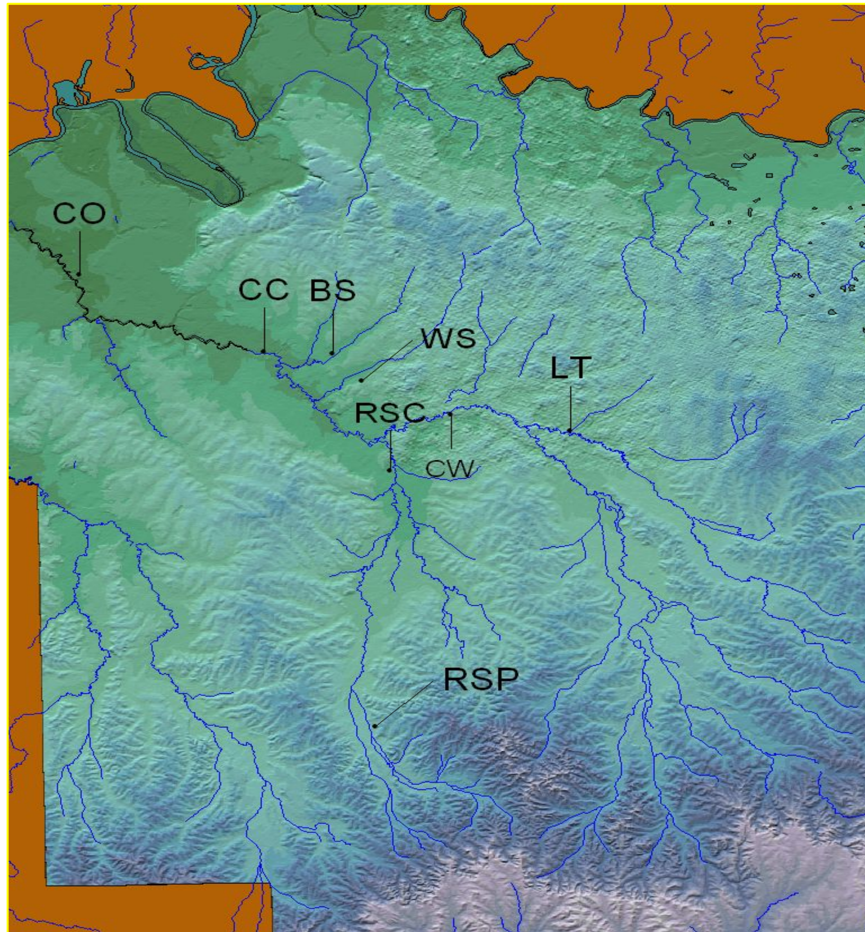


Fig. 2.1 Sampling sites in the Catoma Creek watershed: Catoma Creek at Old Selma Road (CO), Catoma Creek at Court Street (CC), Catoma Creek at Woodley Road (CW), White Slough (WS), Baldwin Slough (BS), Little Catoma Creek (LT), Ramer Creek at Snowdown Chamber Road (RSC), and Ramer Creek at Sprague Junction Road (RSP)

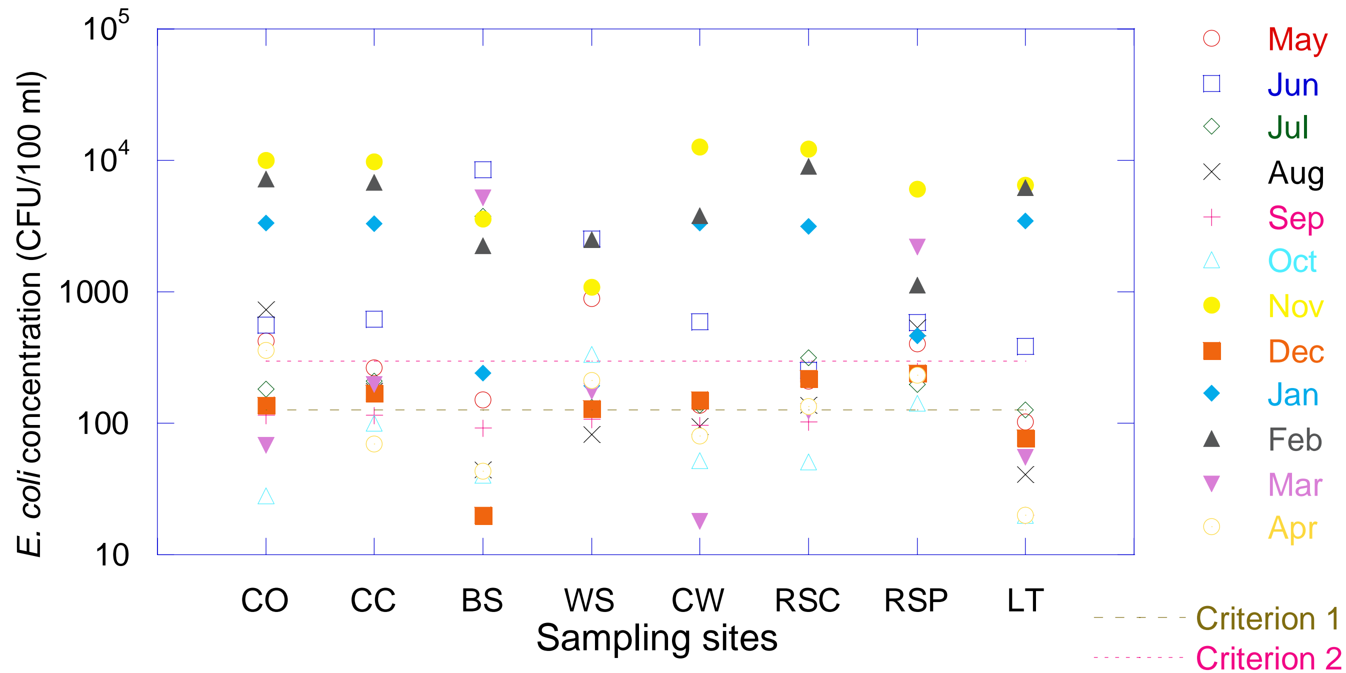


Fig. 2.2 *E. coli* concentrations at Catoma Creek sampling sites from May 2003 to April 2004. EPA criterion 1 (30-day geometric mean): 126 CFU/100 ml and EPA criterion 2 (single sample maximum) : 298 CFU/100 ml

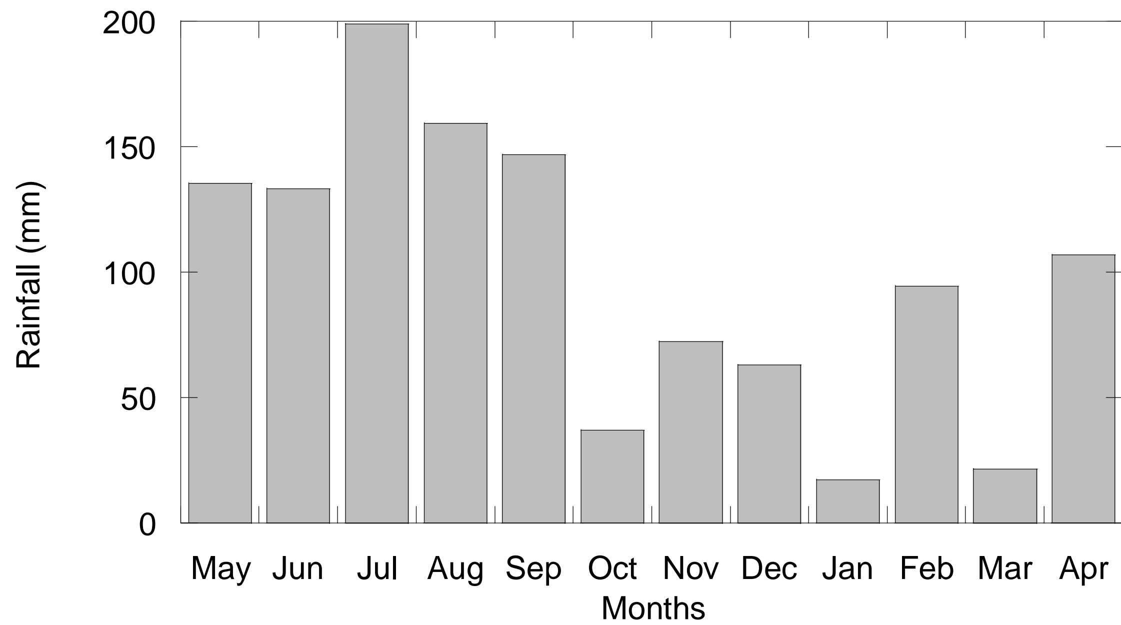


Fig. 2.3 Average monthly rainfall at Catoma Creek watershed
(Source: www.waterdata.usgs.gov)

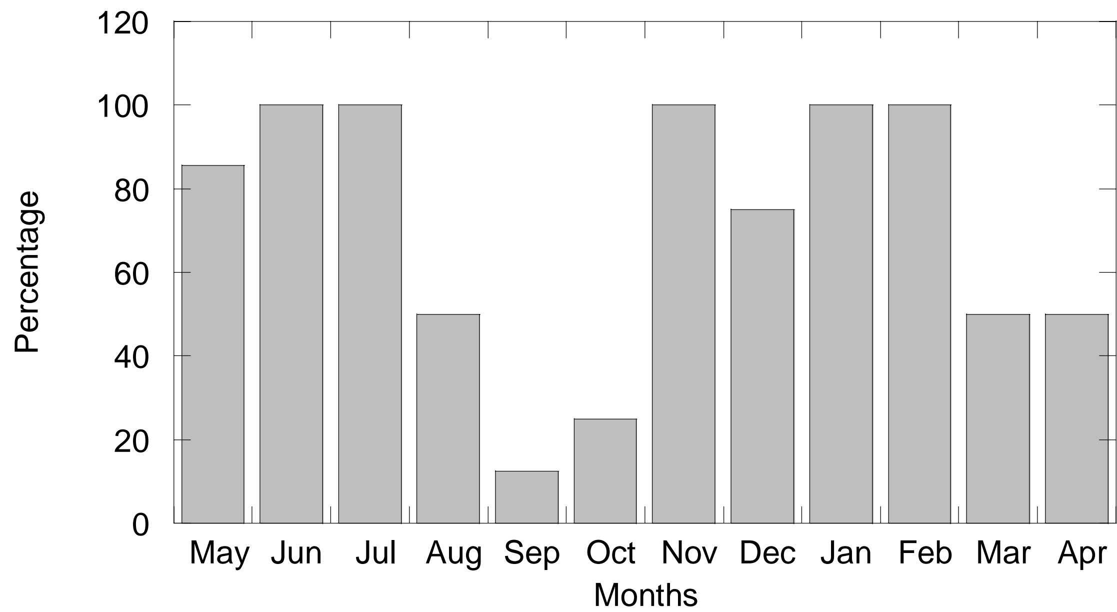


Fig. 2.4 Percentages of water samples having *E. coli* concentrations above the 30-day geometric mean criterion

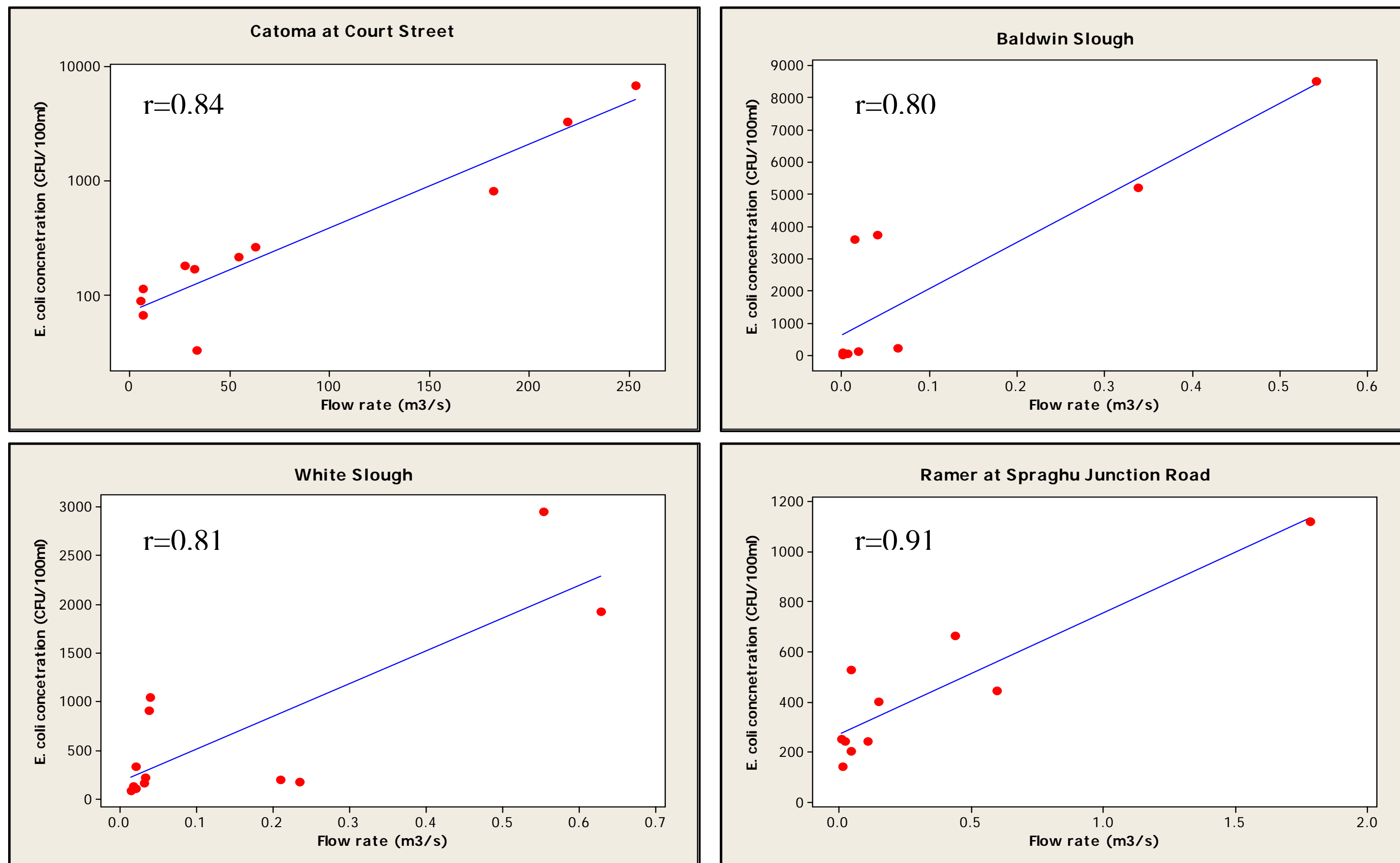


Fig. 2.5 Correlation between flow rate and *E. coli* concentration at selected sites in the Catoma Creek watershed

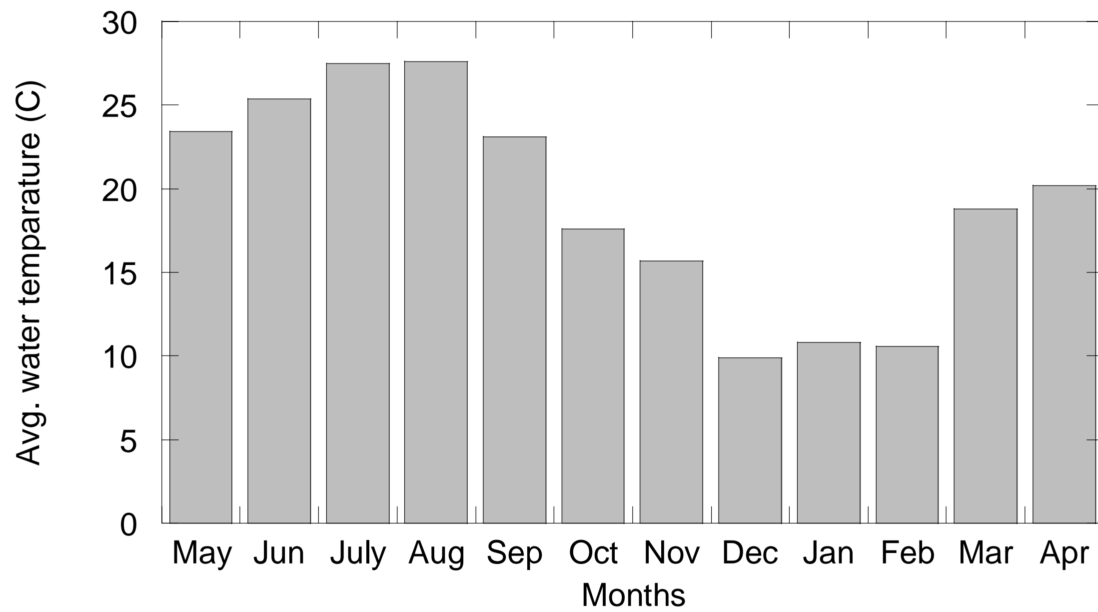


Fig. 2.6 Average water temperature across Catoma Creek sampling sites from May 2003 to April 2004

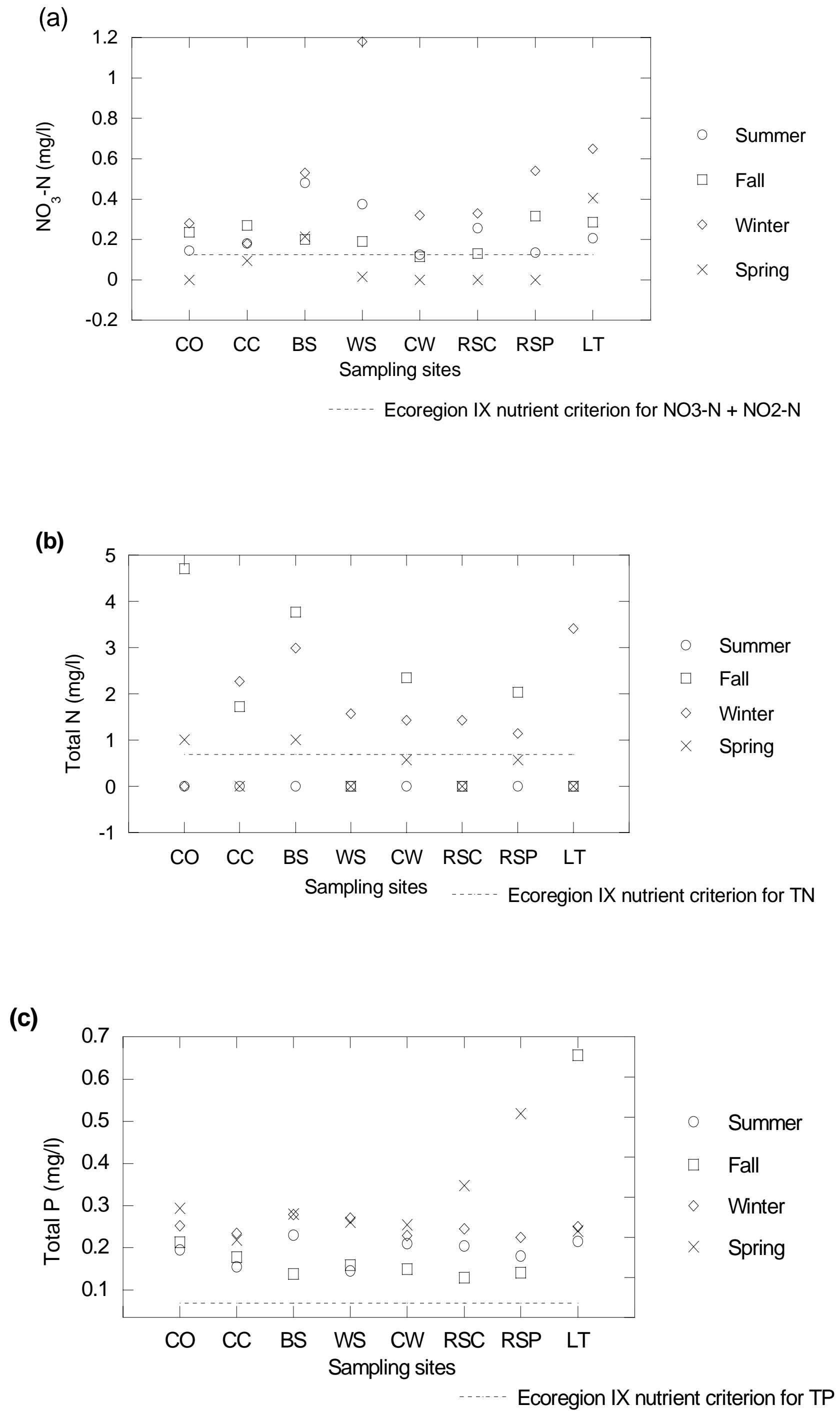


Fig. 2.7 Concentrations of NO_3-N (a), TN (b) and TP (c) at each sampling sites in summer, fall, winter and spring 2003-2004 with respect to Ecoregion IX criteria for NO_3-N + NO_2-N (0.125mg/l), TN (0.692mg/l), and TP (0.036mg/l).

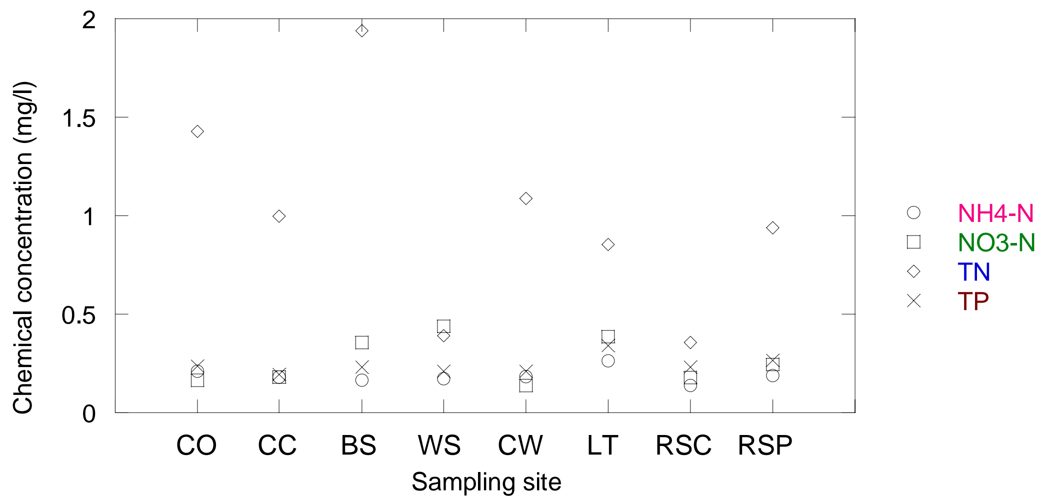


Fig. 2.8 Yearly average NH_4-N, NO_3-N, TN and TP concentrations at each sampling site

III. IDENTIFICATION OF SOURCES OF FECAL CONTAMINATION IN THE CATOMA CREEK WATERSHED

ABSTRACT

High concentrations of fecal indicator bacteria are the most common cause of surface-water impairment in Alabama. A 37 km segment of Catoma Creek in Montgomery Country has been included on the Alabama 303(d) List of impaired water bodies due to elevated concentrations of fecal coliform bacteria and organic enrichment. Fecal contamination can originate from both human and non-human sources, including surface runoff from land application of animal wastes or farm animal feedlots, inadequate septic or sewer systems, improper waste disposal, and wildlife impact. The overall objective of this study was to identify the sources of fecal contamination in the Catoma Creek watershed. A known source library of DNA fingerprints was developed using 582 *E. coli* isolates obtained from humans, dogs, cattle, chickens, horses, wild turkeys, waterfowl, and deer. DNA fingerprints generated using the BOX A1R primer demonstrated great genetic diversity of *E. coli*. Cluster analyses of DNA fingerprint patterns were performed with BioNumerics software using a densitometric curve based matching function (Cosine) and unweighted pair group method with arithmetic average. Jackknife analysis was used to determine cluster/group validity, revealing that the average rate of correct classification for the entire library was 88% and that of the decloned library was 74%. *E. coli* was isolated from monthly water samples over a one-year period at 8 locations in the watershed. The DNA fingerprints (502) obtained from

the water samples were compared against those in the known source library. Results showed that 18% of the *E. coli* isolates were from humans, 14% each from dogs and waterfowl, 4% each from deer and wild turkeys, 2% each from cattle and chickens, 0% from horses, and the remaining 41% unidentified. Further research is needed to improve the representativeness of the library by including more source groups and *E. coli* isolates. Temporal and spatial variations in the *E. coli* populations should also be addressed.

INTRODUCTION

Fecal contamination of surface water creates serious environmental and public problems, which are not limited to a single country. With the aim of finding solutions for these water pollutions, the United States enacted the Clean Water Act in 1972 to achieve the goal of having all the nation's waters both swimmable and fishable. However, after 25 years of implementation, 40% of the waterways in the US cannot be safely used for swimming due to fecal contamination (EPA, 1999). Elevated fecal bacteria counts indicate the potential presence of other pathogenic organisms such as *Shigella* spp., *Salmonella*, hepatitis A virus, and Norwalk group viruses in the water (Dombek et al., 2000). Possible sources of fecal contamination include municipal wastewater discharge, septic leachate, pasture and agricultural land runoff, urban run off, and wildlife.

Accurate identification of the sources of fecal contamination is necessary in order to develop effective pollution control strategies. Bacterial source tracking (BST) is a relatively new method that can be used to identify the sources of fecal

contamination in watersheds. Both genotypic methods, such as host specific molecular markers, tRFLP, PFGE, ribotyping, and rep-PCR, and phenotypic methods such as antibiotic resistant analysis, and carbon source profiling are being used to differentiate the sources of fecal contaminations in the watersheds (Bush et al., 2003; EPA, 2005). However, no single completely satisfactory method has been identified, as all of these methods have both advantages and disadvantages. We selected the rep-PCR (repetitive sequences-based polymerase chain reaction) DNA fingerprinting approach to identify sources of fecal contamination in the Catoma Creek watershed because it is reliable, reproducible, rapid, and highly discriminatory. The *E. coli* genome consists of multiple copies of non-coding, repetitive DNA sequences located at distinct, intergenic positions around the genome. The BOXA1R primer binds these sequences during the polymerase chain reaction and multiple DNA fragments with various lengths are then amplified. Different strains have different DNA banding patterns, and these can be used to differentiate between them at the strain level (Gresshoff, 1997). Since *E. coli* strains are genetically diverse and host specific, these DNA banding patterns can be used to develop a known source DNA fingerprint library and thus to identify the fingerprint patterns of the unknown isolates isolated from the watersheds.

The Catoma Creek watershed is located in Montgomery County, Alabama. A 37-km segment of this creek is included on the Alabama 303(d) List due to elevated concentrations of fecal coliform bacteria and organic enrichment (ADEM 2002). Prior to implementing any remedial measures, it is necessary to first identify the sources of fecal contamination in this watershed. The objectives of this research were: 1) to develop a known source DNA fingerprint library using the rep-PCR method; and 2) to

compare the DNA fingerprint patterns of *E. coli* isolated from the watershed against those in the known source library to identify the sources of fecal contamination in the Catoma Creek watershed.

MATERIALS AND METHODS

Study area

The Catoma Creek watershed is part of the Alabama River basin and covers more than 50% of Montgomery County, Alabama. The drainage area is 932 km^2, with a linear length of 68 km. Forest, agriculture (pasture and row crop), and urban land uses make up 54.5%, 36.2%, and 9.3%, respectively, of the watershed. Average annual rainfall is 1356 mm; average temperature during the winter is 7.6º C and during the summer is 26.7º C. According to the 2000 census, the total population of the county is 223,000 (onlinemontgomery. com). The 37-km segment of Catoma Creek from the Alabama River to Ramer Creek is listed on the 303(d) List of impaired water bodies in Alabama due to nutrient enrichment and elevated fecal coliform concentration. There are several possible sources of fecal contamination in the watershed, including private septic tanks in rural areas, leakage from sewage carrying lines and sewage tanks, urban runoff, wildlife, and runoff from pasture and agricultural land. Eight locations in this watershed were selected to collect water samples (Fig. 2.1). Sampling sites at Catoma Creek at Old Selma Road (CO), Catoma Creek at Court Street (CC) and Catoma Creek at Woodley Road (CW) are on the main stem of the Catoma Creek, while Baldwin Slough (BS), White Slough (WS), Ramer Creek at Sprague Junction Road (RSP), Ramer Creek at Snowdown Chamber Road (RSC) and Little Catoma Creek (LT) are on

tributaries. A known source library was developed using fecal samples collected within Montgomery County.

Collection of water samples and enumeration of *E. coli*

Duplicate water samples were collected using sterilized polyethylene bottles monthly from May 2003 to April 2004. A blank sample was taken randomly at one sampling site each month. Collected water samples were kept on ice, transported to the laboratory and processed within 6 hours. From each sample, three dilutions were filtered through 0.45 µm membrane filters under vacuum. Membrane filters were placed on modified membrane-thermotolerant *Escherichia coli* agar (m-TEC) media (Difco, Detroit, MI) and incubated at 37ºC for 2 hours, and then incubated at 44.5ºC for 24 hours. Isolates giving a magenta/red color were selected (1 colony per plate) and streaked on MacConkey agar (Difco). After overnight incubation at 37ºC, a single colony of dark pink color was selected from each plate. Half of that colony was streaked on Chrom agar (Chromagar Microbiology, Paris, France) and the other half was streaked on MacConkey agar. After overnight incubation at 37ºC, colonies that were dark pink in color on the MacConkey agar and blue on the Chrom Agar were selected to inoculate citrate agar (BBL, Cockeysville, MD), EC broth with 4-methylumbelliferyl-D-glucuronide (EC-MUG) (Difco), 1% trypton (Fisher Biotech, Fair Lawn, N.J), and methyl red–Voges-Proskauer (Difco) broth. Isolates were identified as *E. coli* if they did not use citrate as a substrate, grew at 44.5° C, produced gas and fluorescence in EC-MUG broth, produced indole from tryptophan, and produced an acidic end product

when grown in methyl red-Voges-proskauer broth. *E. coli* isolates were suspended in 50% glycerol /nutrient broth (Difco) and stored at –80ºC for subsequent study.

Stream water temperature was measured on site. Water temperatures above 20°C were considered as warm months and water temperatures below 20°C were considered as cool months. Flow velocities at the BS, WS and RSP sites were measured across the streams at a 3 foot distance using a Flo-Mate portable flow-meter (Marsh-Mc Birney Inc., Frederick, MD) each month. The heights of the water level with respect to each velocity, along with the stream width, were recorded and the mean flow rate was calculated (Mean flow rate = average velocity * stream width * average depth of the flow). USGS flow data (USGS Water Resources of Alabama) was used for the CC sampling site. Flows of the four streams were divided into 3 catogories: low flow, medium flow, and high flow, based on the stream flow rate.

Collection of fecal samples and isolation of *E. coli* from fecal samples

Fecal samples were collected from humans, horses, dogs, cattle, deer, wild turkeys, waterfowl, and chickens within Montgomery County. Fresh fecal samples were collected using BBL culture swabs (BD Biosciences, Sparks, MD), kept on ice until transported to the laboratory and processed within one day. Samples were first streaked on 100 x 15 mm MacConkey plates and incubated overnight at 37°C. Dark pink single colonies (3-6 per plate) were selected and streaked again on MacConkey agar. Further purification and verification of the *E. coli* were performed as described for *E. coli* in water samples. In addition to fecal samples donated by human volunteers, anonymous human *E. coli* isolates were also obtained from a Montgomery hospital.

rep-PCR DNA fingerprinting

The rep-PCR DNA fingerprints of the *E. coli* isolates were obtained using the BOX A1R primer (5'-CTA CGG CAA GGC GAC GCT GAC G–3') (Rademaker amd de Bruijn, 1997) and *E. coli* whole cells as the templates for PCR. PCR was performed according to a protocol modified after Rademaker and de Bruijn (1997) and Dombek et al. (2000). Briefly, the *E. coli* isolates were grown on Plate Count Agar for 18 hours. A portion of a single colony was then removed using a 1 µl sterile inoculation loop and suspended in 100 µl of PCR grade water in a microcentrifuge tube. The 25 µl PCR mixture contained 2 µl of whole cell suspension, 2.5 µl of 10X Promega reaction buffer without $MgCl_2$ (Promega, Madison, WI); 3.0 µl of 25 mM $MgCl_2$ (Promega); 0.2 µl of 100 mM dNTP's (Promega); 0.2 µl of BSA (2% stock) (Invitrogen, Carlibad, CA); 1.0 µl of 10 µM BOX A1R primer (final conc 0.2 pmol/µl) (Invitrogen); 0.4 µl (2 units) of Taq DNA polymerase (Promega), and 15.7 µl of PCR grade water. PCR was performed using a Biometra T-Gradient thermocycler (Whatman, Goettingen, Germany) using the following conditions: initial denaturation at 95ºC for 5 minutes, 35 cycles of denaturation at 94ºC for 1 minute, annealing at 60ºC for 1 minute, and extension at 72ºC for 1 minute, then a final extension at 72ºC for 10 minutes. A negative control containing sterile water and a positive control containing *E. coli* ATCC 25922 were included in each PCR set.

PCR products were mixed with 5 µl of 6X loading dye (Promega) and 10 µl of each reaction mixture was resolved using 1.5% agarose gel (25 cm x 20 cm) in 0.5X TBE buffer. One kb Plus DNA ladder (0.66 µg/well; Invitrogen) was added to the 1st,

10th, 19th, 28th and 36th lanes, a positive control was added to the 2nd lane and a negative control was added to the 35th lane (Fig. 3.1). The gels were electrophoresed at room temperature for 7 hours at 130 V and stained with 0.05% ethidium bromide (Fisher Biotech) in 0.5X TBE buffer for 1 hour. Gel images were captured using the Gel Logic 200 imaging system (Eastman Kodak Co., Rochester, NY) and saved as bip files. Prior to analyzing images with the BioNumeric software, the bip files were converted to 8-bit TIFF format.

Computer based DNA fingerprint analysis

BioNumerics software (version 4.0; Applied Maths, Kortrijk, Belgium) was used for gel normalization, band identification, library development, and source identification. Each gel was normalized by using a 1 kb Plus DNA ladder from 200 to 4000 bp as an external reference standard, which allowed the comparison of multiple gels. When necessary, internal reference standards were used to correct smiling effects. Fingerprint images were added to the database and the library was developed following the method described in the BioNumerics manual. DNA fingerprinting patterns were compared using a densitometric curve-based method with the Cosine coefficient, and dendrograms were developed using the unweighted pair group method with arithmetic averages (UPGMA). A similarity score of 90% was used as the cut off for the same strain types. Jackknife analysis with maximum similarity was used to calculate the rate of correct classification (RCC) of each host group and the average rate of correct classification (ARCC) of the library. ID Bootstrap analysis with a sample size of 100 and 1000 iterations was used to classify isolates based on maximum similarity and to

provide a probability that each isolate was correctly classified. Multivariate analysis of variance (MANOVA), a form of discriminant analysis, was used to determine fingerprint distribution based on the variability among *E. coli* from different hosts, season, stream flow, and sampling sites.

Identification of sources of fecal contamination in the Catoma Creek watershed

DNA fingerprinting patterns obtained from the water isolates of the Catoma Creek watershed were compared against those in the known source library in order to identify the sources of fecal contamination following the method explained in the BioNumerics manual. Water isolate entries were compared against the entire known source library and the decloned library based on maximum similarity. The BioNumeric manual defines the quality quotient of source identification based on the internal heterogeneity of the library units, which is an indication of the level of confidence of the source identification. A quality quotient < 1.5 indicates that source identification is probable, 1.5-2.0 indicates source identification is possible, and > 2.0 indicates that source identification is improbable. In this study, we used a quality quotient of 1.5 as the cut off level for source identification.

RESULTS AND DISCUSSION

1. Construction of known source library

BioNumerics software was used to develop the known source library. The library consisted of 582 *E. coli* DNA fingerprints obtained from 310 fecal samples from humans, chickens, dogs, deer, horses, cattle, waterfowl, and wild turkeys (Table 3.1).

a. Selection of the similarity coefficient

According to the literature, both band-based and curve-based approaches have been used to develop known source libraries. Dombek et al. (2000) used band-based coefficients to develop a rep-PCR DNA fingerprint library from *E. coli* obtained from 7 host groups. Jackknife analysis showed that RCC varied between 78% and 100% and ARCC was 87.5%. Hassan et al. (2005) reported higher RCC and ARCC values for their known source library when using a curve based approach. McLellan (2004) and McLellan et al. (2003) used curve based methods, namely Pearson coefficient and Cosine coefficients, respectively, for cluster analysis of DNA fingerprints of known sources. In our study, curve based Cosine coefficients with UPGMA were used for cluster analysis and dendrogram development. Twenty repeated measures of ATCC 25922 were used to determine the best similarity coefficient for cluster analysis. Cosine coefficient provided an overall maximum similarity of 78.9%. Analyzing the same data using the other curve-based method, Pearson coefficient resulted in an overall similarity of 71.2%. Band based methods, namely the Jaccard and Dice coefficients, showed lower similarities of 34.2% and 50.7%, respectively.Jackknife analysis was performed using Cosine coefficients with maximum similarity. Table 3.2 shows the comparison of Cosine, Pearson, Jaccard and Dice coefficients using Jackknife analysis with maximum similarity for the decloned library. In general, the curve based methods, Pearson and Cosine coefficients, gave higher values for RCC and ARCC than band-based coefficients. When comparing the Jackknife tables developed using maximum and average similarities, both Cosine and Pearson coefficients with maximum similarity gave higher RCC and ARCC values. Based on these results, we used the Cosine

coefficient for library development and subsequent source identification. These results are consistent with the findings reported by Hassan et al. (2005).

b. Cluster analysis of DNA fingerprinting patterns

A typical set of DNA fingerprinting patterns of *E. coli* generated from the BOX A1R primer during the rep-PCR reaction are shown in Figure 3.1. DNA fingerprinting patterns varied from source to source and showed considerable complexity. Cluster analysis showed an overall similarity of 31% for all the DNA fingerprints obtained from known host sources. Cattle and horses had the highest within group similarity, 63%, while humans had the lowest within group similarity, 27%. The number of product bands varied from 7 for a horse isolate to 32 for a deer isolate. A cluster analysis of fingerprint patterns of the entire library revealed no distinct grouping of isolates by source. Human isolates showed the highest tendency to group together, with 96 (54%) of isolates forming one large cluster, while dog isolates showed the lowest tendency to group together, with the largest cluster containing only 8 (13%) of the isolates.

c. DNA fingerprints analysis by using MANOVA

Fig. 3.2 shows the MANOVA plots obtained for the rep-PCR DNA fingerprint patterns of *E. coli*. MANOVA is a form of discriminant analysis used for classifying samples into predefined groups and allows the significance of user-delineated groups to be calculated. Also, this technique makes it possible to determine the characters that are responsible for the separation of the delineated groups (Applied Maths, 2002). First the fingerprint patterns were manually assigned to the correct host groups and a binary band

matching table was generated with 1% optimization and 1% position tolerance. This table was analyzed by MANOVA using a covariance structure. Since eight host groups were specified for a 8-way classification, seven discriminants and P values were calculated. Although rep-PCR DNA fingerprints formed clusters by source group to some extent, several source groups had significant overlap (Fig. 3.2a). These loose clustering patterns indicate the diversity among fingerprints. The first three discriminants accounted for 30%, 21%, and 20% of the total variance. All three together accounted for 71% of the total variation. P values were 0.001%, indicating that the groups were not randomly generated patterns. These results are similar to those obtained by Hassan et al. (2005), who performed discriminants analysis on enterococcal DNA fingerprints obtained from 6 host groups. In their study, DNA fingerprints also failed to show a good separation by source; isolates from one source clustered with isolates from different sources. The first 3 discriminants represented 42%, 21% and 18% of the total variation, and the three together explained 81% of the total variation (Hassan et al., 2005). Dombek et al. (2000) also used MANOVA analysis for rep-PCR DNA fingerprints of *E. coli* from 7 host groups and reported the first 3 discriminants represented 33%, 24.5% and 18.2% of the total variance, together explaining 75.7% of the total variation.

MANOVA plot for 3-way classification (humans, wildlife, and domestic animals) is shown in Fig. 3.2b. When source groups were reduced from 8 to 3, MANOVA showed a good separation for humans, wildlife and domestic animals. The first two discriminants explained 57% and 43% of the total variation. P values for 3-way classification were also 0.001%, indicating that 3-way grouping was valid.

d. Decloning the known source library

E. coli isolates from a single fecal sample often produced identical DNA fingerprint patterns. These identical isolates from a fecal sample, where the fingerprinting patterns had a similarity above 90%, were eliminated, a process referred to as decloning. The resulted decloned library consisted of 414 unique fingerprint patterns (Table 3.1), a 29% reduction from the entire library. Hassan et al. (2005) used 1584 rep-PCR DNA fingerprinting patterns of *E. coli* obtained from humans, dogs, cows, chickens and gulls to developed their know source library. After decloning the library, 38.6% of the fingerprinting patters were removed. In another similar study, 2466 rep-PCR DNA fingerprinting patterns were obtained from 12 host groups and 1535 (62%) showed unique fingerprints. During the decloning, 38% of the isolates were lost (Johnson et al., 2004). Compared with these two studies, we lost fewer isolates during the decloning phase. We used an average of 1.90 *E. coli* isolates per fecal sample to develop our known source library. In comparison, Hassan et al. (2005) used 3.38 *E. coli* isolates per fecal sample and Johnson et al. (2004) used 2.51 *E. coli* isolates per fecal sample to develop their know source libraries. Both of their libraries showed a higher percentage of isolate reduction during the decloning phase. Thus, we can reasonably assume that the number of isolates obtained from a fecal sample is an important factor in determining the number of isolates in the decloned library.

2. Evaluation of the known source library

Since water resource managers use source identification data to implement the total maximum daily loads (TDML) and best management practices (BMPs) for a

watershed, the evaluation of the quality of the library to be used is very important. Evaluation of the performance of our known source library was based on the universal quality measure parameters introduced by the EPA (2005).

a. Reproducibility of DNA fingerprints

Reproducibility of the DNA fingerprinting patterns generated using the BOXA1R primers was tested by repeated measures of a positive control and by replicating about 10% (53) of the isolates in the library. The positive control (*E. coli* ATCC 25922) was included in each set of PCR and gel electrophoresis. Reproducibility was analyzed using cluster analysis, Jackknife analysis, and ID bootstrap analysis. Cluster analysis of the 20 repeated measures of the positive control showed an overall similarity value of 78.9%. Jackknife analysis revealed a RCC of 100% for the 20 repeated measures. ID Bootstrap also correctly classified all the repeated measures with a maximum similarity of 97% and probability of 0.996. Twelve randomly selected human isolates and 5-6 isolates per source from the remaining sources (referred to as test isolates) were used for replication. PCR products of waterfowl, human, deer and 3 dog samples were run in one gel and chicken, horse, cattle, wild turkey and the 3 remaining dog samples were run in a second gel. Overall, 38 out of the 53 test isolates (72 %) formed clusters with same isolates run earlier (Table 3.3). The similarity values varied between 96% for chicken and 73% for dogs. All deer and chicken test isolates, as well as 92% of the human test isolates (11 out of 12), clustered with their previous runs. Eighty percent of cattle (4 out of 5), 67% of wild turkey and horse (4 out of 6), 33% of

waterfowl (2 out of 6) and 17% of dog (1 out of 6) test isolates clustered with their previous runs.

ID Bootstrap analysis with 1000 iterations and a sample size of 75 was also used to analyze these test samples. Of 53 test isolates, 40 of the isolates (75%) were correctly classified (Table 3.3). One hundred percent of deer, 92% of human, 60% of cattle, 83% of horse, 67% of chicken, 83% of wild turkey, 67% of waterfowl and 17% of dog test isolates were correctly classified. Dog showed the lowest similarity 84%, while human showed the highest similarity, 98%, between test samples and their source groups. These results show that the DNA fingerprints in the library were reasonably reproducible. The running condition of PCR or gel electrophoresis did not seem to be a major contributor to fingerprint grouping.

b. Jackknife analysis of the DNA fingerprint patterns from known sources

Jackknife analysis was used to demonstrate the internal stability of the source groups. DNA fingerprint patterns were grouped as human, wildlife and domestic animals (3-way classifications), as well as by species. Table 3.4 (a) shows the results of the Jackknife analyses for the entire library with a 3-way classification. RCC for humans, wildlife and domestic animals were 91%, 90% and 92%, respectively, with an ARCC of 91%. Table 3.4 (b) shows Jackknife analyses for the decloned library, which reveal an ARCC of 84%. The RCC of domestic animals and wildlife were reduced to 77% and 83%, respectively, while the RCC of human samples increased to 92% compared to the entire library.

Table 3.5 shows the RCC of the 8-way classification as human, cattle, dog, chicken, horse, waterfowl, wild turkey, and deer. The ARCC for 8-way classification for the entire library and decloned library were less than that for 3-way classification. The ARCC decreased from 88% for the entire library to 74% after decloning. RCC was the highest for chicken samples, where 98% of fingerprint patterns were identified as chicken, which reduced to 86% once the library was decloned. Dog showed the lowest rate of correct classification, where only 72% of the fingerprints were classified as dog, which reduced to 50% in the decloned library. The RCC of cattle, deer, horse, and waterfowl were 85%, 80%, 96%, and 85% and reduced to 63%, 70%, 91%, and 72%, respectively, in the decloned library. Wild turkey showed the most dramatic reduction of RCC after decloning, where the RCC was reduced to 68% from 94%. However, only 12 fecal samples were available from wild turkey, providing 54 *E. coli* isolates. After decloning the library, 53% fingerprint patterns were lost and this is likely to be the cause of the drastic reduction of RCC. On the other hand, 179 *E. coli* isolates were obtained from 165 individual human samples, but only 7% of the fingerprints were lost during the decloning process and the RCC of the human samples actually increased to 92% from 91% after decloning the library. This indicates that the number of isolates taken from a single fecal sample is an important factor in determining the RCC of that source and the ARCC of the library.

Average rate of correct classification of this library was relatively high compared to others reported in the literature. Dombak et al. (2000) reported an ARCC of 87.5% for their rep-PCR DNA fingerprinting library (without decloning) composed of humans, geese, ducks, sheep, pigs, chicken, and cow isolates. McLellan et al. (2003)

also used the rep-PCR method to develop a DNA fingerprint library. Again without decloning, they obtained their highest RCC for cattle, 88.2%, followed by 83.2% for sewage and 66% for gulls. The ARCC of their library was 79.3%.

Johnson et al. (2004) performed Jackknife analysis on a rep-PCR DNA fingerprint pattern library consisting of 12 host groups. The ARCC for the clonal (entire) library was 82.2% which reduced to 60.9% for the decloned library, a 25.9% reduction. Hassan et al. (2005) used enterococci isolates from 6 host groups to develop a rep-PCR DNA fingerprinting library. Jackknife analysis showed that ARCC for their clonal library was 92% and that of the decloned library was 82%, showing a 11% ARCC reduction after decloning. In our study, a 15.9% ARCC reduction in the 8-way classification library and a 7.7% ARCC reduction in the 3-way classification library were observed after decloning. Clonal isolates in the known source library resulted in overestimates of both the RCC and ARCC. In Jackknife analysis, isolates are divided into host groups and then the individual isolates are removed from the library one at a time, treated as an unknown and assigned into groups based on their similarities. When clonal isolates are present, these isolates are always assigned to the library source where another member of same clone is present. Therefore, Jackknife analysis of a clonal library generally tends to overestimate the RCC (Hassan et al., 2005).

Ribotyping and antibiotic resistant analysis (ARA) are two other library based methods that can be used in MST. Carson et al. (2001) reported 73.6% ARCC for a ribotyping library consisting of human, pig, cattle, horse, dog, chicken turkey and goose isolates. In another study, the ARA method was used to develop 6 libraries for 6 watersheds using 6587 unique *E. coli* isolates obtained from humans, domestic animals

and wildlife. The ARCC for the 6 individual libraries ranged from 65% to 81%, which reduced to 57% after merging all 6 libraries (Wiggins et al. 2003). Booth et al. (2003) obtained an ARCC of 85.3% for an ARA library composed of human, livestock and wildlife samples.

c. ID Bootstrap analysis

ID bootstrap analysis was used to classify the isolates based on maximum similarity. Unlike Jackknife analysis, the ID bootstrap approach provides a probability that each isolate is correctly classified. ID Bootstrap analysis for an 8-way classification showed similar results to the Jackknife analysis for the entire source library, as well as the decloned library (Table 3.6). However, a reduction in the RCC was observed in ID Bootstrap analysis when the probability value of 0.9 was taken into account. No reduction was observed for horse and chicken samples in both the entire and decloned libraries. Dogs, however, showed the highest reduction of RCC after decloning, with reductions of 9% and 26% for the entire library and the decloned library, respectively.

d. Control samples

Positive control and negative control samples were used to measure the performance of each set of PCR products, and to screen for the presence or absence of extraneous microorganisms, respectively. The results of positive control are discussed in previous section.

Sterile water was used as the negative control and one negative control sample was included in each set of PCR products. If PCR product bands were found in a negative control, that set of DNA fingerprints was not used for library construction.

3. Analysis of DNA fingerprint patterns obtained from water isolates

E. coli isolated from water samples collected from the Catoma Creek watershed were used to generate DNA fingerprint patterns using the BOX A1R primer. A total of 502 fingerprint patterns from water samples were used for source identification. DNA banding patterns showed high variability. Total numbers of product bands varied between 5 and 28; 271 unique fingerprint patterns were found. Cluster analysis results showed that the overall similarity of DNA fingerprints from the Catoma Creek watershed was 32.8%, and fingerprints did not cluster by either sampling site (Fig. 3.3.c) or sampling month (data not shown).

a. Fingerprint patterns by season

Fig. 3.3a shows the MANOVA results indicating that fingerprint patterns clustered according to season. The first 3 discriminants accounted for 41%, 36%, and 23% of the total variance, respectively, with a P value of 0.001%. The low P value indicated that there was a tendency of *E. coli* fingerprints to cluster according to the season. These results suggest that sources of fecal pollution in this watershed may exhibit some seasonality.

b. Fingerprint patterns by flow rate

Flow data for the CC site was taken from the USGS stream flow database (USGS Water Resources of Alabama). Stream flow at this site was divided into 3 categories: low flow (<10 m^3/s), medium flow (10-100 m^3/s), and high flow (>100 m^3/s). Since the CC, CO and CW sites are all on the main stem of the Catoma Creek, we assumed that these three sites had similar flow patterns. A portable flow meter was used to measure the flow rates at the BS, WS and RSP sites. Stream flows at these sites were also divided into 3 categories: low flow (<0.01 m^3/s), medium flow (0.01-0.1 m^3/s), and high flow (>0.1 m^3/s). Since RSP and RSC are two locations on the same creek, we again assumed flow rate patterns at both sites were the same. LT site had stagnant water for almost all of the sampling time, therefore we did not use LT site data for this analysis. MANOVA results showed that DNA fingerprint patterns clustered according to stream flow, Fig. 3.3 (b). The total variance explained by the first two discriminants were 59% and 41%, respectively, and the P values were 0.001%. *E. coli* DNA fingerprints seemed to vary with the flow rate.

c. Fingerprint patterns by sites

Fig. 3.3c shows the MANOVA plot of water isolates based on sampling sites. DNA fingerprints did not separate based on the sampling sites. The first 3 discriminants explained 32%, 18%, and 14% of the total variance. However, the P values of these discriminants were 0.413%, 74.98%, and 98.7% respectively, indicating an insignificant relationship between sampling sites and DNA fingerprinting patterns. The fingerprint patterns were thus shared by all sampling sites.

4. Fecal contamination in the Catoma Creek watershed

Of the 502 isolates obtained from the Catoma Creek watershed over the one-year sampling period, 295 (59%) isolates were identified using the decloned library and 290 (58%) of the isolates were identified using the entire library. Thus, there was no significant difference between source identification using the entire or decloned library. The following discussion is based on the results using the decloned library. Humans were the main contributor of fecal pollution in this watershed. Of 502 *E. coli* water isolates, 92 (18.3%) were identified as human. Dogs and waterfowl were the next largest contributors, and were responsible for 14% each of the fecal pollution. Wild turkey and deer showed smaller contributions of 4% each. The percentages of the cattle and chicken contributions were very low, at 2% each. No horse signature was found in any of the water samples. This indicates either that cattle, chickens and horses were not significant sources of fecal contamination in this watershed, or that this library was not large enough to be fully representative of these species. Fig. 3.4 shows the contribution of each source in this watershed. Since dogs had the lower RCC based on Jackknife analysis and dog test isolates were the least reproducible, the results for dog contribution may be less reliable. Source identification using a 3-way classification shows that humans, wildlife and domestic animals represented 18%, 23% and 18% of the water isolates, respectively.

Sadowsky et al. used the HFERP (Horizontal Fluorophorenhanced Rep-PCR) method to identify the sources of fecal contamination in the Vermillion River in Minnesota. Their results showed that sources of fecal contamination included 14% geese, 12% pigs, 12% cats, 10% cows, 9% human, 9% deer, 9% sheep and 9% wild

turkey (EPA, 2005). Booth et al. (2003) used ARA to identify the sources of fecal contamination in the Black Water River in Virginia. Here livestock were the main contributors to the fecal pollution in the watershed, followed by wildlife and humans. Another study using ARA found that dominant sources of fecal contamination in Anacostica River of Maryland / District of Colombia were 31% birds, 25% wildlife, 24% human, and 20% pets (Hagerdorn et al., 2003). A study of Tampa Bay, Florida using ARA and ribotyping methods found that wildlife dominated fecal pollution (Rose et al., 2000). The variation in the results reported by these studies indicates that sources of fecal contamination are different from watershed to watershed. Source identification is thus a vital step that must be taken prior to implementing any remedial measures.

a. Sources of fecal contamination by sampling site

Table 3.7 shows that Baldwin Slough had the highest percentage of source identification; out of 65 water isolates, 46 (71%) were identified. The RSP site had the lowest rate of identification, 39%. CC, CO, WS, BS, and CW are all in residential areas and the average source identification for these sites was 65%, while RSC, RSP, and LT are rural or forest areas and had an average 48% sample identification. This indicates that the known source library should be further expanded by increasing the number of *E. coli* isolates and the number of source groups such as cats, raccoons, beavers, wild hog, etc.

Table 3.7 and Fig 3.5 show that dogs and humans dominated the water isolates at the CC site, waterfowls, dogs and humans dominated at CO, and waterfowls and humans dominated at the BS site, while humans dominated all other sampling sites. The

human contribution was almost evenly distributed over the sampling sites. BS site is a residential area that showed a relatively high percentage (16%) of human signatures. Except for waterfowl, all the source groups showed almost even contributions over the sampling sites. Waterfowl dominated at the BS site, with 16 (25%) isolates, and high percentages were also recorded at CW and CO, while low percentages were found at RSP, RSC and LT.

b. Sources of fecal contamination by season

Sampling months were divided into warm months and cool months based on the water temperature. If the water temperature was below 20ºC, those months were considered to be cool months (October, November, December, January, February, March), and if the water temperature was above 20ºC, those were considered to be warm months (April, May, June, July, August, September). Table 3.8 shows the source contribution for each month, as well as the percent identification in cool and warm months. The percentage of source identification was 10% higher (64% of the samples identified) during the warm months than during the cool months (54%). Humans again dominated in both seasons, with similar proportions of samples were found in the cool months 45 (9%) and the warm months 47 (9%). Dog, cattle, chicken and wild turkey signatures also were found in similar proportions in both seasons. However, more waterfowl signature was identified during the warm months and more deer signature was found during the cool months. A study by Booth et al. (2003) showed that livestock dominated in both cool and warm seasons, although Hagerdorn et al. (2003) also found seasonality of fecal contamination, with birds and wildlife sources

dominating during the low flow warm weather months and human and birds dominating during the cold weather months.

c. Sources of fecal contamination by flow rate

Flow rates at the CC, CO, CW, BS, WS, RSP and RSC sites were categorized into 3 groups: low flow, medium flow and high flow, as described in Section 3.b. *E. coli* isolates obtained under the low flow condition showed a higher percentage of identification than under the high flow condition (Table 3.9). Percent identification during low flow, medium flow, and high flow sampling periods were 75%, 54%, and 51%, respectively. During storms, feces from various sources may be washed into streams and some of the sources may not be represented in the library. The low flow conditions most closely resemble the base flow conditions. Thus, constant sources are likely to play a larger role during these periods such as seepage from septic tanks or leakage from sewage carrying lines, etc.

Table 3.10 shows the proportional contribution of each source group during the low, medium and high flow conditions. Percentages of human isolates identified during low, medium and high flow conditions were 30%, 35%, and 35%, respectively. Higher runoff may be the cause of the increased human signature for the high and medium conditions. However, the difference between low and high flow conditions was small, suggesting that there was a continued human source contributing to the streams during all the flow conditions. This may consist of seepage from septic tanks or leakage from sewer carrying lines. However, further investigation should be conducted to identify these possible causes. Hagerdorn et al. (2003) reported that a high percentage of human

signatures were found under high flow conditions, suggesting sewer overflow as a primary cause.

High percentages of dog and wild turkey signature, 46% and 47%, respectively, were identified under the low flow condition, but there were no marked differences between their signatures found during the medium and high flow conditions. This also shows that there were continues sources for dogs and wild turkeys and they dominated during the low flow conditions. Runoff water in wet months may carry higher fecal loads, resulting in the proportion of the dog and wild turkey contribution being relatively low during the wet months. Cattle showed the highest contribution during high flow conditions, suggesting that their feces were washed into the streams by runoff water. Chicken and waterfowl showed their maximum contribution during the medium flow condition, at 62% and 44%, respectively.

5. Minimum detectable percentage (MDP) and fecal contamination in the Catoma Creek watershed

MDP is used to determine the lower limit at which a source is considered to be a significant contributor to a watershed. Average misclassifications of all sources are added to 4 times the standard deviation and the resulting value is the MDP of any source (Whitlock et al., 2002; Wiggins et al., 2003). According to one study (Wiggins et al., 2003), the MDP for an ARA library consisting of 6587 enterococci was 25% and they suggested that this library would be able to reliably identify sources that are present at average levels above 25%. In our study, the MDP identified in the decloned library was 23.51%. According to these MDP results for the decloned library, there was

no significant source contributing fecal contamination in the watershed as the percentage contribution of each sample was less than 23.51% (Table 3.4). The MDP obtained from the entire library was 11.73%. The entire library shows humans, dogs, and waterfowl to be significant contributors of fecal contamination in this watershed. However, no application of MDP in the rep-PCR method was found in the literature.

SUMMARY

Evaluation of the known source library demonstrated that our library, although small, is comparable to those reported in the literature. Thus, this rep-PCR DNA fingerprinting known source library can be used effectively to identify the sources of fecal contamination in the Catoma Creek watershed. The results showed that *E. coli* isolates obtained from the Catoma Creek watershed originated primarily from humans, dogs, cattle, chickens, waterfowl and wild turkeys. According to the results of this study, no signatures that could be ascribed to horses were identified. Humans were the main contributors of water pollution in this watershed, contributing 18%, while dogs and waterfowl caused 14% each of this pollution. Wild turkeys and deer, 4% each, and cattle and chickens, 2% each, also contributed to this pollution. About 41% of the isolates remained unidentified. In particular, the rural and forest sampling sites showed higher percentages of unidentified isolates, indicating that the known source library should be expanded by increasing the number of samples and number of source groups to further improve the rate of source identification. At present, fecal pollution of the Catoma Creek watershed prevents its usage for body contact recreational activities such

as swimming by local communities. To improve the water quality of this watershed, best management practices should be developed and implemented by the appropriate authorities to reduce the fecal loading in the creek.

Table 3.1 Host groups and DNA fingerprint patterns in the known source library

Host group	No. of fecal samples	No. of *E. coli* isolates	No. of total fingerprints	No. of unique fingerprints
Cattle	20	60	53	33
Chicken	20	60	60	28
Dog	18	60	60	42
Deer	28	61	61	46
Horse	20	54	54	33
Human	165	180	179	167
Waterfowl	21	62	62	40
Wild turkey	12	54	53	25
Total	**310**	**591**	**582**	**414**

Table 3.2 Comparison of different coefficients using Jackknife analysis in the decloned library

Source	Rate of correct classification (%)			
	Cosine	Pearson	Jaccard	Dice
Cattle	64	64	70	70
Chicken	86	82	86	86
Dog	50	48	48	48
Deer	70	72	65	65
Horse	91	91	88	88
Human	92	92	84	84
Waterfowl	78	72	62	62
Wild turkey	68	72	40	40
ARCC	**74**	**74**	**68**	**68**

Table 3.3 Reproducibility of DNA fingerprints using test isolates

Source	No. of test isolates	Cluster analysis		ID Bootstrap analysis	
		Max and min. similarity range	% Formed clusters	Max and min. similarity range	% Correctly identified
Human	12	95-79%	92% (11)	98-85%	92% (11)
Cattle	5	88-75%	80% (4)	92-86%	60% (3)
Dog	6	73%	17% (1)	84%	17% (1)
Chicken	6	96-83%	100% (6)	94-91%	67% (4)
Horse	6	83%	67% (4)	88-85%	83% (5)
Deer	6	94-86%	100% (6)	94-91%	100% (6)
Wild turkey	6	93-87%	67% (4)	96-88%	83% (5)
Waterfowl	6	87-84%	33% (2)	96-93%	67% (4)

The numbers of isolates are listed in parentheses.

Table 3.4 Rates of correct classification for three source groups (a) before decloning and (b) after decloning the library, expressed as percentage.

	Livestock	Wildlife	Human
Livestock	**92**	4.0	5.6
Wildlife	2.8	**90**	3.4
Human	5.0	5.7	**91**
ARCC			**91**

(a)

	Livestock	Wildlife	Human
Livestock	**77**	8.2	5.4
Wildlife	9.1	**83**	3.0
Human	13	9.1	**92**
ARCC			**84**

(b)

Table 3.5 (a) Rates of correct classification for eight source groups before decloning the library, expressed as percentage.

	Cattle	Chicken	Dog	Deer	Horse	Human	Waterfowl	Wild turkey
Cattle	**85**	1.7	1.7	1.6	0	0	1.7	0
Chicken	2.1	**98**	5.0	1.6	0	3.4	1.7	0
Dog	4.0	0	**72**	3.3	0	1.1	1.7	0
Deer	0.2	0	5	**80**	0	0.6	1.7	3.7
Horse	0.2	0	3.3	0	**96**	1.1	0	0
Human	2.1	0	13	8.2	3.7	**91**	8.3	0
Waterfowl	15.9	0	0	3.3	0	1.1	**85**	1.8
Wild turkey	0.2	0	0	1.6	0	1.7	0	**94**
ARCC								**88**

Table 3.5 (b) Rates of correct classification for eight source groups after decloning the library, expressed as percentage.

	Cattle	Chicken	Dog	Deer	Horse	Human	Waterfowl	Wild turkey
Cattle	**63**	3.6	2.4	2.2	0	0	7.7	0
Chicken	3.4	**86**	4.8	2.2	0	3.0	0	0
Dog	3.4	3.6	**50**	4.3	3.0	1.2	2.6	0
Deer	0.4	0	7.1	**70**	0	0	7.7	12
Horse	0.4	0	4.8	2.2	**91**	1.2	0	0
Human	12	3.6	26	8.7	6.1	**92**	10	8
Waterfowl	18	3.6	2.4	8.7	0	1.2	**72**	12
Wild turkey	0.4	0	2.4	2.2	0	1.8	0	**68**
ARCC								**74**

Table 3.6 (a) ID Bootstrap analysis results for the entire library

	% of correct classification without a P value	% of correct classification with a P value above 0.9
Cattle	87%	85%
Chicken	98%	98%
Dog	72%	63%
Deer	80%	75%
Horse	96%	96%
Human	90%	78%
Waterfowl	85%	81%
Wild turkey	94%	91%
ARCC	**88%**	**83%**

Table 3.6 (b) ID Bootstrap analysis results for the decloned library.

	% of correct classification without a P value	% of correct classification with a P value above 0.9
Cattle	64%	60%
Chicken	86%	86%
Dog	50%	24%
Deer	70%	65%
Horse	91%	91%
Human	92%	83%
Waterfowl	72%	68%
Wild turkey	68%	64%
ARCC	**74%**	**68%**

Table 3.7 Number of *E. coli* isolates identified at the eight sampling sites

Sources	CO	CC	WS	BS	CW	RSC	RSP	LT	Total
Human	**11**	**10**	9	15	**14**	**12**	**9**	**12**	92
Cattle	2	1	1	1	1	3	0	1	10
Dog	10	**10**	**10**	8	11	7	8	6	70
Chicken	2	1	1	2	1	1	0	2	10
Horse	0	0	0	0	0	0	0	0	0
Deer	4	4	4	2	3	1	1	3	22
Wild turkey	2	6	0	2	5	2	2	3	22
Waterfowl	**11**	7	9	**16**	13	5	4	4	69
Total fingerprints	67	65	52	65	72	65	61	55	502
Fingerprints identified	42	39	34	46	48	31	24	31	295
% identified	63%	60%	65%	71%	67%	48%	39%	57%	59%

Table 3.8 Number of *E. coli* isolates identified at the eight sampling sites by month

	April	May	June	July	Aug	Sept	**Total-warm months**	Oct	Nov	Dec	Jan	Feb	March	**Total-cool months**
Human	21	7	1	6	2	10	47	4	2	8	8	11	12	45
Cattle	0	2	0	0	2	0	4	2	3	0	0	1	0	6
Dog	11	2	3	5	2	13	36	3	7	4	3	3	14	34
Chicken	0	1	0	4	1	0	6	0	1	0	1	2	0	4
Horse	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Deer	0	1	0	2	3	2	8	4	1	1	3	2	3	14
Wild turkey	0	0	1	1	0	9	11	1	3	1	4	1	1	11
Water fowl	1	15	6	8	10	0	40	6	7	5	4	7	0	29
% identified	**69**	**65**	**39**	**57**	**67**	**77**	**64**	**48**	**67**	**43**	**51**	**56**	**63**	**54**

Table 3.9 Source identification based on flow rate

Flow condition	No. of isolates used	No. of isolates identified	% identified
High	168	86	51%
Medium	168	91	54%
Low	114	85	75%

Note: Data for LT site are not included.

Table 3.10 Source distribution with different flow conditions

Sources	High flow	Medium flow	Low flow
Human	28 (35%)	28 (35%)	24 (30%)
Cattle	5 (56%)	2 (22%)	2 (22%)
Dog	18 (28%)	16 (25%)	30 (47%)
Chicken	3 (38%)	5 (62%)	0
Horse	0	0	0
Deer	6 (30%)	7 (35%)	7 (35%)
Wild turkey	5 (26%)	5 (26%)	9 (47%)
Waterfowl	21 (33%)	28 (44%)	15 (23%)

Note: Data for LT site are not included.

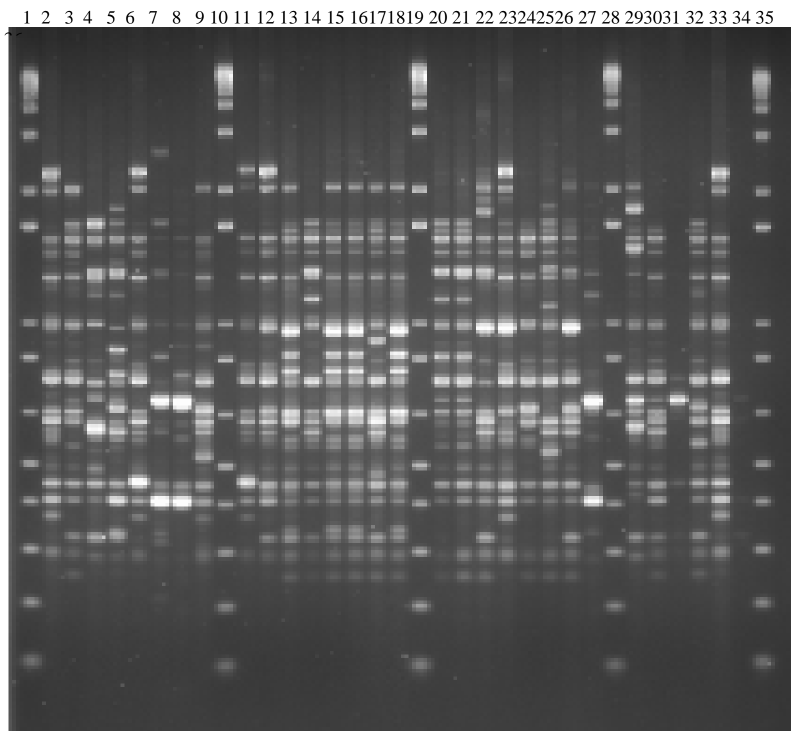
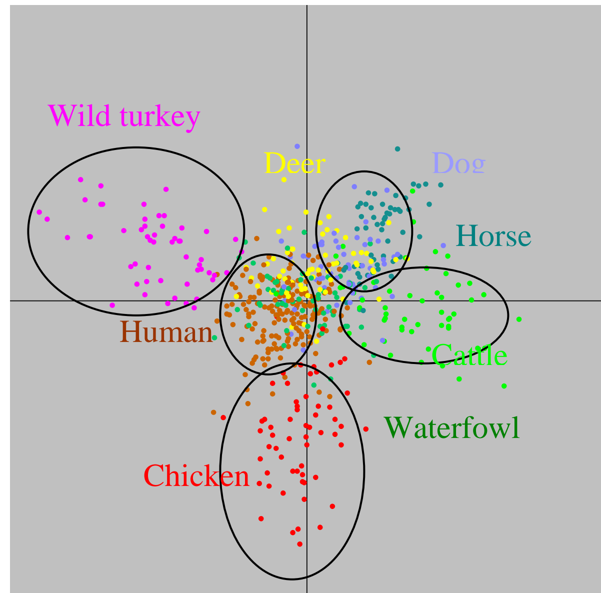


Fig. 3.1 rep-PCR fingerprints generated using the BOX A1R primer**.** Lanes 1, 10, 19, 28, and 35 contain 1-kb Plus DNA ladder. Lanes 2, and 33 contain ATCC 25922, and lane 34 is the negative control. Lanes 3-5 contain dog isolates, lanes 6-9, 11 and 12 chicken, lanes 13-18 horse, lanes 20-25 cattle, and lanes 26, 27, and 29 -32 wild turkey.

(a)



(b)

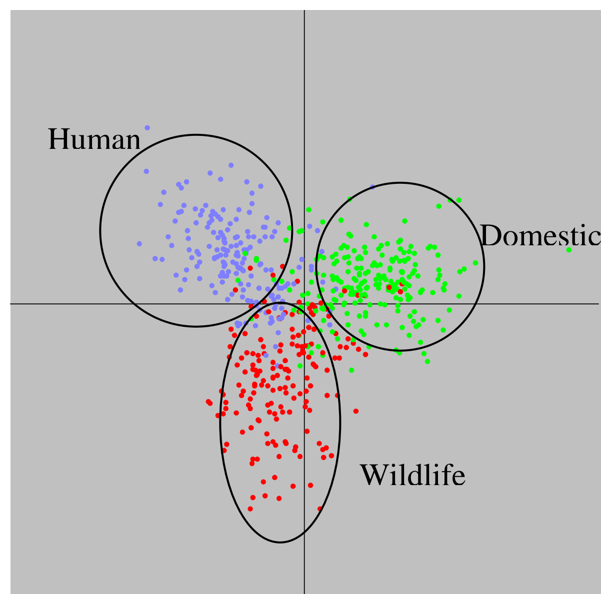


Fig. 3.2 MANOVA plot of rep-PCR DNA fingerprint patterns of *E. coli* from different source groups: (a) 8-way classification and (b) 3-way classification.

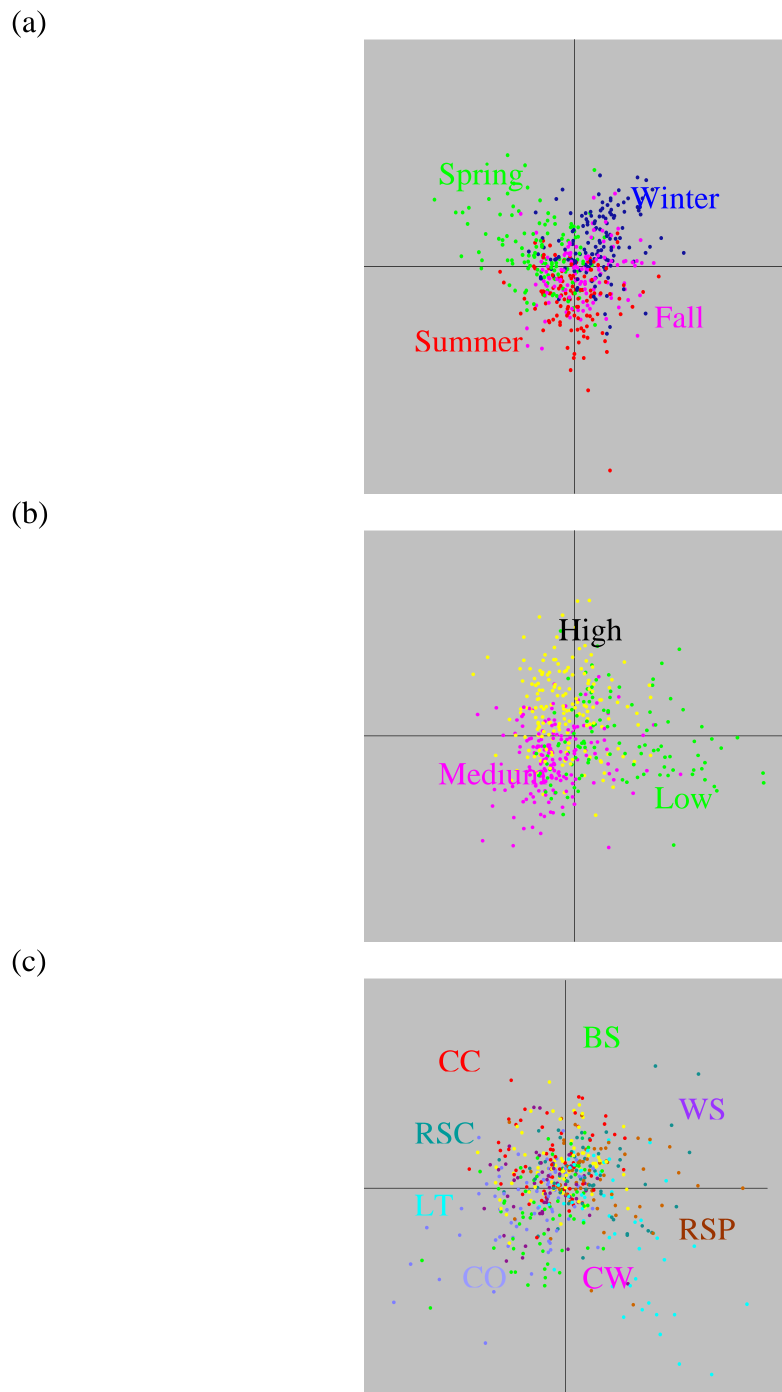


Fig. 3.3 MANOVA plots of rep-PCR DNA fingerprint patterns of *E. coli* isolated from the Catoma Creek watershed by: (a) seasons (b) stream flow rates, and (c) sampling sites.

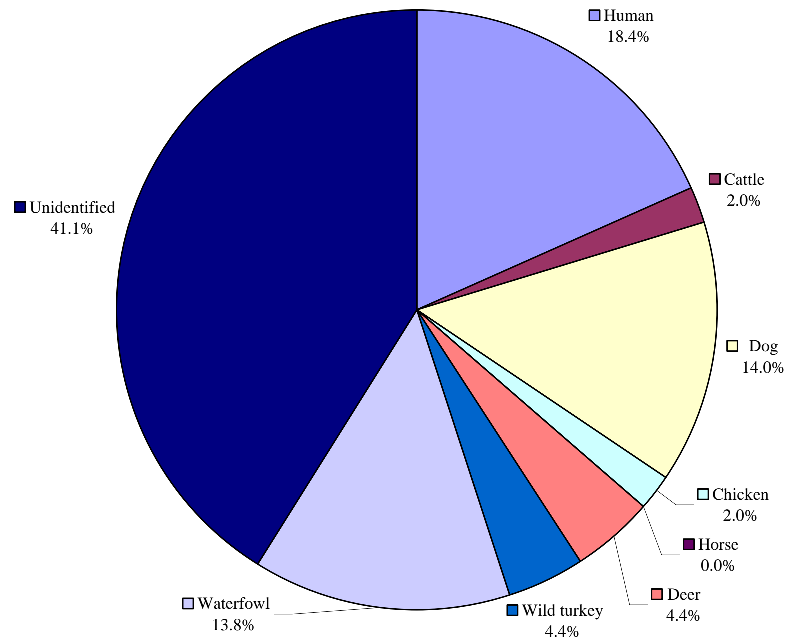


Fig. 3.4 Sources of fecal contamination in the Catoma Creek watershed

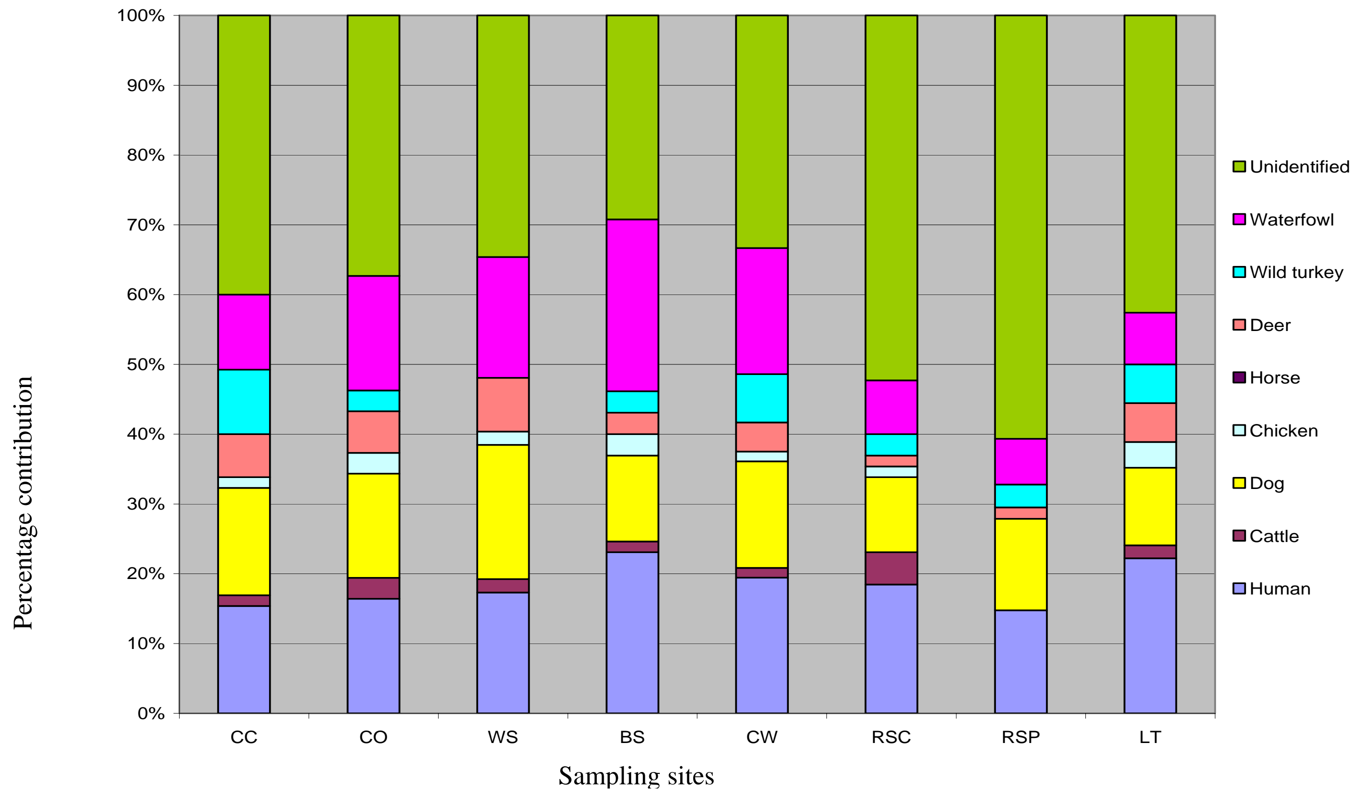


Fig. 3.5 Source distribution at different sampling sites in the Catoma Creek watershed

IV. REFERENCES

ADEM. 2002. Draft TMDL development for Catoma Creek, AL/03150201-080-01 http//www.adem.state.al.us/PublicNotice/Nov/Catoma%20CreekDO%20TMDL%20(Draft).

Ahearn, D. S., R. W. Sheibley, and R. A. Dahlgren. 2005. Effect of river regulation on water quality in the lower Mokelumme River, California. River Research and Application. 21(6): 651-670.

Albert, J. M., J. M. Marr, L. Tenorio, and R. L. Siegrist. 2003. Statistical Evaluation of Bacterial Source Tracking Data Obtained by rep-PCR DNA Fingerprinting of *Escherichia coil*. Environmental Science & Technology. 37: 4554-4560.

An, Y. J., D. H. Kampbell, and G. P. Breidenbach. 2002. *Escherichia coil* and total coliform in water and sediments at lake marinas. Environmental Pollution. 120: 771-778.

Applied Maths. 2002. BioNumeric manual version 3.0. pp 187-197.

Aslam, M., F. Nattress, G. Greer, C. Yost, C. Gill, and L. McMullen. 2003. Origin of Contamination and Genetic Diversity of *Escherichia coil* in Beef Cattle. Applied and Environmental Microbiology. 69:2794-2799.

Bacterial Barcodes (http://www.bacbarcodes.com/underlying_technology.htm)

Boehm, B., J. A. Fuhrman, R. D. Mrse and S. B. Grant. 2003. Tiered Approach for Identification of a human Fecal Pollution Source at a Recreational beach: Case Study at Avalon Bay, Catalina Island, California. Environ. Sci. Technol. 37: 673-680.

Booth, A. M., C. Hagedorn, A. K. Graves, S.C. Hagedorn, and K. H. Mentz. 2003. Sources of fecal pollution in Virginia's Blackwater River. Journal of Environmental Engineering. 129: 547-552.

Boyd, C. E. 2000. Water quality: an introduction. Boston, MA : Kluwer Academic Publishers. 251-283.

Bremner, J. M. 1965. Total nitrogen. In C.A. Black et al. (ed.). Methods of Soil Analysis, Part 2. Agronomy 9: 1149 – 1178. Am. Soc. of Agron., Inc., Madison, WI.

Bush L., W. Kurz, L. Powell, M. Power. G. Putnam, and E. Vandenburgh. 2003. Addressing Microbial Pollution in Costal Waters; A reference for local government. Appendix 1: Technical Source Tracking Methods. 41-47. www.ncnerr.org/ccs/publications/mst_appendix_1.pdf

Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell. 2003. Comparison of Ribotyping and Repetitive Extragenic Palindromic-PCR for Identification of Fecal *Escherichia coil* from Humans and Animals. Applied and Environmental Microbiology. 69:1836-1839.

Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw. 2001. Identification of Fecal *Escherichia coil* from human and Animals by Ribotyping. Applied and Environmental Microbiology. 67: 1503-1507.

Clean Water Act. 1972. http://www.epa.gov/r5water/cwa.htm

Collins, R., S. Elliott and R. Adams. 2005. Overland flow delivery of fecal bacteria to headwater pastoral stream. Journal of Applied Microbiology. 99: 126-132.

Crabill, C., R. Donald, J. Snelling, R. Foust, and G. Southam. 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. Water research. 33(9):2163-2171.

Davis-Colley, R. J., J. W. Nagels, R. A. Smith, R. G, Young, and C. J. Phillips. 2004. Water quality impact of a dairy cow herd crossing a stream. New Zealand Journal of Marine and Freshwater Research. 38 (4): 569-576.

Dombek, P. E., L. K. Johnson., S. T. Zimmerley, and M. J Sadowsky. 2000. Use of Repetitive DNA Sequences and the PCR to Differentiate *Escherichia coil* Isolates from Human and Animal Sources. Applied and Environmental Microbiology. 66: 2572-2577.

Donnison, A., C. Ross, and B. Thorrold. 2004. Impact of land use on the fecal microbial quality of hill country streams. New Zealand Journal of Marine and Fresh Water Research. 38 (5): 845-855.

Enzinger, R. M, and R. C. Cooper. 1976. Role of bacteria and protozoa in the removal of *Escherichia coli* from estuarine waters. Environmental Microbiology. 31: 758-763.

EPA. 2005. Microbial source tracking guide document. U.S. Environmental Protection Agency. EPA/600-R-05-064.

EPA. 2002. Method 1603: *E. coli* in water by membrane filtration using modified membrane- thermotolerant *Escherichia coil* Agar (modified mTEC). EPA-R-02-023

EPA. 2002. List of drinking water contaminants and MCLs. http://www.epa.gov/safewater/mcl.html

EPA 2000. Improved enumeration methods for the recreational water quality indicators: Enterococci and *E. coli*. EPA/821/R-97/004. 35-40.

EPA 2000. Ambient water quality criteria recommendations. www.epa.gov/waterscience/criteria/nutrient/guidelines/rivers/index.html

EPA. 1999. Clean Water Action Plan. http://www.cleanwater.gov/action/c1a.html

EPA 1999. Water quality criteria for ammonia. EPA-822-R-99-014

EPA. 1986. Ambient Water Quality Criteria for Bacteria -1986. http://www.epa.gov/waterscience/beaches/1986crit.pdf

Geldreich, E. E. 1970. Applying bacteriological parameters to recreational water quality. J.Am. Water Works Assoc. 62: 113-120.

Gordon D.M. 1997. The genetic structure of *Escherichia coli* populations in feral house mice. Microbiology. 143: 2039-2046.

Gordon, D. M., and F. FitzGibbon. 1999. The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology. 145: 2663-2671.

Gordon, D.M., and J. Lee. 1999. The genetic structure of entric bacteria from Australian mammals. Microbiology. 145: 2673-2682.

Grandjean, D., S. Fass, D. Tozza, J. Cavard, V. Lahoussine, S. Saby, H. Guilloteau, and J. C. Block. 2005. Coliform culturability in over-versus unsaturated drinking waters. Water Research. 39 (9): 1878-1886.

Gresshoff, P. M. 1997. DNA markers Protocols, Applications, and Overviews. University of Tennessee, Plant Molecular Genetics, Knoxville, Tennessee. 151-169.

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

Hagedon C. 2004. Development of Known Source Libraries. Microbial source tracking using indicator organisms. Workshop WS-20. ASM general meeting 23 May 2004.

Hagerdorn, C., K. Porter, and A. H. Chapman. 2003. Bacterial source tracking to identify sources of fecal pollution in the Potomac and Anacostia Rivers and Rock Creek, Washington, D.C. Final project report. EPA/600-R-05-064.

Hagerdorn C., S.L. Robinson, J.R. Filtz, S.M.Grubbs, T.A. Angier, and R.B. Reneau, Jr. 1999. Determining sources of fecal polluti Virginia watershed with antibiotic resistance patterns in fecal Streptocci. Applied and Environmental Microbiology. 65: 5522-5531.

Hartel, P. G., J. D. Summer, J. L. Hill, V. Collins, J. A. Entry, and W. I. Segars. 2002. Geographic Variability of *Escherichia coil* Ribotypes from Animals in Idaho and Georgia. Environ. Qual. 31: 1273-1278.

Hassan, W. M., S. Y. Wang, and R. D. Ellender. 2005. Methods to increase fidelity of repetitive palindromic PCR fingerprint-based bacterial source tracking efforts. Applied and Environmental Microbiology. 71 (1): 512-518.

Howell, J. M., M. S. Coyne, and P. Cornelius. 1995. Fecal bacteria in agricultural waters of the Bluegrass Region of Kentucky. Journal of Environmental Quality. 24 (3): 411-419.

Hyer, K. E., and D. L. Moyer. 2003. Patterns and sources of fecal coliform bacteria in the three streams in Virginia, 1999-2000. USGS water resource investigating report. 03-4115.

Irvine, K.N., and G.W. Pettibone. 1993. Dynamicsa of indicator bacteria populations in sediment and water near a combined sewer outfall. Environ. Technol. 14: 531-542.

Jagal, P. 1997. Stormwater runoff from typical developed and developing south African urban development: Definitely not for swimming. Water Science and Technology. 35:133-140.

Jamieson, R., D.M Joy, H. Lee, R. Kostaschuk, and R. Gordon. 2005. Transport and deposition of sediment-associated *Escherichia coli* in natural streams. Water Research. 39.12: 2665-2675.

Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky. 2004. Samples size, library composition, and genotypic diversity among natural population of Escherichia coli from different animals influence accuracy of determining sources of fecal pollution. Applied and Environmental Microbiology. 70 (8): 4478-4485.

Koplin D.W., E.T. Furlang, M.T. Meyer, E.M. Thurman, S.D Zaugg, L.B. Barber and H.T Buxton. 2002. Pharmaceutcals, hormones, and other organic wastewater concentrations in the U.S. streams, 1999-2000: A national reconnaissance. Environ. Sci. Technol. 36: 1202-1211.

Leeming, R., A. Ball, N. Ashbolt, and P. Nichols. 1996. Using fecal sterols from humans and animals to distinguish fecal pollution in receiving waters. Wat. Res. 30: 2893-2900.

Leung, K. T., R. Mackereth, Y. C. Tien, E. Topp. 2004. A comparison of AFLP and ERIC-PCR analysis for discriminating *Escherichia coli* from cattle, pig and human sources. FEMS Microbio. Ecol. 47: 111-119.

Matlock, M.D., M. E. Matlock, D. E. Storm, M. D. Smolen, and W. J. Henley. 1998. Limiting nutrient determination in lotic ecosystems using a quantitative nutrient enrichment periphytometer. Journal of American water resources. 34 (5): 1141-1147.

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

McLellan, S. L., A. D. Daniels, and A. K. Salmore. 2003. Genetic Characterization of *Escherichia coli* Population from Host Sources of Fecal Pollution by using DNA Fingerprinting. Applied and Environmental Microbiology. 69: 2587-2594.

McLellan, S. L., A. D. Daniels, and A. K. Salmore. 2001. Clonal population of thermotolerant Enterobacteriaceae in recreational water and their potential interference with fecal *Escherichia coli* counts. Appl. Environ. Microbiol. 67: 4934-4938.

Muirhead, R. W., R. J. Davies-Colley, A. M. Donnison and J. W. Nagels. 2004. Fecal bacteria yields in artificial flood events: quantifying in-stream stores. Water Research. 38(5): 1215-1224.

Myoda S. P., C. A. Carson, J. J. Fuhrmann, B. Hahm, P. G. Hartel, H. Yampara-Iquise, L. Johnson, R.L. Kuntz, C. H. Nakatsu, M. J. Sadowsky, and M. Samadpour. 2003. Comparison of genotypic-based microbial source tacking methods requiring a host origin database. Journal of Water and Health. 01.4 :167-180.

Online Montgomery.com
http://www.onlinemontgomery.com/visitors/demographics.cfm

Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin. 1999. Discriminant Analysis of Ribotype Profiles of *Escherichia coil* for Differentiating Human and Nonhuman Sources of Fecal Pollution. Applied and Environmental Microbiology. 65: 3142-3147.

Parveen, S., Murphree R. L., Edmiston L., Kaspar C. W., Portier K. M., and M. L. Tamplin. 1997. Association of Multiple- Antibiotic –Resistance Profile with Point and Nonpoint Sources of *Escherichia coli* in Apalachicola Bay. Applied and Environmental Microbiology. 63: 2607-2612.

Pupo, G. .M. and B. J. Richardson. 1995. Biochemical genetics of a natural populaton of *Escherichia coli*: seasonal changes in alleles and halotypes. Microblogy 141: 1037-1044.

Rademaker, J. L. W. and F. J. de Bruijn. 2004. Computer-assisted analysis of molecular fingerprint profiles and database construction.Molecular Microbial Ecology manual. 2nd ed., Kluwer Academic publishers : 1397-1446.

Rademaker, J. L. W. and F. J. de Bruijn. 1997. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis, p.151-171. In G. Caetano-Anolles and P. M. Gresshoff (ed.), DNA markers: protocols, applications, and overviews. J. Wiley and Sons, New York, N.Y.

Reynolds, K. A. 2003. Coliform bacteria: A failed indicator of water quality. www.wcp.net/column.cfm?T=T&ID=2349

Ritter, K. J., E. Carruthers, C. A. Carson, R.D. Ellender, V. J. Harwood, K. Kingsley, C. Nakatsu, M. Sadowsky, B. Shear, B. West, J. E. Whitlock, B. A. Wiggins, and J. D. Wilbur. 2003. Assessment of statistical methods used in library-based approaches to microbial source tracking. Journal of Water and Health. 01.4:209-223.

Rodgers, P., C. Soulsby, C. Hunter and J. Petry. 2003. Spatial and temporal bacterial quality of a lowland agricultural stream in northeast Scotland. The Science of the Total Environment. 314: 289-302.

Rose, J. B., J. H. Paul, M. R. McLaughlin, V. J. Harwood, S. Farrah, M. Tamplin, J. Lukasik, M. Flanery, P. Stanek and H. Greening. 2000. Healthy beaches Tempa Bay: Microbilogical monitoring of water quality conditions and public health impacts. Final project report. EPA/600-R-05-064.

Sadowsky, M. 2004. Determination of fecal pollution sources in Minnesota watersheds. Technical report prepared for the Legislative Commision on Minnesota Resources.

Savageau, M. A. 1983. *Escherichia coli* habitats cell types, and molecular mechanism of gene control. Am. Nat 122, 732-744.

Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik. 2002. Microbial Source Tracking: Current Methodology and Future Directions. Applied and Environmental Microbiology. 68: 5796-5803.

Sherer, B. M., J. R. Miner, J. A. Moore, and J. C. Buckhouse. 1992. Indicator bacteria survival in stream sediments. Journal of Environmental Quality. 21(4): 591-595.

Simmons, G. M., D. F. Waye, S. Herbein, S. Myers, and E. Walker. 2000. Estimation nonpoint fecal coliform sources in northern Virginia's four mile run watershed. 248-

267. Proceedings of the Virginia water research symposium, VWRRC special report. SR-19-2000, Black sburg.

Simpson, J. M., W. Jorge, S. Domingo, and D. J. Reasoner. 2002. Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36: 5279-5288.

Sims, G.K., T.R. Ellsworth and R. L. Mulvaney 1995. Microscale determination of inorganic nitrogen in water and soil extracts. Communication in Soil Sci. and Plant. 26.

Smalla, K., H. Heuer, A. Gotz, D. Niermeyer, E. Krogerrecklenfort, and E. Tietze. 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.

Soil Survey. 1960. Montgomery County Alabama. United States Department of Agriculture, Soil Conservation Service; Alabama Department of Agriculture and Industries; Alabama Agriculture Experiment Station. 7: 1-10.

Southern forest resource assessment. 2001.
www.srs.fs.usda.gov/sustain/draft/qual/aqual-22.htm

Souza, C., M. Rocha, A. Valera, and L.E. Eguiarte. 1999. Genetic structure of natural population of *Escherichia coli* n wild hosts on different continents. Appl. Environ. Microbiol. 65: 3373-3385.

Stiles, N., 2003 Fall. Bacterial Source Tracking Uncovers Culprits of Contamination of Watersheds. Small Flows Quarterly. 4:4. 24-29.

Stoeckel, D. M., M.V. Mathes, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, L. O'Brien, T. W, Fenger, M. Samadpour, K. M. Strickler, and B. A. Wiggins. 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. Environ. Sci. Tech. 38: 6109-6117.

Tam, J., T. M. Scott, M. R. Morgan, and J. B. Rose. 2005. Occurrence of alternative fecal indicators and enteric viruses in Michigan Rivers. Journal of Great Lakes Research. 31(1): 22-31.

Todar, K. 2002. Pathogenic *E. coli* University of Wisconsin-Madison Department of Bacteriology Pathogenic. http://www.bact.wisc.edu/bact330/lectureecoli

USGS water resources of Alabama. http://al.water.usgs.gov

Verseveld, H. W., and W. F. M. Roling. 2004. Cluster analysis and statistical comparison of molecular community profile data. Molecular Microbial Ecology manual. 2nd ed. 7.04: 1373-1396.

Whittam, T. S. 1989. Clonal dynamics of *Escherichia coli* in its natural habitat. Antonie Leeuwenhoek 55, 23-32.

Whittam, T. S., H. Ochman, and R. K. Selander. 1983. Geographic components of *Escherichia coil*. Mol. Biol. Evol. 1: 67-83.

Whitlock, J. E., D. T. Jones, and V. J. Harwood. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Res. 36: 4273-4282.

WHO. 2001. Water Immediate Solution for Persistent Problems.
http://www.who.int/inf-pr-2001/en/pr2001-12.html

Wiggins, B. A., P. C. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K Varner. 2003. Antibiotic resistant analysis for representativeness testing of multiwatershed libraries. Applied and Environmental Microbiology. 69 (6): 3399-3405.

V. APPENDIX

Table A.1 Latitudes and Longitudes of the Catoma Creek watershed sampling sites

Sampling site	Latitudes	Longitudes
Catoma Creek at Old Selma Road (CO)	N 32° 20.617’	W 86° 23.521’
Catoma Creek at Court Street (CC)	N 32° 18.442’	W 86° 18.452’
Baldwin Slough (BS)	N 32° 18.166’	W 86° 16.570’
White Slough (WS)	N 32° 17.480’	W 86° 16.130’
Catoma Creek at Woodley Road (CW)	N 32° 16.715’	W 86° 13.152’
Ramer Creek at Snowdown Chamber Road (RSC)	N 32° 15.050’	W 86° 14.634’
Ramer Creek at Sprague Junction Road (RSP)	N 32° 07.807’	W 86° 15.888’
Little Catoma (LT)	N 32° 16.071’	W 86° 09.980’