The Use of Molecular and Genomic Tools to Examine the Population Structure of *Escherichia coli* in the Environment

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Matthew J. Hamilton

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Michael J. Sadowsky

October 2009

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Michael Sadowsky, and my committee chair Dr. Sandra Armstrong, for their mentoring and guidance during my graduate school tenure. I would also like to thank Drs. Lynda Ellis, Francisco Diez-Gonzalez, and Gary Dunny for their assistance and for serving on my committee. I extend special thanks to Dr. Tao Yan who deserves much credit for his work on the high throughput screening methodology.

I also thank Andrew Vail, Dan Norat, Charlie Sawdey, Jessica Eichmiller, Ramya Chandrasekaran, and Dave Sukovich for the rich discussion, even though much of that discussion had little to do with the research presented in this thesis.

Finally, I would like to thank Dr. Siriluck Jitacksorn for being a good friend and all of the past and present students, post-docs, and visiting scholars that helped to define my graduate school experience.

DEDICATION

I would like to dedicate this dissertation to my family: James, Patricia, Sarah, and Ilene and to B. K. and L. K. for their unwavering support.

ABSTRACT

Fecal contamination of water is a widespread public health problem throughout the world. In developing countries, contamination of water by fecal material is a persistent threat to public health, where water-borne diarrheal disease is responsible for a significant number of deaths. The magnitude of contamination is determined by the enumeration of fecal indicator bacteria, including *Escherichia coli*. While *E. coli* is often thought of as a harmless commensal organism in the lower intestine of warm-blooded animals, pathogenic strains capable of causing severe disease and naturalized with the ability to grow in the environment also exist. In this dissertation, I describe several studies concerning *E. coli* in the environment which have implications for future water quality studies and are of interest to other researchers in the field, in addition to water quality regulators, beach managers, and public health professionals.

In the first study, I describe the isolation of DNA markers specific to *E. coli* originating from waterfowl for use in a culture-based, library independent method for microbial source tracking. This system used a DNA subtraction procedure to enrich for waterfowl-specific sequences which were then tested for specificity in colony hybridization assays. These markers were capable of identifying greater than 70% of waterfowl *E. coli*, while cross hybridization with strains from other hosts averaged about 10%.

An additional study focused on the use of the waterfowl *E. coli* DNA marker system and a high throughput, semi-automated robotic system to develop a macroarray colony hybridization assay for large scale quantitative detection of *E. coli* originating

from waterfowl. This methodology was used to study the fecal input from waterfowl in two Minnesota lakes.

The high throughput system robotic system used in the source tracking study was adapted to screen large numbers of *E. coli* isolates obtained from a contaminated beach in California for the presence of virulence genes. The results of this study indicated that potential enteropathogenic *E. coli* (EPEC) strains were consistently detected in the water, which may represent a public health risk to recreational beach users.

Finally, I describe a study examining the spatial and temporal changes in *E. coli* populations present in the water and sediments in a small stream in south central Minnesota using a variety of molecular methods. This study indicated persistent *E. coli* strains are likely impacting the creek and that flow conditions and other environmental factors are likely dominant processes affecting *E. coli* populations in the creek.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

General Introduction

INTRODUCTION

The purpose of the introductory material in this chapter is to give the reader general background information concerning fecal contamination of water, *Escherichia coli* and its use as an indicator of fecal contamination, microbial source tracking, pathogenic *E. coli* strains, and the survival and persistence of *E. coli* in the environment. More detailed introductory material is presented at the beginning of each subsequent chapter.

CONTAMINATION OF WATER WITH FECAL MATERIAL

Fecal contamination of water is a widespread problem affecting much of the United States. According to a 2002 U.S. Environmental Protection Agency (EPA) report, 93,000 river and stream miles are considered impaired due to fecal contamination in the United States alone (175). In developing countries, contamination of water by fecal material is a persistent threat to public health. Water-borne diarrheal disease were responsible for an estimated 1.8 million deaths in 2004, and children were disproportionally represented (186). Water-borne pathogens can also enter the food supply through irrigation of crops with contaminated water (1, 157, 186).

Potential sources of fecal pollution include sewage treatment and septic systems, runoff from livestock feedlots and manure amended agricultural fields, and wildlife (67, 88, 172). Water monitoring for fecal pollution is mandated by the Clean Water Act and indicator bacteria (primarily *E. coli* and enterococci) shed in the feces are used as tools to assess the magnitude of contamination (46). It has been generally accepted that the

presence of fecal indicator bacteria suggests the presence of fecal borne pathogens such as *Salmonella*, *Shigella*, *Campylobacter*, and enteric viruses (44). Several studies support this hypothesis and have shown that elevated counts of indicator bacteria have been correlated to an increased risk of gastrointestinal disease (28, 44, 130, 151, 162). In addition to the public heath risk associated with contaminated water contact, consistent fecal contamination can result in beach and fishery closures which may cause significant societal impacts and economic losses (46, 160). Total Maximum Daily Load (TMDL) determinations are used in efforts to abate bacterial pollution of waterways (173).

E. COLI AS AN INDICATOR OF FECAL CONTAMINATION

Currently, fecal contamination is detected and quantified through the enumeration of indicator bacteria using culture techniques. *Escherichia coli* and *Enterococcus* sp. are often used as fecal indicators in freshwater and marine systems, respectively. In Minnesota, recreational waters exceeding the *E. coli* standard of 126 colony forming units per 100 ml, as a geometric mean of at least 5 samples, are considered impaired and may be subject to beach closures. *E. coli* is a rod shaped Gram-negative bacterium in the family Enterobactericeae within the γ-Proteobacteria. While it has been historically thought of as a harmless, commensal organism found in the gut of warm blooded animals, pathogenic strains causing a variety of infections have also been isolated (91, 115). Viable cells are readily shed from host feces and concentrations may be as high as 10^6 organisms per gram fecal material (143). The use of *E. coli* as an indicator of fecal pollution is based on correlations with public health risk (28, 44) and the assumption that

the organism does not survive or persist in the environment, therefore signaling recent contamination. Several studies, however, have shown that *E. coli* can persist and grow in natural environments, including soil, sand, sediments and in association with algae (25, 27, 50, 63, 76, 77, 156). Results of these studies have obvious implications concerning the use of *E. coli* as an indicator of fecal pollution and will be discussed in more detail later in this chapter.

MICROBIAL SOURCE TRACKING

Total Maximum Daily Load (TMDL) determinations are currently used in an effort to abate fecal contamination of waterways. As required by the federal Clean Water Act, the sources and magnitude of the pollutants must be identified and this information is an important part of the TMDL determination. From a public health perspective, contamination of water with feces from some animals may represent a greater risk to human health than contamination from other animals. For example, cattle, goats, sheep, and other ruminants are known reservoirs for pathogenic strains of *E. coli* and *Salmonella* (78, 88, 91, 115). Additionally, it is generally accepted that human feces is a source of human pathogens, such as enteric viruses (129). In contrast, contamination of waterways with the feces from some animals may represent a low risk to public health if these animals do not frequently harbor pathogens. In light of this information, microbial source tracking studies have implications for beach managers, who are concerned with complying with the law and public health professionals who are concerned about the health risks associated with contact with contaminated water.

Identifying the sources of fecal contamination is a difficult task. Historically, researchers have used phenotypic methods, such as antibiotic resistance profiles of indicator bacteria and fecal streptococci ratios, in an effort to distinguish between fecal pollution resulting from human and livestock sources (52). Unfortunately, these methods suffer from significant inconsistencies and, in many cases, have provided little help in determining actual pollution sources. Most of the current methodologies for source tracking use molecular techniques and can be classified based on the need to culture organisms before analysis and the use of a known source isolate library to identify unknown organisms in the environment (46, 142, 147, 188). Library dependent methods usually require a culturing step and include Horizontal Fluorophore Enhanced Rep-PCR (42, 88), AFLP (102), ribotyping (29, 30, 113) and other DNA fingerprinting techniques. These methods attempt to identify the sources of isolates by comparing fingerprints or banding patterns of environmental isolates to those in a library of known source isolates. These methodologies suffer from the requirements for prohibitively large library of known source isolates to ensure accurate comparisons and source assignments (88). In the case of *E. coli*, it has been estimated that there are between 60,000 and 80,000 genetically distinct strains, hence, an extremely large library is necessary to encompass all of this diversity (88). While several of these assays have been successful in assigning isolates to a host source group using limited libraries, issues such as temporal and geographic variability, low throughput, long processing time, costs, and specificity influence the use and widespread application of these methods (46, 64, 65, 147, 188).

PCR based methods, such as host specific 16S rRNA assays do not require the use of libraries or the need to culture organisms before analysis (13, 38, 39, 100, 148-150).

These assays often detect the presence of host specific *Bacteroides* sp. strains and can utilize real-time PCR technology to generate a quantitative measure of fecal inputs. Some of these techniques have shown much promise in successfully identifying the presence of host specific fecal bacteria (98, 148, 149). Unfortunately, it is not currently possible to relate the results of *Bacteroides*-based PCR assays to standard counts of fecal indicator bacteria. In addition, the percentage of hosts that carry bacteria with the specific 16s rRNA sequence may vary considerably (149). Some of these assays may also suffer from temporal and geographic variability, low throughput, and issues with specificity (46, 147, 188). Library-independent, culture-dependent assays, such as the antiquated fecal streptococci ratio method (52), detection of host specific viruses and bacteriophages (87), and macroarray hybridization assays for strains carrying host specific markers are advantageous to other methods in that libraries are not required and, in the case of the macroarray hybridizations, a quantitative measure of fecal inputs from a specific source can be obtained. These methods, however, may also suffer from specificity issues and processing time, in addition to geographic and temporal effects (46, 147, 188).

It is important to note that the field of microbial source tracking is still relatively new and in the development phase. While several of the methods described above have shown promise, all have their drawbacks and problems. It is possible that in order to successfully identify the sources of fecal bacteria in a given sample, it will be necessary to take a combined analysis approach, utilizing several different microbial source tracking methods (56, 142, 174). For example, while several host-specific 16S rRNA-based PCR methods are available to detect the presence of bacteria from human and bovine feces, there is no primer set to detect bacteria in waterfowl feces and the

macroarray based system described in chapter 3 could be used to gauge inputs from waterfowl. Combined methods have been used in several studies to examine fecal inputs in Santa Monica Bay, CA (118) and in rural Nebraska (176). Additional microbial source tracking research and comprehensive field studies are necessary before these methods can be adapted for use by public health labs and government regulatory agencies concerned with water quality.

PATHOGENIC *E. COLI*

Although most *E. coli* strains are harmless commensal organisms, pathogenic strains do exist and are generally associated with diarrheal disease. Pathogenic *E. coli* strains are often associated with food- and water-borne outbreaks of diarrheal disease, especially in developing countries (91, 115, 186). In recent years, a series of *E. coli* outbreaks in the United States have drawn considerable media attention. These strains make use of genes encoding toxins, adhesins, hemolysins, and other virulence factors to cause disease. Virulence determinants are almost always located on mobile elements such as plasmids, transposons or pathogenicity islands, and likely arose through horizontal gene transfer (91, 115). The exchange of these mobile elements between strains allows for novel combinations of virulence factors and the emergence of previously unseen strains. Clinical disease symptoms arising from *E. coli* infections can range from self-limiting and mild diarrhea to hemorrhagic colitis and death. Children, the elderly, and those with compromised immune systems are at the greatest risk for severe disease.

Pathogenic *E. coli* strains are divided into distinct pathotypes based on the presence of specific virulence genes and mechanisms of pathogenesis. Enteropathogenic *E. coli* (EPEC) are a common cause of infant diarrhea, especially in developing countries, and are noted for the development of attaching/effacing lesions in the lower intestine. These strains are characterized by the presence of the locus of enterocyte effacement (LEE), a pathogenicity island encoding for several virulence factors, including intimin (*eaeA*) (110). As the name suggests, intimin is involved in the intimate attachment of the bacterium to intestinal epithelium and plays a key role in pathogenesis (85). Shiga-like toxin producing strains (STEC) are those strains carrying one or both of the Shiga-like toxin genes, *stx1* and *stx2* and they also usually possess the LEE pathogenicity island. A subset of the STEC pathotype, enterohemorrhagic *E. coli* (EHEC), can cause hemorrhagic colitis and hemolytic uremic syndrome, which can lead to organ failure and death (8). STEC strains are frequently associated with ruminant reservoirs, including cattle, sheep and goats (78, 91, 115). Enterotoxigenic *E. coli* (ETEC) strains are a frequent cause of traveler's diarrhea (17) and these strains can code for two different types of toxins, the heat stable toxin (ST) and heat labile toxin (LT) (120). Other less common pathotypes include diffusely adhering (DAEC), enteroinvasive (EIEC), and enteroaggretive (EAEC) *E. coli* (91, 115). In addition to diarrheagenic strains, pathogenic *E. coli* strains have also been associated with urinary tract infections, neonatal meningitis, and sepsis (155).

E. COLI IN THE ENVIRONMENT

The idea that the lower intestine of warm blooded animals is the only natural habitat for *E. coli* has been challenged by a number of studies reporting on the survival and growth of *E. coli* in the environment. Until recently, it was thought that *E. coli* survived poorly in the environment and this was one of reasons why *E. coli* was chosen as a fecal indicator bacterium. The growth and persistence of *E. coli* in the tropical soils and sediments of Hawaii was first reported by Hardina et al. (63) in 1991. Subsequent studies in tropical and subtropical regions reported on the survival of *E. coli* in soils obtained from Guam (50) and in soils, sediments, and water in Florida (156). More recent reports by Byanhpallahali et al. examined survival and growth of *E. coli* in the temperate soils, sands and sediments of Indiana and several studies by Ishii *et al*. reported on the long term survival of *E. coli* in the northern temperate climates of Minnesota (25, 27, 76, 77). In addition to the ecosystems described above, *E. coli* has also been found to survive and propagate in association with the filamentous macroalgae *Cladophora* (26, 79) and in periphyton (97).

E. coli strains obtained from these environmental sources have been examined using DNA fingerprinting techniques and are genetically distinct from intestinal *E. coli* isolated from various hosts, suggesting that these strains have become naturalized to these environments (77). It is likely that naturalized *E. coli* were originally deposited into the environment in the feces from host animals and adapted to natural environments by making use of versatile energy and nutrient acquisition systems. *E. coli* are heterotrophic bacteria and only need simple carbon, nitrogen and phosphorus sources, plus sulfur and other trace elements, for growth. Díaz et al. 2001, reported that *E. coli* can obtain carbon

through the degradation of various aromatic compounds, a common trait of soil microbes (37). Additionally, *E. coli* is a facultative anaerobe explaining its persistence in the hypoxic sediments of lakes and streams and in soils. *E. coli* can also grow across a wide range of temperatures (7.5 - 49^{o} C) and displays long term survival at freezing temperatures (9, 20, 51, 75, 89, 152). Survival and growth in the environment is likely limited by various environmental stresses, including desiccation (15, 20, 23, 27, 121, 156), organic matter content (164, 184), soil texture differences (36, 121), salinity (163), UV light exposure (180), predation (18, 22), and temperature extremes (121, 156, 184).

Naturalized *E. coli* populations are of particular interest to water quality researchers and regulators. Fecal indicator counts can be artificially inflated by run-off containing naturalized *E. coli* from soils and beach sands and through the resuspension of sediments and the inoculation of water with algal and peridriphyton associated bacteria. This may lead to unnecessary beach closures. Studies by Hardina et al. (63) and Fujioka et al. (50) reported that soil was a significant source of the *E. coli* detected in creeks and streams in tropical regions. Ishii et al. (76) also reported that beach sands and sediments were a source and a sink for *E. coli* isolated from water at a popular swimming beach at a Lake Superior beach in Duluth, MN. In light of these studies and others, researchers are currently examining alternative bacterial species for use as indicators of fecal pollution, in addition to developing rapid, molecular-based methodologies to measure fecal pollution in contaminated waters.

SUMMARY OF THESIS

All of the work described in this thesis focuses on *E. coli* and while all the chapters are related, several different topics concerning *E. coli* in the environment are investigated. The studies presented in this thesis have implications for future water quality studies and are of interest to other researchers in the field, in addition to water quality regulators, beach managers, and public health professionals. Chapter 2 describes the identification of DNA markers specific to *E. coli* originating from geese and ducks for use in a culture-based, library independent method for microbial source tracking. This system used a DNA subtraction procedure to enrich for waterfowl-specific sequences which were then tested for specificity in colony hybridization assays. Chapter 3 focuses on the use of the waterfowl *E. coli* DNA marker system and a high throughput, semi-automated robotic system to develop a macroarray colony hybridization assay for large scale quantitative detection of *E. coli* originating from ducks and geese. This methodology was used to study the fecal input from waterfowl in several Minnesota lakes. In Chapter 4, I discuss the use of the high throughput macroarray screening technology, developed in Chapter 3, for detecting the presence of several virulence factor genes in a large numbers of *E. coli* isolates obtained from a contaminated beach in California. Strains carrying virulence factor genes were characterized by virulence subtype and phylogenetic group analyses and the populations of potentially pathogenic strains were examined by HFERP. Lastly, Chapter 5 examines the spatial and temporal changes in *E. coli* populations present in the water and sediments in a small stream in south central Minnesota using a variety of molecular methods.

CHAPTER 2

Development of Goose- and Duck-specific DNA Markers to Determine Sources of *Escherichia coli* in Waterways.

OVERVIEW

The contamination of waterways with fecal material remains a persistent threat to public health. Identification of the sources of fecal contamination is a vital component for abatement strategies and in the determination of total maximum daily loads (TMDLs). While phenotypic and genotypic techniques have been used to determine potential sources of fecal bacteria in surface waters, most methods require the construction of large known-source libraries, and often fail to adequately differentiate among environmental isolates originating from different animal sources. In this study, I used pooled genomic tester and driver DNAs in suppression subtractive hybridizations to enrich for host source-specific DNA markers for *E. coli* originating from locally-isolated geese. Seven markers were identified. When used as probes in colony hybridization studies, the combined marker DNAs identified 76% of the goose isolates tested, and cross-hybridized, on average, with 5% of the human *E. coli* strains, and less than 10% with strains obtained from other animal hosts. In addition, the combined probes identified 73% of the duck isolates examined, suggesting that they may be useful for determining the contribution of waterfowl to fecal contamination. However, the hybridization probes mainly reacted with *E. coli* obtained from geese in the upper Midwest U.S., indicating regional specificity of the identified markers. Coupled with high throughput, automated, macro- and microarray screening, these markers may provide a quantitative, cost-effective, and accurate library-independent method to determine sources of genetically diverse *E. coli* for use in source tracking studies. However, future efforts to generate DNA markers specific for *E. coli* must be done using isolates obtained from geographically diverse animal hosts.

INTRODUCTION

The contamination of waterways by pathogenic microorganisms is, and remains, a persistent threat to public health (151, 162). These waterborne pathogens can be transmitted though drinking water systems, water-related recreational activities, and by consumption of shellfish (6, 32). The contamination of waterways with fecal material has generally been regarded as the major contributor of waterborne pathogens and Total Maximum Daily Loads (TMDL) determinations are currently being used to abate this type of pollution and restore waterways for their designated uses. Identification of the sources of fecal contamination is a vital component of TMDL determinations, providing information about the type, magnitude, and location of pollutant inputs (173). Sources of fecal coliform bacteria in the environment include: runoff from feedlots, manure-amended agricultural land, wildlife, malfunctioning septic systems, urban runoff, sewage discharge, and soil-borne bacteria (77, 88).

Phenotypic and genotypic techniques have been used to determine potential sources of fecal bacteria found in surface waters (12, 13, 38, 42, 56, 67, 88, 111, 124, 145-147, 151, 162), and *Escherichia coli* and *Enterococcus* sp. strains are the most widely used bacteria for these studies. The majority of these methodologies require the construction of known-source libraries to differentiate among environmental isolates originating from different animal sources (154). However, since the size of the host source libraries is often limited, many consisting of about 35 to about 2500 isolates (88), they do not allow for adequate determination of potential sources of environmental *E. coli* and *Enterococcus* isolates. Moreover, the utility of known source libraries is further challenged by the lack of representation, due to temporal and geographic variation in

bacterial genotypes within and between animal species (54, 64, 84, 146), the presence of multiple strains within a single animal (111), host animal diet variation (65), the presence of soil- and algal- borne indicator organisms (25, 77), transient inhabitants of GI tracts, and a great degree of genetic diversity among microorganisms used for source tracking analyses (88, 111).

Based on these shortcomings, investigators have evaluated the use of library-independent methods to define sources of fecal bacteria in the environment. These methods, which avoid issues of library size and isolate diversity, use both growth-dependent as well as growth-independent technologies. Enteric viruses have also been investigated for use in the growth- and library independent analyses of fecal pollution sources. These studies revealed that viruses from various animal sources display some level of host specificity (87, 105, 129), and molecular assays have been developed to examine the usefulness of viruses in microbial source tracking studies (47, 86). Several studies have reported on the development of 16S rRNA gene-based genetic markers for the growth- and library independent analysis of *Bifidobacterium* and *Bacteroides-Prevotella* for source identification purposes (12, 13). Recently, Dick and coworkers reported the effective use of a microplate subtractive hybridization method to define host-specific 16S-based genetic markers for *Bacteroides* sp. strains(39) . Similarly, Scott and coworkers (145) reported the isolation of a host-specific marker gene from *Enterococcus faecium*, encoding a putative virulence factor (*esp*), that allows determination of sources of enterococci in waterways. While these methods show great promise as microbial source tracking tools, they may be limited by the inability to obtain high throughput data, and the expense and limitations associated with the use of PCR in environmental

samples. In addition, both systems do not allow correlation to fecal coliform or *E. coli* counts that are commonly obtained by government agencies for fresh water systems.

In this study, I report on the development and validation of host source-specific genetic markers for *E. coli* strains originating from Canada geese (*Branta canadensis*). These markers were shown to be useful in determining sources of fecal pollution in Lake Superior, and are useful for high throughput studies. Instead of randomly screening for host source-specific genes, I took a genomic comparison approach by using suppression subtractive hybridization (SSH) to define host source-specific markers. The SSH technique has been found to be useful to examine genetic diversity in *E. coli* (112) , identify genetic differences between closely related strains (5, 112), examine pathogenicity determinants in *E. coli* (81), and to develop probes to examine natural bacterial communities (109). More importantly, the SSH approach has been found to be an effective tool for the development of strain- and host source-specific marker probes (4, 62, 72, 82, 106).

MATERIAL AND METHODS

Escherichia coli strains

The *E. coli* strains used in suppression subtractive hybridizations (SSH), and subsequent specificity analyses, were obtained from a previous library of unique isolates obtained from the feces of 12 known animal host sources (cats, chickens, cows, deer, dogs, ducks, Canada geese, goats, horses, pigs, sheep, and turkeys), and humans (42, 88). All *E. coli* isolates were obtained in Minnesota and Wisconsin from 1998-2005. Unique

strains were defined as those isolates from a single host animal that had DNA fingerprint similarity coefficients less than 92%. Horizontal fluorophore-enhanced rep-PCR (HFERP) DNA fingerprints (88) from *E. coli* strains obtained from goose and human sources were analyzed for genetic relatedness using Pearson's product-moment correlation coefficient, with 1% optimization, and dendrograms were generated using the unweighted pair group method with arithmetic means (UPGMA). Based on these analyses, five strains from geese (Go66, Go90, Go126, Go172, and Go206) and humans (Hu51, Hu130, Hu132, Hu188, and Hu252), that showed maximum differences in genetic relatedness, were chosen for suppression subtractive hybridization studies and subsequent probe development. An additional 200 unique *E. coli* isolates were obtained, on multiple days in 2005, from the water column two meters off-shore in Lake Superior harbor in Duluth, MN as previously described (77). Twenty-seven of these strains were presumptively identified as having originated from geese based on HFERP DNA fingerprint comparisons and bootstrap analyses done using known source fingerprint libraries (77, 88). To determine if marker DNAs were capable of hybridizing with goose isolates from other geographic areas 172, 100, 73, and 14 *E. coli* isolates were also obtained from Canada geese in Delaware, West Virginia, Wisconsin, and Indiana, respectively.

Isolation of environmental *E. coli*

Offshore lake water samples were collected, using standard procedures (21), from Lake Phalen (St. Paul, MN), an urban lake frequented by Canada geese. Water samples (2 L) were filtered through 0.45 µm Nuclepore® polycarbonate membranes (Whatman,

Florham Park, NJ). Bacteria on membrane surfaces were resuspended in phosphate buffered saline, pH 7.0, using a sterile magnetic stir bar and vortexing to facilitate suspension of bacterial cells. A total of 1,152 *E. coli* isolates were isolated from the concentrated samples as previously described (42), and stored at –80^{o} C before use.

Suppression subtractive hybridizations

SSH was done using the CLONTECH PCR-Select™ Bacterial Genome Subtraction kit (BD Biosciences CLONTECH, Mountain View, CA) according to the manufacturer's instructions. Genomic DNA from the five goose and human *E. coli* strains was prepared using a cesium chloride, density-gradient centrifugation method as previously described (140). Two µg aliquots of genomic DNAs from the five goose and human *E. coli* strains were separately pooled together, and used as tester and driver DNAs, respectively. Prior to the subtraction procedure, 2 µg aliquots of each pooled sample were digested to completion with *RsaI*. SSH was repeated using PCR amplified secondary subtraction products as tester DNA to further enrich for tester-specific fragments. To create a library of potential DNA inserts that were specific for geese, the final subtraction products were cloned into pGEM-T vector using a T/A cloning procedure (Promega, Madison, WI). One hundred ninety-two clones were randomly selected and stored frozen at -80^{o} C in 50% glycerol until use.

Identification of DNA sequences specific for goose *E. coli*

The library of cloned potential goose-specific DNA fragments was screened for hybridization specificity using a dot-blot procedure as described by Schleicher & Schuell,

Keene, NH (http://www.schleicher-schuell.com/bioscience). Cloned insert DNAs were amplified by PCR using nested primers 1 (5'- TCGAGCGGCCGCCCGGGCAGGT- 3') and 2R (5' -AGCGTGGTCGCGGCCGAGGT- 3') provided in the CLONTECH SSH kit. PCR reactions were carried out using the following conditions: 94^{o} C for 2 min; followed by 25 cycles of 94^{o} C for 30 sec, 68^{o} C for 30 sec, and 72^{o} C for 1 min, and a final elongation step of 2 min at 72^{o} C. PCR products (0.5 µg) were spotted onto duplicate Nytran® SuPerCharge nylon membranes (Schleicher & Schuell, Keene, NH) using a dot-blot vacuum manifold (Gibco- BRL, Gaithersburg MD) and the Minifold® spotting protocol (Schleicher & Schuell, Keene, N.H.). Membranes were baked at 80^{o} C for 2 h, and prehybridized overnight at 42^{o} C in 6X SSC, 10X Denhardt's solution, 1% SDS, and 100 µg denatured herring sperm DNA per ml (141). Aliquots (125 ng) of *RsaI*-digested pooled genomic DNAs from the five human or goose *E. coli* strains were labeled with ^{32}P-αCTP using a random primer labeling kit (Invitrogen, Carlsbad, California) according to the manufacturer's protocol. Probes were hybridized for 18 h at 46^{o} C. Membranes were washed under high stringency as previously described (141). Images were captured using a STORM 840 densitometer (Molecular Dynamics, Piscataway, NJ). Presumptive goose-specific DNA inserts were identified by visual differences in hybridization intensity.

Plasmids from presumptive goose-specific clones were isolated using the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Insert DNA was amplified by PCR using nested primers 1 and 2R as described above, electrophoresed on 2% agarose gels, and DNAs transferred to Nytran®

SuPerCharge nylon membranes as described (141). Membranes were probed with the *RsaI*-digested, pooled, genomic DNAs as described.

DNA sequencing and analysis

Confirmed goose-specific DNA inserts were sequenced, in both directions, using pUC/M13 universal forward 5´-CGCCAGGGTTTTCCCAGTCACGAC-3´ and reverse 5´-TCACACAGGAAACAGCTATGAC-3´ sequencing primers. Sequencing reactions were performed using BigDye® (Applied Biosystems, Foster City, CA) sequencing chemistry at the Biomedical Genomics Center, University of Minnesota, St. Paul, MN. Translated sequences were analyzed using the BLASTX algorithm at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) using GenBank and the *E. coli* databases.

Colony hybridization for probe evaluation and environmental applications

The specificity of subtracted DNA inserts was evaluated by colony hybridizations to 48 cat, 96 chicken, 96 cow, 96 deer, 96 dog, 81 duck, 135 goose, 42 goat, 78 horse, 210 human, 96 pig, 60 sheep, and 96 turkey *E. coli* isolates (88). An additional 27 *E. coli* strains isolated from Lake Superior Harbor in Duluth, MN, 1,152 isolates from Lake Phalen (St. Paul, MN) and 359 isolates from Canada geese obtained in Delaware, West Virginia, Wisconsin, and Indiana were also evaluated by colony hybridization. Probe specificity was also evaluated using blind samples consisting of 96 randomly selected isolates obtained from geese, horses, pigs, sheep, and humans. *E. coli* strains from animal and human sources were inoculated from frozen stocks onto Nytran® SuPerCharge (20 cm^2) membranes (Schleicher & Schuell, Keene, NH) using a 48 pin

multiple inoculator. Membranes were placed onto the surface of 22 x 22 cm LB (141) agar plates (Genetix, UK) and incubated at 37^{o} C for approximately 5 h. Colonies were lysed and DNA on membranes was processed as described (141). Membranes were prehybridized at 68^{o} C overnight in a solution containing 6X SSC, 10X Denhardt's, and 100 µg denatured herring sperm DNA per ml. Probes from insert DNAs (50 ng) were labeled using the Random Primer DNA Labeling System (Invitrogen, Carlsbad, California), according to the manufacturer's protocol. Membranes were hybridized overnight at 68^{o} C in a solution containing 6X SSC, 10X Denhardt's, and 100 µg denatured herring sperm DNA per ml. Blots were finally washed in 0.1X SSC at 65^{o} C, and imaged as described below.

Quantitative image analysis

Quantitative image analysis was used to determine positive and negative signals on colony hybridization membranes. Images were captured using a STORM 840 densitometer (Molecular Dynamics, Piscataway, NJ) and analyzed using ScanAlyze version 2.50 software (http://rana.lbl.gov/EisenSoftware.htm). The normalized intensity of each spot was calculated by subtracting the median intensity of background from the mean intensity of each spot. Normalized spot intensities were plotted using Sigma Plot version 8.0 software (Systat Software, Point Richmond, CA) and a cut-off value was assigned based on normalized mean intensities of negative control spots, plus three times the standard deviation.

Generation of a Go066 gene library and sequencing of the goose associated adhesin gene

Genomic DNA from goose *E. coli* strain Go066 was used to construct a gene library in the pCUGIBAC1 (107) vector as previously described (107, 141). Genomic DNA was partially digested with BamHI to produce a mean fragment size of ~12 kb and the resulting fragments were cloned into the *E. coli* strain DH10B. The library was screened using colony hybridization with a ^{32}P labeled 332 bp fragment of a goose specific gene, designated GB2, as described above. A primer walking technique was used to obtain the DNA sequence flanking the GB2 goose specific fragment. Sequence data from individual reactions was assembled into a contiguous sequence using the CAP3 sequence assembly program (73). Sequence data was used in blast homology searches with the BLASTX and BLASTP algorithms and the GenBank database.

Cloning the goose specific adhesin gene

Primers GB2-5F2 (5'-TACGGCAGTGAGAGCAGAGA-3') and GB2-3R3 (5'-CGCGACTGTCAGATATTCA-3') were developed to amplify a 3.1 KB fragment from Go066 genomic DNA containing the open reading frame of the goose associated adhesin gene. The PCR product was cloned into the pGEM-T vector, making plasmid pMJH001 (Appendix 2). The pMJH001 plasmid was doubly digested with BstZ17I and Bsu36I to liberate an internal fragment of the goose associated adhesin gene. The resulting linearized plasmid was religated with a PCR-amplified kanamycin resistance cassette from pACYC177 (136) in the opposite orientation to create plasmid pMJH002 (Appendix 2). Plasmids pMJH001 and pMJH002 were transformed in the *E. coli* strain JM109 by

electroporation (141). A vector only control strain was also created by transforming JM109 cells with pGEM-5Z(+) (Promega, Madison, WI).

SDS-PAGE Electrophoresis

E. coli JM109 strains carrying pMJH001, pMJH002, pGEM-5Z(+) were grown overnight in 50 ml LB medium containing 100 μg/ml ampicillin. The control JM109 strain, not containing a plasmid, was also grown in LB medium. Cells were lysed by sonication and the insoluble protein fraction was collected by centrifugation. The insoluble protein fraction was washed several times in 10 ml of phosphate buffer (20 mM sodium phosphate, 0.5 NaCl, pH 7.4) and resuspended in 1 ml of running buffer (1% SDS, 50 mM Tris, pH 7.4). Aliqouts (5 or 10 μl) were run on a 10% polyacrylamide gels at 90 V for 2 hours. Gels were stained with Coomassie Blue for 1 hour and destained with dH_2O for 10 minutes before visualizing the gels.

Protein Sequencing

The 98 kD band found in the insoluble protein fraction in strains carrying pMJH001 was excised from the gel, purified, and digested with trypsin. Peptides were sequenced using mass spectrometry on an LCQ-1 mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as previously described (153). Peptide sequencing work was performed at the Center for Mass Spectrometry and Proteomics, University of Minnesota, St. Paul, MN.

Clumping Assay

The ability of strains carrying the pMJH001 plasmid to clump together was analyzed by visual inspection of cultures and by a spectrophotometric assay performed in triplicate, as previously described (177).

Nucleotide sequence accession number

The sequences obtained in this study were deposited in GenBank under accession numbers: DQ300500 to DQ300502, and DQ300504 to DQ300507.

RESULTS

Isolation of goose *E. coli* specific DNA fragments

Following SSH, 192 putative goose-specific DNA clones were randomly selected to create a DNA subtraction library. The cloned insert DNAs were initially screened using a dot-blot protocol to determine hybridization specificity. Twenty clones demonstrated increased hybridization intensity when probed with labeled *RsaI*-digested, genomic DNAs from the five pooled *E. coli* strains from geese, relative to that seen with the pooled *E. coli* genomic DNAs from humans. The hybridization specificity of DNA inserts from these clones was further evaluated by Southern hybridization using the same probes as those used in the dot blot hybridizations. Southern hybridization analyses indicated that 17 cloned insert DNAs were goose-specific. Southern hybridization analyses of a representative group of eight goose-specific insert DNAs are shown in Figure 2.1.

Analysis of insert specificity

While Southern hybridization analyses confirmed that several of the cloned DNA fragments hybridized specifically to goose genomic DNAs, this specificity analysis was limited only to probes derived from the goose and human *E. coli* strains used in the initial SSH procedure. To examine hybridization specificity of the clones in more detail, colony hybridization experiments were done to identify cloned insert DNAs that hybridized with many *E. coli* strains from geese, and only a few strains from humans. A library consisting of 135 and 210 unique *E. coli* isolates from geese and humans, respectively, was cultured on nylon membranes, and individually probed with 14 of the ^{32}P-labeled PCR amplified insert DNAs from the confirmed goose-specific clones. The remaining three other cloned insert DNAs were not further evaluated since they were duplicates of existing clones. A representative image of a colony hybridization membrane is shown in Figure 2.2. DNAs from the five goose and human *E. coli* strains that were used in SSH were used as references for determining positive and negative hybridization signals, respectively, and quantitative image analysis was performed to determine the pixel intensities of the individual colony spots (Figure 2.3). The cut-off value was determined as the mean intensity of the five human strains plus three times the standard deviation. Based on these analyses, 7 of 14 (50%) goose-specific DNA inserts (GA9, GB2, GD5, GE3, GE11, GF5, and GG11) demonstrated specific hybridization with goose *E. coli* strains relative to that seen with strains isolated from humans (Table 2.1). The insert DNAs hybridized to 20.7 to 48.1% of the 135 unique goose strains tested. In contrast, the tested insert DNAs cross-hybridized with 1 to 10% of the 210 *E. coli* strains from humans. Insert DNAs GB2 and GE11 hybridized to the greatest number of goose isolates, 48.1%. Together the seven

probes hybridized with about 76% of *E. coli* strains from geese and cross-hybridized, on average, with 5% of the human *E. coli* strains. These hybridization experiments were repeated twice, in triplicate, to verify results.

Host specificity determination

Since the identified goose-specific marker DNAs will ultimately be used to examine *E. coli* in natural habitats, it is important to determine whether the probes cross-hybridize with *E. coli* from other host animal species. To examine this, I hybridized each ^{32}P-labeled insert DNA probes to 891 unique *E. coli* strains isolated from cats, chickens, cows, deer, ducks, goats, horses, humans, pigs, sheep, and turkeys. Results, summarized in Figure 2.4, show that the probes hybridized to 76% of the geese isolates examined. Similarly, the probes cross-hybridized to 73% of the duck isolates. In contrast, the probes only cross-hybridized with a limited number *E. coli* isolates from other host species, with the greatest cross-hybridization occurring with *E. coli* from turkeys (14.6%) and chickens (12.5%). These results indicated that the greatest degree of cross-hybridization occurred with *E. coli* isolated from avian hosts. The mean, false-positive, cross-hybridization frequency of the probes to *E. coli* from other host sources was about 9%.

Hybridization specificity was also evaluated by using a blind sample consisting of 96 isolates, comprised of 19 goose, 20 horse, 20 pig, 20 sheep, and 17 human *E. coli* strains. The seven probes evaluated (GA9, GB2, GD5, GE3, GE11, GF5, and GG11) hybridized with 14 of 19 goose strains (73.7%), and only 6 of 77 (7.8%) of the strains from other animals (data not shown).

Environmental *E. coli* and geographic analyses

To examine the correlation between results obtained using the new markers described in this study and other methods, I isolated about 200 *E. coli* strains from Duluth harbor water and analyzed them firstly by using the HFERP DNA fingerprinting technique, and subsequently by hybridization using combined ^{32}P-labeled insert DNAs GB2 and GE11. Of the 200 *E. coli* isolates examined, 27 strains (13.5%) were identified as likely originating from geese using the HFERP DNA fingerprinting technique, a comprehensive known source DNA fingerprint library, and ID bootstrap analysis with a P ≥ 0.9 (27). When screened by colony hybridization to a pooled GB2/GE11 insert DNA probe, 22 of 27 strains hybridized with the probes. This corresponded to an 81.5% agreement between HFERP classification and marker probe analysis using the GB2/GE11 screening system presented here. The applicability of DNA marker technology was also demonstrated by screening randomly-selected, environmental, *E. coli* isolates from Lake Phalen, a local, urban, lake frequently impacted by Canada geese. Out of the 1,152 isolates examined, 301 isolates (26.1%) tested positive with the markers (GB2/GE11).

To determine if the DNA markers identified *E. coli* from geese obtained from other geographic regions of the U.S., I hybridized probes GB2 and GE11 to an additional 359 goose isolates obtained from Delaware, Indiana, Wisconsin, and West Virginia. Results of this experiment demonstrated that only 24% of the isolates hybridized to the marker DNAs (data not shown). Probes GB2 and GE11 hybridized to 20, 28, 38, and 20% of the goose *E. coli* strains from Delaware, Indiana, Wisconsin, and West Virginia, respectively.

Nucleotide Sequencing and BLAST searches

The seven confirmed goose specific DNA inserts were sequenced, in both directions, and translated sequences were subjected to BLASTX analyses using *E. coli* protein databases at NCBI. The sequenced inserts were 332 to 885 bp in length. Results of BLASTX homology searches are summarized in Table 2.2. The GB2 and GE11 inserts, that each hybridized to about 48% of the *E. coli* strains from geese, were 93% identical to each another at the nucleotide level, and when translated had significant amino acid homology, 65% and 66% amino acid identity, respectively, to the C-terminal fragment of the AIDA-I adhesin-like protein of *E. coli* O157:H7 (GenBank accession BAB33785). The GD5 insert showed 89% amino acid identity to a fragment of the TraT complement resistance protein of *E. coli* (accession AAT85681), and insert GF5 was 98% identical to ORF5 in *E. coli*, with no significant matches to any entries in the database. Other matches with less than 50% amino acid identity to proteins in the database included two Type III secretion machinery proteins from *E. coli* O157:H7 (accessions AAG57987 and BAB37142), and a NikB nickase (accession NP_052661).

Cloning the full length goose associated adhesin gene

Sequence analysis of several gene library clones carrying the GB2 marker showed the marker gene was part of a 2,685 bp open reading frame which encodes a hypothetical protein of 98 kD. A conserved domain search showed that the middle region contained an pertactin-like (PL2) passenger domain, part of the autotransporter superfamily of proteins (69). The C-terminal region contains beta barrel domains also consistent with

the autotransporter superfamily (69, 123). The N-terminal region did not have any conserved domains (Figure 2.5).

SDS-PAGE and Protein Sequence analysis

SDS-PAGE analysis of the insoluble protein fraction from *E. coli* strains containing pMJH001 showed a prominent 98 kD band that was not seen in strains carrying pMJH002 or the vector only control (Figure 2.6). Protein sequencing results confirmed the 98 kD protein corresponds to the goose associated adhesin (Figure 2.7).

Autoaggregation Assays

Strains carrying the pMJH001 plasmid autoaggregated and precipitated at the bottom of the culture tube after incubation (Figure 2.8). The spectrophotometric assay confirmed these results and showed the supernatant from strains carrying the pMJH001 plasmid absorbed less light than strains carrying pMJH002 or a vector only control. The optical density (600 nm) of the culture after settling for 2.5 hours was 0.398 ($\sigma = 0.04$) for pMJH001 compared to 1.22 ($\sigma = 0.05$) for pMJH002. The optical density for the vector only control was 1.29 ($\sigma = 0.04$).

DISCUSSION

In this study, SSH was successfully used to identify seven DNA markers with high levels of hybridization specificity for *E. coli* strains originating from geese. Combined, the marker DNAs were capable of identifying about 76% of the goose and

73% of the duck *E. coli* strains examined. In contrast, the probes on average cross-reacted with about 10% of the *E. coli* isolates from other host species. As my goal was to identify sequences specific to goose strains, I adapted the standard SSH protocol by using pooled genomic DNAs from multiple goose strains as the tester, and five human isolates as the driver. By using pooled genomic DNAs, rather than DNA from a single strain, it was expected that more genetic diversity among the goose *E. coli* isolates could be uncovered, and that the subtraction products obtained would more likely be present in other goose isolates than in *E. coli* strains from humans. Thus, the method employed was expected to enrich for sequences found in all of the pooled tester genomes, rather than fragments present in only a single genome. This hypothesis was shown to be true by finding highly similar, but not identical, DNA sequences in inserts GB2 and GE11. An additional clone with 100% identity to GE11 was also identified using this approach, but it was not used in further analyses (data not shown).

One downside of using multiple tester DNAs is reduced subtraction efficiency, due to the increased complexity introduced into the reaction. Generally, genome subtractions yield greater than 25% tester-specific sequences after screening (CLONTECH, Mountain View, CA), compared to the approximately 9% efficiency that was observed in this study. However, reduced efficiency was not found to be an issue using the screening procedures I employed, and for my purposes, increased hybridization specificity and the ability to identify more isolates are the most important parameters. Seven goose-specific insert DNAs demonstrated increased hybridization with strains isolated from geese relative to isolates obtained from humans. While these insert DNAs each hybridized with less than half of the goose isolates tested, showing strain genetic

diversity in goose *E. coli*, combined the inserts identified 76 and 73% of *E. coli* from goose and ducks, respectively. Consequently, subsequent field studies will need to be done using pooled insert DNAs as hybridization probes.

When translated, the nearly identical insert DNAs GB2 and GE11 shared 65% amino acid identity, to the C-terminal portion of the AIDA-I adhesin-like protein of *E. coli* strain O157:H7. This result suggests that inserts GB2 and GE11 are fragments of an unidentified adhesin-like gene. This notion is further supported by the sequencing results obtained from a gene library clone carrying the GB2 marker, which showed the full length gene sequence shared considerable homology to the AIDA-I adhesin-like protein and other autotransporter adhesins. Laboratory strains carrying a high copy number plasmid with the full length gene were also shown to auto-aggregate, a trait shown to be mediated by other adhesins, including those from the autotransporter superfamily (11, 90, 96, 128, 138). As adhesins mediate the attachment of bacteria to host tissues (167), it seems plausible that this putative gene may mediate the attachment of *E. coli* to the goose intestinal tract. Attachment to the host intestinal epithelium is the necessary first step in gut colonization (168) and therefore this gene may convey preferential colonization of the goose host. It is also possible, that since this gene conveys an ability to auto-aggregate, that it may also play a role in colony or biofilm formation in the host (53, 138). If validated by experimental *in vivo* colonization data, other adhesin genes that participate in host-specific colonization may also represent ecologically meaningful markers that can be targeted for microbial source tracking purposes.

Since the combined seven DNA inserts hybridized with 76% of goose isolates, I examined whether the probes cross-hybridized to isolates from cats, chickens, cows, deer,

ducks, goats, horses, humans, pigs, sheep, and turkeys. Interestingly, the seven probes cross-hybridized with 73% of the *E. coli* isolates from ducks, and with 14.6 and 12.5% of isolates from turkeys and chickens, respectively, but only about 10% of the *E. coli* strains from other hosts. However, results from preliminary studies indicated that the GB2 and GE11 probes cross-hybridized to 11 and 8.8% of gull and tern isolates, respectively. Presumably, these results are due to the close genetic relationship between chickens, ducks, geese, and turkeys, and may indicate that the intestinal tracts of some avian species can be colonized by the same *E. coli* strain. Alternately, it may reflect the cosmopolitan nature of some *E. coli* strains (178), a transient intestinal population structure (66), a lack of host-specificity in this subgroup of *E. coli*, or the presence of multiple adhesins mediating colonization (167).

In a recent study by Soule et al., several *Enterococcus* DNA markers were identified and used to develop host-specific PCR primers (158). Although many of the identified markers were not detected in *Enterococcus* isolates from non-targeted host species, they often failed to detect greater than 10% of targeted isolates. Other host-specific markers identified by Soule et al. detected between 27 and 45% of isolates from targeted host species, but also detected between 1.6 and 7.1% of non-targeted isolates (42), which was similar to the results of the individual marker probes presented in my study (Table 1 and data not shown). Several other studies also identified host-specific markers for PCR-primer design (13, 38, 39, 100, 148-150). Most of the identified markers in these studies were detected in 100% of the samples from targeted host species and 0% of samples from non-targeted hosts, suggesting the markers are more specific than those presented in my study (13, 39). However, diluted fecal samples rather than

individual isolates were used for their analyses, preventing any direct comparison with the method described here.

Results obtained from screening water isolates from Lake Superior with the combined GB2/GE11 probe compared favorably with results obtained using the HFERP DNA fingerprinting method of assigning isolates to host source groups. Of the 27 isolates assigned to goose sources by HFERP, 22 (81.5%) demonstrated a positive hybridization signal with the GB2/GE11 probe. While the library-dependent HFERP method was previously shown to correctly identify about 70% of waterfowl isolates in a known source library (88), and far fewer environmental isolates, the method described here is vast improvement in accurately and quantitatively determining the host origin of environmental isolates. Moreover, the library-independent hybridization-based marker method suffers from fewer false positive and negative reactions than did HFERP and other techniques that have been recently evaluated, except for a host-specific PCR analysis (56). The applicability of this DNA marker technology was also evaluated by screening *E. coli* isolates from Lake Phalen, a local urban lake frequently impacted by Canada geese. Results of this analysis indicated that 26% of the 1152 isolates examined hybridized to the GB2/GE11 probes. These data further illustrates that the identified DNA markers can be used for environmental isolates. Considerably greater numbers of environmental isolates will most likely be found if hybridizations are done using the seven combined markers. Large-scale field studies using the combined seven probes will be done in the summer of 2006 to assess the impact of geese on Lake Superior beaches.

To assess whether the DNA markers allowed detection of geese *E. coli* from different geographic regions, I obtained isolates from the East and Midwest U.S. Results

of my studies indicated that the combined GB2 and GE 11 probes identified only 24% of the isolates examined. While the percentage identification most likely will increase if all seven marker probes are used, my results suggest that waterfowl *E. coli* strains are geographically distributed. Since the library I used was made from goose *E. coli* strains isolated in Minnesota, it is not surprising that the greatest percentage of strains identified were isolated in Wisconsin. Consequently, future efforts using SSH to generate DNA markers specific for animal hosts should be done using tester strains originating from several regions of the U.S.

The development of microbial source tracking techniques has in the past focused on library-dependent methods (147, 154). However, these methods suffer from the need to develop and maintain large reference libraries for comparison to environmental isolates. Additionally, geographic and temporal variability in isolates, transportability issues, the inability to assign many environmental isolates to source groups, large library sizes needed to adequately capture genetic diversity, and high false positive and negative assignments, make these methods difficult to implement at a large and economically-feasible scale (56, 88). In contrast, library-independent methods that screen for host-specific and ecologically-meaningful genes alleviate many of these issues. These genes most likely would not be influenced by geographic and temporal variability, as they would exist stably in bacterial isolates recently obtained from a specific host source. While library-independent marker gene approaches have recently been investigated as source-tracking tools with members of the genus *Bifidobacterium* and the *Bacteroides-Prevotella* group (12, 13, 38), these organisms are rarely quantified in routine analyses of fecal bacteria in waterways. Conversely, *E. coli* is becoming one of the most frequently

monitored indicators of fecal contamination of freshwater systems, and thus, source-tracking information obtained using the markers reported here can be easily coupled with existing and new fecal count data for TMDL analyses and abatement strategies. A library-independent marker gene method has also been recently developed for *Enterococcus* species (145), allowing similar analyses for saltwater environments.

Since waterways are most often contaminated by fecal bacteria originating from several different sources, rather than a single animal host species, it is frequently necessary to screen large numbers of isolates for accurate determination of host sources (111). The development of host-specific DNA fragments for screening by colony hybridization represents a cost-effective quantitative method for the simultaneous analysis of many bacterial isolates. Moreover, these methods can be easily adapted for automated, rapid, and high throughput macro- and microarray screening strategies, reducing the time and expense of analyzing thousands of isolates needed for large-scale and accurate source tracking studies.

In summary, my results provide evidence that SSH is an effective tool for the identification of ecologically meaningful marker DNAs that have the ability to identify a large number of genetically diverse *E. coli* isolates originating from geese. While my initial studies indicate that these markers can be effectively used as hybridization probes to determine the source of environmental *E coli* isolates, more extensive field testing is needed to before large scale microbial source tracking studies can be initiated. Nevertheless, I believe that the SSH approach will allow us to identify additional markers for *E. coli* strains from humans and other animals, allowing us to obtain a more comprehensive analysis of sources of fecal contamination in waterways. Coupled with

high throughput, automated, macro- and microarray screening, these markers may provide a cost-effective, quantitative, and accurate method to determine sources of genetically diverse *E. coli* for use in water quality analyses and TMDL determinations.

ACKNOWLEDGEMENTS

This work was supported, in part, by a training grant (2T32-GM008347) from the National Institutes of Health.

I would like to thank Satoshi Ishii, Sam Myoda, Cindy Nakatsu, Don Stoeckel, and Greg Kleinheinz for providing *E. coli* isolates and John Ferguson for help with the blind studies, cluster analyses, and library maintenance. I would also like to thank Todd Markowski and Bruce Witthun for their help with peptide sequencing and Janis Frias for her assistance with SDS-PAGE.

This work was previously published in Applied and Environmental Microbiology (Appl. Environ. Microbiol. 2007. 73:890-896). Permission to reprint this manuscript was obtained from the publisher, American Society for Microbiology, under license number 2216580925616.

.

Table 2.1

Goose-specific marker DNAs isolated using suppression subtractive hybridization.

Marker DNA	Percent of *E. coli* isolates from **geese** hybridizing to marker DNA Probes (134)[1]	Percent of *E. coli* isolates from **humans** hybridizing to marker DNA Probes (209)
GA9	27.4	1.9
GB2	48.1	3.3
GD5	30.4	9.1
GE3	23.6	7.2
GE11	48.1	4.8
GF5	20.7	10.0
GG11	31.1	1.0

[1]Value in parentheses refer to the total number of strains examined by colony hybridization.

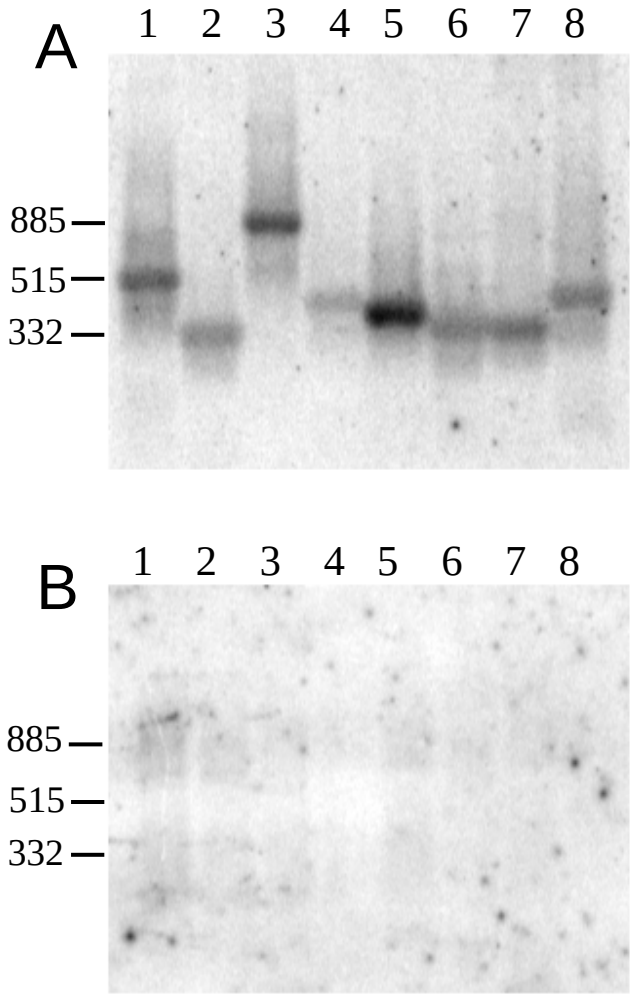
Table 2.2

Properties of insert marker DNAs showing hybridization specificity to *E. coli* isolated from geese.

Insert DNA	Length (bp)	Protein homolog in database	GenBank accession	AA Identity[1]	E value
GA9	515	Type III secretion apparatus protein (*E. coli* O157:H7 EDL933)	AAG57987	61/161 (37%)	1.00E-26
GB2	332	AIDA-I adhesin-like protein (*E. coli* O157:H7 RIMD 0509952)	BAB33785	81/123 (65%)	2.20E-40
GD5	885	TraT (*E. coli* plasmid R1)	AAT85681	112/132 (85%)	2.00E-57
GE3	380	NikB (*E. coli* O157:H7 RIMD 0509952)	NP_052661	30/88 (34%)	2.00E-05
GE11	336	AIDA-I adhesin-like protein (*E. coli* O157:H7 RIMD 0509952)	BAB33785	81/123 (66%)	1.00E-40
GF5	346	ORF5: no significant homology proteins in database (*E. coli* B171)	AAB36834	57/58 (98%)	2.00E-27
GG11	427	Type III secretion protein *EprH* (*E. coli* O157:H7 RIMD 0509952)	BAB37142	31/101 (32%)	1.00E-11

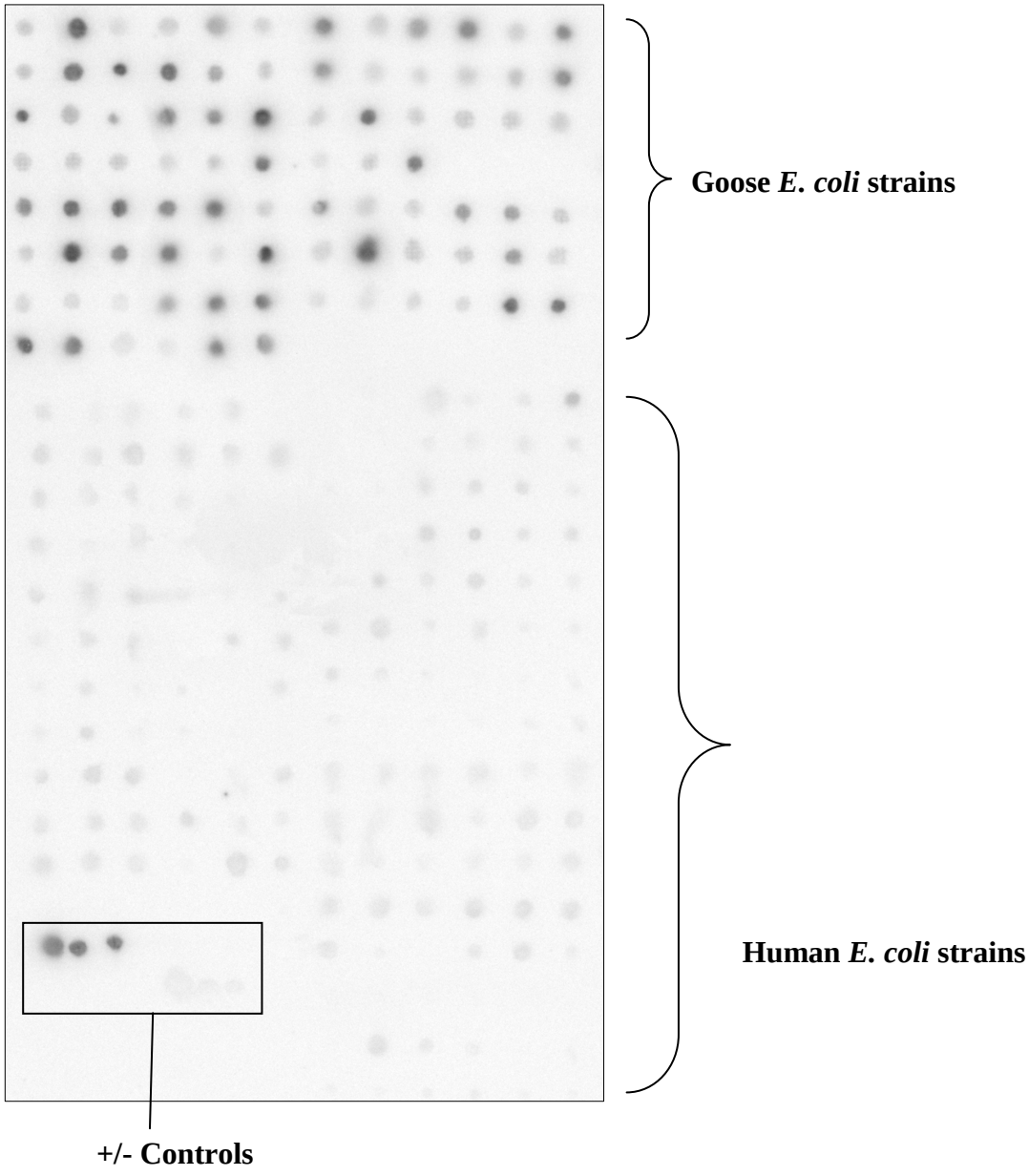
[1] Values refer to number of amino acids with identity per total number examined. Values in parentheses refer to percent identity with database entry.

Figure 2.1



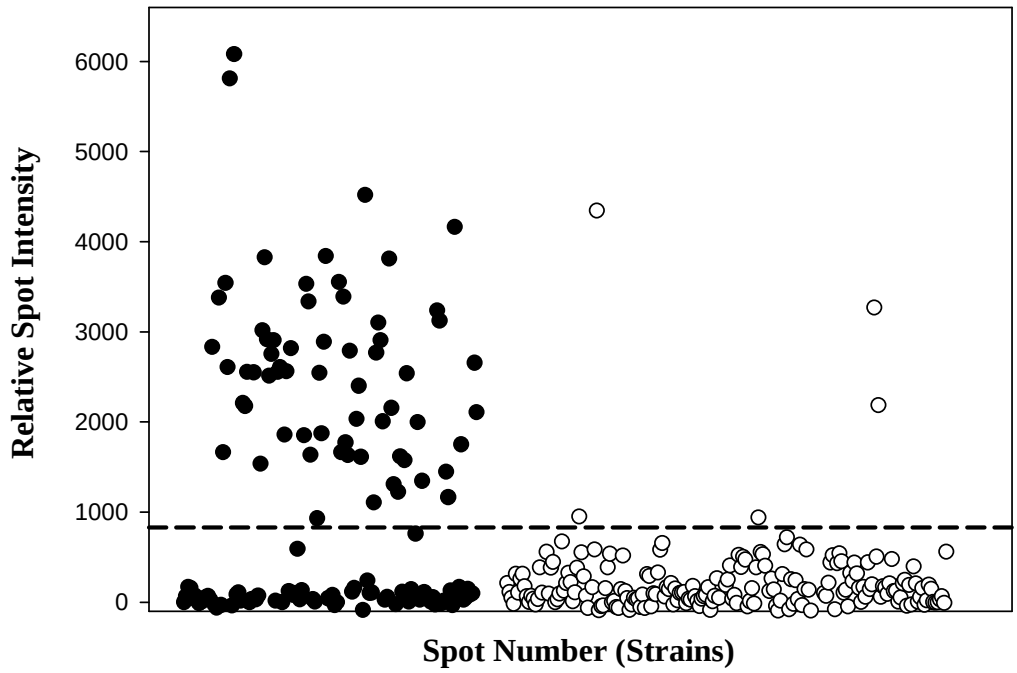
Southern hybridization of eight SSH-derived, PCR-amplified, insert DNAs to ^{32}P-labeled *RsaI*-digested pooled genomic DNAs from *E. coli* isolates obtained from geese (A) and humans (B). Panels A and B show duplicate membranes probed with the respective genomic DNAs.

Figure 2.2



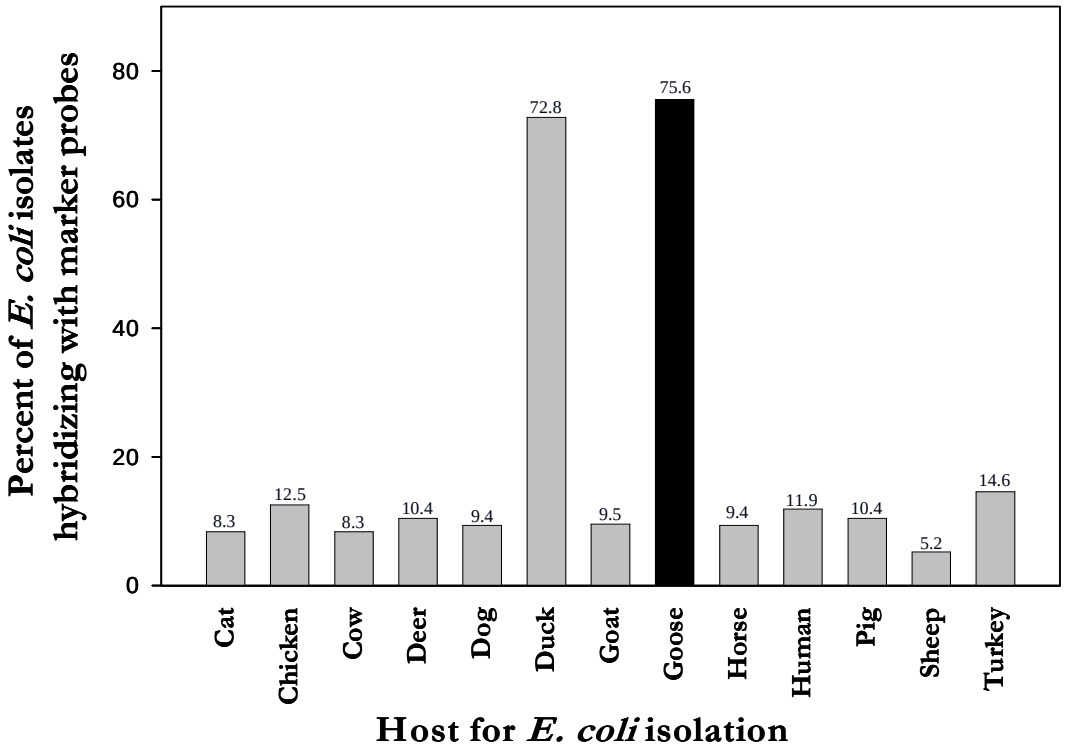
Colony hybridization of ^{32}P-labeled GE11 insert DNA to 134 and 209 unique *E. coli* isolates obtained from geese and humans, respectively. Positive and negative control strains are boxed.

Figure 2.3



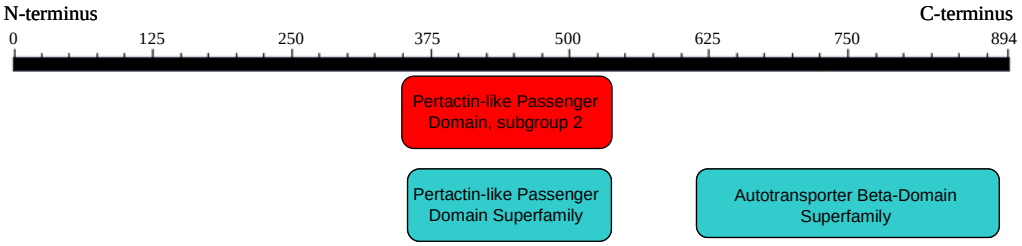
Pixel intensities from colony hybridization membrane containing 134 and 209 unique *E. coli* strains isolated from geese (●) and humans (○), respectively. The membrane was probed with ^{32}P-labeled DNA from marker insert GE11. A cut-off value for determining positive and negative signals is indicated (---).

Figure 2.4



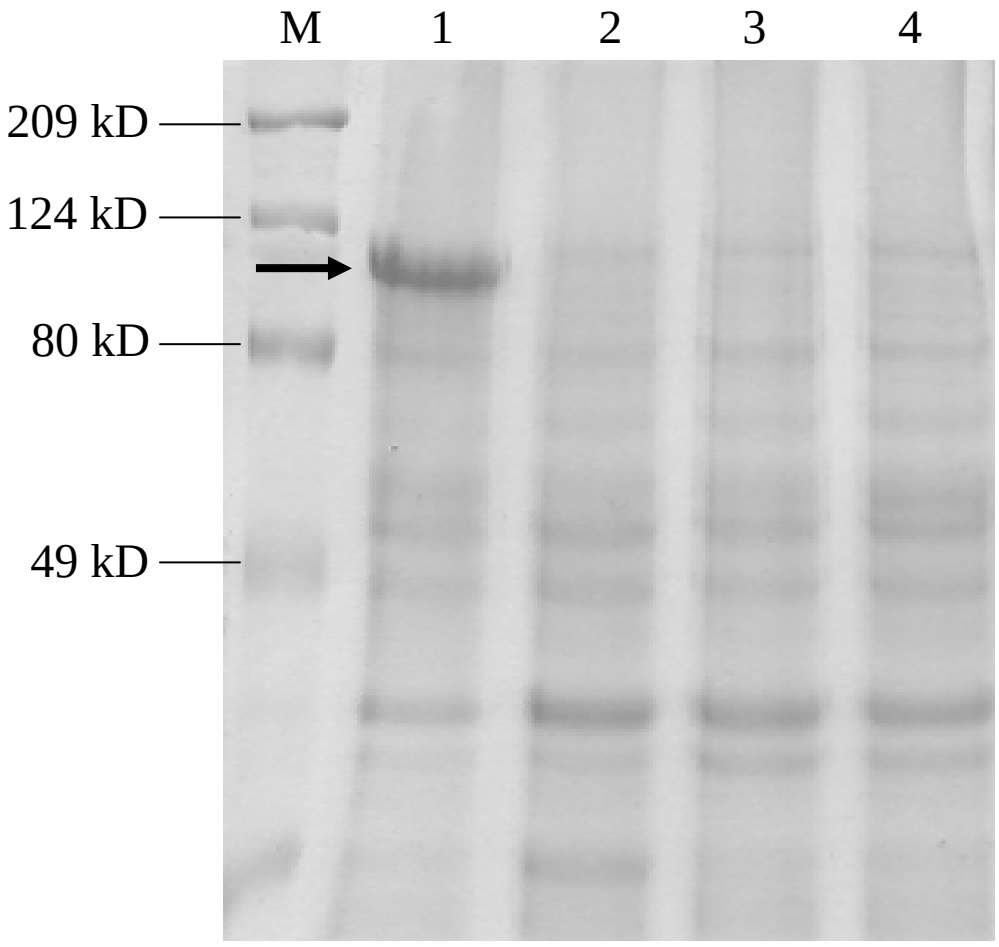
Percentages of *E. coli* strains hybridizing to ^{32}P-labeled, pooled, insert GB2/GE11 marker DNAs obtained by colony hybridization and pixel intensity analysis. Values above each bar are hybridization percentages.

Figure 2.5



Conserved protein domain search of the amino acid sequence of the putative goose associated adhesin using the GenBank database. Specific hits are shown in red, superfamily hits are shown in blue.

Figure 2.6



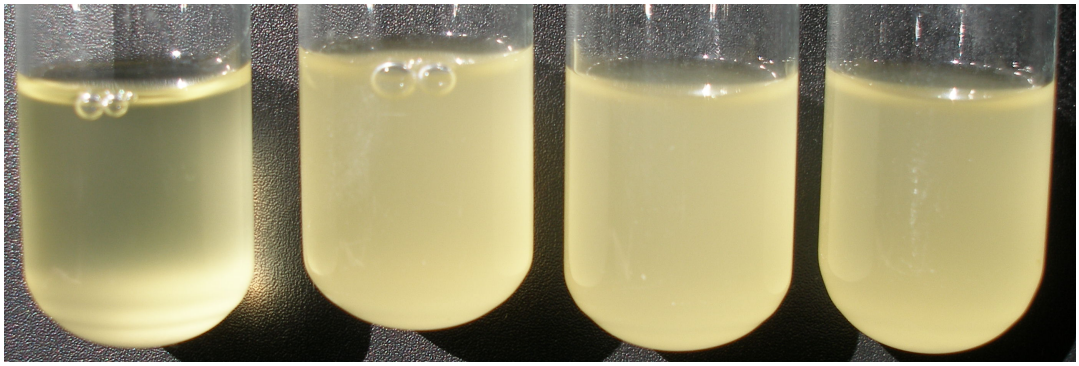
SDS-PAGE gel of the insoluble protein fraction from JM109 *E. coli* strains carrying (1) pMJH001 (full length goose associated adhesin gene), (2) pMJH002 (mutant), (3) a vector only control, and (4) no plasmid. A molecular weight marker is shown lane M and the 98 kD protein corresponding to the goose associated adhesin is denoted by the arrow.

Figure 2.7

MKHHNRQIMKFSALYSVMILSGMSFPFCSIAAGGMNTTQYTFNSEYNLTDGDTIK[FTAQPNVSTAK][L
SISGSGILNIDGK][NSLLTFDNSGVK][GQGDMLIAGK][LNITNGGGFNAFGTEPR]MQVGVGSEGSV
IVSGTESYLTVPYIVVGNTAK[GNMSISDGAAVTLNALK][VGYK[ADGDVLVSGK]][GTSLTVSSDSD
TIR][LGFYDANGR]LTVNNNATVTAPYIVVGNDFNNHSGELIIGSGVSEPASEAGIINANEIKFRNKGI
LTLNHTNNDFNLTSDLSSSGESITTGVFSGDYGLVQAVAGTTILSGNNEK[YNGSINIENGAGIVVSEQK
]NLGTSVVTDNGLLTIDTTTDWQLTNDVSGGGNFRKTGSGSLTVGNNAAWTGQTDIDAGTLILGNAGAPV
MLASSQVNIAKDGILTGFGGVSGNVTNSGTLDLRADAPGNVLTVGGNYTGNNGTLLMNTVLGDDSSATDK
LVIKGDASGQTRVEVTKAGGTGAQTLNGIELIHVEGNADSAEFVQAGR[IAAGAYDYTLGR][GQGSNSG
NWYLTSGK]NTPEPTPTPDPDSKPEPAPGGYDNDLRPEAGSYTANMAAVNTMFVTRLHER[LGPMQYTDI
MTGETK][NTSMWMR]HEGGHNR[WRDGTGQLK]TQGNRYVVQLGGDIAQWGWGETDRWHLGVMAGYGNE
HNNTDSVRTGYRSKGSVNGYSTGLYATWFASDETHNGAYLDTWAQYGWFDNHVK[GDGLPGESWK][SK[
GLTASLETGYTWK]]AGEFSGSHGSLNEWYVQPQAQVVWMGVKADEHR[ESNGTRVENTGDGNVR]TRLG
VKTWIKGHNRMDDGKSREFRPFVEVNWLHNTREFGTR[MNGVTVHQDGAR]NIGEVK[AGVEGQINDR][
LNLWGNVGVQAGDK][GYSDTSAMLGVK]YTF

Results of peptide sequencing by mass spectroscopy. Twenty four unique peptides were identified, representing 284/894 amino acids (32% coverage), and are shown in brackets. A total of 143 spectra were analyzed and peptides shown below were highlighted to represent the number of spectra associated with each peptide. Green highlighted peptides were detected in >10 spectra. Yellow peptides were detected between 5 and 10 times, and blue represents those peptides detected in < 5 spectra. Based on the peptide sequences, the probability that the goose associated adhesion gene was detected was 100%.

Figure 2.8



1 **2** **3** **4**

Qualitative auto-aggregation assay with *E. coli* JM109 strains carrying (1) pMJH001 (full length goose associated adhesin gene), (2) pMJH002 (mutant), (3) vector only control, and (4) no plasmid.

CHAPTER 3

High Throughput and Quantitative Procedure for Determining Sources of *Escherichia coli* in Waterways using Host-specific DNA Marker Genes

OVERVIEW

Escherichia coli is currently used as an indicator of fecal pollution and to assess water quality. While several genotypic techniques have been used to determine potential sources of fecal bacteria impacting waterways and beaches, they do not allow for the rapid analysis of a large number of samples in a relatively short period of time. In this chapter, I report that the DNA markers identified in Chapter 2 were useful for the development of a high-throughput and quantitative macroarray hybridization system to determine numbers of *E. coli* bacteria originating from geese and ducks. The procedure developed uses a QBot robot system for picking and arraying colonies and allows for simultaneous analysis of up to 20,736 *E. coli* colonies from water samples, with minimal time and human input. Statistically significant results were obtained by analyzing 700 *E. coli* colonies per water sample, allowing for the analysis of approximately 30 sites per macroarray. Macroarray hybridization studies done on *E. coli* collected from water samples obtained from two urban Minnesota lakes and one rural South Carolina lake indicated that geese/ducks contributed up to 51% of the fecal bacteria in the urban lake water samples, and the level was below the detection limit in the rural lake water sample. This technique, coupled with the use of other host source-specific gene probes, holds great promise as a new quantitative microbial source tracking tool to rapidly determine the origins of *E. coli* in waterways and on beaches.

Note: This work was done in collaboration with Dr. Tao Yan.

INTRODUCTION

The contamination of waterways with human feces represents a significant risk to public health due to the possible presence of human enteric pathogens (46, 67). The frequent occurrence of fecal bacteria in waterways and recent changes in government regulations have prompted increased interest in developing methods to determine sources of fecal bacteria. For large watersheds, numerous potential fecal sources are present, and contamination may be due to feedlot runoff, manure-amended agricultural fields, wildlife, leaking septic systems, and sewage discharge (39, 46, 47). Identifying fecal pollution sources and apportioning their contribution to the total fecal load provide essential information for implementing cost-effective remediation strategies and for establishing total maximum daily loads.

Over the past decade, numerous methodologies have been developed for microbial source tracking (MST) (12, 13, 42, 57, 102, 125, 126, 134, 145, 148-150, 158, 182, 183). The underlying assumption of all MST techniques is that host-specific phenotypic or genotypic differences in microbial populations originating with different animal and human sources exist, and that their detection in environmental samples can be used to determine the host origin of fecal microorganisms. These methods generally target specific populations or lineages (genotypes or ecotypes) of the fecal bacteria *Escherichia coli* (182), *Enterococcus* sp. strains (145, 182), *Bifidobacterium* (134), or members of the *Bacteroides-Prevotella* group (13).

Many of the commonly used MST methodologies require the construction of libraries of known-source fecal bacteria (42, 125, 182). However, given the high degree of genetic diversity among fecal indicator bacteria, reference libraries need to be very

large in order to allow adequate determination of potential sources of fecal bacteria in environmental samples (88). In addition, several studies have shown that most known-source libraries lack representativeness, mainly due to the presence of transient (temporary) inhabitants in the gastrointestinal tracts of different host sources, multiple strains within a single animal, and temporal and geographic variation in bacterial isolates (genotypes) within and between animal species (61, 88, 174, 183). Together, these problems and limitations account for low average rates of correct classification among library entries and for the inability to correctly identify the majority of environmental isolates (56, 113, 147, 154, 161, 183). Some of these limitations are compounded by the inability to adequately analyze a sufficient number of environmental isolates. For example, the rep-PCR DNA fingerprinting technique, which has been used extensively to examine sources of fecal bacteria impacting beaches and waterways (88), suffers from limitations in throughput, allowing for the analysis of only about 400 *E. coli* isolates per week. Other genotypic methods are more labor intensive and are therefore not good candidates for high-throughput methods.

Due to the limitations, shortcomings, and problems associated with the use of known-source libraries for MST, many investigators have evaluated the use of library-independent methods to determine sources of environmental fecal bacteria. These technologies, using culture-dependent and -independent approaches, avoid problems associated with limitations of library size and isolate diversity issues. To date, these methods have focused on the use of enteric viruses (47, 86, 87, 105) and the use of host-specific PCR-based markers for *Bifidobacterium* (134), *Bacteroides-Prevotella* ((12, 13,

38, 39, 100), *Bacteroidales* (149, 150), *Enterococcus faecium* (145), *Methanobrivibacter smithii* (170), and toxin genes possessed by *E. coli* (92, 93).

Although these approaches and methods hold promise for MST studies, they are not designed to obtain high-throughput data, are labor intensive, and costly for a large number of samples. Moreover, many methods suffer from general limitations imposed by PCR, especially those associated with the use of environmental samples (61). In addition, aside from a marker-based, library-independent method developed for analysis of *Enterococcus faecium* (145), the other systems thus far examined depend on detection of organisms other than the fecal indicator bacteria that are most frequently used by state and local agencies to assess water quality. In chapter 2, I reported on the identification of host source-specific genetic markers for *E. coli* strains originating with geese and ducks (61). These markers were shown, by using dot blot and colony hybridization analyses, to be useful in determining sources of fecal pollution in Lake Superior harbor, and I have speculated that such a hybridization-based marker system would be useful for high-throughput studies.

Robot-assisted high-throughput technologies have revolutionized biology and allow for large-scale genomic and biochemical analyses of micro- and macro-organisms. There is little doubt that these evolving technologies will have significant positive impacts on environmental and water quality studies, especially where high precision and accuracy are needed for the analysis of a large numbers of samples (83, 139).

This chapter discusses the development and evaluation of a high-throughput, semiautomated, quantitative procedure for determining sources of *E. coli* in waterways.

The method I developed uses a QBot robot for colony picking and arraying, as well as gene probes specific for *E. coli* originating from geese/ducks (61). The method allows for the simultaneous analysis of up to 20,736 *E. coli* colonies from water samples with minimal time and human input. Assay sensitivity and specificity were examined with artificially inoculated lake water, and field studies were done in two urban lakes and one rural lake. The results indicated that as few as 32 *E. coli* colonies originating from geese/ducks could be detected in a background of relatively high concentrations of other *E. coli* in water samples. Coupled with the use of other host source-specific gene probes, this technique may hold great promise as a new quantitative microbial source tracking tool to rapidly determine the origins of *E. coli* in surface waters.

MATERIALS AND METHODS

Escherichia coli reference strains and molecular markers

Reference strains used in this study were previously isolated from the feces of geese (Go66, Go90, Go126, Go172, and Go206) or humans (Hu51, Hu130, Hu132, Hu188, and Hu252) using selective and differential growth media (42, 61, 88). Genomic DNAs from these strains were previously used in suppression subtractive hybridizations to identify goose/duck-specific molecular marker probes. When combined, these probes identified 76 and 73% of the goose and duck isolates tested, respectively (61). Two of the marker gene probes, GB2 and GE11, each hybridized with about 50% of the 135 goose isolates tested and on average cross-hybridized with less than 4% of *E. coli* from humans and most other animal hosts.

Sample collection and concentration

Offshore lake water samples were collected using standard procedures as described previously (21). Water samples used in spiking experiments were collected from Lake Como (St. Paul, MN) in August 2005. Geese were not present at this lake at the time of sampling. The water samples used for field evaluation were collected in September 2005 from two urban lakes suspected of being contaminated by goose feces, St. Louis Bay in Lake Superior (Duluth, MN) and Lake Calhoun in Minneapolis, MN. In addition, Lake Hartwell (Clemson, SC), a relatively pristine lake reportedly free of fecal contamination from geese, was sampled in September 2006. Bacteria in water samples were concentrated by membrane filtration at 25°C using 0.45-µm (47-mm-diameter) Nuclepore polycarbonate membranes (Whatman, Florham Park, NJ); multiple membranes were successively used to facilitate filtration and circumvent membrane clogging. Filters were washed with 10 ml of phosphate-buffered saline (PBS) (20 mM sodium phosphate, 15 mM sodium chloride, pH 7.2), and bacterial cells were released by gentle agitation for 10 min using a sterile magnetic stir bar. The solution and membranes were transferred to sterile plastic culture tubes and vortexed for 10 min to further dislodge bacterial cells from membranes.

Enumeration of *E. coli*

The number of *E. coli* cells in water samples was determined by using the modified mTEC membrane filtration method (174), except that 5-bromo-4-chloro-3-indolyl-ß-D-glucuronic acid (X-Gluc) (500 µg/ml) was used as the chromogenic

indicator. Concentrated cell suspensions were diluted in sterile PBS prior to enumeration.

Spiked samples

Bacterial cells in approximately 20 liters of Lake Como water were concentrated by membrane filtration as described above and suspended in 20 ml of PBS. The suspensions contained about 3,000 CFU of *E. coli* per ml, as determined by using the membrane filtration method and modified mTEC agar medium. *E. coli* strain Go66 was grown overnight at 37°C in LB medium and centrifuged at 10,000 x *g* for 10 min at 4°C, and the pellet was washed and resuspended in PBS. Washed Go66 cells were added to the concentrated lake water suspension to obtain 0, 20, 100, 200, 400, or 800 spiked cells per ml. Triplicate samples were prepared for each treatment.

Field studies

The relative contributions of geese to fecal contamination in Lake Superior (St. Louis Bay, Duluth, MN) and Lake Calhoun (Minneapolis, MN) were determined by picking and arraying 1,000 *E. coli* isolates from each water sample onto membranes as described above. Ducks and geese are frequent inhabitants of both lakes. Samples (183) of near-shore water were concentrated to 5 ml as described above. Colony hybridizations (141) were done using the ^{32}P-labeled combined GB2 and GE11 probes and analyzed as described below. Fecal *E. coli* counts in water samples were determined, prior to concentration, by using the modified mTEC membrane filtration method as described above.

Automated isolation and picking of *E. coli* colonies

Approximately 2,000 *E. coli* cells from each environmental or spiked water sample were spread plated onto the surface of modified mTEC agar medium containing X-Gluc (500 µg/ml) in 22-by-22-cm Qtray plates (Genetix, Boston, MA). Plates were incubated at 35°C for 2 h, followed by incubation at 44.5°C for 22 h or until colonies reached 1 to 2 mm in diameter. Plates were stored at 4°C overnight to allow further development of blue pigment in colonies. This allowed differentiation of *E. coli* from other coliform and gram-negative bacteria.

The automated picking of blue *E. coli* colonies was done using a Genetix QBot robotic system (Boston, MA), running Qsoft XP version 6000.11 software. Colony-bearing agar plates were illuminated from below using incandescent lighting, and grayscale digital pictures were taken from the top. Translucent blue paper, which separates the hybridization membranes (Genetix, Boston, MA), was placed under Qtray plates to enhance contrast and facilitate picking of blue-colored *E. coli* colonies. The color of the blue contrast paper was found to be 107:151:167 (red:green:blue), as determined by using the RGB mode of Photoshop CS version 8.0 software (Adobe Systems, San Jose, CA). The translucent paper had 22.5% transmittance at 600 nm, as determined using a Beckman DU-70 spectrophotometer (Fullerton, CA). The parameters used by the QBot for colony identification included colony diameter, roundness, axis ratio, proximity, and overlaps, which were determined on a plate-by-plate basis. Mean pixel intensities of colonies were calculated using grayscale analysis; the upper-threshold value was set to 255, and the lower-threshold values ranged from 100 to 140. Selected blue colonies were picked and placed, in random order, in 384-well microplates containing mTEC

medium supplemented with 4.4% glycerol, but without X-Gluc. Microplates were incubated at 44.5°C for 12 h and stored at -70°C before use. The taxonomic identities of 384 randomly selected *E. coli* isolates were verified by growth and reaction on selective and differential laboratory media and by biochemical tests, as described previously (42).

Robotic arraying of isolates

Aliquots (0.5 µl) of *E. coli* isolates in each microplate well were spot inoculated onto positively charged, 22-by-22-cm Performa nylon membranes (Genetix, Boston, MA) using the QBot robotic platform. The membranes were wetted with sterile water, and excess water was subsequently removed using dry sterile Whatman 3MM filter paper to prevent colonies from running together. Membranes were divided into 6 sections, each section contained 384 subunits, and each subunit consisted of 9 individual spots. Using this gridding format, each membrane was arrayed with 20,736 *E. coli* colonies, although it is possible to array 36,864 colonies on the membrane using an alternate grid pattern. Replication and standardization was performed by spotting each isolate in three different sections of the membrane and placing one of the positive and negative reference *E. coli* strains described above in each subunit. Arrayed membranes were placed on the surfaces of mTEC agar plates (without X-Gluc), incubated in an inverted position at 44.5°C for ~ 8 to 10 h, and stored at 4°C overnight, or until used.

Colony hybridizations

Colony hybridizations were done to determine the reactivities of the *E. coli* isolates to the mixed goose/duck-specific DNA probes GB2 and GE11 as described

previously (61). Extra caution was taken during the cell lysis step to prevent mixing of the lysed colonies. Goose/duck-specific DNA probes were labeled with [^{32}P]dCTP using the Random Primer DNA labeling system (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Membranes were hybridized overnight at 68°C and washed under high stringency at 65°C in 0.1x SSC (1x SSC is 0.15 M NaCl and 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate.

Quantitative image analysis

Quantitative image analysis was done as described previously to determine positive and negative signals on colony hybridization membranes (61). Images were captured using a STORM 840 densitometer (Molecular Dynamics, Piscataway, NJ) and analyzed using ScanAlyze version 2.50 software (http://rana.lbl.gov/EisenSoftware.htm). The normalized intensity of each spot was calculated by subtracting the median background intensity from the mean intensity of each spot, and data were plotted using Sigma Plot version 8.0 software (Systat Software, Point Richmond, CA). A cutoff value was assigned based on normalized mean intensities of negative control spots plus three times the standard deviation.

Equations for estimating relative fecal contributions

Targeted goose- or duck-derived environmental *E. coli* can have various levels of reaction to host-specific DNA marker probes, and reactions of *E. coli* from other host animal sources can also vary (61). Therefore, for reliable host source determination and quantification, a variety of probes targeting different host source *E. coli* strains are

required. The relative contributions of fecal *E. coli* strains from different host sources [$C_{\text{host}(i)}$] can be estimated based on the following equations.

Equation 1

$$\sum_{i=1}^{n} C_{host(i)} = 1$$

$$P_{probe(1)} = C_{host(1)} S_1 + \sum_{i=2}^{n} C_{host(i)} K_{(probe(1), host(i))}$$

$$\Lambda$$

$$P_{probe(n)} = C_{host(n)} S_n + \sum_{i=1}^{n-1} C_{host(i)} K_{(probe(n), host(i))}$$

where $P_{probe(i)}$ is the percentage of environmental *E. coli* isolates hybridizing with the DNA probes specific for host *i*; $S_{(i)}$ is the sensitivity of probes toward *E. coli* from the target host; and $K_{[\text{probe}(i),\text{host}(i)]}$ are the rates of cross-hybridization with *E. coli* strains from other host sources. However, since only one set of goose/duck-specific probes was used in this study, simplifying assumptions were made to treat *E. coli* strains from other host sources as one entity, and thus, an average cross-hybridization rate was used. Equation 2 was derived and used to estimate the contribution of fecal *E. coli* from geese/ducks in field studies.

Equation 2

$$C_{goose/duck} = (P_{probe} - K_{(probe,animals)}) / (S_{goose/duck} - K_{(probe,animals)})$$

For the GB2/GE11 mixed probes, the $K_{(\text{probe,animals})}$ and $S_{\text{goose/duck}}$ values were previously determined to be 5.1 and 48.1%, respectively (61).

Statistical analysis

Statistical significance of data was determined by using the analysis of variance (ANOVA) and Tukey's honestly significant difference (HSD) routines, at an α value of 0.01, using R program version 2.0.1 (http://www.r-project.org/). Standard statistical analysis parameters were obtained by using Excel 2002 (Microsoft, Redmond, WA). Method linearities for different sample sizes were determined by plotting the number of probe-positive isolates against the number of *E. coli* cells spiked into water samples, and *r* values from linear regression were determined.

RESULTS AND DISCUSSION

Most commonly used MST methodologies suffer from limitations due to requirements for the use of known-source libraries and the inability to adequately analyze and identify sufficient numbers of environmental isolates. Moreover, many of the methods do not allow for high-throughput data acquisition and analysis, suffer from being labor intensive and cost prohibitive for large-scale studies, and do not allow for correlation to data obtained from fecal indicator bacteria that are most frequently used by state and local agencies to assess water quality in freshwater systems (61, 88, 147). Based on these shortcomings, I developed a high-throughput, library-independent, and semiautomated procedure to quantify goose/duck-derived *E. coli* contributing to the fecal contamination of beaches and waterways.

Robot-assisted isolation of environmental *E. coli*

The automated picking of colonies by using the QBot system drastically decreased the labor intensiveness associated with obtaining environmental *E. coli* isolates, relative to manual picking methods. Using this system, 20,736 colonies can be scanned, picked from *E. coli*-bearing agar plates, and inoculated into master microtiter plates in approximately 6.5 h. Only limited human intervention is required for this procedure, mainly for adjusting the robot scanning parameters (colony diameter, roundness, axis ratio, and proximity) to select target *E. coli* colonies. This, however, needs to be done only once for each initial batch of Q-tray plates in each series. While some of the instruments involved in this procedure (i.e., the colony-picking robot) are not present in most individual laboratories, many researchers have access to these resources in shared genomic facilities located across the United States.

The efficacy of *E. coli* isolation and picking was determined by the specificity of modified mTEC agar in differentiating *E. coli* colonies from other waterborne microbes, by uniformity in spread plating, and by cell growth characteristics. Modified mTEC agar medium has previously been shown to be effective in the isolation of *E. coli* from water, soil, and fecal samples (77, 88). Preferential *E. coli* colonies used in this study were 1 to 2 mm in diameter and dark blue in color. Smaller colonies were found to increase colony picking error due to a limited target size, as were larger colonies that had white edges, most likely due to depletion of X-Gluc in the area surrounding colonies. The parameters of roundness, axis ratio, and proximity specified reduced the risk of picking non-*E. coli* colonies; typical values used in this study were 0.9, 0.8, and 0.3 mm, respectively. In addition, the blue-colored background sheet used in our studies greatly

facilitated robotic picking of colonies. Without this blue background, there was insufficient contrast to allow accurate colony identification.

The accuracy of colony scanning and picking was tested by examining the taxonomic identities of 384 randomly selected isolates. Microbiological and biochemical assays indicated that 99.7, 100, 98.7, and 100% of the isolates were correctly identified as *E. coli* by their reactions on ChromaAgar, MacConkey agar, and the methyl red and indole tests, respectively. By combining all of these tests, only a limited number of false-positive or -negative isolates would be chosen using the robot picking system. The overall performance of the automated picking system was comparable to that of non-automated procedures, of which the false-positive rates for environmental water samples were reported as ≤1% (174).

Automated colony arraying

The application of robotic automation also allowed high-density colony arraying, owing to size uniformity of the gridding pins. This allowed for highly accurate spot positioning and subsequent uniform colony growth, due to small deviations in inocula applied to membranes. This also benefited subsequent hybridization and data analyses. The contact area of the individual arraying pin was 0.24 mm^2, and the area of each spot was 2.2 mm^2, allowing for the spotting of 20,736 colonies on a single 22-by-22-cm membrane. This large number of colonies allowed for proper replication of unknown isolates and inclusion of replicated positive and negative controls, even when a larger number of isolates were being processed. An example of typical hybridization reactions is shown in Figure 3.1. This is similar to macroarray analysis used to identify

genomic clones containing specific DNA sequences (132). The inclusion of multiple positive and negative control organisms and replications of unknowns on the membrane increase the statistical power of the results (101) and reduce the number of false-positive and -negative results that are common to many MST methods (135).

In the procedure developed here, both negative and positive marker strains were placed in each section of the membrane, and at least one marker (negative or positive) was placed in each subunit of a section. This allowed establishment of a local binary classification scheme (positively or negatively hybridizing colonies) on a section-by-section basis. Quantitative differences in pixel intensities between positively and negatively hybridizing colonies could be easily distinguished by image analysis (Figure 3.2). The effectiveness of this classification scheme using controls placed in each subunit of the array gave average rates of correct classification of 90 to 97% and of 83 to 92% for the six negative and positive control strains, respectively. The relatively high average-rate-of-correct-classification values reported here suggest that this method will be effective in correctly classifying environmental isolates.

Method detection limit

Spiked-sample experiments were done to determine the method detection limit. The *E. coli* strain Go66 was added to a concentrated environmental water sample to prepare six treatments of spiked samples, which were subsequently analyzed using the developed procedure. A total of 1,000 *E. coli* colonies were isolated from each sample in each treatment. The number of *E. coli* isolates that hybridized to the combined probes in each treatment is presented in Table 3.1. One-way analysis of variance indicated that

the number of *E. coli* isolates reacting to the probes in the spiked samples was significantly different ($P = 3.95 \times 10^{-6}$) than that for the control. Tukey's HSD analysis was performed to compare the spiked samples with the nonspiked control samples, and significant differences ($\alpha \leq 0.01$) were observed only for spiked samples receiving 400 or 800 Go66 *E. coli* cells. This analysis indicated that the minimum detection limit for the automated screening procedure was 108 spiked *E. coli* cells for each 1,000 isolates examined (Table 3.1). The method detection limit, however, was directly related to the existing background level of indigenous *E. coli* from geese/ducks in water samples, which was 21 CFU per 100 ml, and the background total number of *E. coli* cells in the water sample. Thus, the detection limit will most likely improve when using sample water containing less *E. coli* from these animals. For example, a smaller number of colonies would need to be screened in the more pristine Lake Hartwell water, since when sampled, it had a background of only 72 *E. coli* CFU/100 ml, and only 0.28% of these reacted with our probes.

Influence of sample size on method linearity

To determine the minimum number of colonies that needed to be examined in order to obtain statistically significant detection results, 1,152 *E. coli* isolates were randomly picked from spiked samples and hybridized with the goose/duck-specific probes using the developed screening procedure. Since *E. coli* isolates were picked in a random order for each spiked sample, the first 100, 300, 500, 700, and 1,000 isolates could be used individually to determine the effects of sample size on method linearity. For this analysis, *r* values were determined by linear regression of hybridizing isolates

versus the number of spiked Go66 *E. coli* cells (Table 3.2). Results of these analyses showed that method linearity increased with larger sample sizes. When 700 isolates were analyzed, satisfactory method linearity was achieved ($r = 0.95$), whereas the larger sample size (1,000 isolates) did not further improve method linearity. Moreover, ANOVA analyses, followed by significance testing using Tukey's HSD test, confirmed that at least 700 environmental isolates should be screened in order to obtain statistically ($P < 0.01$) meaningful results. It should be noted, however, that while this number may appear to be large, it represents only about 3% of the hybridization capacity of the membrane, and the robot system used can pick 700 isolates in about 13 minutes. Thus, the developed procedure allows the analysis of approximately 30 water samples (sites) per membrane analyzed.

Field applications

To evaluate the utility of the developed protocol under more realistic conditions, I performed host source-specific macroarray hybridizations on *E. coli* collected from water samples obtained from two urban Minnesota lakes where geese and ducks were suspected to contribute to fecal pollution. Of the 1,000 *E. coli* isolates obtained from each lake sample, 206 from Lake Superior and 276 from Lake Calhoun reacted positively with the goose/duck-specific probes (Table 3.3). Based on the use of Equation 2, the relative contributions of fecal *E. coli* from geese/ducks were estimated to be 34% and 51% in Lake Superior (St. Louis Bay) and Lake Calhoun, respectively. These high numbers are not unexpected, since Canada geese and ducks frequent these lakes many times during the year, and control measures are currently in place to control

their numbers in Lake Calhoun, an urban lake 4.5 miles from downtown Minneapolis. This is in agreement with results from previous studies showing that migratory waterfowl substantially increase fecal counts in freshwater and saltwater systems (71, 74) and that beaches and water in Lake Superior are frequently contaminated with fecal bacteria originating from geese (Satoshi Ishii, personal communication). In contrast, only 1 of 360 *E. coli* isolates (0.28%) obtained from Lake Hartwell in Clemson, SC, reacted with the probes. This is consistent with reports that this lake is relatively pristine and is not frequented by migratory goose populations.

Summary and conclusions

These studies are the first to use a robotics-based approach to quantitatively determine sources and sinks of fecal *E. coli* in the environment. If implemented on a large scale, this method should allow for the rapid determination of the contribution of geese/ducks to fecal loading of beaches and waterways. As currently configured, the system allows for the high-throughput quantitative analysis of approximately 20,700 *E. coli* isolates in a relatively short time, with minimal human intervention. However, it is possible to array up to 36,864 colonies on a single growth/hybridization membrane using an alternate grid pattern, thus allowing analysis of even more water samples at once. The high-throughput analysis realized by this method is different from that obtained using host-specific PCR methods (39, 100, 145, 150) in that a large number of individual isolates can be analyzed for each sample. In contrast, the latter method analyzes each water sample using a single PCR. Additionally, since *E. coli* was used in these analyses, the results can be directly correlated with fecal indicator bacterial counts

commonly used by state and federal agencies to assess contamination in impacted lakes, rivers, and beaches. This is not possible using PCR-based methods directed against bacteria that are not currently used as fecal indicators (12, 13, 38, 39, 149, 150). Although the availability of colony-picking robots has steadily increased in recent years as genomic research expands, access may still be limited for some researchers. The probe method described here can still be used in the absence of a colony-picking robot, by hand-picking colonies and using a multiprong replicator to transfer cells to membranes. However, the use of this manual method will result in reduced throughput and accuracy relative to those of the robotic system.

While these studies used hybridization probes that were specific for geese/ducks in the central Midwest (61), the method can be easily adapted for use with any marker gene system developed to examine bacterial populations impacting waterways and beaches. Development of marker DNA probes specific for *E. coli* from fecal host sources other than geese/ducks is still needed. With the availability of an array of probes specific for *E. coli* originating from other major fecal host sources impacting watersheds, the number of simplifying assumptions made in this study can be considerably reduced, and the overall reliability and quantitative nature of the method can be improved in the future.

ACKNOWLEDGEMENTS

I wish to extend special thanks to Tao Yan who made a significant contributions to this work. I would also like to thank Carl Rosen for his assistance with statistical

analyses and Satoshi Ishii and Nick Hahn for their technical assistance. I am also grateful to John Ferguson for *E. coli* library maintenance.

This work was supported, in part, by a training grant 2T32-GM008347 from the National Institutes of Health.

This work was previously published in Applied and Environmental Microbiology (Appl. Environ. Microbiol. 2006. 72:4012-4019). Permission to reprint this manuscript was obtained from the publisher, American Society for Microbiology, under license number 2216581003429.

Table 3.1

Number of isolates hybridizing with the goose/duck-specific hybridization probes.

Number of spiked Go066 *E. coli* cells (CFU)[a]	Number of hybridizing isolates[b]	Q values for Tukey's HSD test[c]
None	71.0 ± 13.5	-
20 (6)	54.7 ± 17.1	-0.68
100 (29)	92.3 ± 26.1	1.25
200 (57)	104.0 ± 24.1	1.93
400 (108)	207.0 ± 40.4	7.99
800 (195)	280.7 ± 42.8	12.32

[a]Values in parentheses refer to the number of spiked cells per 1,000 indigenous *E. coli*.
[b]Values are means of triplicate samples ± standard errors on mean. The ANOVA F ratio and p value were 26.9 and 3.95×10^{-6}, respectively.
[c]Q values ≥ 6.1 are statistically significant at $P < 0.01$.

Table 3.2

Influence of sample size on method linearity.

Sample size	Method linearity (R)[a]	ANOVA (P value)
100	0.74	4.6×10^{-4}
300	0.88	5.2×10^{-4}
500	0.89	4.5×10^{-4}
700	0.95	2.4×10^{-6}
1000	0.94	4.0×10^{-6}

[a]R values determined by linear regression of hybridizing isolates versus spiked Go066 *E. coli* cells.

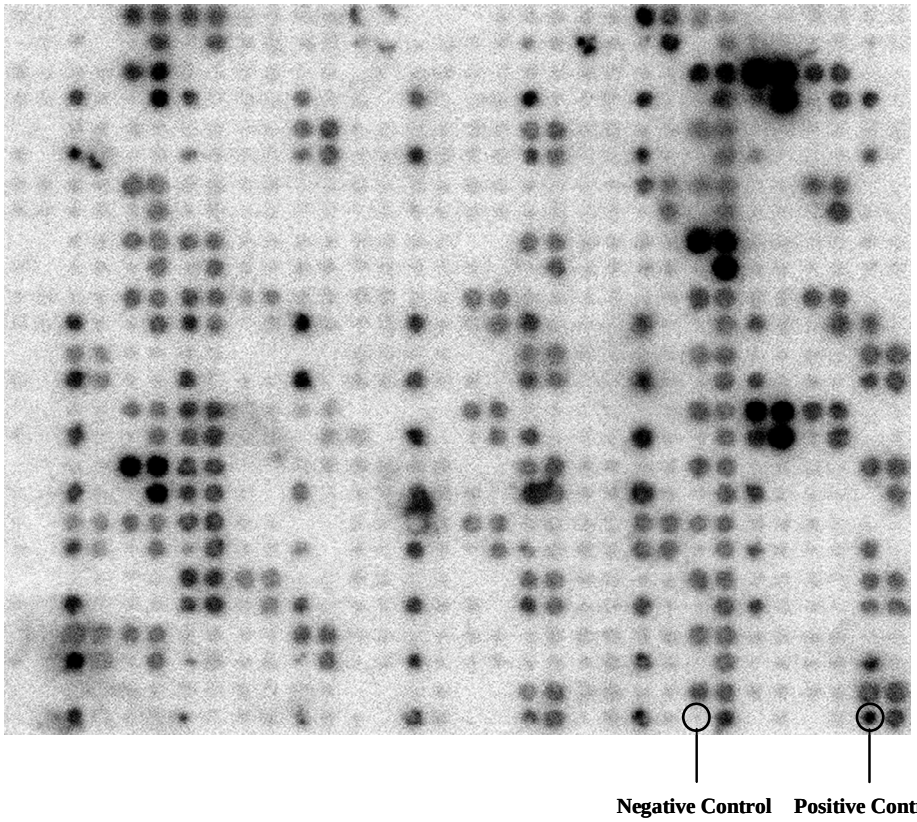
Table 3.3

Contribution of geese/ducks to fecal loading in two Minnesota lakes.

Water source	Number of isolates hybridizing to goose/duck-specific DNA probes[a]	Estimated contribution of *E. coli* from geese/ducks ($C_{geese/ducks}$)[b]
Lake Superior	206	34%
Lake Calhoun	276	51%
Lake Hartwell	1	BDL[c]

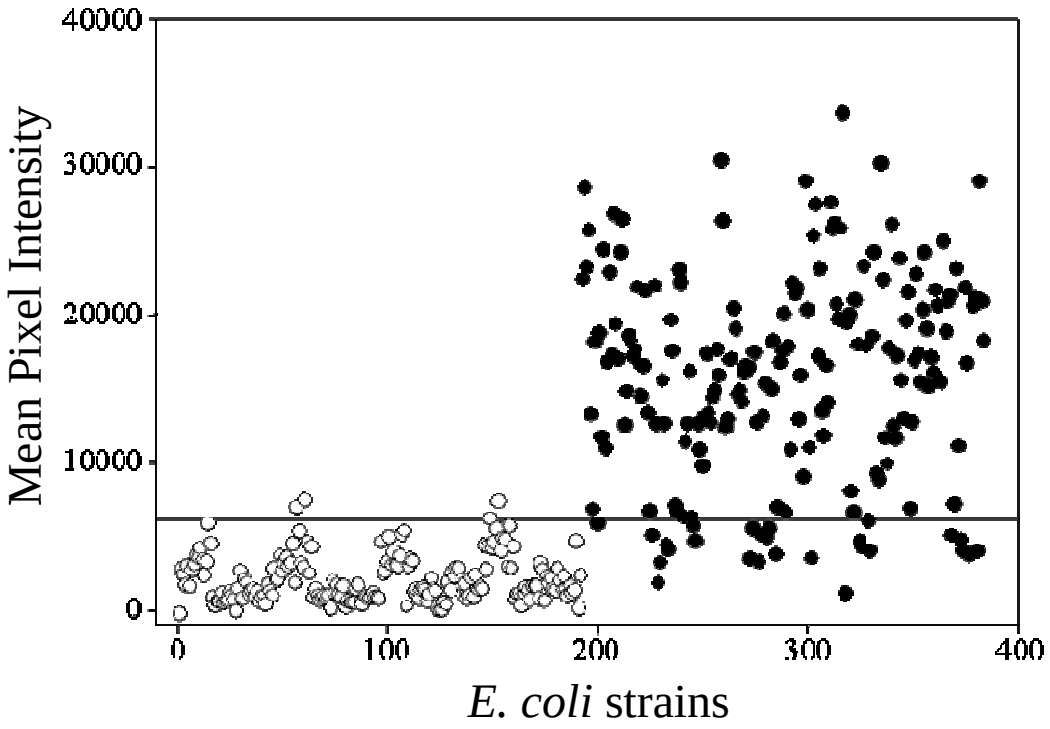
[a]One thousand isolates per lake sample from Lake Superior and Lake Calhoun were analyzed, whereas 360 isolates were analyzed from the Lake Hartwell sample.
[b]Estimated contribution of geese/ducks to fecal loading was determined using Equation II as described in the methods section.
[c]BDL = Below detection limit.

Figure 3.1



Colony hybridization of ^{32}P-labeled GB2/GE11 goose/duck-specific marker gene probes to *E. coli* isolates obtained from geese and humans. *E. coli* strains Go66 and Hu132 served as positive and negative controls, respectively.

Figure 3.2



Pixel intensities from macroarray hybridization membrane containing five *E. coli* strains isolated from geese (●) and five strains from humans (○). The *E. coli* strains from geese and humans were replicated 192 times. The membrane was probed with ^{32}P-labeled GB2/GE11 goose/duck-specific marker gene probes. The solid line indicates a cutoff value for determining positive and negative signals.

CHAPTER 4

Large Scale Analysis of Virulence Gene Profiles of Beach Water-Associated *Escherichia coli* strains

OVERVIEW

Contamination of recreational waters with *E. coli* and enterococcus species is a widespread problem resulting in beach closures and loss of recreational activity. While *E. coli* is frequently used as an indicator of fecal contamination, and has been extensively studied in waterways, few studies have examined the presence of potentially pathogenic *E. coli* strains in beach waters. In this study, we used a combination of high-throughput, robot-assisted colony hybridization and PCR-based analyses to determine the genomic composition and frequency of virulence genes present in *E. coli* isolated from beach water at Avalon Beach on Santa Catalina Island, CA. A total of 24,493 *E. coli* isolates were collected between August through September 2007 and from July through August 2008, respectively, from two sites at popular swimming beach in Avalon, CA. All isolates were examined for the presence of shiga-like toxins (*stx1*/*stx2*), intimin (*eaeA*), and enterotoxins (ST/LT). Of the 24,493 isolates examined, 3.6% contained the *eaeA* gene, indicating that these isolates were potential EPEC strains. On five dates, however, >10% of the strains were potential EPEC, suggesting that incidence of virulence genes at this beach has a strong temporal component. No STEC or ETEC isolates were detected and only eight (<1.0%) of the potential EPEC isolates were found to carry the EAF plasmid. The potential EPEC isolates mainly belonged to *E. coli* phylogenetic groups B1 or B2 and carried the beta intimin subtype. DNA fingerprint analyses of the potential EPEC strains indicated that the isolates belonged to several genetically diverse groups, although clonal isolates were detected. While the presence of virulence genes alone cannot be used to determine the pathogenicity of strains, results from this study show that potential EPEC strains can be

found in recreational beach water and their presence needs to be considered as one of the factors used in decisions concerning beach closures.

INTRODUCTION

Contamination of recreational waters with *Escherichia coli* and *Enterococcus* species is a common problem resulting in beach closures and loss of recreational activity. Because these organisms are most frequently found in the intestinal tract of warm-blooded animals and are shed with feces, these bacteria have traditionally been used as indicators of fecal contamination. While it has been generally accepted that the presence of indicator organisms in recreational waters suggests the presence of fecal-borne pathogens (28, 44, 130), several studies have shown that these bacteria can persist in the environment in the absence of a host, and can become naturalized to beach sand, soil and algae (79, 188). Although *E. coli* is often thought of as a harmless, commensal, organism, some strains have been shown to be capable of causing human disease (91, 115), and fecal material from some animals contains high frequencies of *E. coli* strains harboring virulence genes (78). Despite the increasing evidence that *E. coli* strains from several animal hosts contained virulence genes, and some have been shown to cause serious or fatal diseases in humans (115), few studies have determined whether *E. coli* strains isolated from marine recreational waters contain virulence genes and are potentially pathogenic strains.

Pathogenic *E. coli* infections generally cause diarrhea and other gastrointestinal disease (91, 115), although some strains have been found to cause extraintestinal infections (40, 108). Diarrheagenic strains are classified into several groups based on the mechanisms of pathogenesis and on the presence of various virulence factors or determinants. These groups include diffusely adhering (DAEC), enteroaggregative

(EAEC), enteroinvasive (EIEC), enteropathogenic (EPEC), enterotoxigenic (ETEC), and shiga-like toxin producing (STEC) *E. coli* (115).

STEC are defined as *E. coli* strains expressing either of the shiga-like toxin genes *stx1* or *stx2* or other toxins sharing significant homology to the shiga toxin gene originally identified in *Shigella dysenteriae* (91, 127). The common reservoirs for STEC strains are ruminants and swine (41, 49, 59, 78). Strains such as *E. coli* 0157:H7 belong to a subset of the STEC group, known as enterohemorrhagic *E. coli* (EHEC), which cause severe and potentially fatal diseases in humans, including hemolytic uremic syndrome and hemorrhagic colitis (115). EHEC strains may also carry other virulence determinants, including the locus of enterocyte effacement (LEE) pathogenicity island (91, 115). While the LEE region is not required to cause human disease, it is associated with enhanced virulence and encodes several virulence factors, including intimin (*eaeA*). Intimin is necessary for the development of attaching and effacing lesions (91, 115).

ETEC strains are defined by the presence of at least one of the heat stable (ST) and the heat labile (LT) enterotoxin genes (103). These strains are often associated with infantile diarrhea, particularly in the developing world where the disease is considered endemic (17, 91, 115). ETEC strains are also a common cause of traveler's diarrhea, where visitors from the developed world may lack immunity to these strains (17). Both of these toxins function to increase chloride ion secretion into the intestinal lumen, resulting in osmotic diarrhea (91, 115).

EPEC strains are a common cause of human diarrheal diseases in developing countries, particularly among children less than 2 years of age (104, 169). EPEC strains

have been isolated from many animal host species, including humans, cats, cows, dogs, deer, ducks, geese, and horses (78). These strains are defined by the presence of the LEE pathogenicity island and the absence of the shiga-like toxin genes (91, 115). While the LEE region contains several virulence genes, studies done on volunteers using mutant strains indicated the *eaeA* gene plays a prominent role in pathogenesis (43), and detection of this gene is used to screen for EPEC isolates (78, 127).

Thirteen different intimin subtypes have been defined and may be associated with specific reservoirs for *eaeA*$^{+}$ strains (78, 131). EPEC strains may also possess the EAF (EPEC adherence factor) plasmid, which encodes the bundle forming pilus and carries several regulatory genes which enhance, but are not required for the development of human disease (58, 169). Strains carrying pEAF are referred to as typical EPEC, while those strains without the plasmid are referred to as atypical EPEC.

Most *E. coli* strains have been assigned to one of six major phylogenetic groups (A, B1, B2, C, D, and E) based on their evolutionary origins (33, 45). The association of STEC strains with certain phylogenetic groups has been tentatively established. Escobar et al. (45) determined the majority of STEC strains they examined from clinical samples belonged to phylogenetic group B1. However, while a tentative a link has been established between intimin subtype and phylogenetic group (45, 133), not all EPEC strains have a specific genetic background, and they may be distributed among several phylogenetic groups (45, 78).

Detection of EPEC and STEC strains has previously been done by using multiplex PCR, using primers specific to *eaeA* and *stx1*/*stx2*, respectively (78, 127). While these analyses are useful for screening relatively small numbers of isolates, they

suffer from the high costs and low throughput associated with PCR. We previously reported the successful use of high-throughput, semi-automated, robotic technology to detect the presence of host source-specific genes in environmental *E. coli* isolates (187). In the studies presented here, we adapted this technology to quickly screen large numbers of *E. coli* isolates from marine recreational water for the presence of virulence determinants and confirmed positive hybridization results by PCR.

The objectives of this present study were to: 1) examine the distribution and frequency of potentially pathogenic *E. coli* stains isolated from beach water at a popular swimming beach in Avalon, CA; 2) characterize the identified potential pathogenic strains by virulence profile testing and phylogenetic analyses; and 3) determine the genetic relatedness of the potentially pathogenic strains by using horizontal fluorophore-enhanced rep-PCR (HFERP) DNA fingerprinting analyses.

MATERIALS AND METHODS

Sample Collection

Water samples were obtained in 2007 and 2008 at either 8:00 am or 12:00 pm from two sites (B and C) at Avalon beach, Santa Catalina, CA between 8/18/07 and 9/9/07, as previously described (21). The two sampling sites described in this study are identified as sites B and C (Figure 4.1). Water samples (4 L) were collected at each time point and bacteria were concentrated by membrane filtration using 0.45 μm (142 mm dia.) Supor hydrophilic polyethersulfone membranes (Pall Corporation, East Hills, NY). After filtration, membranes were cut into four equal sections and each section

was placed in a 50 ml conical tube containing 10 ml of sterile phosphate buffered saline (20 mM sodium phosphate, 15 mM sodium chloride, pH 7.2) amended with 0.1% hydrolyzed gelatin and 10 g of sterile 3 mm glass beads. Bacteria were removed from the filter surface by gentle agitation for 30 min using a wrist action shaker. Filters were removed, 50% glycerol was added to obtain a final concentration of 10%, and the samples were stored frozen at -80^{o}C until used.

E. coli reference strains

The *E. coli* reference strains used as positive and negative controls for colony hybridization with the *stx1* and *stx2* probes were Pig206 and Deer090, respectively (42, 88). Deer090 and Pig206 strains, respectively, served as positive and negative controls for hybridizations with the *eaeA* DNA probe. The ETEC strain 1362 and Pig206 were used as positive and negative controls for hybridizations with the enterotoxin gene probes, respectively. The *E. coli* strain O157:H7 (ATCC 43895) was used as positive control for PCR-based assays for virulence genes and for amplifying DNA used as probes for eaeA, stx1, and stx2. The ETEC strain 1362 was used for amplifying DNA used as probes for *STa* and *LT-I*. The *E. coli* strain H120 was used a positive control for PCR reactions to detect the EAF plasmid (42, 78, 88). The *E. coli* strain Pig 294 was used as control for HFERP DNA fingerprint analysis (42, 88).

Isolation and picking of *E. coli*

E. coli strains were isolated from filter washings as previously described (187). One to three ml aliquots of filter washings were spread-plated on to the surface of

modified mTEC medium in 22-by-22-cm Q-tray bioassay plates (Genetix Boston, MA). Modified mTEC was made as described, except that 500 μg of 5-bromo-4-chloro-3-indolyl-β-d-glucuronic acid (X-Gluc) per ml was used as the chromogenic indicator (174, 188). Plates were incubated at 35^{o} C for 2 h, and then at 44.5^{o} C for 22 h. After incubation, plates were stored at 4^{o} C overnight to facilitate development of blue pigment in colonies, allowing for the differentiation of *E. coli* from other fecal coliforms. Well-isolated blue colonies were picked by hand or by using a Q-Bot robotic system (Genetix, Boston, MA) into 384 well microplates (Genetix, Boston, MA) containing HMFM medium. Microplates were incubated at 37^{o} C overnight, and stored at -80^{o} C until use. A total of 12,000 and 12,493 individual *E. coli* isolates were obtained in 2007 and 2008, respectively. A random sample of 1024 strains were confirmed as *E. coli* by using a series of microbiological and biochemical tests including, the indole and methyl red tests, the inability of isolates to grow on citrate, β-d-glucuronidase activity using EC-MUG broth (Difco), and their color reaction on ChromAgar and MacConkey agar (42, 88).

Automated arraying of isolates

Automated arraying of isolates onto positively charged, 22-by-22-cm Performa II nylon membranes (Genetix, Boston, MA) was done using a QBot robot system (Genetix, Boston, MA) as previously described (187). Membranes were arrayed according to a pattern whereby each membrane was divided into 6 sections, each section contained 384 subunits, and each subunit consisted of 4 individual spots. One spot per subunit, contained either a positive or negative control strain. Using this

format, each membrane was arrayed with a maximum of 6,912 *E. coli* isolates obtained from water samples. Arrayed membranes were placed on the surface of LB agar plates, incubated at 37^{o} C for 8-10 hr, and stored at 4^{o} C overnight.

Colony hybridization screening for potentially pathogenic *E. coli*

Colony hybridizations were done to determine the reactivity of the *E. coli* isolates to DNA probes for *stx1*, *stx2*, and *eaeA* as previously described (61, 187). PCR reactions to amplify probe DNA were done using primers pairs stx1F and stx1R, stx2F and stx2R, and eaeAF and eaeAR for *stx1*, *stx2*, and *eaeA*, respectively (127). Probe DNA for the heat stable (*STa*) and heat labile (*LT-I*) toxin genes was amplified using primer pairs JW14/JW7 and LTPr1/JW11, respectively (159). Probes were labeled with [^{32}P]dCTP using the Random Primer DNA Labeling system (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Probes for the shiga-like toxin genes and enterotoxin genes were pooled before labeling. Membranes were hybridized overnight at 68° C and washed under high stringency at 68° C in 0.1× SSC (1× SSC is 0.15 M NaCl and 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate (187). Membranes were air dried after washing, wrapped in plastic film, and exposed to storage phosphor imaging screens (GE Healthcare, Chalfont St. Giles, UK) overnight. All colony hybridizations were done in triplicate. Colony hybridization images were captured using a STORM 840 densitometer (GE Healthcare, Chalfont St. Giles, UK) and quantitative image analysis was done using ScanAlyze version 2.51 software (http://rana.lbl.gov/dowloads/ scanalyze/scanaylze2_vers_2_51.exe). Positive and negative hybridization signals were determined as previously described (187). Strains

showing positive hybridization signals on any of the triplicate membranes were selected for subsequent confirmation by PCR.

PCR screening for potentially pathogenic *E. coli* and subtype analysis

Strains showing positive hybridization signals when analyzed using gene probes were confirmed using a multiplex PCR approach with primer pairs stx1F/stx1R, stx2F/stx2R, eaeAF/eaeAR (127) and with primer pairs JW14/JW7 and LTPr1/JW11 (159). Template DNA was extracted from cells grown overnight in LB medium as previously described (77), diluted 10-fold in distilled H_2O, and 1-2 μl of lysate was added to each PCR reaction. Isolates shown to carry the *eaeA* gene were also tested for the presence of the EAF virulence plasmid using primer pairs EAF1 and EAF25 (48). Subtype analysis of the *eaeA* gene was done using a PCR-RFLP technique with the primers EaeVF, EaeR, EaeZetaVR, and EaeIotaVR as previously described (131). Amplified products were digested with restriction enzymes AluI, CfoI, or RsaI to distinguish among 14 *eae* subtypes (131). DNA fragments were separated by electrophoresis using 2% SeaKem LE agarose gels (FMC, Rockland, ME) at 90 V for 2-3 hours. Gels were stained with 0.5 μg ethidium bromide per ml and visualized with UV light.

Phylogenetic group classification and HFERP DNA fingerprinting

E. coli strains were classified into one of four major phylogenetic groups (A, B1, B2, and D) using a multiplex PCR protocol as previously described (33). Strains with the Clermont (+ - +) genotype (phylogenetic group D) were tested for the presence of

ibeA to determine if these strains actually belonged to phylogenetic group B2 (55). HFERP DNA fingerprint analyses were done using the BOXA1R primer to compare the genetic relatedness of isolates as previously described by Johnson et al. (88).

Statistical analysis

G-tests were used to determine if *E. coli* carrying virulence factor genes were evenly distributed among samples. Statistical analysis of DNA fingerprint data was done using Bionumerics software (version 2.1) (Applied Math, Kortrijk, Belgium). Dendrograms were constructed using the curve-based, Pearson's product-moment correlation coefficient and clustering was done using the unweighted pair group method with arithmetic means (UPGMA) (88). Multivariate analysis of variance (MANOVA) was used to cluster *E. coli* strains isolated from different sites and years (27, 42, 77)

RESULTS

Isolation of *E. coli* from contaminated beach water

Water samples were collected at 8:00 am or 12:00 pm from two different sites, B and C, at Avalon beach, Santa Catalina, CA between 8/18 - 9/16/07 and 7/5 – 8/31/2008 (Figure 4.2, Table 4.1). Bacteria present in the samples were concentrated by membrane filtration and spread plated onto the surface mTEC agar to allow for the differentiation of *E. coli*, which have blue colored colonies, from other fecal coliforms. A total of 12,000 and 12,493 well isolated individual blue colonies from 2007 and 2008, respectively, were selected for further analysis. The frequency of obtaining *E. coli* from water samples varied considerably by date (Table 4.1) and was directly related to *E. coli*

counts from these days. A series of microbiological and biochemical assays were done on 1,024 randomly-selected isolates to confirm the strains were *E. coli*. Assay results showed that 99.8% of the isolates were *E. coli* as determined by their reactions to the methyl red and indole tests, ChromAgar, MacConkey agar, citrate agar, and their β-d-glucuronidase activity, respectively (data not shown). These results are consistent with other reports concerning the ability of mTEC medium to selectively isolate *E. coli* from water samples (174).

Screening for potentially pathogenic *E. coli*

A hybridization-based method incorporating an automated robotic system for arraying colonies on nylon membranes was used for screening large numbers of environmental *E. coli* isolates for the presence of virulence genes. A representative image of a colony hybridization membrane probed with the *eaeA* fragment is shown in Figure 4.3. Quantitative image analysis was used to distinguish between positive and negative hybridization signals. Positive signals were defined as those with pixel intensities greater than the mean intensity of negative control spots plus three times the standard deviation. Multiplex PCR reactions were run on all isolates with positive hybridization signals. Results showed that 875 of the 24,493 isolates (3.6%) examined were positive for *eaeA* while no isolates were found to carry either the STEC toxin genes *stx1* or *stx2* or the ETEC toxin genes *LT-I* or *STa* (Table 4.1).

E. coli isolates carrying the *eaeA* virulence factor gene were found at both beach sites, but were found at a greater frequency at site B compared to site C (4.8% and 1.5%, respectively). The percentage of *eaeA* positive isolates for a given sample varied

from 0 to 11.8%, with the greatest number present on 8/16/2008 at site B (Figure 4.2). G-test analysis indicated that the *eaeA* positive isolates were not evenly distributed among samples ($P < 0.001$). When G-test analysis was done to compare individual samples collected at common dates and times and at each of the two sampling sites, results indicated that the distribution of EPEC was not even ($P < 0.001$) and that the frequency of *eaeA*+ isolates was greater at site B. However, samples collected at 8am on 8/18/07 were evenly distributed between site B and C. Samples collected at 8 am were not compared to samples collected at 12pm because of the lack of isolates obtained from 12pm samples.

E. coli isolates confirmed to carry *eaeA* were also examined for the presence of the EAF virulence plasmid by PCR (48). Of the 875 potential EPEC isolates tested, eight isolates (0.9%) were found to carry the EAF plasmid (data not shown). Intimin subtyping of all *eaeA* positive *E. coli* strains showed that 87.5% of the isolates carried the β intimin subtype (Figure 4.4, Table 4.2). Subtypes α-2, ζ, η, θ, ι, κ, ν, and ξ were found at lower frequencies.

Phylogenetic group classification

Phylogenetic analysis was used to assign the *eaeA* positive isolates to phylogenetic groups A, B1, B2, and D. Preliminary group assignment was based on multiplex PCR detection of the *chuA* and *yjaA* genes and a chromosomal DNA region known as TSPE4.C2 (33) and a PCR reaction to detect the presence of *ibeA* (55). Of the 875 isolates examined, the greatest number (70.5%) belonged to phylogenetic group B1, followed by those in group B2 (25.0%). Strains in groups A and D were

represented at lower frequencies, 3.2 and 1.3%, respectively. The relationship between phylogenetic group classification and intimin subtype is shown in Figure 4.4.

Population structure of beach water isolates

The genetic relatedness of the *eaeA*$^{+}$ *E. coli* isolates was determined using HFERP DNA fingerprinting. Dendrograms created using data from all *eaeA*-postive isolates showed that isolates were distributed into several distinct and genetically diverse groups containing clonal and closely related strains. The relative similarity among these strains ranged from 3.42 to 100% (Figure 4.5). The genetic relatedness of isolates from 2007 and 2008 was also examined and the relative similarity of isolates ranged from 6.43 to 100% and 2.35 to 100%, respectively. Clonality was defined as strains having a similarity value of 92% or greater (88). Based on this criterion, 128 different *eaeA*-positive *E. coli* strains were isolated during this study. The two largest groups of clonal isolates, designated as groups I and II contained 211 and 137 isolates, respectively. Interestingly, group I strains were obtained from sites B and C from several different dates in 2007 and from site B on two dates in 2008. In contrast, group II strains only contained isolates from sites B and C isolated from several dates in 2007 and were not found in 2008 samples. The remaining 126 strains were represented by 64 or less isolates. Eighty-two of the 128 (64%) strains found in this study were represented by a single isolate. A dendrogram showing the relatedness of strains represented by 10 or more isolates is shown in Figure 4.6. MANOVA analysis showed that isolates from 2007 and 2008 generally clustered separately into two large groups (Figure 4.7). Within these clusters, isolates from sites B and C clustered into overlapping groups

(Figure 4.7). The first and second discriminants accounted for 91% of the variation indicating the strains clustered well to their respective groups.

DISCUSSION

The water samples used in this study were collected from two popular swimming beach sites in Avalon, CA. Both of these sites suffer from fecal contamination and fecal indicator bacteria counts at these sites frequently exceeded standards. A recent Natural Resources Defense Council report named these sites as two of the ten most highly contaminated beaches in CA, based on the frequency of samples exceeding standards (116). In this study, we focused on the isolation of *E. coli* from the contaminated marine water. *E. coli* has historically been used as an indicator of fecal contamination, but is generally not examined in marine environments (171). Enterococci species are used as the indicator for fecal contamination in marine systems because *Enterococcus* counts have a stronger correlation to human disease rates (28, 44). Despite this fact, *E. coli* can readily be isolated from marine waters at Avalon, CA and potentially pathogenic strains have been detected. Thus, *E. coli* in marine water may impact human health as wave action may facilitate ingestion of contaminated water.

Initial screening of the 24,493 *E. coli* isolates collected from marine water samples was done using an automated arraying system and methodology which was previously successful in examining large numbers of environmental *E. coli* isolates for the presence of other genes (187). The results obtained in this study further validate this methodology. Of the 24,493 isolates examined, 875 (3.6%) were found to be potential

EPEC strains based on detection of *eaeA* gene. Since the majority (>99%) of the potential EPEC strains described in this study did not harbor the EAF plasmid, they were classified as atypical EPEC (91, 115). There is some controversy regarding the pathogenicity of atypical EPEC (aEPEC) and it has been suggested that these strains arose from STEC that have lost bacteriophages carrying *stx* genes or EPEC strains that have lost the *bfpA*-encoding EAF plasmid (16, 104). A recent study by Tennant et al. (165) showed that 80% of the EPEC strains they examined from clinical and water samples were genetically distinct, carried adhesins that may serve as replacements for the lack of *bfpA* and suggested that different populations of aEPEC may have varying degrees of pathogenicity.

Historically, most EPEC cases in industrialized countries were associated with typical strains, but more recently atypical EPEC strains were linked to outbreaks of human disease affecting both adults and children all over the world (2, 3, 19, 68, 114, 117, 189). A study by Scotland et al. (144) using clinical isolates collected in the UK, showed that 90% of EPEC cases were caused by atypical strains. More recently, Nguyen et al. 2006 (117) showed that patients infected with atypical EPEC strains were more likely to experience diarrhea lasting longer than 2 weeks, increasing the risk for serious illness and death. Taken together, aEPEC strains in contaminated recreational water sources may represent a public health risk to recreational beach users.

Interestingly, no STEC or ETEC strains were found in the 24,493 isolates screened. This result is similar to that reported by Higgins et al. (70) where the EPEC virulence gene *tir* was detected in water samples at a significantly greater frequency than shiga-like toxin genes. The location of the sampling sites may in part explain the

lack of STEC isolates, as the common reservoirs for these strains, such as ruminants and swine (41, 49, 59, 78), are likely not contributing to the fecal load at the beach sites examined. At this site, potential input sources likely include humans, pets, and wildlife. Also, the lack of ETEC strains may not be surprising as these strains are often associated with the developing world (91, 115).

The potential EPEC strains identified in this study were found in every sample examined but were not evenly distributed according to the G-test ($P < 0.001$) and varied from 0 and 11.8% of the *E. coli* isolated from a given sample. Overall, the potential EPEC strains were found at a greater frequency at site B compared to site C (5.3 and 1.9%, respectively). Statistical analysis indicated the EPEC strains were also not evenly distributed among samples collected at common dates and times at different sites, with the notable exception of samples collected from sites B and C on the morning of 8/18/08. The reason for the uneven distribution is not known, but may be related to the proximity of the sampling sites to point sources of fecal contamination, currents present at the sampling sites, and wind and wave action. It is also possible that nonpoint source run-off originating from humans, pets, and wildlife may have impacted site B to a greater extent than site C.

Results of this study also showed that there was temporal distribution of the *E. coli* strains. More than 99% of the *E. coli* isolates screened in this study were collected from samples obtained at 8 am, and only 3 of the 875 potential EPEC strains were isolated from the 12 pm samples. Additional isolates from 12:00 pm samples were not collected due to low *E. coli* counts in samples (data not shown). Inactivation of fecal indicator bacteria as result of exposure to sunlight has been shown to be a major factor

in the persistence of this bacterium in the environment (35) and may help to explain the reduced number of culturable organisms in samples collected at 12 pm.

Phylogenetic analyses, intimin subtyping assays, and HFERP DNA fingerprinting revealed a diverse population of potential EPEC strains. The strains mainly belonged to phylogenetic groups B1 and B2, with frequencies of about 70 and 25%, respectively. Strains belonging to groups A and D were also found, but at much lower frequencies. Nine intimin subtypes (α-2, β, ζ, η, θ, ι, κ, ν, and ξ) were identified among the potential EPEC isolates by PCR-RFLP analysis. The β, ι , and ζ subtypes were most represented and present in frequencies of 87.5, 6.2, and 3.1% respectively. Strains possessing the β intimin subtype were assigned to all four phylogenetic groups, although almost 77% of the strains were assigned to the B1 group. This result is consistent with previous reports by Reid et al. (133) and Ramachandran et al. (131) who showed that the majority of strains comprising the β intimin subtype belonged to phylogenetic group B1 and that the β intimin subtype was found at higher frequencies than other subtypes in clinical isolates from humans, respectively. The ι intimin subtype was found only in B1 and B2 strains, with approximately 75% of the isolates belonging to group B2. The other intimin subtypes were detected at frequencies of <1.0%. Ishii et al. (78) reported on the presence of the ι subtype in isolates obtained from ducks and geese and the κ subtype among *E. coli* from domestic dogs and cats. Humans, pets and birds are likely the main contributors to fecal loading at the sampling sites described in this study and, in light of the studies described above, may explain the prevalence of the β, ι, and κ subtypes. Additional studies are necessary to conclusively determine the sources of the potential EPEC strains at the Avalon Beach site.

Dendrograms generated using HFERP fingerprint data also showed that the potential EPEC strains belonged to several groups, consisting of clonal and closely related strains. The 875 potential EPEC isolates examined represented 128 different strains as defined by Johnson et al. (88) and were genetically diverse with relative similarities ranging to less than 5%. However, about 40% of the potential EPEC isolates belonged to one of the two most represented strains. The remaining 60% of the isolates were classified into 126 different strains which were found only once or in smaller groups and, in most cases, were closely related to other strains. Clustering analysis done using MANOVA confirmed the results of the dendrogram analysis and showed that isolates cluster relatively well by year into overlapping groups of isolates from each site. Taken together, these results suggest that a few different potential EPEC strains are predominantly isolated from both sites, at least during the summer months, and that most other strains exist transiently. Also, the population of potential EPEC strains may shift in successive years. The reason that some strains were detected at much greater frequencies over a range of dates than other strains is not clear, but may be due to continual deposition as the result of an unknown reservoir or through persistence in the environment.

The presence and detection of potential EPEC strains in environmental samples taken from beach sand and water has been previously reported, although these studies examined freshwater environments and fewer isolates were examined (60, 76, 99). To our knowledge, this study is the largest examination of *E. coli* isolated from contaminated water for virulence genes completed to date. Although the presence of virulence determinants alone cannot be used to determine the pathogenicity of strains,

results from this study show that potential EPEC strains can be found in water obtained from contaminated beaches. The ability of the strains described in this study to cause human disease has yet to be determined. Screening of the potential EPEC strains for other virulence factor genes, serotype testing, and other assays may provide further evidence to support this hypothesis. A quantitative measure of the health risk associated with exposure to contaminated water containing these strains also needs to be established through epidemiological studies.

ACKNOWLEDGEMENTS

I would like to thank the Southern California Coastal Waters Research Project (SCCWRP) staff, especially John Griffith, Yiping Tsao, Ben Ferraro, and Nick Miller, for their assistance with sample collection and processing. I also thank Nick Hahn and the High-Throughput Biological Analysis facility at the University of Minnesota for assistance with automated arraying, Daniel Norat and Chris Brandsey for help with colony-picking, and John Ferguson for helping with cluster analyses.

This work was funded, in part, by grants from the Minnesota Agricultural Research Station and SCCWRP and by training grant 2T32-GM008347 from the National Institutes of Health.

Table 4.1

Number of *E. coli* isolates obtained from water at Avalon Beach screened and found positive for the *eaeA* gene.

Strains positive for *eaeA*				
Date	**Site**	**Time**	**# screened**	**# positive**
8/18/2007	B	8:00	384	2
8/19/2007	B	8:00	384	29
8/23/2007	B	8:00	96	5
8/24/2007	B	8:00	2304	256
8/25/2007	B	8:00	384	41
8/26/2007	B	8:00	2760	22
9/9/2007	B	8:00	864	36
8/18/2007	B	12:00	192	1
8/24/2007	B	12:00	96	2
8/18/2007	C	8:00	1920	7
8/23/2007	C	8:00	288	32
8/24/2007	C	8:00	768	35
8/26/2007	C	8:00	1560	12
	2007 Total		**12000**	**480 (4.0%)**
7/5/2008	B	8:00	3245	1
7/6/2008	B	8:00	1517	67
8/3/2008	B	8:00	673	14
8/15/2008	B	8:00	762	53
8/16/2008	B	8:00	768	91
8/31/2008	B	8:00	1061	122
7/6/2008	C	8:00	383	13
8/3/2008	C	8:00	2260	16
8/16/2008	C	8:00	1824	18
	2008 Total		**12493**	**395 (3.2%)**
	Grand Total		**24493**	**875 (3.6%)**

Table 4.2

Frequency of nine different intimin (*eaeA*) subtypes in the 875 potential EPEC strains isolated in this study.

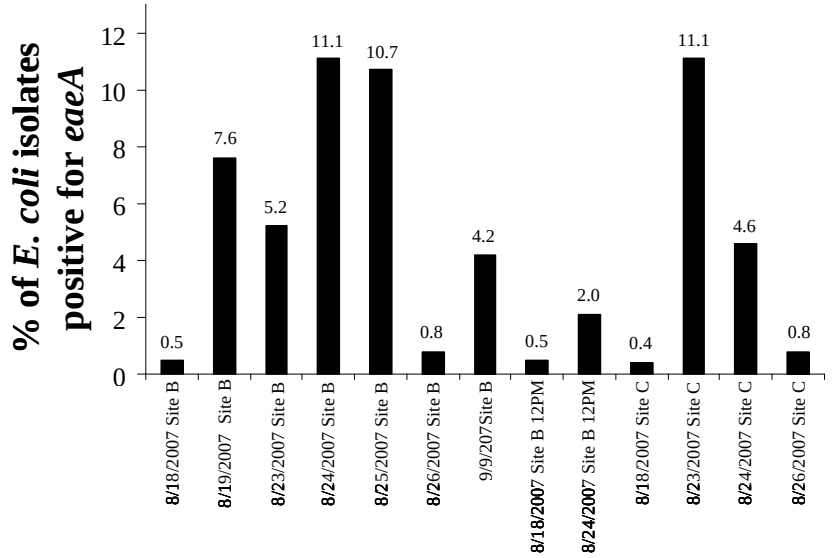
Intimin Subtype	Number positive
α-2	7 (0.8%)
β	766 (87.5%)
ζ	6 (0.7%)
η	1 (0.1%)
θ	8 (0.9%)
ι	54 (6.2%)
κ	3 (0.3%)
ν	3 (0.3%)
ξ	27 (3.1%)

Figure 4.1

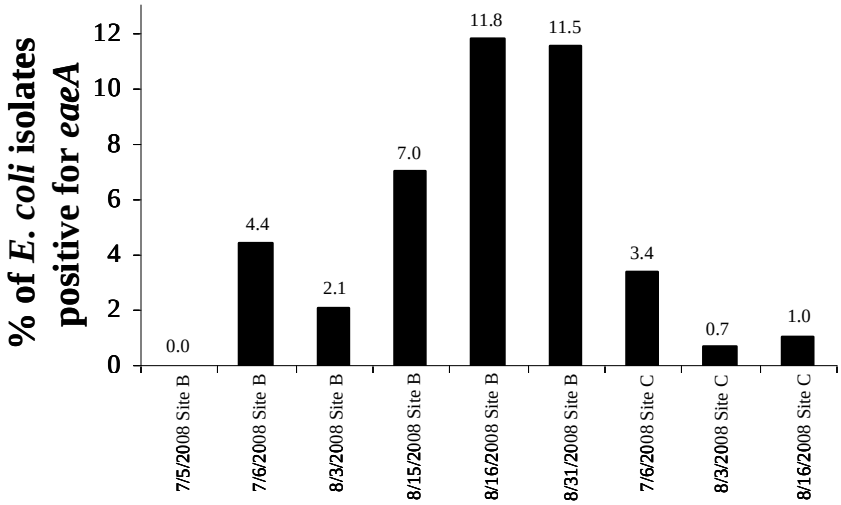
Photograph of the sample sites at Avalon Beach, CA. Sites B and C are noted.

Figure 4.2

A.



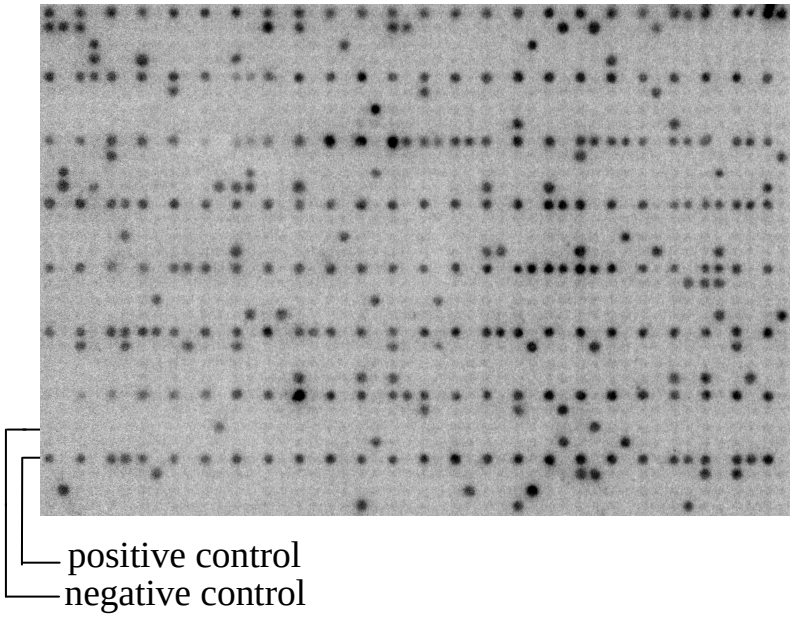
B.



.

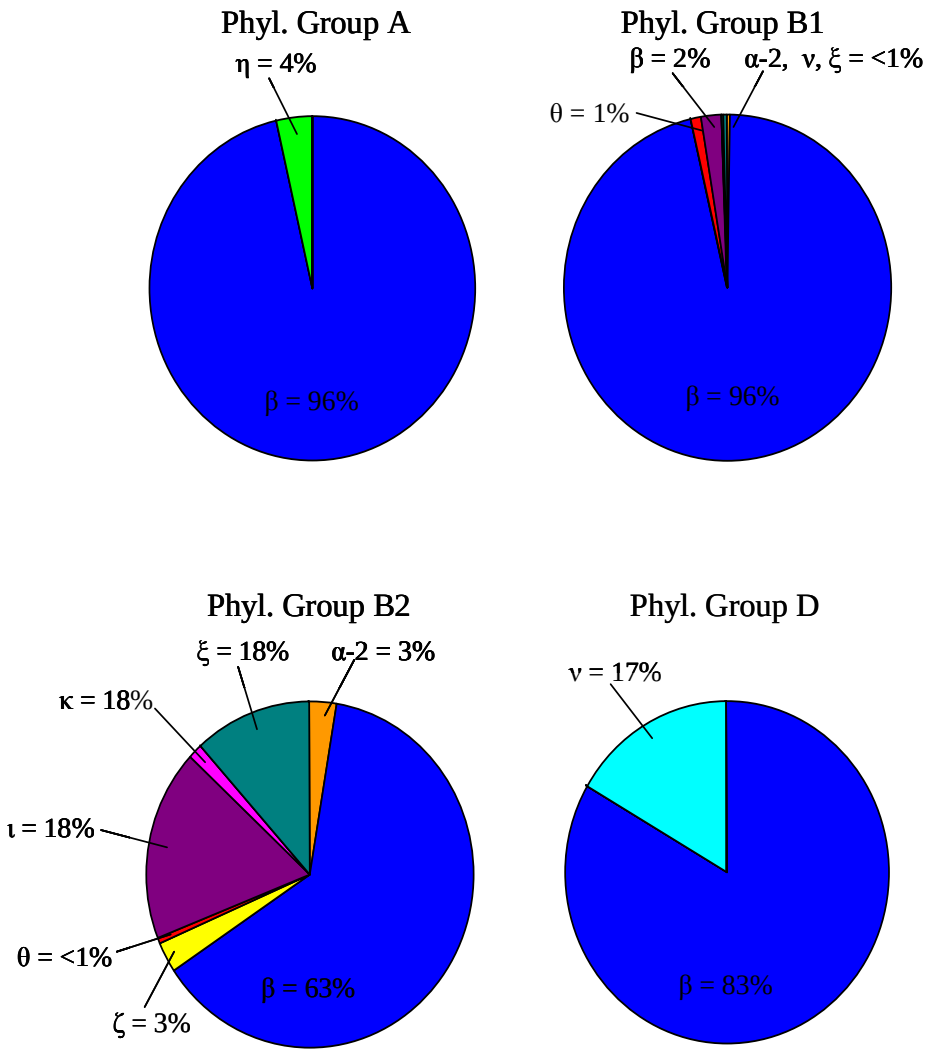
Frequency of *E. coli* carrying the *eaeA* (intimin) gene from water samples collected at Avalon beach. Samples collected in 2007 and 2008 are shown in panels A and B, respectively. All samples were collected at 8AM unless noted. Values above the bars are percentages.

Figure 4.3



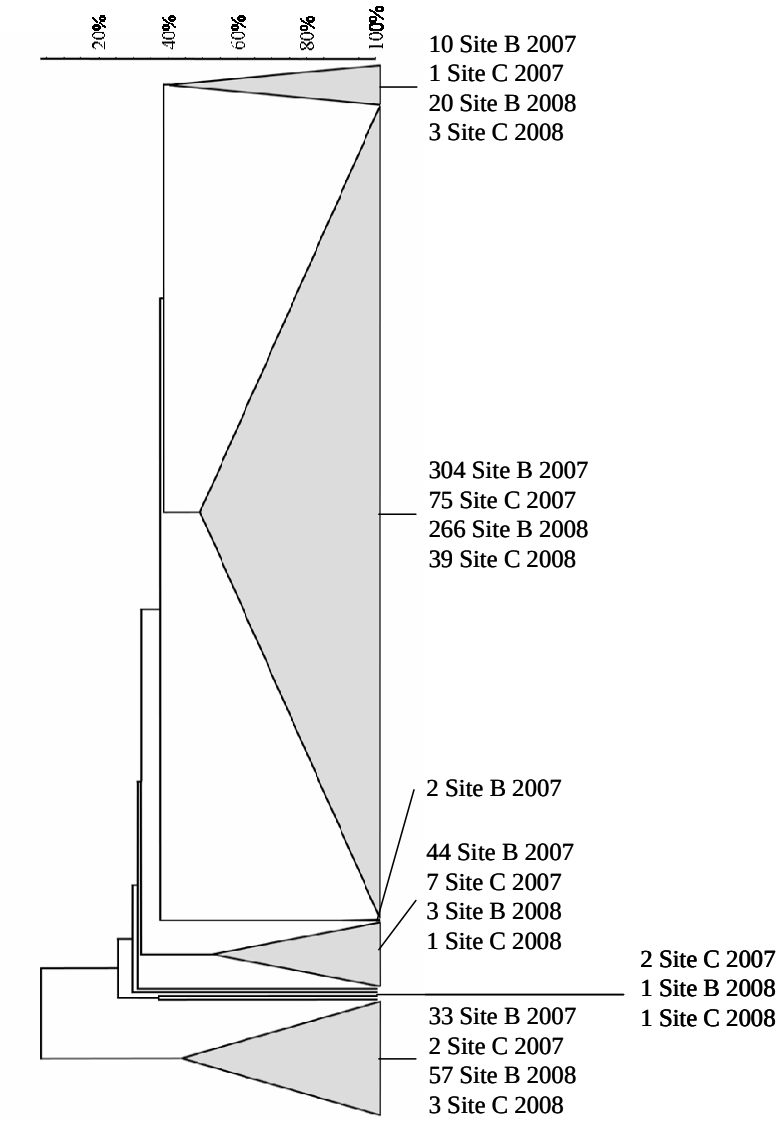
A representative image of a colony hybridization membrane probed with ^{32}P-labeled, PCR amplified fragment of the *eaeA* gene. Positive and negative control strains from a single subunit are shown.

Figure 4.4



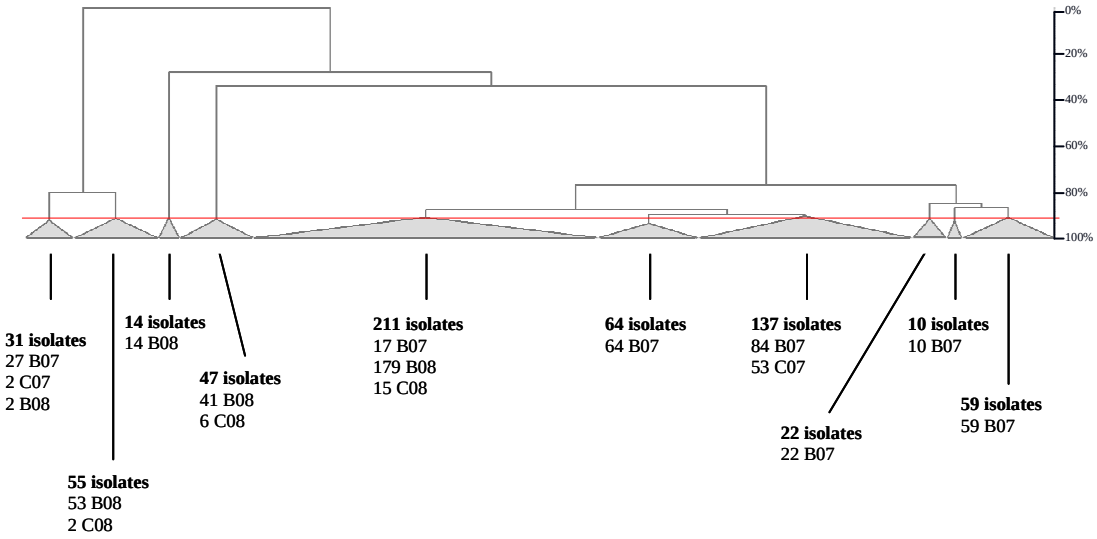
Intimin subtype of potential EPEC strains separated by phylogenetic groups A (n = 28), B1 (n = 617), B2 (n = 224), and D (n = 6). Intimin subtype types α-2 ( ), β ( ), ζ ( ), η ( ), θ ( ), ι ( ), κ ( ), ν ( ), and ξ ( ) are shown.

Figure 4.5



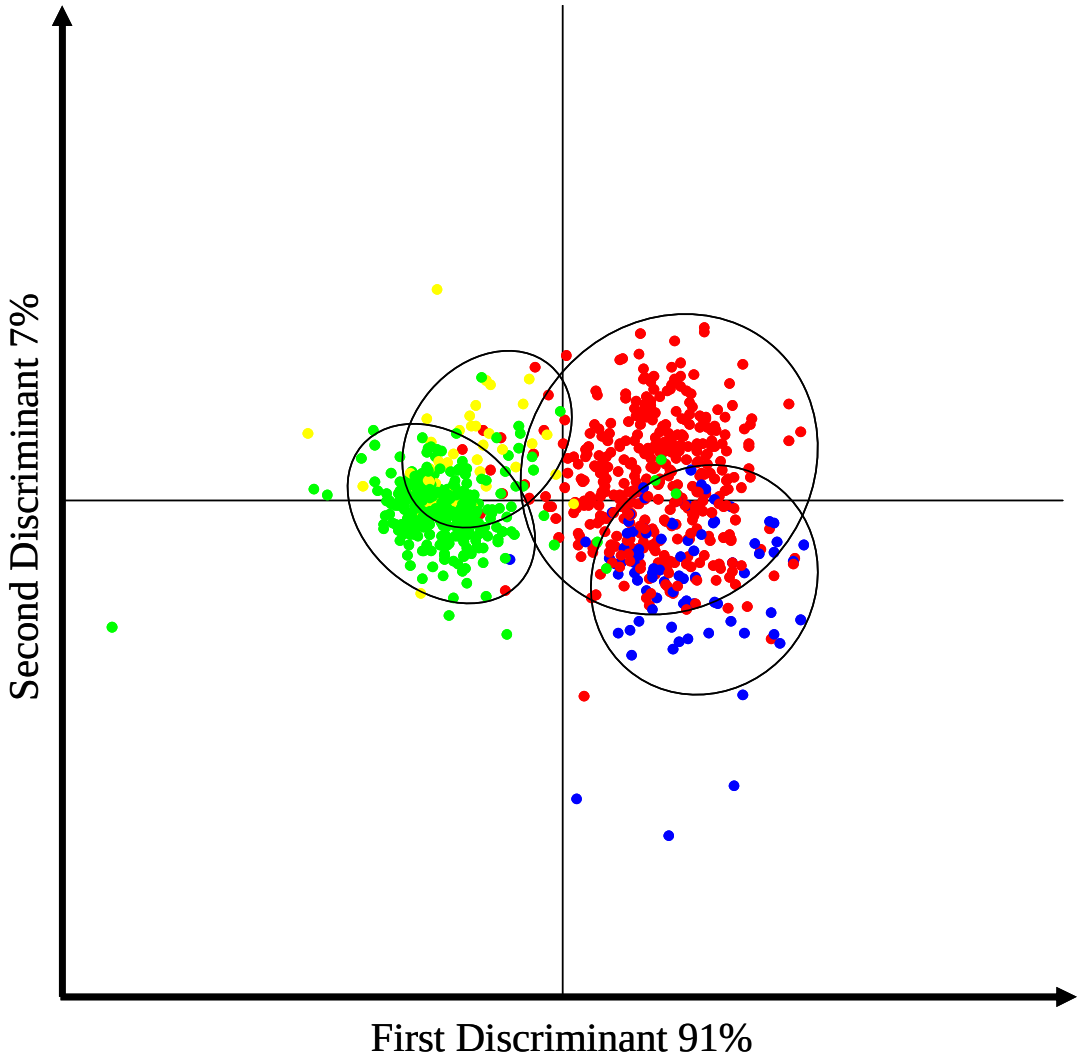
Dendrogram generated from HFERP fingerprint data obtained from potential EPEC isolates from Avalon Beach, CA. The dendrogram is collapsed at 40% similarity due to size constraints.

Figure 4.6



Dendrogram generated from HFERP DNA fingerprint data obtained from strains represented by 10 or more clonal isolates. The dendrogram is collapsed at 92% (—).

Figure 4.7



MANOVA analysis of all potential EPEC isolates from Avalon Beach, grouped by year and sample site. Site B and C isolates collected in 2007 are shown as blue (●) and red (●) circles, respectively. Site B and C isolates collected in 2008 are shown as green (●) and yellow (●) circles, respectively.

CHAPTER 5

Spatial and Temporal Distribution of *Escherichia coli* Populations in the Seven Mile Creek Watershed

OVERVIEW

Contamination of water with feces is a widespread public health problem. Water quality monitoring programs frequently use *Escherichia coli* as an indicator of fecal pollution, especially in freshwater systems. Several studies have reported the presence of naturalized *E. coli* strains that persist and potentially grow in the environment. This confounds the use of this bacterium as an indicator organism. In this study, I examined the fecal inputs and spatial and temporal distribution of *E. coli* in water and sediments of the Seven Mile Creek (SMC), a small man-made waterway in Nicollet County, MN. Results of this study indicated that *E. coli* counts varied considerably across sites and by dates and were likely affected by seasonal parameters, such as temperature. Host specific PCR assays indicated that cattle were the major contributors to the fecal loading of the SMC, although swine and poultry fecal markers were also sporadically detected. HFERP DNA fingerprint analysis indicated that the *E. coli* populations present in SMC were very diverse, but consisted of both transient and persistent strains. Persistent strains appeared to be naturalized to the environment, particularly in the sediments. Multivariate analysis of variance (MANOVA) showed that water and sediment isolates from a given year clustered together suggesting mixing of *E. coli* strains in the sediment and water column. *E. coli* populations, however, shifted from year to year. Isolates obtained during times of high flow conditions clustered together and low flow isolates clustered into distinct groups, suggesting that mixing and transport between sites occurs during high flow conditions. Taken together, results of this study suggest that both newly acquired and indigenous *E. coli* strains are present in

the SMC and this has obvious implications for water quality monitoring programs and for TMDL determinations.

INTRODUCTION

Fecal contamination of waterways is a widespread public health problem in the United States and throughout much of the world. Sources of fecal contamination include sewage and septic systems, livestock feedlots, wildlife, and run off from urban and agricultural land (76, 88). Currently, water monitoring programs assess the level of contamination at a given site by determining numbers of through fecal indicator bacteria (FIB), such as *Escherichia coli* and *Enterococcus* sp. The presence of these microorganisms in waterways is hypothesized to be due to contamination from human feces and the likely presence of fecal pathogens, such as *Salmonella*, *Shigella*, and gastrointestinal viruses. Thus, in the state of Minnesota and in most freshwater systems, *E. coli* are frequently used as the indicator organism for fecal contamination. Several studies have shown that elevated counts of FIB correlate with an increased risk of gastrointestinal disease after contact with contaminated water (28, 34, 44, 181).

Historically, *E. coli* have generally been thought of as a harmless commensal organisms found in the lower gut of warm blooded animals, although several pathogenic *E. coli* strains, such as EPEC or STEC, have been isolated (115). The use of *E. coli* as an indicator organism relies partly on the notion that it does not persist or grow in the environment, thus signaling the presence of recent contamination (46). Several studies, however, have reported on the persistence and potential growth of *E. coli* in water, sediment, and soil ecosystems, and in association with macrophytic algae (24-26, 50, 76, 77, 79, 95, 97, 122, 156, 179). Results from several of these studies have shown that specific strains of *E. coli* have become naturalized to or indigenous to these environments, and these strains may contribute to elevated *E. coli* counts through

the inoculation of water by runoff of *E. coli* from soils and sands, and through sediments mixing into the water column (76, 80, 179). The ability of naturalized strains to persist, grow and their impact on FIB counts has obvious effects on TMDL determinations and other water quality management plans.

Both seasonal and weather-related factors likely play large roles in the population dynamics of naturalized *E. coli,* as the ability to persist and grow in natural environments is affected by changes in temperature, moisture, salinity, organic matter content and predation (14, 23, 163, 164, 166). Several studies have shown that naturalized *E. coli* strains can survive winter temperature extremes and freeze thaw cycles, as well as the summer months in northern temperate climate soils (25, 77). The transport of naturalized strains from soils and to water and the resuspension of stream sediment-borne *E. coli* during high flow periods has also been documented (80).

In light of the studies described above, I examined the spatial and temporal changes in *E. coli* populations present in sediments and water in the Seven Mile Creek, a small constructed waterway in Nicollet County, MN. The purposes of this study was to: i) examine the spatial and temporal distribution of *E. coli* in water and sediments of the SMC, ii) determine sources of fecal contamination in the SMC by using PCR-based assays, iii) examine the persistence and transport of these bacteria in the SMC, and iv) use DNA fingerprinting analyses to examine the population structure of *E. coli* within water and sediments to determine if these bacteria are likely growing in this environment.

MATERIALS AND METHODS

Study Site Description

The site chosen for this study was the Seven Mile Creek (SMC) in Nicollet County, MN (Figure 1). The 24,551 acre watershed consists of 86% agricultural land, which is mostly used for corn and soybean production. About 20% of the watershed receives manure fertilizers each year. Other potential sources of fecal bacteria within the SMC watershed include 24 animal feedlots for cattle and swine. Most of the drainage entering the creek itself comes from three constructed ditches and two tile systems. The drainage ditches within watershed have intermittent flow and often do not contribute water to the creek after July. Water and sediment samples were collected from four sites within the SMC watershed, designated as SM1 - SM4 (Figure 1).

Sample Collection and Processing

Samples were collected at all four sites from July through October in 2008, and from April through June in 2009. Water samples were collected as previously described (21). Sediment samples were collected 2 cm below the water column. Samples were shipped on ice to the laboratory after collection for processing. Bacteria present in the samples were concentrated by membrane filtration prior to enumeration and DNA extraction. Approximately 200 to 600 ml of water from each sample was filtered through a 0.45 μm (47mm) cellulose ester membrane (Millepore, Billerica, MA). Sediment samples (10 g) were suspended in 95 ml of 0.1 M sodium phosphate buffer amended with 0.1% hydrolyzed gelatin (94) in a milk dilution bottles containing 10 g of 3 mm glass beads. Samples were vigorously agitated using a wrist action shaker for 30

minutes to release bacteria bound to sediment particles. Samples were allowed to settle for 30 min and 25-80 ml of the upper phase was filtered though 0.45 µm membranes as described above. Water and sediment filtrations were done in triplicate, or greater. Membrane filters for DNA extraction were stored frozen at -80C until further use. Filters for colony isolation were placed into a conical centrifuge tube containing 10 ml of phosphate buffered saline and 5 g of 3 mm glass beads. Bacteria were removed from the filters by gentle agitation for 30 min using a wrist action shaker. Glycerol was added to a final concentration of 10%, and bacterial preparations were stored frozen at -80C until use.

Enumeration of *E. coli*

Counts of *E. coli* in water and sediment samples were done using the Colilert® Quanti-tray 2000®, a most probably number based analysis, according to the manufacturer's instructions. Count data are expressed as MPN / 100 ml or MPN / g sediment (dry weight) for water and sediment samples, respectively.

***E. coli* isolation**

E. coli were isolated from concentrated water and sediment filter wash samples using modified mTEC medium as previously described (174, 188). Approximately 2-3 ml of each filter wash sample was spread-plated onto the surface of mTEC medium in 20 x 20cm Q-tray bioassay plates (Genetix, Boston, MA). Each plate contained 250 ml of modified mTEC medium. Plates were incubated at 37^{o} C for 2 hr, then at 42^{o} C for 16 hr. After incubation, plates were stored at 4^{o} C overnight to facilitate the

development of blue pigment in colonies, allowing for the differentiation of *E. coli* from other fecal coliform bacteria. Twenty-four well isolated blue colonies were hand-picked into 96 well microtiter plates containing 150 µl of Hogness modified freezing medium (HMFM). In some cases, less than 24 isolates were collected due to low numbers of *E. coli* present. *E. coli* were not collected from sediment samples obtained in April and May 2009 because *E. coli* were not detected in these samples. Microtiter plates were incubated at 37^{o} C overnight then stored frozen at -80C before use.

DNA Extraction

Total DNA was extracted from frozen membrane filters using a PowerSoil DNA kit (MO BIO, Carlsbad, CA) with slight modifications. Filters were finely chopped using a sterile razor blade and used as the input material for DNA extraction. Total DNA was eluted using 100 µl of sterile, nuclease free dH_2O. DNA concentrations were determined using an Eppendorf BioPhotometer (Eppendorf, New York, NY). DNA samples were diluted to 3 ng/µl in sterile, nuclease free dH_2O for use as templates in PCR reactions.

PCR analyses of environmental DNA samples

PCR analyses to done to detect the presence of DNA from *Bacteriodes* sp. and *Brevibacterium* sp. strains. The primer pairs HF183F/Bac708R, PF163F/Bac708R, LA35F/LA35R, and CowM3F/CowM3R were used to detect the presence of *Bacteroides* associated with humans, swine, poultry, and cattle, respectively (13, 38, 149). The primer pair LA35F/LA35R were used to detect the presence of poultry-

specific *Brevibacterium* sp. (V. J. Harwood, unpublished data). All PCR reactions were performed as previously described. Each reaction used 15 ng of environmental DNA as template. Positive control reactions were performed using template DNA isolated from human sewage or the feces of swine, chickens, or cattle. Negative controls consisted of PCR reactions without added template DNA. All samples were also tested using a general Bacteroides primer pair, AllBac296F/Bac708R, to test for PCR inhibitors present in the extracted DNA sample (100).

HFERP DNA fingerprinting

Horizontal, fluorophore enhanced, repetitive element PCR was performed using BOXA1R primers as previously described (88). Gel images were captured using a Typhoon 8600 Variable Mode Imager (GE Healthcare, Chalfont St. Giles, UK) and analyzed using Bionumerics (version 2.1) software (Applied Maths, Kortrijk, Belgium) as previously described (42, 88).

Statisitical Analysis

Dendrograms were constructed from the HFERP fingerprint data using the curve-based, Pearson's product-moment correlation coefficient and the unweighted pair group method with arithmetic means (UPGMA) clustering method. Isolates sharing 92% similarity or greater were defined as clonal (88). The Shannon index of diversity was calculated as previously described (27). The HFERP fingerprint data was also used to generate a binary band-matching character table which was analyzed using multivariate analysis of variance (MANOVA) to cluster isolates (27, 42, 77).

Clustering of isolates was also done using Jackknife analysis. ID bootstrap analysis (at $P = 0.9$), done using a Bionumerics script (http://www.applied-maths.com/bn/scripts/bnscripts.htm), was performed to identify the potential sources of *E. coli* isolates in SMC (42, 77).

RESULTS

E. coli counts

The number of *E. coli* in water a sediments samples at the Seven Mile Creek site was evaluated from mid-summer to fall in 2008 and from spring to mid-summer in 2009. The *E. coli* counts for individual water and sediment samples varied greatly by site and by date (Figure 5.2). Relative to current numbers, *E. coli* counts were consistently elevated in 2008. Of the 48 water samples collected in 2008, 34 (70.8%) exceeded the state standard of 126 CFU/100 ml. While *E. coli* were detected in all sediment samples at varying levels through the summer months, several samples from fall 2008 did not contain culturable *E. coli*. Moreover, samples from spring 2009 generally had lower counts than those from summer 2008 and in 2009, the number of *E. coli* in water and sediment samples began to increase in late spring and early summer. About 29% (20 of 68) of the water samples from 2009 exceeded the state standard, and all of these samples were obtained from May to August. *E. coli* was not detected in several of the sediment samples, especially those from earlier dates in 2009. Monthly averages for all count dates were calculated for each sample and are shown in Table 5.1.

Host Source-Specific PCR

PCR analysis was used to examine water and sediment DNA samples from the Seven Mile Creek for the presence of *Bacteroides* and *Brevibacterium* strains originating from human, bovine, swine, and poultry sources. This was done using host-source-specific DNA primers targeting these animal sources. Results of this analysis indicated that most samples (63%) from mid to late summer 2008 were positive for the bovine specific *Bacteriodes* marker, as were 100% of the samples from spring to summer in 2009. In contrast, the human fecal marker gene was not detected in any sample tested, and the markers indicative of poultry and swine fecal material were detected infrequently in the 2008 and 2009 samples. All samples were positive for a universal *Bacteroides* marker indicating that PCR inhibitors were not responsible for negative results. Host specific PCR results are summarized in Table 5.2.

HFERP DNA Fingerprinting to Determine Diversity of *E. coli* in the Watershed

DNA fingerprint analysis of *E. coli* strains collected from each sample was used to examine genetic diversity and to determine if growth of cells in water and sediments was involved in exceedances of state and national *E. coli* limits and persistence of strains. Analysis of the dendrogram generated from fingerprint data indicated that the *E. coli* populations present at the SMC were very diverse (the dendrogram is not shown due to size constraints). Overall, fingerprint similarity ranged from 5.21 to 100%, and the isolates could be divided into three large, diverse groups containing 24, 44, and 1028 isolates. The 1096 isolates examined were comprised of 335 different strains. Of these, 208 (62%) were represented by a single isolate, and the remaining 127 strains

were represented by between 2 to 50 isolates. The Shannon diversity index for all isolates was 5.04. The Shannon diversity indices for water and sediment samples isolates from 2008 and 2009 were 4.43, 4.06, 4.05, and 3.41, respectively.

Several strains were found in samples from both 2008 and 2009 and across different sampling sites and types, suggesting that these strains persist at the sites. A dendrogram showing the genetic relatedness of one of these strains is shown in Figure 5.3. DNA fingerprint analysis of the SMC isolates were compared to those obtained from a library of known source *E. coli* from Minnesota using ID bootstrap analysis. None of the SMC isolates were identified as originating from a specific source using this method.

MANOVA analysis of fingerprint data obtained from water and sediment isolates from both 2008 and 2009 was used to determine if isolates found at one site and during one year could be found at other sites during the same year and into the next season. Results in Figure 5.4 showed that isolates clustered well by year into overlapping groups of water and sediment isolates. The MANOVA explained 93% of the variation in two dimensions, showing that the isolates were fairly tightly clustered by dates and sites.

A MANOVA analysis was also used to compare water isolates obtained from high flow conditions (those in 2008) to the low flow conditions found in 2009. This analysis showed that high flow isolates clustered together regardless of site, while low flow isolates clustered into distinct groups containing isolates from a single site, with the exception of low flow isolates from sites 3 and 4, which clustered together (Figure 5.5). Jacknife analysis of the HFERP fingerprint data obtained from high and low flow

isolates provided a similar result, where high flow isolates were often assigned to other high flow groups (Table 5.3).

DISCUSSION

In this study, I examined the population structure of *E. coli* strains present water and sediment samples from the SMC. My results indicated that the water at the SMC is highly contaminated with *E. coli*, especially during the summer and fall months. Sediment samples had the highest *E. coli* counts in the summer months although count data indicated significant variability across sites and dates. Temperature changes may explain the failure to detect *E. coli* in several sediment samples in fall 2008 and spring 2009. In fall 2008, the average daily temperatures were around 15^{o} C near the SMC, whereas summer 2008 average temperature were consistently above 20^{o} F. Average daily temperatures were less than 4^{o} C for the earliest 2009 sample dates and increased to around 20^{o} C for the summer months. These results are similar to those previously reported, where soil *E. coli* counts from several sites in northern Minnesota were highest in the summer months and decreased over the winter, presumably due to temperature effects (77). The inherent variability of environmental samples and the patchy distribution of *E. coli* in the environment (10, 27, 77, 119, 185) were also a likely contributors to count variability. It is also important to note that the *E. coli* enumeration method used in this study requires the use of a culture technique and, therefore, isolates in a viable but not culturable (VBNC) state would not be detected. Previous studies have suggested that the VBNC state is important for bacteria to survive

extreme or unfavorable conditions (137) and may account for the survival of *E. coli* during low temperature conditions (77).

Host specific PCR assays for fecal bacteria showed that cattle were major contributors to the fecal loading of SMC (Table 5.2). More than 60% of the summer 2008 samples and 100% of 2009 samples were positive for a bovine-specific marker gene. The presence of several large scale cattle feedlots and manure amended agricultural fields throughout the watershed are likely sources for bovine fecal pollution which would reach the creek through run-off. The bovine marker was detected infrequently in the fall of 2008, which may be related to low water levels and a lack of run-off during this time.

Several samples were also positive for swine in both 2008 and 2009, although poultry specific markers were only detected in a few samples from 2009. Humans likely do not contribute to fecal loading at the SMC as evidenced by the lack of any positive samples. Since the vast majority the SMC watershed is agricultural land or otherwise undeveloped, the lack of the human-specific marker gene is not surprising. It is probable that wildlife and other sources also impact the fecal loading of SMC, although due to a lack of validated source tracking tools, we did not test for fecal bacteria from other sources.

Despite the prevalence of the bovine-specific fecal marker gene in the samples, none of the 1,096 *E. coli* isolated from SMC water and sediment samples was identified as originating from bovine using DNA fingerprint data and ID bootstrap analysis (at $P = 0.9$). In fact, none of the tested *E. coli* isolates matched with any of the known source *E. coli* isolates in a library obtained from cats, dogs, chickens, cows, deer, ducks, geese,

goats, humans, pigs, sheep, and turkeys. This suggests that the library used is likely not representative of *E. coli* strains found at this site (188). Failure to assign any of the *E. coli* collected in this study may be a result of temporal or geographic limitations of the known source library used for this analysis which was largely collected from 2000 to 2005 from across Minnesota (42, 88). A reference library containing *E. coli* isolates from livestock and wildlife obtained within the SMC watershed would be better than the current library, and may serve to identify some of the isolates as originating from sources within the watershed.

In addition to direct input of fecally-derived strains, *E. coli* inputs into SMC may also include soil and other environmental strains that enter the system via run-off from rain events. It has been previously shown that run-off from soil may contribute up to 19% of *E. coli* detected in some waterways (77). Unfortunately, rain events occurred sporadically through the sampling seasons and did not track well to our sampling dates; therefore, it is extremely difficult to gauge inputs from these sources. Obtaining additional samples collected immediately after run-off reaches the creek would be helpful to determine if the *E. coli* isolates collected after rain events are different than those present during periods without rain.

Results from HFERP DNA fingerprinting revealed the presence of very diverse *E. coli* populations at the SMC study site. The Shannon diversity indices for water and sediment isolates from 2008 and 2009 were much higher than those reported from the analysis of the ribotypes of *E. coli* isolated from cattle, horses, and humans (7). More than 60% of the strains were only detected once, suggesting that transient strains make up a large proportion of the *E. coli* found in SMC water and sediments. However,

several strains were detected in multiple samples, across months and years, and at different sites, suggesting that some strains in SMC can persist for extended periods of time or grow at the SMC sites. In some cases, especially in sediment samples, clonal isolates from a single strain were found many times in the same sample (Figure 5.3). Some sediment samples consisted almost entirely of isolates belonging to a single strain. This result suggests that these isolates were likely growing in the sediment or water, as the isolation method does not have an enrichment step that would allow for single isolates to multiply. Growth in the sediments and water of stream and lake ecosystems has been reported in the past (27, 76), and the results reported here support the idea that *E. coli* can become naturalized to these environments.

The diversity amongst *E. coli* isolates from SMC was also seen using a MANOVA analysis comparing isolates from water and sediment from both years of this study. Results of this analysis (Figure 5.4) showed that isolates clustered well with their respective groups and water and sediment isolates from a given year were more closely related to each other than to isolates from different years. The apparent similarity between sediment and water *E. coli* and the detection of identical strains from both sediment and water sources further supports the idea that mixing of sediments with the water column can contribute to elevated *E. coli* counts (80). Furthermore, these results indicate that while the total population of *E. coli* shifts from year to year, overlapping clusters provide further evidence that some strains can persist for extended periods.

An additional analysis of fingerprint data using MANOVA showed that isolates from high flow conditions clustered together (Figure 5.5) regardless of sampling site and those from low conditions clustered into distinct groups. Jackknife analysis of the

fingerprint data yielded similar results, where high flow isolates were assigned to the correct group at lower frequencies than those from low flow conditions. These results suggests that mixing of the water column and transport of these isolates downstream during high flow conditions was a dominant process affecting *E. coli* populations in SMC. This result is similar to that of Jamieson et al. which also showed the resuspension of tracer *E. coli* strains from sediments and downstream transport (80). During low flow conditions, water became stagnant at the sampling sites and evaporation caused the sites to no longer be physically connected to one another, which could have allowed for differentiation of *E. coli* populations at the respective sites. Interestingly, isolates from sites SM3 and SM4 clustered together during low flow conditions. The reason for this is not clear, but this result may be related to similar inputs, as these sites are relatively close to each other in the watershed.

In conclusion, in this study I report that the *E. coli* populations present in SMC are dynamic and are made up of both transient and persistent strains. Persistent strains are likely naturalized to the sites and may grow in the water and/or sediments. Dispersal of these strains in the water may increase *E. coli* counts in the absence of new fecal inputs, which has obvious implications for water quality monitoring and TMDL determinations. Seasonal effects such as temperature and flow conditions likely play important roles in the survival and transport of *E. coli* from site to site. Futures studies and continued sampling are necessary to conclusively determine the inputs and examine the long term dynamics of *E. coli* at the Seven Mile Creek watershed.

ACKNOWLEDGEMENTS

I would like to thank Ramya Chandrasekaran, Chris Brandsey, and Dan Norat for help with sample processing and John Ferguson for his assistance with statistical analyses. Scott Matteson and the field crew at Minnesota State University – Mankato deserve many thanks for sample collection and for providing the *E. coli* count data. I would also like to thank Jack Bovee and the Nicolett County, MN Soil and Water Conservation Board for the collection of physical data from the SMC and Adam Birr at for his help in organizing this project. I would also like to acknowledge the Minnesota Department of Agriculture for financial support of this project.

Table 5.1

Monthly average *E. coli* counts for water and sediment samples (shaded) collected from SMC. Average values for water samples and sediment samples are expressed as MPN / 100 ml water and MPN / 1 g sediment (dry weight), respectively.

	Average Monthly *E. coli* counts[1]							
	July 2008	**August 2008**	**Sept. 2008**	**October 2008**	**April 2009**	**May 2009**	**June 2009**	**July 2009**
SM1 - Water	235.2 (184.1)	319.3 (155.6)	282.2 (309.4)	123.5 (51.9)	41.2 (34.3)	66.2 (62.2)	520.9 (471.3)	280.2 (193.6)
SM2 - Water	192.9 (36.6)	1469.9 (1097.6)	210.4 (38.0)	100.3 (113.2)	3.4 (3.8)	24.4 (19.1)	165.2 (156.9)	1034.1 (1008.5)
SM3 - Water	118.8 (32.9)	460.3 (730.0)	58.9 (42.4)	139.8 (128.3)	5.9 (7.2)	138.5 (171.9)	186.9 (187.6)	928.6 (549.5)
SM4 - Water	771.5 (62.9)	196.1 (157.0)	1584.0 (1130.5)	479.4 (629.6)	4.5 (6.5)	22.7 (17.7)	276.0 (274.9)	82.7 (77.0)
SM1 - Sediment	1292.8 (1596.1)	130.9 (224.2)	779.2 (1137.9)	843.1 (1363.2)	17.0 (6.5)	1.2 (0.5)	6.1 (2.6)	552.6 (955.4)
SM2 - Sediment	4885.6 (6860.7)	54.3 (42.6)	8.5 (3.3)	2.8 (3.1)	8.0 (14.4)	2.8 (2.2)	220.0 (430.9)	92.8 (122.6)
SM3 - Sediment	27.8 (21.2)	22.9 (29.2)	68.9 (127.7)	1.0 (0.0)	15.9 (33.3)	8.7 (11.5)	17.8 (19.4)	310.5 (414.7)
SM4 - Sediment	15.9 (0.1)	22.8 (34.8)	15.2 (12.1)	5.8 (8.3)	0.1 (0.1)	8.1 (11.6)	1.5 (1.0)	8.9 (5.5)

[1] Values in parentheses refer to the standard deviation.

Table 5.2

Results of host species specific PCR assays for fecal bacteria originating from bovine, human, poultry and swine sources.

	Number of samples positive for host-specific markers[1]							
	Sediment				Water			
	Bovine Marker	**Human Marker**	**Poultry Marker**	**Swine Marker**	**Bovine Marker**	**Human Marker**	**Poultry Marker**	**Swine Marker**
7/23/2008	4 (100)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	0 (0)	0 (0)
8/13/2008	3 (75)	0 (0)	0 (0)	0 (0)	3 (75)	0 (0)	0 (0)	0 (0)
8/27/2008	1 (25)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
9/5/2008	0 (0)	0 (0)	0 (0)	1 (25)	0 (0)	0 (0)	0 (0)	0 (0)
9/22/2008	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
10/7/2008	0 (0)	0 (0)	0 (0)	0 (0)	1 (25)	0 (0)	0 (0)	1 (25)
4/1/2009	4 (100)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	0 (0)	0 (0)
4/15/2009	4 (100)	0 (0)	2 (50)	0 (0)	4 (100)	0 (0)	0 (0)	0 (0)
4/29/2009	4 (100)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	0 (0)	0 (0)
5/12/2009	4 (100)	0 (0)	0 (0)	1 (25)	4 (100)	0 (0)	0 (0)	1 (25)
5/26/2009	4 (100)	0 (0)	0 (0)	1 (25)	4 (100)	0 (0)	0 (0)	1 (25)
6/8/2009	4 (100)	0 (0)	0 (0)	1 (25)	4 (100)	0 (0)	0 (0)	1 (25)
6/23/2009	4 (100)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	0 (0)	0 (0)
7/7/2009	4 (100)	0 (0)	0 (0)	1 (25)	4 (100)	0 (0)	0 (0)	1 (25)
7/21/2009	4 (100)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	0 (0)	1 (25)

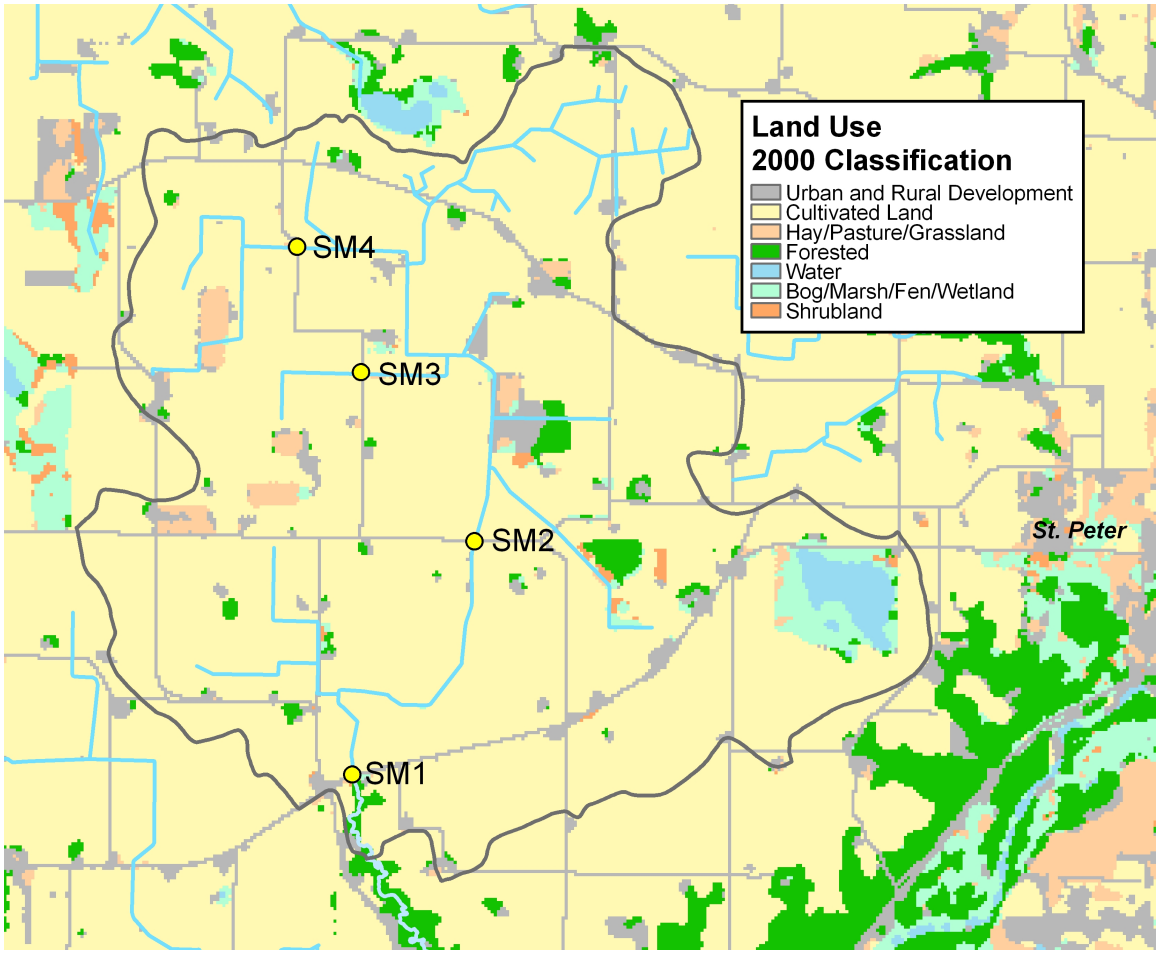
[1] Values in parentheses are percentages.

Table 5.3

Jackknife analysis of water isolates from high and low flow conditions at SMC. Values are expressed as percentages.

		Maximum similarities (%)							
		High Flow				Low Flow			
Group Assignment		**SM1**	**SM2**	**SM3**	**SM4**	**SM1**	**SM2**	**SM3**	**SM4**
High Flow	**SM1**	79.2	20.8	8.3	4.3	0	0	4.2	4.2
	SM2	12.5	33.3	29.2	8.7	0	0	0	0
	SM3	4.2	12.5	41.7	13	0	0	0	0
	SM4	0	8.3	8.3	56.5	8.3	4.2	0	0
Low Flow	**SM1**	0	12.5	8.3	4.3	91.7	0	4.2	4.2
	SM2	0	0	0	8.7	0	95.8	4.2	4.2
	SM3	0	4.2	0	4.3	0	0	83.3	4.2
	SM4	4.2	8.3	4.2	0	0	0	4.2	83.3

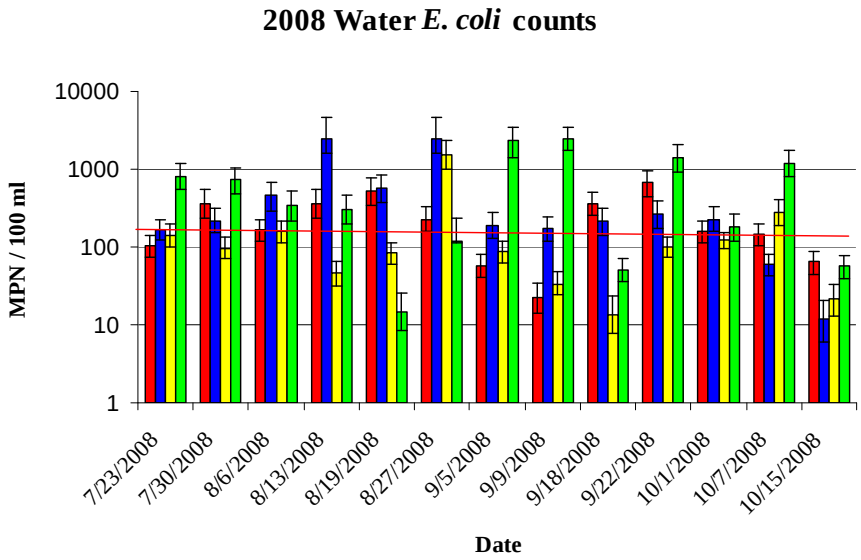
Figure 5.1



GIS land use map of the Seven Mile Creek watershed. Samples sites, SM1-4 are denoted by yellow circles. The creek flows from SM4 towards SM1.

Figure 5.2

A.


2008 Water E. coli counts
MPN / 100 ml
10000
1000
100
10
1
7/23/2008
7/30/2008
8/6/2008
8/13/2008
8/19/2008
8/27/2008
9/5/2008
9/9/2008
9/18/2008
9/22/2008
10/1/2008
10/7/2008
10/15/2008
Date

B.

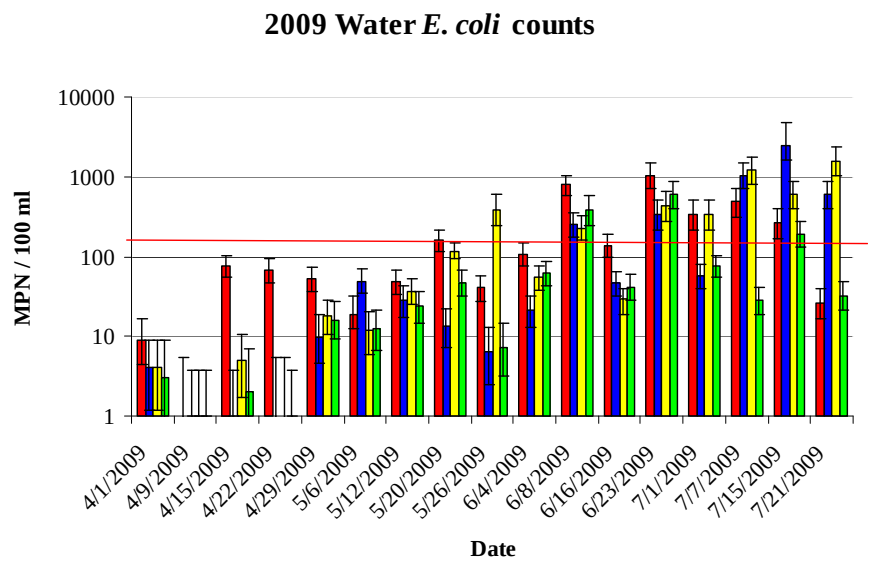
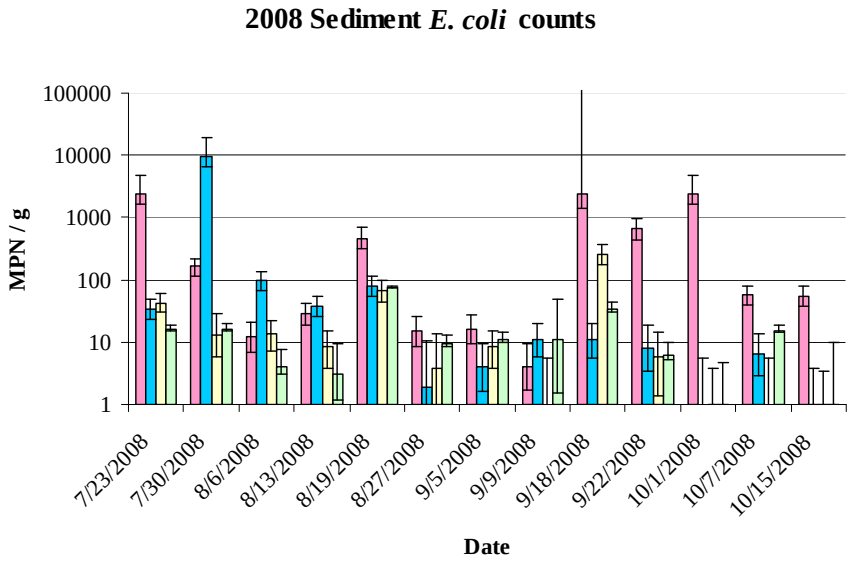
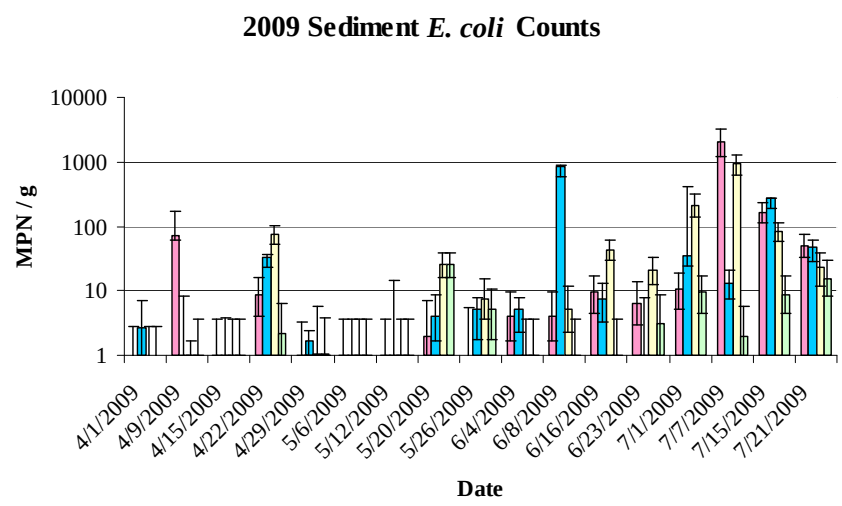

2009 Water E. coli counts
MPN / 100 ml
10000
1000
100
10
1
4/1/2009
4/9/2009
4/15/2009
4/22/2009
4/29/2009
5/6/2009
5/12/2009
5/20/2009
5/26/2009
6/4/2009
6/8/2009
6/16/2009
6/23/2009
7/1/2009
7/7/2009
7/15/2009
7/21/2009
Date

Figure 5.2 - continued

C.

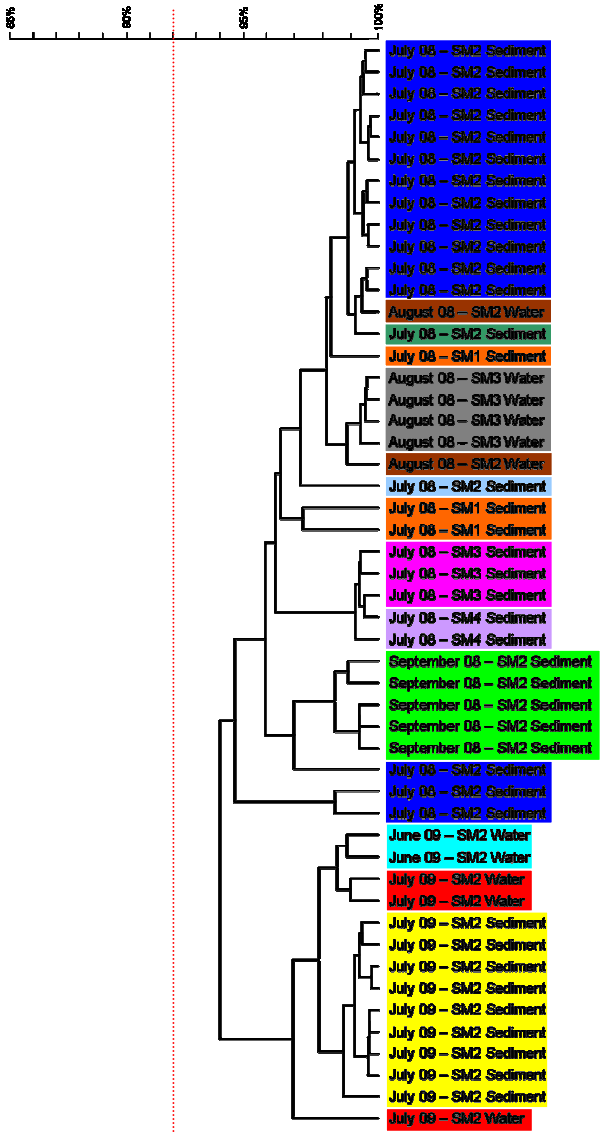


D.



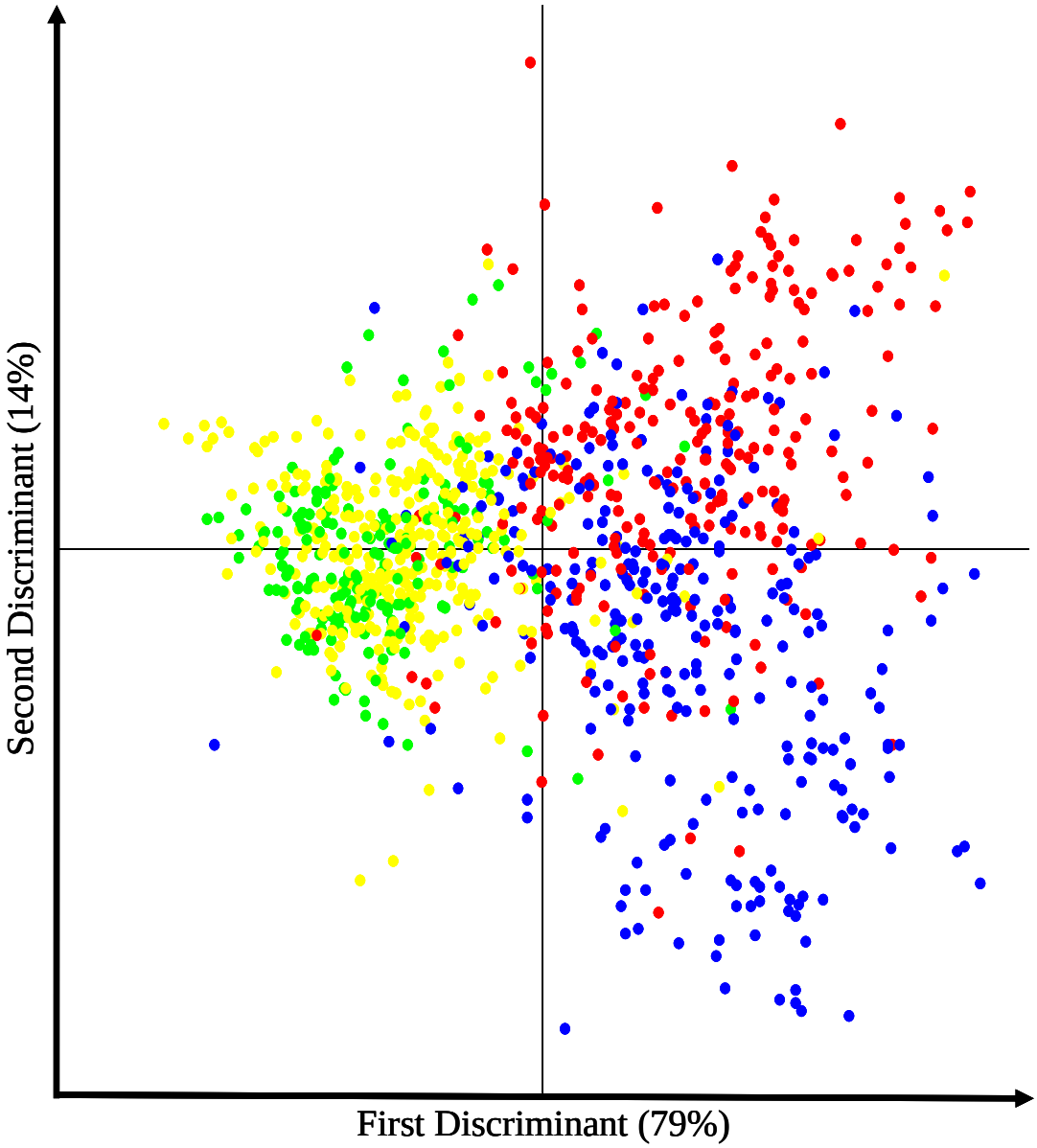
E. coli counts in water and sediment samples collected from July to October 2008 and April to July 2009. Water from 2008 and 2009 are shown separately in panels A and B, and sample sites SM1, SM2, SM3, and SM4 are represented as (■), (■), (■), (■), respectively. Sediment samples are shown in panels C and D and sample sites SM1, SM2, SM3, and SM4 are represented as (■), (■), (■), (■), respectively. The red line (-) represents the state standard of 126 CFU / 100 ml.

Figure 5.3



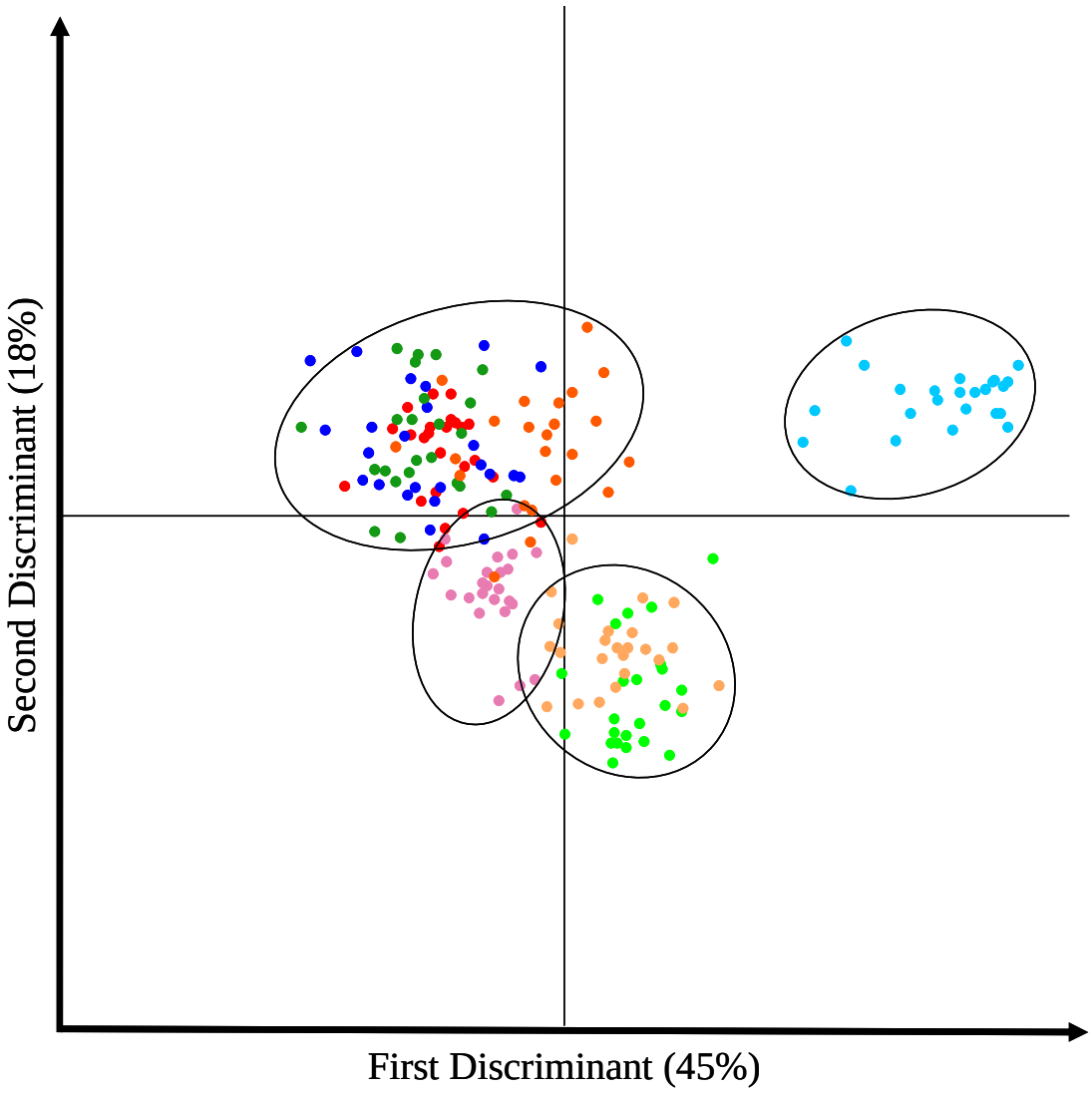
Dendrogram generated using HFERP DNA fingerprint data from isolates identified as a single strain. Similarity ranged from 95.67 to 99.83%. Sample date, location and type for individual isolates is shown on the right. Isolates from a single sample are grouped by color.

Figure 5.4



MANOVA analysis of all SMC *E. coli* isolates, grouped by year and sample type. 2008 water and sediment isolates are shown as (●) and (●), respectively. Water and sediment isolates from 2009 are represented as (●) and (●), respectively.

Figure 5.5



MANOVA analysis of high and low flow SMC *E. coli* isolates from water in 2008, grouped by flow condition and by site. High flow isolates from sites SM1, SM2, SM3, and SM4 are shown as (●), (●), (●), and (●), respectively. Low flow isolates from sites SM1, SM2, SM3, and SM4 are shown as (●), (●), (●), and (●), respectively.

CHAPTER 6

Conclusions and Future Recommendations

CONCLUSIONS

The overall goal of this thesis was to examine presence, genetic structure, virulence, and host origin source of *E. coli* in the environment through the development of a DNA marker system. In this thesis I determined the contribution of waterfowl to the fecal loading of contaminated waterways, investigated the presence of pathogenic *E. coli* strains at a California beach, and examined the populations of *E. coli* in the water and sediments of a small creek in south central Minnesota.

In Chapter 2, a genome subtraction method was used to identify DNA markers specific to *E. coli* originating from ducks and geese. These markers were tested using a library of known source isolates and, when combined, could correctly identify almost 75% of the goose and duck isolates tested. False positive rates were, on average, around 10%. One of the markers was found in nearly 50% of goose isolates and few isolates from other hosts. This marker corresponded to a putative gene homologous to several autotransporter adhesins. It is possible that this gene encodes an adhesin that allows for preferential colonization of the goose intestinal tract. Experimental evidence showed the gene conveyed an auto-aggregation ability, but the ecological role of the gene is yet unknown.

In Chapter 3, *E. coli* DNA marker genes specific to geese and ducks were used to examine several Minnesota lakes for the presence of waterfowl associated *E. coli* and estimates of the contribution of these hosts on total fecal loads. Through this work, we developed a high throughput colony hybridization screening assay utilizing a robotic system for arraying large numbers of colonies. The results of this study suggested that waterfowl are a major contributor of fecal loads in the lakes tested. This study also

showed that our screening assay is a cost effective and efficient method for examining large numbers of isolates for the presence of DNA marker genes.

Chapter 4 describes the examination of large numbers of *E. coli* isolates from contaminated beach water in California for the presence of the virulence genes *eaeA* (intimin), *stx1*, *stx2*, *STa* and *LT-I* and the EAF plasmid. The results of this study showed that *eaeA*$^{+}$ potential EPEC strains were present in beach water and, on several dates over 10% of the *E. coli* isolates present at the beach were potential EPEC. No *stx1/2* carrying isolates were detected, indicating that STEC strains were not present at the sampling sites. Almost all of the potential EPEC isolated in this study were atypical strains and belonged to several genetically diverse groups. Results of this study indicated potential EPEC strains were consistently detected at Avalon Beach, which may be a potential public health risk to recreational beach users.

In chapter 5, I examined *E. coli* populations in the sediment and water of Seven Mile Creek (SMC) in south central Minnesota. These studies showed that the *E. coli* counts varied considerably over time and are likely due to seasonal effects such as temperature and water flow. The populations were made up of both transient and persistent strains which are likely growing in the environment, particularly in the sediments. Sediment and water isolates are mixed suggesting that mixing of the sediments with the water column has a major influence on the structure of *E. coli* populations in inland waterways. The results of this study have implications on water quality monitoring programs and total maximum daily load (TMDL) determinations.

FUTURE DIRECTIONS

The research presented in this thesis has opened up several avenues for future studies of *E. coli* in the environment and of fecal contamination of waterways. These studies would increase the understanding of specific interactions between *E. coli* and waterfowl as a host, pathogenic strains in beach water, and the population dynamics of *E. coli* in the stream ecosystem.

While I was successful in identifying waterfowl specific DNA markers and these markers were useful for estimating the contribution of geese and ducks to fecal loads in several Minnesota lakes, the contribution from other hosts remains unknown. A combined approach using PCR based assays for human-, livestock-, and wildlife-specific fecal bacteria could be used to determine the impact of other hosts in the lakes tested. Additionally, the potential public health impacts of waterfowl fecal loads could be examined by a large scale study of potential pathogens carried by these hosts. This information could be used to better gauge when beach closures are necessary in lakes where waterfowl are the predominant contributors of fecal contamination.

Determining the role of the goose specific adhesion gene in colonization of the goose intestinal tract is also an exciting direction for future studies. While preliminary results indicate the gene encodes for auto-aggregation ability, colonization studies with wild type and mutant strains could be used to determine if the gene conveys an increased ability to attach to goose intestinal epithelium. This could be done using an *in vitro* assay with excised goose intestinal tissue, or it could be done *in vivo* using antibiotic-tagged or GFP-marked strains. Perhaps of more interest, an *in vivo* competition assay could be done to compare the adhesion abilities of a wild type strain

versus the mutant. The results of these studies would provide a better understanding of the ecological role of this gene and possibly explain why it is commonly found in waterfowl hosts and infrequently in other hosts.

Another interesting direction for future research is to further examine the potential EPEC isolates obtained from contaminated beach water in southern California. While these isolates were shown to carry the intimin virulence factor, the ability of these strains to cause disease is unknown. Although true virulence testing would be difficult as it requires human volunteers or an animal model for infection, screening for additional virulence factors, serotyping, and other tests could help to determine the likelihood that these strains could cause disease. Also, the library of nearly 25,000 beach water *E. coli* isolates could be tested for other virulence genes. These analyses would help to provide a better assessment of the public health risk associated with contaminated water contact.

The Seven Mile Creek study described in Chapter 5 is still in progress and sampling will continue, until at least October 2009. When the data collection is finished, a more complete analysis of the population dynamics will be possible. Also, samples can be tested for additional host specific DNA markers to determine the impacts of other fecal pollution sources, such as wildlife. Another interesting direction would be to isolate *E. coli* directly from feedlots and from manure-amended fields for comparison to SMC isolates. These results could help to explain why the ID bootstrap analysis failed to identify any of the SMC isolates. Finally, long term sampling at SMC to examine persistent strains would be helpful to determine the impact of naturalized strains at the SMC and the potential of these strains to elevate *E. coli* counts.

REFERENCES

1. **Ackers, M. L., B. E. Mahon, E. Leahy, B. Goode, T. Damrow, P. S. Hayes, W. F. Bibb, D. H. Rice, T. J. Barrett, L. Hutwagner, P. M. Griffin, and L. Slutsker.** 1998. An outbreak of *Escherichia coli* 0157:H7 infections associated with leaf lettuce consumption. J. Infect. Dis. **177:**1588–1593.

2. **Afset, J. E., E. Anderssen, G. Bruant, J. Harel, L. Wieler, and K. Bergh.** 2008. Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic *Escherichia coli* strains from a case-control study using multilocus sequence typing and DNA microarray analysis. J. Clin. Microbiol. **46:**2280-2290.

3. **Afset, J. E., K. Bergh, and L. Bevanger.** 2003. High prevalence of atypical enteropathogenic *Escherichia coli* (EPEC) in Norwegian children with diarrhea. J. Med. Microbiol. **52:**1015-1019.

4. **Agron, P. G., R. L. Walker, H. Kinde, S. J. Sawyer, D. C. Hayes, J. Wollard, and G. L. Andersen.** 2001. Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar enteritidis. Applied and environmental microbiology **67:**4984-4991.

5. **Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A. Lukyanov, E. D. Sverdlov, and D. E. Berg.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc. Natl. Acad. Sci. (USA) **95:**13108-13113.

6. **Alexander, L. M., A. Heaven, A. Tennant, and R. Morris.** 1992. Symptomatology of children in contact with sea water contaminated with sewage. J. Epidemiol. Commun. Health **46:**340-344.

7. **Anderson, M. A., J. E. Whitlock, and V. J. Harwood.** 2006. Diversity and distribution of *Escherichia coli* genotypes and antibiotic-resistant phenotypes in feces of humans, cattle, and horses. Appl. Environ. Microbiol. **72:**6914-6922.

8. **Andreoli, S. P., H. Trachtman, D. W. Acheson, R. L. Siegler, and T. G. Obrig.** 2002. Hemolytic uremic syndrome: epidemiology, pathophysiology, and therapy. Pediatr. Nephrol. **17:**293–298.

9. **Ansay, S. E., K. A. Darling, and C. W. Kaspar.** 1999. Survival of *Escherichia coli* 0157:H7 in ground-beef patties during storage at 2, -2, 15 and then -2^{o}C, and -20^{o}C. J. Food Prot. **62:**1243-1247.

10. **Atlas, A. M., and R. Bartha.** 1998. Microbial ecology: Fundamentals and applications, 4th ed. Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

11. **Benz, I., and M. A. Schmidt.** 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. Mol. Microbiol. **6:**1539-1546.

12. **Bernhard, A. E., and K. G. Field.** 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. **66:**1587-1594.

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

14. **Berry, C., B. J. Lloyd, and J. S. Colbourne.** 1991. Effect of heat shock on recovery of *Escherichia coli* from drinking water. Water Sci. Technol. **24:**85-88.

15. **Berry, E. D., and D. N. Miller.** 2005. Cattle feedlot soil moisture and manure content: II. Impact on *Escherichia coli* 0157. J. Environ. Qual. **34:**656-663.

16. **Bielaszewska, M., B. Middendorf, R. Köck, A. W. Friedrich, A. Fruth, H. Karch, M. A. Schmidt, and A. Mellmann.** 2008. Shiga toxin-negative attaching and effacing *Escherichia coli*: Distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. Clin. Infect. Dis. **47:**208-217.

17. **Black, R. E.** 1990. Epidemiology of traveler's diarrhea and relative importance of various pathogens. Rev. Infect. Dis. **12:**S73-S79.

18. **Bogosian, G., L. E. Sammons, P. J. Morris, J. P. O'Neil, M. A. Heitkamp, and D. B. Weber.** 1996. Death of the *Escherichia coli* K-12 strain W3110 in soil and water. Appl. Environ. Microbiol. **62:**4114-4120.

19. **Bokete, T. N., T. S. Wittam, R. A. Wilson, C. R. Clausen, C. M. O'Callahan, S. L. Moseley, T. R. Fritsche, and P. I. Tarr.** 1997. Genetic and phenotypic analysis of *Escherichia coli* with enteropathogenic characteristics isolated from Seattle children. J. Infect. Dis. **175:**1382-1389.

20. **Bollman, J., A. Ismond, and G. Blank.** 2001. Survival of *Escherichia coli* 0157:H7 in frozen foods: impact of the cold shock response. Intl. J. Food Microbiol. **64:**127-138.

21. **Bordner, R., J. A. Winter, and P. Scarpino.** 1978. Microbiological methods for monitoring the environment, water, and wastes. Report No. EPA 600/8-78-017. U.S. Environmental Protection Agency. Washington, D. C.

22. **Brettar, I., and M. G. Hofle.** 1992. Influence of ecosystematic factors on survival of *Escherichia coli* after large-scale release into lake water mesocosms. Appl. Environ. Microbiol. **58:**2201-2210.

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

24. **Byappanahalli, M. N., and R. S. Fujioka.** 1998. Evidence that tropical soil environment can support the growth of *Escherichia coli*. Water Sci. Technol. **38:**171-174.

25. **Byappanahalli, M. N., D. A. Shively, M. B. Nevers, M. J. Sadowsky, and R. L. Whitman.** 2003. Growth and survival of *E. coli* and enterococci populations in the macro-alga *Cladophora* (Chlorophyta). FEMS Microbiol. Ecol. **46:**203-211.

26. **Byappanahalli, M. N., R. L. Whitman, D. A. Shively, J. Ferguson, S. Ishii, and M. J. Sadowsky.** 2007. Population structure of *Cladophora*-borne *Escherichia coli* in nearshore water of Lake Michigan. Water Res. **41:**3649-3654.

27. **Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky, and S. Ishii.** 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. Environ. Microbiol. **8:**504-513.

28. **Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin.** 1982. Swimming-associated gastroenteritis and water quality. Am. J. Epidemiology **115:**606-616.

29. **Carson, C. A., B. L. Shear, M. R. Ellershiek, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Appl. Environ. Microbiol **67:**1503–1507.

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

31. **Cascales, E., S. K. Buchanan, D. Duché, C. Kleanthous, R. Lloubès, K. Postle, M. Riley, S. Slatin, and D. Cavard.** 2007. Colicin Biology. Microbiol. Mol. Biol. Rev. **71:**1092-2172.

32. **Centers for Disease Control and Prevention.** 1998. Outbreak of *Vibrio parahaemolyticus* infections associated with eating raw oysters -- Pacific Northwest, 1997. Morb. Mortal. Wkly. Rep. **47:**457-462.

33. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66:**4555-4558.

34. **Corbett, S. J., G. L. Rubin, G. K. Curry, and D. G. Kleinbaum.** 1993. The health effects of swimming at Sydney beaches. Am. J. Public Health. **83:**1701-1706.

35. **Davies-Colley, R. J., R. G. Bell, and A. M. Donnison.** 1994. Sunlight inactivation of enterococci and fecal coliforms within sewage effluent diluted in seawater. Appl. Environ. Microbiol.**:**2049-2058.

36. **Desmarais, T. R., H. M. Solo-Gabriele, and C. J. Palmer.** 2002. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. Appl. Environ. Microbiol. **68:**1165-1172.

37. **Díaz, E., A. Ferrandez, M. A. Prieto, and J. L. Garcia.** 2001. Biodegradation of aromatic compunds by *Escherichia coli*. Microbiol. Mol. Biol. Rev. **65:**523-569.

38. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. S. Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005. Host distributions of uncultivated fecal *Bacteroidales* reveal genetic markers for fecal source identification. Appl. Environ. Microbiol. **71:**3184-3191.

39. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. Appl. Environ. Microbiol. **71:**3179-3183.

40. **Dietzman, D. E., G. W. Fischer, and F. D. Schoenknecht.** 1974. Neonatal *Escherichia coli* septicemia-bacterial counts in blood. J. Pediatr. **85:**128–130.

41. **Djordjevic, S. P., V. Ramachandran, K. A. Bettelheim, B. A. Vanselow, P. Holst, G. Bailey, and M. A. Hornitzky.** 2004. Serotypes and virulence gene profiles of Shiga toxin-producing *Escherichia coli* strains isolated from feces of pasture-fed and lot-fed sheep. Appl. Environ. Microbiol. **70:**3910-3917.

42. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572–2577.

43. **Donnenberg, M. S., C. O. Tacket, S. P. James, G. Losonsky, J. P. Nataro, S. S. Wasserman, J. B. Kaper, and M. M. Levine.** 1993. Role of the *eaeA* gene in experimental enteropathogenic *Escherichia coli* infection. J. Clin. Invest. **92:**1412–1417.

44. **Dufour, A. P.** 1984. Health effects criteria for fresh recreational waters. Report No. EPA-600/1-84-004. United States Environmental Protection Agency. Washington, D. C.

45. **Escobar-Páramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. L. Bouguenec, and E. Denamur.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol. Biol. Evol. **21:**1085-1094.

46. **Field, K. G., and M. Samadpour.** 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. **41:**3517-3538.

47. **Fong, T. T., D. W. Griffin, and E. K. Lipp.** 2005. Molecular assays for targeting human and bovine enteric viruses in coastal waters and their application for library-independent source tracking. Appl. Environ. Microbiol. **71:**2070-2078.

48. **Franke, J., S. Franke, H. Schmidt, A. Schwarzkopf, L. H. Wieler, G. Baljer, L. Beutin, and H. Karch.** 1994. Nucleotide sequence analysis of enteropathogenic *Escherichia coli* (EPEC) adherence factor probe and development of PCR for rapid detection of EPEC harboring virulence plasmids. J. Clin. Microbiol. **32:**2460-2463.

49. **Fratamico, P. M., L. K. Bagi, E. J. Bush, and B. T. Solow.** 2004. Prevalence and characterization of Shiga toxin-producing *Escherichia coli* in swine feces recovered in the National Animal Health Monitoring System's Swine 2000 Study. Appl. Environ. Microbiol. **70:**7173-7178.

50. **Fujioka, R., C. Sian-Denton, M. Borja, J. Castro, and K. Morphew.** 1999. Soil: The environmental source of *Escherichia coli* and enterococci in Guam's streams. *J. Appl. Microbiol.* **85:**83S-89S.

51. **Gagliardi, J. V., and J. S. Karns.** 2002. Persistance of *Escherichia coli* 0157:H7 in soil and on plant roots. Environ. Microbiol. **4:**89-96.

52. **Geldreich, E. E.** 1976. Fecal coliform and fecal *Streptococcus* density relationships in water discharges and receiving waters. CRC Crit. Rev. Environ. Contam. **6:**349–369.

53. **Girard, V., and M. Mourez.** 2006. Adhesion mediated by autotransporters of Gram-negative bacteria: Structural and functional features. Res. Microbiol. **157:**407-416.

54. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiol. **147:**1079–1085.

55. **Gordon, D. M., O. Clermont, H. Tolley, and D. E.** 2008. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. Environ. Microbiol. **10:**2484-2496.

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

57. **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.

58. **Gunzburg, S. T., N. G. Tornieporth, and L. W. Riley.** 1995. Identification of enteropathogenic *Escherichia coli* by PCR-based detection of the bundle-forming pilus gene. J. Clin. Microbiol. **33:**1375-1377.

59. **Gyles, C. L.** 2007. Shiga toxin-producing *Escherichia coli*: An overview. J. Anim Sci. **85:**E45-E62.

60. **Hamelin, K., G. Bruant, A. El-Shaarawi, S. Hill, T. A. Edge, S. Bekal, M. J. Fairbrother, J. Harel, C. Maynard, L. Masson, and R. Brousseau.** 2006. A virulence and antimicrobial resistance DNA microarray detects a high frequency of virulence genes in *Escherichia coli* isolates from Great Lakes recreational waters. Appl. Environ. Microbiol. **72:**4200-4206.

61. **Hamilton, M. J., T. Yan, and M. J. Sadowsky.** 2006. Development of goose- and duck-specific DNA markers to determine sources of *Escherichia coli* in waterways. Appl. Environ. Microbiol. **72:**4012-4019.

62. **Harakava, R., and D. W. Gabriel.** 2003. Genetic differences between two strains of *Xylella fastidiosa* revealed by suppression subtractive hybridization. Applied and Environmental Microbiology **69:**1315-1319.

63. **Hardina, C. M., and R. S. Fujioka.** 1991. Soil: The environmental source of *E. coli* in enterococci in Hawaii's streams. Environ. Toxicol. Water Qual. **6:**185-195.

64. **Hartel, P. G., J. D. Summer, J. L. Hill, J. Collins, J. A. Entry, and W. I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273–1278.

65. **Hartel, P. G., J. D. Summer, and W. I. Segars.** 2003. Deer diet affects ribotype diversity of *Escherichia coli* for bacterial source tracking. Water Res. **37:**3263–3268.

66. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli*. Annu. Rev. Genet. **18:**31-68.

67. **Harwood, V. J., J. Whitlock, and V. H. Withington.** 2000. Classification of the antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical Florida waters. Appl. Environ. Microbiol. **66:**3698-3704.

68. **Hedberg, C. W., S. J. Savarino, J. M. Besser, C. J. Paulus, V. M. Thelen, L. J. Myers, D. N. Cameron, T. J. Barrett, J. B. Kaper, and M. T. Osterholm.** 1997. An outbreak of foodborne illness caused by *Escherichia coli* O39:NM, an agent not fitting into the existing scheme for classifying diarrheogenic *E. coli*. J. Infect. Dis. **176:**1625-1628.

69. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370-378.

70. **Higgins, J. A., K. T. Belt, J. S. Karns, J. Russell-Anelli, and D. R. Shelton.** 2005. *tir*- and *stx*-positive *Escherichia coli* in stream waters in a metropolitan area. Appl. Environ. Microbiol. **71:**2511-2519.

71. **Hill, G. A., and D. J. Grimes.** 1984. Seasonal study of a freshwater lake and migratory waterfowl for *Campylobacter jejuni*. Can. J. Microbiol. **30:**845-849.

72. **Hsieh, W.-J., and M.-J. Pan.** 2004. Identification *Leptospira santarosai* serovar shermani specific sequences by suppression subtractive hybridization. FEMS Microbiol. Lett. **235:**117-124.

73. **Huang, X., and A. Madan.** 1999. CAP3: A DNA sequence assembly program. Genome Res. **9:**868-877.

74. **Hussong, D., J. M. Damare, R. J. Limpert, W. J. Sladen, R. M. Weiner, and R. R. Colwell.** 1979. Microbial impact of Canada geese (*Branta canadensis*) and whistling swans (*Cygnus columbianus columbianus*) on aquatic ecosystems. Appl. Environ. Microbiol. **37:**14-20.

75. **Ingraham, J. L., and A. G. Marr.** 1996. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, D. C.

76. **Ishii, S., D. L. Hansen, R. E. Hicks, and M. J. Sadowsky.** 2007. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. Environ. Sci. Technol. **41:**2203-2209.

77. **Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. App. Environ. Microbiol. **72:**612-621.

78. **Ishii, S., K. P. Meyer, and M. J. Sadowsky.** 2007. Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of *Escherichia coli* strains from diverse human and animal sources. Appl. Environ. Microbiol. **73:**5703-5710.

79. **Ishii, S., T. Yan, D. A. Shively, M. N. Byappanahalli, R. L. Whitman, and M. J. Sadowsky.** 2006. *Cladophora* (Chlorophyta) spp. harbor human bacterial pathogens in nearshore water of Lake Michigan. Appl. Environ. Microbiol. **72:**4545-4553.

80. **Jamieson, R. C., D. M. Joy, H. Lee, R. Kostaschuk, and R. J. Gordon.** 2005. Resuspension of sediment-associated *Escherichia coli* in a natural stream. J. Environ. Qual. **34:**581-589.

81. **Janke, B., U. Dobrindt, J. Hacker, and G. Blum-Oehler.** 2001. A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. FEMS Microbiol. Lett. **199:**61-66.

82. **Janssen, P. J., B. Audit, and C. A. Ouzounis.** 2001. Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. Nucleic Acids Research **29:**4395-4404.

83. **Jarvis, L. J., R. H. Dowdy, D. L. Wyse, and D. D. Buhler.** 1991. Automation of atrazine and alachlor extraction from soil using a laboratory robotic system. Soil Sci. Soc. Am. J. **55:**561-562.

84. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J. Environ. Qual. **32:**305–309.

85. **Jerse, A. E., J. Yu, B. D. Tall, and J. B. Kaper.** 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. (USA) **87:**7839–7843.

86. **Jiang, S., R. Noble, and W. Chu.** 2001. Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. Appl. Environ. Microbiol. **67:**179-184.

87. **Jiménez-Clavero, M. A., C. Fernández, J. A. Ortiz, J. Pro, G. Carbonell, J. V. Tarazona, N. Roblas, and V. Ley.** 2003. Teschoviruses as indicators of porcine fecal contamination of surface water. Appl. Environ. Microbiol. **69:**6311-6315.

88. **Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Appl. Environ. Microbiol. **70:**4478-4485.

89. **Jones, T., C. O. Gill, and L. M. McMullen.** 2004. The behaviour of log phase *Escherichia coli* at temperatures that fluctuate about the minimum for growth. Lett. Appl. Microbiol. **39:**296-300.

90. **Kaiser, P. O., T. Riess, C. L. Wagner, D. Linke, A. N. Lupas, H. Schwarz, G. Raddatz, A. Schäfer, and V. A. Kempf.** 2008. The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. **10:**2223-2234.

91. **Kaper, J. B., J. P. Nataro, and H. L. T. Mobley.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2:**123-140.

92. **Khatib, L. A., Y. L. Tsai, and B. H. Olson.** 2002. A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. **59:**97-104.

93. **Khatib, L. A., Y. L. Tsai, and B. H. Olson.** 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. **63:**231-238.

94. **Kingsley, M. T., and B. B. Bohlool.** 1981. Release of *Rhizobium* spp. from tropical toils and recovery for immunofluorescence enumeration. Appl. Environ. Microbiol. **42:**241-248.

95. **Kinzelman, J., S. L. McLellan, A. D. Daniels, S. Cashin, A. Singh, S. Gradus, and R. Bagley.** 2004. Non-point source pollution: determination of replication versus persistence of *Escherichia coli* in surface water and sediments with correlation of levels to readily measurable environmental parameters. J. Water Health. **2:**103-114.

96. **Klemm, P., L. Hjerrild, M. Gjermansen, and M. A. Schembri.** 2004. Structure-function analysis of the self-recognizing antigen 43 autotransporter protein from *Escherichia coli*. Mol. Microbiol. **51:**283-296.

97. **Ksoll, W. B., S. Ishii, M. J. Sadowsky, and R. E. Hicks.** 2007. Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. Appl. Environ. Microbiol. **73:**3771-3778.

98. **Lamendella, R., J. W. S. Domingo, D. B. Oerther, J. R. Vogel, and D. M. Stoeckel.** 2007. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene. FEMS Microbiol. Ecol. **59:**651–660.

99. **Lauber, C. L., L. Glatzer, and R. L. Sinsabaugh.** 2003. Prevalence of pathogenic *Escherichia coli* in recreational waters. J. Great Lakes Res. **29:**301-306.

100. **Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Sayler.** 2006. Development of *Bacteroides* 16S rRNA gene Taqman-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl. Environ. Microbiol. **72:**4214-4224.

101. **Lee, M. L. T., F. C. Kuo, G. A. Whitmore, and J. Sklar.** 2002. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. (USA) **97:**9834-9839.

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

103. **Levine, M. M.** 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemmorrhagic, and enteroadherent. J. Infect. Dis. **155:**377-389.

104. **Levine, M. M., and R. Edelman.** 1984. Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: Epidemiology and pathogenesis. Epidemiol. Rev. **6:**31–51.

105. **Ley, V., J. Higgins, and R. Fayer.** 2002. Bovine enteroviruses as indicators of fecal contamination. Appl. Environ. Microbiol. **68:**3455-3461.

106. **Liu, L., T. Spilker, T. Coenye, and J. J. LiPuma.** 2003. Identification by subtractive hybridization of a novel insertion element specific for two widespread *Burkholderia cepacia* genomovar III strains. Journal of Clinical Microbiology **41:**2471-2476.

107. **Luo, M., Y. H. Wang, D. Frisch, T. Joobeur, R. A. Wing, and R. A. Dean.** 2001. Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (Fom-2). Genome **44:**154-162.

108. **Manges, A. R., J. R. Johnson, B. Foxman, T. T. O'Bryan, K. E. Fullerton, and L. W. Riley.** 2001. Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. N. Engl. J. Med. **345:**1007-1013.

109. **Mau, M., and K. N. Timmis.** 1998. Use of subtractive hybridization to design habitat-based oligonucleotide probes for investigation of natural bacterial communities. Appl. Environ. Microbiol. **64:**185-191.

110. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl Acad. Sci. (USA) **92:**1664–1668.

111. **McLellan, S. L., A. D. Daniels, and A. K. Salmore.** 2003. Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution using DNA fingerprinting. Appl. Environ. Microbiol. **69:**2587–2594.

112. **Mokady, D., U. Gophna, and E. Z. Ron.** 2005. Extensive gene diversity in septicemic *Escherichia coli* strains. J. Clin. Microbiol. **43:**66-73.

113. **Moore, D. F., V. J. Harwood, D. M. Ferguson, D. J. Lukasik, P. Hannah, M. Getrich, and M. Brownell.** 2005. Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed. J. Appl. Microbiol. **99:**618-628.

114. **Nataro, J. P.** 2006. Atypical enteropathogenic *Escherichia coli*: typical pathogens? Emerg. Infect. Dis. **12:**696.

115. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142-201.

116. **Natural Resources Defense Council.** 2008. Testing the Waters: A Guide to Water Quality at Vacation Beaches. Natural Resources Defense Council. New York, NY.

117. **Nguyen, R. N., L. S. Taylor, M. Tauschek, and R. M. Robins-Browne.** 2006. Atypical enteropathogenic *Escherichia coli* infection and prolonged diarrhea in children. Emer. Infect. Dis. **12:**597-603.

118. **Noble, R. T., J. F. Griffith, A. D. Blackwood, J. A. Fuhrman, J. B. Gregory, X. Hernandez, X. Liang, A. A. Bera, and K. Schiff.** 2006. Multitiered approach using quantitative PCR to track sources of fecal pollution affecting Santa Monica Bay, California. Appl. Environ. Microbiol. **72:**1604–1612.

119. **Nunan, N., K. Ritz, D. Crabb, K. Harris, K. Wu, J. W. Crawford, and I. M. Young.** 2001. Quantification of the *in situ* distribution of soil bacteria by large-scale imaging sections of undisturbed soil. FEMS Microbiol. Ecol. **37:**67-77.

120. **O'Brien, A. D., and R. K. Holmes.** 1996. Protein toxins of *Escherichia coli* and *Salmonella*, p. 2788-2802. *In* F. C. Neidhardt, R. C. III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd Edition ed. ASM Press, Washington, D. C.

121. **Ogden, I. D., D. R. Fenlon, A. J. A. Vinten, and D. Lewis.** 2001. The fate of *Escherichia coli* 0157 in soil and its potential to contaminate drinking water. Intl. J. Food Microbiol. **66:**111-117.

122. **Olapade, O. A., M. M. Depas, E. T. Jensen, and S. L. McLellan.** 2006. Microbial communities and fecal indicator bacteria associated with *Cladophora* mats on beach sites along Lake Michigan shores. Appl. Environ. Microbiol. **72:**1932-1938.

123. **Oomen, C. J., P. v. Ulsen, P. v. Gelder, M. Feijen, J. Tommassen, and P. Gros.** 2004. Structure of the translocator domain of a bacterial autotransporter. EMBO J. **23:**1257-1266.

124. **Parveen, S., N. C. Hodge, R. E. Stall, S. R. Farrah, and M. L. Tamplin.** 2001. Genotypic and phenotypic characterization of human and nonhuman *Escherichia coli*. Water Res. **35:**379-386.

125. **Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kaspar, K. M. Portier, and M. L. Tamplin.** 1997. Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola Bay. Appl. Environ. Microbiol. **63:**2607-2612.

126. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin.** 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65:**3142-3147.

127. **Paton, A. W., and J. C. Paton.** 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli hlyA*, rfbO111, and rfbO157. J. Clin. Microbiol. **36:**598-602.

128. **Paton, A. W., P. Srimanote, M. C. Woodrow, and J. C. Paton.** 2001. Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. Infect. Immun. **69:**6999-7009.

129. **Pina, S., M. Puig, F. Lucena, J. Jofre, and R. Girones.** 1998. Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses. Appl. Environ. Microbiol. **64:**3376-3382.

130. **Prüss, A.** 1998. Review of epidemiological studies on health effects from exposure to recreational water. Int. J. Epidemiol. **27:**1-9.

131. **Ramachandran, V., K. Brett, M. A. Hornitzky, M. Dowton, K. A. Bettelheim, M. J. Walker, and S. P. Djordjevic.** 2003. Distribution of intimin subtypes among *Escherichia coli* isolates from ruminant and human sources. J. Clin. Microbiol. **41:**5022-5032.

132. **Reents, H., R. Münch, T. Dammeyer, D. Jahn, and E. Härtig.** 2006. The Fnr regulon of *Bacillus subtilis*. J. Bacteriol. **188:**1103-1112.

133. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature. **406:**64-67.

134. **Rhodes, M. W., and H. Kator.** 1999. Sorbitol-fermenting bifidobacteria as indicators of diffuse human fecal pollution in estuarine watersheds. J. Appl. Microbiol. **87:**528-535.

135. **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. J. Water Health **1:**209-223.

136. **Rose, R. E.** 1988. The nucleotide sequence of pACYC177. Nucleic Acids Res. **16:**356.

137. **Roszak, D. B., and R. R. Colwell.** 1987. Survival strategies of bacteria in the natural environment. Microbiol. Rev. **51:**365-379.

138. **Roux, A., C. Beloin, and J. M. Ghigo.** 2005. Combined inactivation and expression strategy to study gene function under physiological conditions: Application to identification of new *Escherichia coli* adhesins. J. Bacteriol. **187:**1001-1013.

139. **Sadowsky, M. J., W. C. Koskinen, J. Seebinger, B. L. Barber, and E. Kandeler.** 2006. Automated robotic assay of phosphomonoesterase activity in soils. Soil Sci. Soc. Am. J. **70:**378-381.

140. **Sadowsky, M. J., R. E. Tully, P. B. Cregan, and H. H. Keyser.** 1987. Genetic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype-specific nodulation of soybeans. Appl. Environ. Microbiol. **53:**2624-2630.

141. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

142. **Santo Domingo, J. W., D. G. Bambic, T. A. Edge, and S. Wuertz.** 2007. *Quo vadis* source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. Water Res. **41:**359-352.

143. **Savageau, M. A.** 1983. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. Am. Nat. **122:**732?744.

144. **Scotland, S. M., H. R. Smith, T. Cheasty, B. Said, G. A. Willshaw, N. Stokes, and B. Rowe.** 1996. Use of gene probes and adhesion tests to characterize *Escherichia coli* belonging to enteropathogenic serogroups isolated in the United Kingdom. J. Med. Microbiol. **44:**438-443.

145. **Scott, T. M., T. M. Jenkins, J. Lukasik, and J. B. Rose.** 2005. Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. Environ. Sci. Technol. **39:**283-287.

146. **Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik.** 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. **69:**1089–1092.

147. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002. Microbial source tracking: current methodology and future directions. Appl. Environ. Microbiol. **68:**5796–5803.

148. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. D. Siciliano.** 2005. Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. Environ. Microbiol. **7:**249-259.

149. **Shanks, O. C., E. Atikovic, A. D. Blackwood, J. Lu, R. T. Noble, J. Santo Domingo, S. Seifring, M. Sivaganesan, and R. A. Haugland.** 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. Appl. Environ. Microbiol **74:**745-752.

150. **Shanks, O. C., J. W. S. Domingo, R. Lamendella, C. A. Kelty, and J. E. Graham.** 2006. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Appl. Environ. Microbiol. **72:**4054-4060.

151. **Sharma, S., P. Sachdeva, and J. S. Virdi.** 2003. Emerging water-borne pathogens. Appl. Microbiol. Biotechnol. **61:**424-428.

152. **Shaw, M. K., A. G. Marr, and J. L. Ingraham.** 1971. Determination of the minimal temperature for growth of *Escherichia coli*. J. Bacteriol. **105:**683-684.

153. **Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Anal. Chem. **68:**850-858.

154. **Simpson, J. M., J. W. S. Domingo, and D. J. Reasoner.** 2003. Microbial source tracking: State of the science. Environ. Sci. Technol. **36:**5280–5288.

155. **Smith, J. L., P. M. Fratamico, and N. W. Gunther.** 2007. Extraintestinal pathogenic *Escherichia coli*. Foodborne Pathog. Dis. **4:**134-163.

156. **Solo-Gabriele, H. M., M. A. Wolfert, T. R. Desmarais, and C. J. Palmer.** 2000. Sources of *Escherichia coli* in a coastal subtropical environment. Appl. Environ. Microbiol. **66:**230–237.

157. **Solomon, E. B., S. Yaron, and K. R. Matthews.** 2002. Transmission of *Escherichia coli* 0157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization. Appl. Environ. Microbiol. **68:**397-400.

158. **Soule, M., E. Kuhn, F. Loge, J. Gay, and D. R. Call.** 2006. Using DNA microarrays to identify library-independent markers for bacterial source tracking. Appl. Environ. Microbiol. **72:**1843–1851.

159. **Stacy-Phipps, S., J. J. Mecca, and J. B. Weiss.** 1995. Multiplex PCR assay and simple preparation method for stool specimens to detect enterotoxigenic *Escherichia coli* DNA during course of infection. J. Clin. Microbiol. **33:**1054-1059.

160. **Stewart, J., J. S. Domingo, and T. J. Wade.** 2007. Fecal Pollution, Public Health, and Microbial Source Tracking. ASM Press, Washington, D. C.

161. **Stoeckel, D. M., M. V. Mathes, V. Melvin, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. L. O. Brien, T. W. Fenger, M. Samadpour, K. Strickler, M. Kriston, and B. A. Wiggins.** 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. Environ. Sci. Technol. **38:**6109-6117.

162. **Szewzyk, U., R. Szewzyk, W. Manz, and K. H. Schleifer.** 2000. Microbiological safety of drinking water. Annu. Rev. Microbiol. **54:**81-127.

163. **Tassoula, E. A.** 1997. Growth possibilities of *E. coli* in natural waters. Int. J. Environ. Stud. **52:**67-73.

164. **Tate III, R. L.** 1978. Cultural and environmental factors affecting the longevity of *Escherichia coli* in histosols. Appl. Environ. Microbiol. **35:**925-929.

165. **Tennant, S. M., M. Tauschek, K. Azzopardi, A. Bigham, V. Bennett-Wood, E. L. Hartland, W. Qi, T. S. Whittam, and R. M. Robins-Browne.** 2009. Characterisation of atypical enteropathogenic *E. coli* strains of clinical origin. BMC Microbiol. **9:**117-128.

166. **Terzieva, S. I., and G. A. McFeters.** 1991. Survival and injury of *Escherichia coli*, *Campylobacter jejuni*, and *Yersinia enterocolitica* in stream water. Can. J. Microbiol. **37:**785-790.

167. **Torres, A. G., J. B. Kaper, A. G. Torres, X. Zhou, and J. B. Kaper.** 2003. Multiple elements controlling adherence of enterohemorrhagic *Escherichia coli* O157:H7 to HeLa cells. Infect. Immun. **71:**4985-4995.

168. **Torres, A. G., X. Zhou, and J. B. Kaper.** 2005. Adherence of diarrheagenic *Escherichia coli* strains to epithelial cells. Infect. Immun. **73:**18-29.

169. **Trabulsi, L. R., R. Keller, and T. A. T. Gomes.** 2002. Typical and Atypical Enteropathogenic *Escherichia coli*. Emer. Infect. Dis. **8:**508-513.

170. **Ufnar, J. A., S. Y. Wang, J. M. Christiansen, H. Yampara-Iquise, C. A. Carson, and R. D. Ellender.** 2006. Detection of the *nifH* gene of *Methanobrevibacter smithii*: a potential tool to identify sewage pollution in recreational waters. J. Appl. Microbiol. **101:**44-52.

171. **United States Environmental Protection Agency.** 1986. Bacteriological Water Quality Criteria for Marine and Fresh Waters. United States Environmental Protection Agency. Washington, D. C.

172. **United States Environmental Protection Agency.** 2000. National Water Quality Inventory: 1998 Report to Congress. United States Environmental Protection Agency. Washington, D. C.

173. **United States Environmental Protection Agency.** 2001. Protocol for developing pathogen TMDLs. United States Environmental Protection Agency. Washington, D. C.

174. **United States Environmental Protection Agency.** 2002. Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar. U. S. Environmental Protection Agency. Washington, D. C.

175. **United States Environmental Protection Agency.** 2002. The 2000 National Water Quality Inventory EPA-841-R-02-001. United States Environmental Protection Agency. Washington, D. C.

176. **Vogel, J. R., D. M. Stoeckel, R. Lamendella, R. B. Zelt, J. W. S. Domingo, S. R. Walker, and D. B. Oerther.** 2007. Identifying fecal sources in a selected catchment reach using multiple source-tracking tools. J. Environ. Qual. **36:**718–729.

177. **Waters, C. M., and G. M. Dunny.** 2001. Analysis of functional domains of the *Enterococcus faecalis* pheromone-induced surface protein aggregation substance. J. Bacteriol. **183:**5659-5667.

178. **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.

179. **Whitman, R. L., and M. B. Nevers.** 2003. Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. Appl. Environ. Microbiol. **69:**5555-5562.

180. **Whitman, R. L., M. B. Nevers, G. C. Korinek, and M. N. Byappanahalli.** 2004. Solar and temporal effects on *Escherichia coli* concentration at a Lake Michigan beach. Appl. Environ. Microbiol. **70:**4276-4285.

181. **Wiedenmann, A., P. Kruger, K. Dietz, J. M. Lopez-Pila, R. Szewzyk, and K. Botzenhart.** 2006. A randomized controlled trial assessing infectious disease risks from bathing in fresh recreational waters in relation to the concentration of *Escherichia coli*, intestinal enterococci, *Clostridium perfringens*, and somatic coliphages. Environ. Health Perspect. **114:**228-236.

182. **Wiggins, B. A.** 1996. Discriminant analysis of antibiotic resistance patterns in fecal *Streptococci*, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. **62:**3997-4002.

183. **Wiggins, B. A., P. W. 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. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. **69:**3399-3405.

184. **Williams, A. P., L. M. Avery, K. Killham, and D. L. Jones.** 2005. Persistence of *Escherichia coli* O157 on farm surfaces under different environmental conditions. J. Appl. Microbiol. **98:**1075-1083.

185. **Wollum, I. A. G., and D. K. Cassel.** 1984. Spatial variability in *Rhizobium japonicum* in two North Carolina (USA) soils. Soil Sci. Soc. Am. J. **48:**1082-1086.

186. **World Health Organization.** 2004. Water, Sanitation, and Hygiene Links to Health: Facts and Figures. World Health Organization. Geneva, Switzerland.

187. **Yan, T., M. J. Hamilton, and M. J. Sadowsky.** 2007. High-Throughput and Quantitative Procedure for Determining Sources of *Escherichia coli* in Waterways by Using Host-Specific DNA Marker Genes. Appl. Environ. Microbiol. **73:**890-896.

188. **Yan, T., and M. J. Sadowsky.** 2007. Determining sources of fecal bacteria in waterways. Environ. Monit. Assess. **129:**97-106.

189. **Yatsuyanagi, J., S. Shioko, Y. Miyajima, K. Amano, and K. Enomoto.** 2002. Characterization of atypical enteropathogenic *Escherichia coli* strains harboring the *astA* gene that were associated with a waterborne outbreak of diarrhea in Japan. J. Clin. Microbiol. **40:**294–297.

APPENDIX 1

Suppression Subtractive Hybridization Studies to Identify DNA Markers for *E. coli* Originating From Human and Cattle Sources

STUDY DESCRIPTION

In Chapter 2, I described the use of subtractive suppressive hybridization (SSH) to identify DNA markers specific for goose and duck *E. coli*. These markers were successful in identifying about 73 and 76% of *E. coli* isolates in our goose and duck library, respectively, and showed that the SSH was a promising method for identifying host specific markers. The eventual goal was to develop a suite of DNA markers that were capable of identifying *E. coli* originating from various sources in order to quantitatively determine the fecal loading of a contaminated waterway. To this end, I performed several SSH experiments to identify DNA markers specific for human and cattle *E. coli*.

In my initial studies to identify goose and duck specific markers, I used pooled genomic DNA from five strains originating from geese (tester strains) and subtracted five *E. coli* strains obtained from humans (driver strains) to identify those fragments specific to the geese *E. coli*. In a concurrent experiment, I used the same five human strains as tester DNAs and the five goose strains as driver DNAs (Hu-SSH1, Table A.1) to isolate human *E. coli* specific fragments. The SSH reactions were performed as described in Chapter 2, and a total of 192 subtraction library clones were picked to screen for specificity. These analyses identified 11 unique DNA marker fragments that were specific to the goose *E. coli* tester DNA pool used in the SSH reactions. Insert DNAs from the 11 cloned markers were subsequently used as probes in colony hybridization experiments to *E. coli* strains in the known source library used in Chapter 2, representing *E. coli* from 12 animal hosts and humans (42). While hybridization analyses indicated that seven of the marker DNAs identified a high percentage of

human *E. coli* isolates, they also cross hybridized with a significant numbers of *E. coli* strains from geese that were not included in the driver DNA pool and with strains from other animal sources. Conversely, four of the marker DNAs hybridized with few isolates from several animal hosts, but also reacted with a low percentage of human isolates. These results indicated that the strains used as the source of tester DNAs in the SSH experiment did not adequately encompass the genetic diversity of human *E. coli* strains, and likely that the strains used as driver DNAs lacked the diversity necessary to prevent significant cross hybridization with strains not included in the SSH experiment. It is important to note that the DNA fragments identified in this experiment were not found in the strains used for the driver, indicating that the subtraction experiment worked efficiently.

Additional SSH reactions were completed using pooled DNAs from 60 human *E. coli* strains as tester DNAs and pooled DNAs from 50 cow strains as driver (Hu-SSH2), and 60 human strains as tester DNA and 20 animal strains as driver (Hu-SSH3). Additionally, SSH reactions were performed to identify DNA markers for cow *E. coli* using 5 cow strains as tester DNAs and 20 non-cow strains as driver DNAs (Co-SSH1), 50 cow strains as tester DNAs and 60 human strains as driver DNAs (Co-SSH2), and 50 cow strains as tester DNAs with 20 non-cow strains as driver DNAs (Co-SSH3). The strains used for genomic DNA preparations are shown in Table A.1. The SSH reactions were performed as described in Chapter 2 and additional SSH reactions, done using the same tester and driver pools from Hu-SSH2, Hu-SSH3, Co-SSH2, and Co-SSH3 were also done (designated as HuSSH2*, Hu-SSH3*, Co-SSH2*, and Co-SSH3*, respectively) were repeated using minor modifications to the protocol to account for the

increased complexity of the reactions. This included increasing the primary and secondary hybridization times to 2 and 24 hours, respectively. Lastly, an additional SSH reaction was done using the products from Hu-SSH2 as tester DNAs and the driver DNAs from Hu-SSH3 (designated as Hu-SSH4*). Subtraction library construction, screening, and the subsequent colony hybridization analyses using the known source *E. coli* library were done as described in Chapter 2.

These additional SSH reactions yielded varying numbers of DNA fragments that were specific to the strains used as tester, and were not found in the strains used as driver (Table A.2). Thus, in all instances the SSH reactions were successful. In many cases, identical subtraction products were obtained in independent reactions with different tester and driver DNA pools. Unfortunately, these potential markers suffered from the same setbacks as described above, where significant cross hybridizations and/or low percentages of specific source isolates were observed to hybridize with the potential marker DNA fragments. The exception to this was a single human-specific marker that identified about 12% of human isolates and only 2, 5 and 2% of the chicken, goat, and sheep *E. coli* isolates, respectively. This probe failed to hybridize to *E. coli* from cats, cattle, deer, dogs, ducks, geese, horses, pigs, or turkeys. When sequenced, this marker DNA fragment had 100% nucleotide identity to a portion of the *E. coli colN* gene, which encodes for colicin-N, a bacteriocin produced by, and toxic to, some strains of *E. coli* (31). This marker gene fragment was identified in multiple isolates from several different SSH experiments, suggesting that the subtraction method was efficient at identifying tester-specific fragments, but the high levels of genetic

diversity amongst human *E. coli* strains likely prevented the identification of common marker carried by a high percentage of human strains.

The reason why the SSH method worked well in identifying goose and duck *E. coli* markers, but was largely unsuccessful in identifying markers for human or cow *E. coli* is not completely clear. The fact that identical subtraction products were obtained from several independent SSH reactions with different tester and driver DNAs indicates that the subtraction procedure was working properly. It is likely that the genetic diversity of *E. coli* in humans and cows is too great, and that a marker common to large percentages of *E. coli* from these hosts may be rare in the population examined. It is also possible that total populations of *E. coli* in geese and ducks are genetically closely related, and the common DNA sequences are shared between many isolates from these hosts.

Table A.1

Strains used in SSH experiments to identify human and bovine-specific *E. coli* markers

SSH Name	DNA Categories	Strains used for DNA pools
Hu-SSH1	Tester	Hu051, Hu130, Hu132, Hu188, Hu252
	Driver	Go066, Go090, Go126, Go172, Go206
Hu-SSH2	Tester	Hu001, Hu007, Hu018, Hu045, Hu055, Hu058, Hu078, Hu085, Hu090, Hu096, Hu100, Hu112, Hu119, Hu121, Hu124, Hu152, Hu153, Hu154, Hu155, Hu158, Hu160, Hu161, Hu163, Hu166, Hu177, Hu185, Hu199, Hu204, Hu205, Hu208, Hu212, Hu217, Hu220, Hu221, Hu226, Hu229, Hu243, Hu245, Hu248, Hu250, Hu254, Hu256, Hu257, Hu259, Hu262, Hu263, Hu266, Hu267, Hu270, Hu272, Hu277, Hu278, Hu280, Hu284, Hu286, Hu287, Hu297, Hu301, Hu303, Hu309
	Driver	K019, K023, K040, K045, K051, K054, K055, K059, K065, K074, K078, K083, K090, K092, K096, K098, K104, K109, K121, K124, K125, K127, K133, K141, K156, K158, K172, K175, K177, K179, K180, K181, K183, K185, K188, K190, K192, K203, K208, K210, K227, K251, K267, K274, K276, K301, K315, K316, K319, K324
Hu-SSH3	Tester	Hu001, Hu007, Hu018, Hu045, Hu055, Hu058, Hu078, Hu085, Hu090, Hu096, Hu100, Hu112, Hu119, Hu121, Hu124, Hu152, Hu153, Hu154, Hu155, Hu158, Hu160, Hu161, Hu163, Hu166, Hu177, Hu185, Hu199, Hu204, Hu205, Hu208, Hu212, Hu217, Hu220, Hu221, Hu226, Hu229, Hu243, Hu245, Hu248, Hu250, Hu254, Hu256, Hu257, Hu259, Hu262, Hu263, Hu266, Hu267, Hu270, Hu272, Hu277, Hu278, Hu280, Hu284, Hu286, Hu287, Hu297, Hu301, Hu303, Hu309
	Driver	K010, K051, K088, K104, K155, Go66, G090, G126, G172, G206, P026, P038, P044, P069, P104, C136, D105, CT058, H010, S105
Co-SSH1	Tester	K005, K090, K205, K238, K277
	Driver	C009, C061, C053, C104, C142, G030, H057, H089, Hu072, Hu149, Hu175, Hu231, Hu275, P013, P028, P050, P089, P106, S058, S067
Co-SSH2	Tester	K019, K023, K040, K045, K051, K054, K055, K059, K065, K074, K078, K083, K090, K092, K096, K098, K104, K109, K121, K124, K125, K127, K133, K141, K156, K158, K172, K175, K177, K179, K180, K181, K183, K185, K188, K190, K192, K203, K208, K210, K227, K251, K267, K274, K276, K301, K315, K316, K319, K324
	Driver	Hu001, Hu007, Hu018, Hu045, Hu055, Hu058, Hu078, Hu085, Hu090, Hu096, Hu100, Hu112, Hu119, Hu121, Hu124, Hu152, Hu153, Hu154, Hu155, Hu158, Hu160, Hu161, Hu163, Hu166, Hu177, Hu185, Hu199, Hu204, Hu205, Hu208, Hu212, Hu217, Hu220, Hu221, Hu226, Hu229, Hu243, Hu245, Hu248, Hu250, Hu254, Hu256, Hu257, Hu259, Hu262, Hu263, Hu266, Hu267, Hu270, Hu272, Hu277, Hu278, Hu280, Hu284, Hu286, Hu287, Hu297, Hu301, Hu303, Hu309
Co-SSH3	Tester	K019, K023, K040, K045, K051, K054, K055, K059, K065, K074, K078, K083, K090, K092, K096, K098, K104, K109, K121, K124, K125, K127, K133, K141, K156, K158, K172, K175, K177, K179, K180, K181, K183, K185, K188, K190, K192, K203, K208, K210, K227, K251, K267, K274, K276, K301, K315, K316, K319, K324
	Driver	C009, C061, C053, C104, C142, G030, H057, H089, Hu072, Hu149, Hu175, Hu231, Hu275, P013, P028, P050, P089, P106, S058, S067

Table A.2

Number of subtraction clones screened, unique potential markers identified, and host-specific marker DNAs. The markers identified in SSH reactions Hu-SSH2, Hu-SSH3, Hu-SSH3*, and Hu-SSH4* were identical fragments with sequence identity to the *E. coli colN* gene.

SSH Reaction	No. of Subtraction library clones screened	No. of unique potential markers	No. of host specific markers
Hu-SSH1	192	11	0
Hu-SSH2	576	10	1
Hu-SSH3	384	20	1
Hu-SSH3*	384	11	1
Hu-SSH4*	384	15	1
Co-SSH1	192	12	0
Co-SSH2	576	14	0
Co-SSH3	384	9	0
Co-SSH4	384	16	0
Co-SSH2*	384	10	0
Co-SSH3*	384	9	0

APPENDIX 2

Nucleotide Sequences

Nucleotide sequence of plasmid pMJH001.

GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGAATTCGATTTACGGCAGTGAG
AGCAGAGATAGCGCTGATGTCCGGCAGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGA
GGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTT
GGCAGCATCACCATAGCATCAATTGATCGGCAGGAGCAATCAATAACAACTATTGCGATTGTTTTTTATC
TATCATTCATCAACAATTGATCTGCAGTGTCGATCAATGGCGTTTTTCTGCTACAATACCACCATATAAA
ACAATGAGTTTTATATGTCATGATTAGTGTATTTTTTCGTCTGTATATTCAGAGAGTAAATGTATATTAA
TGAGGCGCTTCGATGAAACACCATAATAGACAAATAATGAAGTTTTCAGCTTTATACTCAGTAATGATTC
TGTCCGGAATGTCTTTTCCGTTTTGTTCCATTGCAGCAGGAGGAATGAATACAACACAGTATACATTTAA
TAGCGAGTATAATCTTACTGATGGTGATACCATCAAATTCACAGCTCAACCAAATGTGAGTACAGCAAAA
CTATCTATAAGTGGCAGTGGAATCCTGAATATTGATGGGAAAAACAGCCTGTTAACCTTTGATAATAGTG
GTGTCAAAGGACAGGGAGATATGCTTATTGCTGGTAAACTAAATATTACCAATGGGGGCGGTTTTAATGC
ATTTGGTACTGAACCAAGAATGCAGGTAGGCGTAGGTTCTGAAGGCAGCGTTATAGTTTCAGGAACGGAA
AGCTATCTGACTGTACCATATATAGTAGTAGGTAACACGGCCAAAGGTAATATGTCCATCAGTGATGGTG
CTGCTGTTACTCTGAATGCCCTGAAAGTGGGATATAAGGCAGATGGCGATGTTCTGGTATCAGGGAAAGG
AACATCTCTGACTGTAAGTTCTGATTCAGACACAATTCGTCTTGGTTTTTACGATGCCAATGGCAGACTA
ACTGTTAATAATAATGCTACCGTTACAGCACCATACATTGTTGTAGGTAATGATTTCAATAATCATTCAG
GAGAACTCATCATTGGTTCTGGCGTATCAGAACCTGCCTCTGAAGCCGGTATTATTAATGCAAATGAAAT
AAAATTCCGTAATAAGGGAATTTTAACACTCAATCACACCAATAATGATTTTAATCTGACGTCAGATTTA
TCAAGTAGTGGAGAATCAATTACAACAGGAGTATTTTCAGGAGATTATGGGCTTGTTCAGGCCGTCGCCG
GGACTACTATTCTTTCCGGCAATAATGAAAAATATAATGGTTCCATTAATATAGAAAATGGAGCTGGTAT
TGTTGTATCTGAACAGAAAAACCTGGGTACCTCAGTTGTTACTGATAATGGCTTGTTAACCATTGATACT
ACAACAGACTGGCAACTTACGAATGATGTCAGTGGTGGCGGTAATTTCCGTAAAACCGGCTCTGGTTCAC
TGACCGTTGGCAATAATGCTGCCTGGACCGGGCAGACTGATATTGATGCAGGCACACTGATTCTGGGTAA
TGCAGGCGCCCCTGTCATGCTTGCCAGCAGCCAGGTCAATATTGCAAAGGATGGTATTCTTACTGGCTTT
GGTGGTGTCTCCGGGAATGTGACCAACAGCGGTACTCTTGACCTGAGAGCTGATGCTCCGGGTAATGTGC
TGACTGTTGGCGGCAACTACACCGGTAATAACGGCACGCTGCTCATGAACACCGTTCTGGGAGATGACAG
CTCTGCAACGGATAAACTGGTGATTAAAGGTGATGCATCCGGCCAGACCCGTGTGGAAGTCACTAAGGCC
GGCGGTACAGGTGCACAGACACTCAATGGTATTGAACTGATTCATGTCGAAGGTAACGCCGACAGTGCTG
AATTTGTTCAGGCCGGTCGTATAGCTGCCGGTGCTTATGACTACACGCTGGGGCGTGGTCAGGGAAGCAA
CAGTGGTAACTGGTATCTGACCAGCGGTAAAAATACGCCGGAACCAACACCGACGCCTGACCCGGACTCT
AAACCTGAACCTGCACCGGGTGGTTATGATAATGATCTTCGTCCGGAAGCCGGTTCCTATACGGCCAATA
TGGCTGCAGTAAACACCATGTTCGTGACCCGTCTTCATGAACGTCTGGGGCCGATGCAGTACACCGATAT
CATGACCGGTGAGACGAAAAATACCAGCATGTGGATGCGCCACGAAGGGGGGCATAACCGCTGGCGTGAT
GGTACAGGCCAGCTGAAAACACAGGGTAACCGTTATGTTGTTCAACTGGGGGGAGACATTGCACAGTGGG
GCTGGGGAGAAACTGACCGCTGGCACCTTGGTGTTATGGCCGGTTACGGTAACGAGCATAACAACACGGA
CTCTGTGCGCACCGGATACCGCTCAAAAGGCAGTGTGAACGGATACAGCACAGGTCTGTATGCCACCTGG
TTTGCCAGTGATGAAACACATAACGGGGCTTATCTTGATACGTGGGCACAGTACGGCTGGTTTGACAACC
ATGTGAAAGGTGATGGACTGCCGGGCGAGTCCTGGAAATCAAAAGGCCTGACTGCTTCGCTGGAAACTGG
TTATACCTGGAAAGCAGGTGAGTTCAGCGGCAGCCATGGCAGCCTGAATGAATGGTATGTTCAGCCTCAG
GCACAGGTTGTCTGGATGGGTGTAAAAGCGGATGAACACCGTGAAAGCAACGGTACCCGTGTTGAGAACA
CCGGTGACGGTAACGTCCGTACCCGTCTGGGTGTAAAAACCTGGATTAAGGGACACAACAGAATGGACGA
CGGTAAATCCCGTGAGTTCCGTCCGTTCGTGGAGGTGAACTGGCTGCACAACACCCGTGAATTTGGTACC
CGCATGAACGGCGTGACGGTACATCAGGACGGTGCCCGCAATATCGGGGAAGTGAAAGCCGGTGTTGAAG
GGCAGATAAATGACCGTCTGAATCTGTGGGGTAATGTGGGTGTTCAGGCTGGTGACAAGGGGTACAGCGA
CACCTCAGCCATGCTGGGTGTGAAGTACACTTTCTGAGTACGGTGAATATCTGACAGTCGCGGAATCACT
AGTGCGGCCGCCTGCAGGTCGCCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTA
TAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACT
CACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA
ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGC
TGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA

ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTC
TTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGG
TCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCC
TGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT
TCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTG
GTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAA
AGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT
CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTT
TTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
GAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAAT
TGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAG
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTT
GGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA
TGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGG
AACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACC
CGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGC
GATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTG
GGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTA
TA

pGEM-5z(+) vector DNA is shown in yellow. The open reading frame corresponding to the putative goose specific adhesin gene is shown in red. Insert DNA flanking the ORF is shown with no highlight.

Nucleotide sequence of plasmid pMJH002.

GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGAATTCGATTTACGGCAGTGAG
AGCAGAGATAGCGCTGATGTCCGGCAGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGA
GGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTT
GGCAGCATCACCATAGCATCAATTGATCGGCAGGAGCAATCAATAACAACTATTGCGATTGTTTTTTATC
TATCATTCATCAACAATTGATCTGCAGTGTCGATCAATGGCGTTTTTCTGCTACAATACCACCATATAAA
ACAATGAGTTTTATATGTCATGATTAGTGTATTTTTTCGTCTGTATATTCAGAGAGTAAATGTATATTAA
TGAGGCGCTTCGATGAAACACCATAATAGACAAATAATGAAGTTTTCAGCTTTATACTCAGTAATGATTC
TGTCCGGAATGTCTTTTCCGTTTTGTTCCATTGCAGCAGGAGGAATGAATACAACACAGTATACACAACC
AATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATT
ATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGG
ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCT
CGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAG
CTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCA
ACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAAT
TACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATC
AGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCA
GGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCT
CATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCC
ATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCA
GCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCC
TTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTA
ACATCAGAGATTTTGCCTCAGGCACAGGTTGTCTGGATGGGTGTAAAAGCGGATGAACACCGTGAAAGCA
ACGGTACCCGTGTTGAGAACACCGGTGACGGTAACGTCCGTACCCGTCTGGGTGTAAAAACCTGGATTAA
GGGACACAACAGAATGGACGACGGTAAATCCCGTGAGTTCCGTCCGTTCGTGGAGGTGAACTGGCTGCAC
AACACCCGTGAATTTGGTACCCGCATGAACGGCGTGACGGTACATCAGGACGGTGCCCGCAATATCGGGG
AAGTGAAAGCCGGTGTTGAAGGGCAGATAAATGACCGTCTGAATCTGTGGGGTAATGTGGGTGTTCAGGC
TGGTGACAAGGGGTACAGCGACACCTCAGCCATGCTGGGTGTGAAGTACACTTTCTGAGTACGGTGAATA
TCTGACAGTCGCGGAATCACTAGTGCGGCCGCCTGCAGGTCGCCATATGGGAGAGCTCCCAACGCGTTGG
ATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG
TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAG
GCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACA
AAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGC
TGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTC
CAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA
TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCAC
CGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTAT
ATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC
AGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT

TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG
CCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGT
TCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCAC
CCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGC
CGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAG
GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGA
GAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATC
AGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAG
GGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCG
AAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGG
TGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGA
ACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCAC
GCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGC
TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG
TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGT
GAATTGTAATACGACTCACTATA

pGEM-5z(+) vector DNA is shown in yellow. The open reading frame corresponding to the putative goose specific adhesin gene is shown in red. The *kanR* gene inserted in the reverse orientation of the ORF to disrupt the gene is shown in green. Insert DNA flanking the ORF is shown with no highlight.