Evaluation, Development and Improvement of Genotypic, Phenotypic and Chemical Microbial Source Tracking Methods and Application to Fecal Pollution at Virginia's Public Beaches

Jerold W. Dickerson Jr.

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Crop and Soil Environmental Sciences

Charles Hagedorn III, Chair
Katharine F. Knowlton
Ann M. Stevens
Tamim Younos
Carl E. Zipper

July 29, 2008

Blacksburg, VA

Keywords: microbial source tracking, antibiotic resistance analysis, enterococci, fecal pollution, recreational waters, fluorometry

Evaluation, Development and Improvement of Genotypic, Phenotypic and Chemical Microbial Source Tracking Methods and Application to Fecal Pollution at Virginia's Public Beaches

Jerold W. Dickerson Jr.

Abstract

The microbial source tracking (MST) methods of antibiotic resistance analysis (ARA) and fluorometry (to detect optical brighteners in detergents) were used in the summers of 2004 and 2005 to determine the origins of fecal pollution at beaches with a past history of, or the potential for, high enterococci counts and posted advisories. At Hilton and Anderson beaches, ARA and fluorometry in the summer of 2004 detected substantial human-origin pollution in locations producing consistently high counts of *Enterococcus spp.* Investigations by municipal officials led to the fluorometric detection and subsequent repair of sewage infrastructure problems at both beaches. The success of these mitigation efforts was confirmed during the summer of 2005 using ARA and fluorometry, with the results cross-validated by pulsed-field gel electrophoresis (PFGE). Results at other beaches indicated that birds and/or wildlife were largely responsible for elevated enterococci levels during 2004 and 2005. The application of fluorometry proved difficult in opens waters due to high levels of dilution, but showed potential for use in storm drains.

An additional study developed and tested a new library-based MST approach based on the pattern of DNA band lengths produced by the amplification of the 16S-23S rDNA intergenic spacer region, and subsequent digestion using the restriction endonuclease MboI. Initial results from small known-source libraries yielded high average rates of correct classification (ARCC). However, an increase in the library size was accompanied by a reduction in the ARCC of the library and the method was deemed unsuccessful, and unsuitable for field application.

A final study focused on the potential for classification bias with disproportionate source category sizes using discriminant analysis (DA), logistic regression (LR), and k-nearest neighbor (K-NN) statistical classification algorithms. Findings indicated that DA was the most robust algorithm for use with source category imbalance when measuring both correct and incorrect classification rates. Conversely k-NN was identified as the most sensitive algorithm to imbalances with the greatest levels of distortion obtained from the highest k values.

Conclusions of this project include: 1) application of a validation set, as well as a minimum detectable percentage to known-source libraries aids in accurately assessing the classification power of the library and reducing the false positive identification of contributing fecal sources; 2) the validation of MST results using multiple methods is recommended for field applications; 3) fluorometry displayed potential for detecting optical brighteners as indicators of sewage leaks in storm drains; 4) the digestion of the 16S-23S rDNA intergenic spacer region of *Enterococcus* spp. using MboI does not provided suitable discriminatory power for use as an MST method; and 5) DA was the least, and k-NN the most, sensitive algorithm to imbalances in the size of source categories in a known-source library.

Acknowledgments

I would like to thank:

- First and foremost, Dr. Charles Hagedorn III for his immeasurable help and guidance throughout my time at Virginia Tech. And for sparking my interest in the field of environmental microbiology.
- My committee: Dr. Katharine Knowlton, Dr. Ann Stevens, Dr. Tamim Younos, Dr. Carl Zipper for their time reading and evaluating my work, as well their patience in dealing with several defense date postponements.
- Annie Hassall, Mike Saluta, Sue Herbein, and Dr. Brooks Crozier for their work and participation on this project. Without the help of these dedicated scientists this project would not have been possible.
- Greg Touchton, Tim Wade, Justin Evanylo, Paul Younos for their support as part of the Hagedorn lab.
- Michelle Monti and the Virginia Department of Health, Division of Zoonotic and Environmental Epidemiology for providing the funding for this research.
- Danny Barker of the Hampton Road Sanitation District and the Virginia Department of Conservation and Recreation for assistance in the collection of sewage and fecal samples.
- And finally, my parents for their constant love and encouragement.

Attribution

Several colleagues and coworkers aided in the research presented in several of the chapters of this dissertation. Brief descriptions of their contributions are included here.

Prof. Charles Hagedorn III - (Department of Crop and Soil Environmental Science, Virginia Tech) is the primary Advisor and Committee Chair. Dr. Hagedorn provided guidance and expertise from decades of research in the field of environmental microbiology. And as one of the founders of Microbial Source Tracking, he imparted insight into the direction of the project and its place in the advancement of the field. In addition, Dr. Hagedorn was invaluable in his assistance establishing the direction of all chapters, as well as the completion and eventual publications of Chapters IV and V.

Annie Hassall - (Department of Crop and Soil Environmental Science, Virginia Tech) was involved in all laboratory research conducted in Chapters IV and V. Annie aided in the implementation of laboratory protocols as the manager of the Hagedorn laboratory.

Prof. J. Brooks Crozier - (Department of Biology, Roanoke College) and his entire laboratory were instrumental in the development of the molecular source tracking method developed in Chapter V. Dr. Crozier provided knowledge of molecular microbiology surpassing that of all others on this project. Without his help this method and resulting publication would not have been possible.

Table of Contents

List of Tables

List of Figures

List of Pictures

List of Abbreviations

AFM	Average Frequency of Misclassification
ARA	Antibiotic Resistance Analysis
ARCC	Average Rate of Correct Classification
BEACH	Beaches Environmental Assessment and Coastal Health
BMP	Best Management Practice
BP	Base Pair
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
CFU	Colony Forming Units
CHEF	Contour-clamped Homogeneous Electric Field
CICEET	Cooperative Institute for Coastal and Estuarine Environmental Technology
CRIS	Current Research Information System
CUP	Carbon Utilization Profiling
CWA	Clean Water Act
CZM	Coastal Zone Management
DA	Discriminant Analysis
DDI	Distilled De-Ionized
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide Triphosphate
dsDNA	Double-Stranded DNA
ECC	Estimate of Correct Classification
ERIC	Enterobacterial Repetitive Intergenic Consensus
esp	Enterococcal Surface Protein
FB-28	Fluorescent Brightener 28
FC	Fecal Coliforms
FIB	Fecal Indicator Bacteria
FIGE	Field Inversion Gel Electrophoresis
FRET	Fluorescence, Resonance Energy Transfer
FS	Fecal Streptococci
GC	Gas Chromatography
HBSA	Human Bifid Sorbitol agar
HPLC-MS	High Performance Liquid Chromatography
IGS	Intergenic Spacer
ISR	Intergenic Spacer Region
k-NN	k-Nearest Neighbor
LH-PCR	Length Heterogeneity PCR
LR	Logistic Regression
MAR	Multiple Antibiotic Resistance

MC	Method Comparison
MDP	Minimum Detectable Percentage
mRNA	Messenger RNA
MS	Mass Spectrometry
MST	Microbial Source Tracking
NOAA	National Oceanic and Atmospheric Administration
NRDC	National Resources Defense Council
OB	Optical Brightener
PCR	Polymerase Chain Reaction
PFGE	Pulsed-Field Gel Electrophoresis
PU	Palindromic Unit
QA/QC	Quality Assurance/Quality Control
q-PCR	Quantitative PCR
RCC	Rate of Correct Classification
rDNA	Ribosomal DNA
rep-PCR	Repetitive Sequence PCR
RFLP	Restriction Fragment Length Polymorphism
RGE	Rotating Gel Electrophoresis
RNA	Ribonucleic Acid
Rnase	Ribonuclease
rRNA	Ribosomal RNA
SCCWRP	Southern California Coastal Water Research Project
SD	Storm Drain
SDS	Sodium Dodecyl Sulfate
ssDNA	Single-Stranded DNA
T-RFLP	Terminal Restriction Fragment Length Polymorphism
T_a	Annealing Temperature
T_m	Melting Temperature
TAFE	Transverse-Alternation Field Gel Electrophoresis
TMDL	Total Maximum Daily Loads
TSA	Tryptic Soy Agar
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
VDH	Virginia Department of Health
VS	Validation Set
WERF	Water Environment Research Foundation

Evaluation, Development and Improvement of Genotypic, Phenotypic and Chemical Microbial Source Tracking Methods and Application to Fecal Pollution at Virginia's Public Beaches

Jerold W. Dickerson Jr.

I. Introduction

Water contaminated with feces has the potential to negatively affect both human health and the economies of local communities. It has long been known that water containing a high number of fecal bacteria is likely to contain pathogens (Stevenson 1953) and, exposure to this water can cause outbreaks of illness. Economically, locations reputed for having unclean water can lead to declining visitors, and potentially severe economic losses in tourist-driven areas (Rabinovici et al., 2004). Similarly, businesses that rely on clean water, such as the shellfish industry, will also suffer financial losses if beds are forced close due to fecal contamination. In the interest of protecting human health, various federal agencies have published limits for the number of fecal bacteria permitted in fresh and marine waters (USEPA, 2002). Based on these guidelines, states have worked to develop standards for the number of fecal bacteria permissible in its waters, typically varying with use (e.g., recreation). In addition, states and local communities routinely monitor fresh and marine waters to ensure that these standards are upheld.

Environmental waters which fail to meet state standards for fecal bacteria are out of compliance with the 1972 Federal Clean Water Act, and are eligible for funds for a Total Maximum Daily Load (TMDL) analysis. A TMDL establishes the maximum pollutant load a water body can receive and still meet water quality standards. Although a TMDL often includes the analyses of multiple pollutants (e.g. metals, sediments), one frequent component is the identification of the animal sources contributing to an elevated bacterial load. Specific knowledge of fecal sources allows for the most efficient use of best management practices (BMPs) in reducing fecal counts and aids in efforts to bring waters into compliance with state regulatory standards. These TMDL analyses are currently the single greatest force driving the development of Microbial Source Tracking (MST) methods in the United States.

The underlying principle of MST is that specific markers or strains of bacteria are uniquely associated with a single, or related group (e.g. livestock or wildlife), of animal species (Scott et al., 2002). Beginning in the mid-nineties with the development of several genotypic (DNA-based) methods to identify the host origin of *Escherichia coli* (Samadpour and Chechowitz, 1995; Simmons et al., 1995), MST has expanded considerably, procuring of a wide array of genotypic and phenotypic (based on characteristics expressed by the bacterium) methods. Most MST research on genotypic methods has concentrated on ribotyping (Parveen et al., 1999), pulsed-field gel electrophoresis (Simmons et al., 1995), and various polymerase chain reaction (PCR)-based (Dombek et al., 2000) methods, while most research into phenotypic methods has focused on antibiotic resistance analysis (ARA; Wiggins et al., 1999) or carbon utilization profiling (Hagedorn et al., 2003). Other methods, still largely in development, do not rely on bacteria at all, but instead detect chemicals associated with bacterial contamination. For example, human fecal pollution has been identified in natural waters with the chemical method, fluorometry, which measures levels of optical brighteners (OBs) found in laundry detergents thereby identifying human wastewater contamination (Hagedorn et al., 2005).

The adverse health and economic effects of water contaminated with feces, as well as the need for TMDL analyses, has meant that MST is consistently identified as a high priority need by a variety of stakeholders. On a federal level, the U. S. National Oceanic and Atmospheric Administration (NOAA) and United States Environmental Protection Agency (USEPA) have consistently supported MST research. The Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET), supported by NOAA, identified the development of "novel and cost-effective technologies and methods to identify sources of microbial contaminants" as one of its four funding priorities in FY 2005. The USEPA, in conjunction with

other groups such as the Southern California Coastal Water Research Project (SCCWRP) or the Water Environment Research Foundation (WERF), has, and continues to offer, workshops on MST methods. Because MST research has led to the development of advanced methods to detect bacterial pathogens, it belongs to one of the three priority areas identified by the American Society for Microbiology at the United States Department of Agriculture – Cooperative State Research, Education, and Extension (CSREES) Stakeholders Workshop (USDA-CSREES, 2002) and USDA also supports a Current Research Information System (CRIS) Multistate project (S-1022). On a state level, both Water Research Institutes and Coastal Zone Management (CZM) Programs consistently fund research on MST. The CZM Program in Georgia identified "refinement of bacterial analysis of coastal water" as one of its four funding priorities in FY 2005-2006. MST is considered Delaware's number one research priority for environmental waters (S. Myoda, pers. comm.). The results of these combined forces have driven MST researchers across the U.S. to explore a wide-array of methods for identifying fecal sources.

Although all MST methods have their advantages and disadvantages with respect to reproducibility, discriminatory power, ease of interpretation, ease of performance, and cost (Scott et al., 2002), studies detailing the refinement of old and development of new methods continue to be reported. Two country-wide method comparison (MC) studies have been undertaken in the U.S. to date, with preparation for a third currently underway. The first study was sponsored by the Southern California Coastal Water Research Project (SCCWRP) and tested the ability of both library-dependent and library-independent methods to correctly identify sources present in blind, spiked water samples. In the second study, sponsored by the United States Geological Survey (USGS), researchers were provided with individual isolates to identify the animal host using only library-dependent methods. In both studies correct classification rates

were generally below levels reported in the literature, with false positive identifications high amongst most methods (Stewart et al., 2003; Stoeckel et al., 2004). Both studies concluded that more research is needed to refine methods, and because much of the fecal pollution is detected in poor communities unable to afford many MST technologies, decreasing method costs should remain a priority. Finally, from a practical standpoint, there is evidence that nearly 70% of water quality infractions are single-day events (Leecaster and Weisberg, 2001). Considering that the public is typically notified 24 to 48 hours after samples are collected, most warnings to the public about fecal contamination are already out-of-date when they are issued. Thus, MST methods producing more rapid results continue to be sought by regulators. The results of a third, European-based MST study were reported in 2006 (Blanch et al., 2006). This study, sponsored by the European Union focused solely on library-independent methods, with no single method emerging as superior in attempts to identify human or livestock fecal pollution.

An U.S. EPA MST Guidance document (USEPA, 2005) published after both U.S.-based MC studies provided primarily a synopsis of established source-tracking work (both laboratory and field) for new inductees into the field. Highlights of the document include sections on method options, limitations, and field applications. Although some specific guidelines were established, not included were the recommendations of the MC studies favoring the use of a challenge (or validation) set of isolates over the traditional correct classification rates of library isolates, as a guide for determining how well a library represents the diversity of a given bacterial species in a specific watershed; and thus how successful it will be at identifying isolates collected from environmental water samples.

A second factor driving the development of MST in coastal and Great Lakes states has been the passage of amendments to the Clean Water Act (CWA) known as the federal BEACH

(Beaches Environmental Assessment and Coastal Health) Act in October 2000. The federal BEACH Act included three major provisions: a) CWA amendment to section 303(i) requires that states have to adopt the standards of the EPA's Ambient Water Quality Criteria for Bacteria - 1986, or standards that offered equal protection to human health, by April 10, 2004; b) CWA section 304 (a) must be revised by the EPA to include new studies on waterborne pathogens or indicators by October 10, 2005, and states have three years to adopt these newly revised standards; c) CWA section 406 was added to allow the EPA to provide states with grants to establish monitoring and public notification programs within public beaches (USEPA, 2002). The bacterial standards adopted by Virginia are the same as those recommended by the EPA's *Ambient Water Quality Criteria – 1986*. Using the fecal indicator bacteria (FIB), *Enterococcus spp*., the maximum allowable concentrations for beach waters are: a geometric mean of 35 colony-forming units per 100 milliliters (cfu/100 mL) for two or more samples taken in a calendar month, or a single sample maximum of 104 cfu/100 mL. A sample found to exceed this standard requires local health departments to post notice of a swimming-advisory and alert the media concerning the beach from which the sample was obtained.

The goal of this study, beginning in 2004, was to use monitoring and MST to evaluate Virginia's public beaches as they were rebuilt following destruction from Hurricane Isabel in September 2003. Specific objectives were to: a) identify beaches with persistent water quality issues; b) determine the sources of fecal contamination at these problematic beaches; c) develop and assess the utility of fluorometric detection of OBs as an indicator of human fecal pollution; d) examine the reliability of ARA with the application of a blind validation set, a minimum detectable percentage, and the cross-validation of results with PFGE; e) evaluate the success of mitigation efforts at the two beaches following detection and repair of infrastructure problems; f)

develop and test a novel MST method to discriminate between fecal sources through the amplification, digestion and visualization of the 16S-23S rDNA intergenic spacer region in *Enterococcus* spp., and; g) develop a comparative study of the effect of imbalanced source categories on the classification bias using several classification algorithms.

Conclusions of this project found that: a) multiple public beaches produced advisories in 2004, warranting further testing in 2005; b) while most beaches indicated that birds and/or wildlife were largely responsible for elevated enterococci levels, Hilton and Anderson beaches were marked by the discovery and successful remediation of sewage leaking into beach waters; c) application of a validation set, as well as a minimum detectable percentage to known-source libraries aids in accurately assessing the classification power of the library and reducing the false positive identification of contributing fecal sources; d) the validation of MST results using multiple methods is recommended for field applications; e) fluorometry displayed potential for detecting optical brighteners as indicators of sewage leaks in storm drains; f) the digestion of the 16S-23S rDNA intergenic spacer region of *Enterococcus* spp. using MboI does not provided suitable discriminatory power for use as an MST method; and g) DA was the least, and k-NN the most, sensitive algorithms to imbalances in the size of source categories in a known-source library.

References

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

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

Hagedorn, C., J.B. Crozier, K.A. Mentz, A.M. Booth, A.K. Graves, N.J. Nelson, and R.B. Reneau Jr. 2003. Carbon source utilization profiles as a method to identify sources of faecal pollution in water. Appl. Environ. Microbiol. 94:792-799.

Hagedorn, C., M. Saluta, A. Hassall, and J. Dickerson. 2005. Fluorometric detection of optical brighteners as an indicator of human sources of water pollution. Part I. Description and detection of optical brighteners. Crop and Soil Environmental News, November 2005. [online] http://www.ext.vt.edu/news/periodicals/cses/2005-11/part1.html. last accessed 27 March 2008.

Leecaster M.K. and S.B. Weisberg, 2001. Effect of sampling frequency on shoreline microbiology assessments. Mar. Pollut. Bull. 42:1150-1154.

Parveen, S., K.M. Portier, K. Robinson, L. Eminston, 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.

Rabinovici, S.J.M., R.L. Bernknopf, A.M. Wein. 2003. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Tech. 38:2737-2745.

Samadpour, M. and Chechowitz, N. 1995. Little Soos Creek microbial source tracking: a survey. Seattle, WA: University of Washington, Department of Environmental Health.

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.

Simmons, G.M., Jr., S.A. Herbein, and C.M. James. 1995. Managing nonpoint fecal coliform sources to tidal inlets. Wat. Res. Update 100: 64-74.

Stevenson, A.H. 1953. Studies of bathing water quality and health. J. Amer. Public Health. Assoc. 43:529-538.

Stewart, J.R., R.D. Ellender, J.A. Gooch, S. Jiang, S.P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons learned from a methods comparison study. J. Wat. Health. 1:225-231.

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

USEPA. 2002. Implementation Guidance for Ambient Water Quality Criteria for Bacteria. U.S. Environmental Protection Agency, Washington, D.C. May 2002 Draft.

USEPA. 2005. Microbial Source Tracking Guide Document. U.S. Environmental Protection Agency, Washington, D.C. EPA/600-R-05-064.

Wiggins, B. A., Andrews, R. W., Conway, R. A., Corr, C. L., Dobratz, E. J., Dougherty, D. P., Eppard, J. R., Knupp, S. R., Limjoco, M. C., Mettenburg, J. M., Rinehardt, J. M., Sonsino, J., Torrijos, R. L., and Zimmerman, M. E. 1999. Identification of sources of fecal pollution using discriminant analysis: Supporting evidence from large datasets. Appl. Environ. Microbiol. 65:3483-3486.

II. Review of the Literature

Library-Dependent Source Tracking

Two major classes of microbial source tracking (MST) methods have received application in surface waters. Known as library-dependent and library-independent methods, each approach currently possesses both strengths and weaknesses for source identifications. Library-dependent methods rely on the construction of a host-origin database, or library, of isolates from known fecal sources. Bacterial isolates collected from these known fecal sources are assayed to provide a collection of possible 'fingerprint' patterns allowing for a direct comparison with the fingerprints of isolates of unknown origin using a variety of statistical classification algorithms. The rate at which isolates within the library are correctly classified, using the models generated from those same isolates is referred to as the rate of correct classification (RCC) for a single category, and the average rate of correct classification (ARCC) for the entire library.

Phenotypic Methods

Antibiotic Resistance

MST methods employing antibiotic resistance were some of the earliest developed, and have been tested in a variety of different watershed environments (Wiggins, 1996; Hagedorn et al., 1999; Parveen et al., 1997; Harwood et al., 2000). The basis of antibiotic resistance analysis (ARA), multiple antibiotic resistance (MAR) and the Kirby-Bauer disk diffusion MST variations is that enteric bacterial strains from different host organisms have acquired and retain unique patterns of antibiotic resistance as selected for by the differential exposure of their host-animal to antibiotics or other xenobiotic compounds. Growth of these bacterial isolates in the presence of

varying antibiotics and/or concentrations will produce a 'fingerprint' pattern unique to a specific, or group of, host animals. Use of antibiotics in humans and domestic animals, as well as addition to animal feed has likely selected for resistant strains, although resistance is still seen in wild animals as well (Butaye et al., 2001; Sayah et al., 2005). Following early work by Kaspar et al., 1990, studies have confirmed the existence of quantifiable and consistent differences in the antibiotic resistance patterns of streptococci isolated from the feces of a variety of host organisms (Wiggins, 1996; Wiggins et al., 1999). While enterococci have received the most application in antibiotic resistance studies, both *Escherichia coli* and fecal coliforms (FCs) have been employed as well.

All current antibiotic resistance methods are culture-based, requiring the physical isolation of the organism, followed by inoculation on a variety of antibiotics and/or antibiotic concentrations. Individual isolates are generally assessed as to their susceptibility, or lack there of, to between five and fourteen specific antibiotics. Although typically scored by hand, at least one study has attempted to develop an automated version (Ebdon et al., 2004) using a PhenePlate™ system plate reader (PhPlate AB, Stockholm, Sweden). Following evaluations of growth patterns among isolates, results are generally converted to binary data (1 or 0) representing growth or no growth, respectively. Data are characterized using a statistical algorithm, typically discriminant analysis (Wiggins, 1996), to generate a model which is used to identify the fingerprint patterns of isolates of unknown origin collected from environmental water samples. Several protocol variations employing antibiotic resistance have been evaluated as MST tools.

ARA has been one of the most commonly used library-dependent MST methods and has undergone widespread field applications. In ARA, isolates are cultured in microtiter plates, and

replica plated onto between 20 and 40 media plates containing different antibiotics and concentrations, allowing for the simultaneous evaluation of up to 48 isolates. ARA has consistently demonstrated effectiveness in small watersheds (Hagedorn et al., 1999; Graves et al., 2002; Booth et al., 2003) and on modest budgets (Crozier et al., 2002), and has shown success in moderately-sized and/or mixed-used watersheds (Burnes, 2003; Carroll et al., 2004; Dickerson et al., 2007a). However, questions concerning the ability of the method to discriminate between sources in large, urban watersheds have been raised (Moore et al., 2005).

In one of the first field applications, Hagedorn et al., 1999 assembled a known-source library of 7,058 fecal streptococci (FS) isolates from known human, livestock, and wildlife sources in Montgomery County, VA, achieving an average rate of correct classification (ARCC) of 87% for the three-way classification split. Application of this library to Page Brook in Clarke Co., VA identified cattle with unrestricted stream access as the dominant contributing source. An average reduction of 94% in FC counts was achieved by fencing streams and installing watering stations within pastures.

In additional Virginia studies, Graves et al., 2002 identified the appearance of enterococci of human origin in Spout Run, a small Virginia stream, as it passed through the non-sewered community of Millwood; with the human signature fading in the stream with increasing distance from the town. And Booth et al., 2003, using a library of 1,451 enterococci isolates, employed ARA as a means to identify livestock as the dominant fecal source in the Blackwater River, in Virginia, as part of a Total Maximum Daily Load (TMDL) project. Within the coastal region of Virginia, Dickerson et al., 2007a used ARA with discriminant analysis and logistic regression to identify human fecal pollution at two Virginia beaches, and monitor successful remediation efforts over two consecutive summers.

Within the tropical waters of Florida, Harwood et al., 2000 used ARA to classify fecal pollution from a variety of animal sources in coastal waters using libraries with ARCCs of 63.9% and 62.3% for FC and enterococci, respectively. A study in Stevenson Creek in Clearwater, FL by Whitlock et al., 2002, using an ARA library of 2398 FCs, identified wildlife as the dominant fecal source when coliform counts were high and human when counts were low.

In mixed-use watersheds, Burnes 2003 used six antibiotics and a library of 1125 FCs isolates to identify chickens and livestock as the major contributors to fecal contamination in a mixed-use Georgia watershed. Choi et al., 2004 used ARA to identify birds as the dominant fecal source (vs urban runoff sources) in at Huntington Beach, CA, achieving a library ARCC above 80%.

Internationally, Carroll et al., 2004 used eight different antibiotics to generate fingerprints for a library of 717 *E. coli* isolates from humans, wildlife, domestic animals, and livestock in the Gold Coast region, Queensland State, Australia. Results indicated that sources in rural areas were primarily non-human, while isolates classified as human in origin increased dramatically in urban areas where septic tanks predominated. Edge and Hill, 2005 identified birds as a primary fecal contributor to the waters of Bayfront Park in Hamilton, Ontario using an *E. coli* library with an ARCC of 68% for bird and wastewater sources.

In MAR, isolates are replica plated on multiple plates, each containing a single concentration of an antibiotic. Early in method development, Parveen et al., 1997, using 10 different antibiotics, concluded that *E. coli* MAR profiles were capable of discriminating between point and nonpoint sources of pollution in isolates collected from Apalachicola Bay, FL. Kelsey et al., 2003 used MAR analyses of *E. coli* isolates to determine that the fecal pollution in a South Carolina estuary was not human in origin, even in areas surrounding clusters of active

septic tanks. In another South Carolina study, Webster et al., 2004 used MAR to successfully differentiate human- and wildlife-derived *E. coli*. Vantarakis et al., 2006 used MAR with cluster and discriminant analysis to separate 128 *E. coli* isolates collected from humans and animals in southwestern Greece, achieving an ARCC of 99.2%. In a recent study employing multiple MST methods, *E. coli* MAR was used in conjunction with human-specific molecular markers for *Bacteroides* and *Enterococcus faceium* at two Lake Erie beaches in Ohio. Results among all methods were consistent in determining that mixed bird and wastewater fecal sources were the likely contributors to the fecal impairment of beach waters (Francy et al., 2007).

Use of the Kirby-Bauer tests, where antibiotic susceptibility is assessed by zones of clearing around antibiotic impregnated disks on a plate inoculated with a specific bacterial strain, has only been reported in the SCCWRP MC study (Harwood et al., 2003). Initial results found this method poorly identified fecal sources in blind water samples, and further testing was suspended.

The widespread application of antibiotic resistance patterns in field studies can be contributed largely to the relative simplicity with which each isolates can be fingerprinted, the ability to employ multiple recognized fecal indicator bacteria (FIB), the lack of specialized equipment required, and the ability to rapidly fingerprint a large number of isolates. All of these factors have become increasingly desirable as researchers are beginning to discover the diversity of bacterial strains present even within small watersheds (Stewart et al., 2003), and recognizing the need to collect large numbers of isolates and samples to achieve a representative population, avoiding errors generated by undersampling (Choi et al., 2004). In two of the only attempts to assess the temporal stability of fingerprint patterns used for library construction, Wiggins et al.,

2003 and Dickerson et al., 2007a used ARA to prove the stability of patterns at least a one- and two-year period, respectively.

Questions of method effectiveness have been raised in both U.S.-based MC studies, as well as individual studies. Unfortunately, most have based conclusions on results obtained from a relatively small collection of isolates. Guan et al., 2002 created MAR profiles of 319 *E. coli* isolates, and found that only 46% of the livestock isolates, 95% of the wildlife isolates, and 55% of livestock isolates were correctly classified. Samadpour et al., 2005 determined ARA was less successful at discriminating between sources than ribotyping based on a collection of 120 *E. coli* isolates. In addition, the reproducibility and stability of antibiotic resistance patterns has been questioned (Parveen et al., 1999), but remains unconfirmed. Moore et al., 2005 found problems repeating ARA patterns in a subset of isolates fingerprinted over several days. This was a problem reported in most MST methods, including ARA, in the USGS MC study (Stoeckel et al., 2004).

Carbon utilization profiling

Use of metabolic patterns, also known as carbon utilization profiling (CUP), in MST is based on the concept that enteric bacteria are adapted to utilize a selection of unique carbon and nitrogen sources for energy production and growth based on the diet of their host animal. The automated commercial Biolog™ system (Biolog, Inc) was adapted by Hagedorn et al., 2003a, for MST, and used to type enterococci from fecal sources of known origin. Biolog™ has had a primarily medical application (Holmes et al., 1994) but has also received use in the fields of soil ecology (Wünsche et al., 1995), aquatic microbiology (Choi et al., 1999) and plant pathology (Fisher et al., 1992). Utilizing 30 carbohydrate sources from the 95 contained within the plate, a

library of 365 *Enterococcus spp*. isolates was constructed. With discriminant analysis, an ARCC of 92.7% was achieved using a two-way classification (human vs. non-human) and an ARCC of 81.9% using a four-way split (human, wildlife, livestock, and pets). Testing both *E. coli* and enterococci CUP in the SCCWRP MC studies yielded ARCCs of 73 and 93% respectively. However, both methods also falsely identified sources at a high rate (66%, *E. coli* and 51%, enterococci).

A more recently application of the Biolog™ CUP method was tested on both FC and FS collected from four animal sources in Oklahoma and North Texas (Evenson and Strevett, 2006). Using discriminant analysis, the RCCs for poultry, cattle, pigs, and humans were 47%, 64%, 70%, and 71%, respectively for coliforms, and 74%, 87%, 83%, and 86% respectively for streptococci. Combining the patterns from the known-source coliforms and streptococci however, allowed for a significant improvement in classification rates, yielding an ARCC of 93%.

Graves et al., 2007 used the Biolog™ system to identify enterococci at the species level in both dry and fresh livestock manure, and determined that for MST purposes, no significant differences existed in species survival rates within the environment. Thus, the frequency of enterococci species likely to enter waterways would be consistent with those found in collected manure samples and used for the construction of known-source libraries.

A similar CUP method using the PhenePlate™ system (PhPlate AB, Stockholm, Sweden) was applied in fecal samples from the United Kingdom (Wallis and Taylor, 2003). This study tested 1,766 enterococci isolates from six sources (sheep, donkey, dog, cattle, sewage, and birds) using eleven different carbohydrates. Using the Simpson Diversity Index, researchers determined that enterococci isolated from sewage wastewaters showed much higher levels of

diversity than those from non-sewage sources. An additional study into MST potential was conducted using the PhenePlate™ system in Spain (Manero et al., 2002). Measurements of enterococcal diversity were calculated from the fecal wastewaters of human (hospital), pig (lagoon slurry) and urban sewage (considered mixed origin). Although overlap in the 528 enterococci strains analyzed were detected among all three wastewater types, several strains believed to be unique to humans (16 strains) and pigs (5 strains) were also isolated.

The fingerprinting of 4057 enterococci and 3728 *E. coli* collected from ten animal sources (horses, cattle, sheep, pigs, ducks, chickens, deer, kangaroos, dogs, and human septage) using the PhenePlate™ system in Australia provided 530 and 526 unique phenotypes respectively (Ahmed et al., 2005a). Of these unique patterns, 295 *Enterococcus spp.* and 273 *E. coli* were found to be unique to a specific host source. In a subsequent field test, these patterns were applied to 791 enterococci and 550 *E. coli* collected from 27 water samples at five sites along a local creek. Strong agreement in source identifications were observed between the two bacterial methods, as 10 and 13% of isolates were identified as human in origin, and 61 and 54% were matches for non-human patterns for enterococci and *E. coli* respectively; with the remaining isolates producing either unknown, or shared patterns. In an additional field test, using the PhenePlate™ system *E. coli* fingerprints detected septic leakage into small creeks in southern Australia (Ahmed et al., 2005b).

The advantages of using CUP includes the elimination the human "judgement" factor in interpreting results (a potential problem in ARA), as positive wells are determined using an automated plate reader. In addition after the purchase of initial equipment (plate reader and software) the methods are both rapid and relatively inexpensive. The disadvantages include the

less favorable performance in the SCCWRP and USGS MC studies, and the overall limited amount of research to date devoted to the development of this method.

Genotypic Methods

Intergenic Spacer Regions (IGS or ISR)

A few MST studies have attempted to utilize the intergenic spacer (IGS) regions in *E. coli* or *Enterococcus spp.* to search for source-specific strains. Typically present in multiple copies the ribosomal DNA (rDNA) operon is usually arranged 16S-IGS-23S-IGS-5S in bacteria. The indicator organisms *E. coli* and *Enterococcus spp*. have seven (Condon et al., 2002) and at least five copies (Sechi and Daneo-Moore, 1993), respectively, of these nearly identical sequences present within their genomes. The absence of selection pressures in the IGS region, as opposed to the highly conserved nature of the bordering rDNA, has proven useful as a target site for the molecular subtyping of a variety of pathogenic bacteria (Guertler and Stanisich, 1996; Stubbs et al., 1999; Graham et al., 1996; Chun et al., 1999; Riffard et al., 1998), providing a simpler (in both equipment needed and level of training required) and more cost-effective assay than more traditional MST methods such as pulsed-field gel electrophoresis (PFGE; Bedendo and Pignatari, 2000) or ribotyping (Carson et al., 2003). PCR amplification (see Appendix section B) of these IGS regions can be performed using primers targeting conserved regions of the flanking 5S, 16S or 23S rRNA genes.

A recent study reported success using *E. coli* 16S-23S ribosomal DNA IGS regions to discriminate between humans, cows, and chickens (Buchan et al., 2001) finding 84 unique banding patterns out of 132 isolates used. Using 267 *E. coli* isolates collected from sewage, horses, cows, gulls and dogs, Seurinck et al., 2003 found only two potential source-specific

fingerprints, (one for cow and one for sewage) of 87 unique patterns obtained. Achieving an ARCC of only 67.2% for a 5-way classification split, the approach has received only limited further MST use.

Although unsuccessful as a potential MST method, amplification of the 16S-23S rDNA IGS region in *Enterococcus spp*., and subsequent digestion (see Appendix section D) using MboI by Dickerson et al., 2007b, demonstrated the dangers in basing method success on the use of a small number of isolates, as ARCCs decreased considerably, from initial high values, when the number of fecal samples and isolates increased. Despite mixed results, research into this method continues, as Crozier, at Roanoke College (Virginia) and Graves at N.C. State are currently pursuing a library-based MST method using the IGS region of *E. coli* with the restriction enzyme RsaI.

Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) has become a valuable instrument in bacterial genomic research (see Apendix section F). As one of the most commonly used genotyping methods in epidemiological studies (Yan et al., 1991; Schoonmaker et al., 1992; de Moissac et al., 1994), PFGE allows for analyses of the DNA fragments generated by the digestion of an entire microbial genome with a rare cutting restriction endonuclease. Within MST, PFGE has show success in both the identification of host sources and the reproducibility of the method. Dr. George Simmons of the Department of Biology at Virginia Tech was the first researcher to adapt PFGE for use as a MST tool. In 1992, he identified raccoons as the major *E. coli* source polluting shellfish beds along the Cherrystone Inlet in Eastern Virginia (Simmons et al., 1995). After the trapping and removal of a substantial number of raccoons by local game officials,

bacterial counts decreased, substantiating Simmons' results. In addition, the application of Simmon's PFGE method in a 2000 TMDL project in Four-Mile Run (Virginia) identified fecal isolates from water samples as dominantly waterfowl, raccoon, and human in origin (Simmons et al., 2002). In another Virginia field application, a 420-isolate library was constructed by Dickerson et al., 2007a, to serve as an additional line of evidence (with ARA and fluorometry) in the evaluation of the remediation of two sewage-polluted public beaches in Virginia. PFGE results agreed with those of ARA and fluorometry, confirming remediation success.

McClellan et al., 2003 demonstrated that PFGE provided greater resolution in distinguishing between closely related *E. coli* strains that did repetitive extragenic palindromic PCR (rep-PCR). Implying that extensive host sampling will be required to encompass the strain diversity of *E. coli* within the environment. Parveen et al., 2001 had less success with PFGE, concluding, based on the testing only 32 *E. coli* isolates, that none of the three restriction enzymes tested (SfiI, NotI, and XbaI) provided patterns unique to a specific host organism. Another unsuccessful attempt applied the restriction endonuclease XbaI to 86 *E.coli* isolates, failing to produce significant source separations (Lasalde et al., 2005), confirming the importance of enzyme selection.

PFGE method disadvantages include the initial cost of equipment, time-required to run samples, and the sensitivity of the method, which may require the construction of even larger databases than necessary for other library-based MST methods (USEPA, 2005). Method performance during both U.S.-based MC studies was better than most of the methods attempted. In the SCCWRP-sponsored study, PFGE using XbaI correctly identified all sources in 5 of the 12 water samples, and identified at least one of the sources in all of the remaining samples (Myoda et al., 2003). In the USGS-sponsored study, PFGE was the only method which successfully

replicated results for a subset of 24 isolates used to construct the known-source library. In spite of the relative success of PFGE in the MC studies, the method has largely fallen out of favor with MST researchers, due primarily to the expense and limited number of isolates which can be fingerprinted in a reasonable amount of time (~24 per week)

Ribotyping

Ribotyping is one of the more popular library-dependent genotypic MST methods (see Appendix section G); and one of only two methods (with rep-PCR) included in both U.S.-based MC studies in multiple forms (different restriction enzymes, Myoda et al., 2003; Stoeckel et al., 2004). Ribotyping is similar to PFGE however instead of using the entire genome the highly conserved genes encoding the 16S rRNA are analyzed (Parveen et al., 1999). Ribotyping produces a genetic fingerprint for an isolate by amplifying and digesting rDNA in a process that involves gel electrophoresis, Southern blotting, and hybridization with probes targeting rRNA genes. Like PFGE, this method has received considerable use for pathogen identification in epidemiological studies (Kostman et al., 1992; Graves et al., 1994). As a library-dependent method, ribotyping, requires the creation of a database of isolates using the banding patterns generated from the digested rDNA.

Application of ribotyping to MST studies has focused almost exclusively on the FIB, *E. coli*, with known-source libraries ranging in size from 120 isolates (Samadpour et al., 2005) to a combined database of sources from across the U.S. of over 5000 (Scott et al., 2004). Although the compiled database represents the largest ever reported for the method, no other ribotyping library has reached a size over 500 isolates, limited largely by both temporal and fiscal constraints of the method itself.

In the early developmental stages, Parveen et al., 1999, applied discriminant analysis to 238 *E.coli* isolates, detecting 41 human and 61 non-human ribotype patterns and achieving an ARCC of 82% for a human/non-human classification split. Carson et al., 2001, using the restriction endonuclease *Hin*dIII, examined 287 known-source isolates from humans and seven non-human sources (cattle, pigs, horses, dogs, chickens, turkeys, and geese). The study demonstrated that classification accuracy could be increased by reducing the number of source classes with, the highest ARCC (97.1%) achieved with a human/non-human split. In a follow-up study, using discriminant analysis, Carson et al., 2003 achieved an ARCC of 72.9% for the same eight-animal split and 87% for a human/non-human split for 482 *E. coli* isolates (136 human and 346 non-human).

In an early field application within a Washington state creek, cows and dogs were identified as the primary fecal sources based on results from a 421-isolate library (Samadpour and Chechowitz, 1995). Farag et al., 2001 used *E. coli* isolates from Cascade Creek in Grand Teton National Park to match the riboprint patterns of bird, deer, canine, elk, rodent, and human.

In a successful demonstration of large multiwatershed libraries, a database of *E. coli* ribotype patterns was generated from ~3800 isolates of animal origin and ~1500 human-derived isolates from across the United States (Scott et al., 2004). Using a collection of known-source samples from a South Carolina watershed, as a validation set, 92% of isolates were correctly identified by the known-source library using discriminant analysis. When applied to 515 *E. coli* isolated from South Carolina water samples, 88% of isolates were classified as animal in origin.

In addition to the relative expense and time involving in generating ribotyping library, questions of method reproducibility have been raised. In a study by Lefrense et al., 2004, the reproducibility of ribotyping was measured both within a single laboratory and between

laboratories. While banding patterns obtaining from a single laboratory were generally consistent, attempts to reproduce patterns in multiple laboratories were less successful.

As a part of method development, ribotyping has been used a tool for assessing the geographic and temporal variability of host strains. Scott et al., 2003, using *Hin*dIII and discriminant analysis, determined that significant geographic variation existed when examining 317 *E. coli* isolates from farm animals across the state of Florida. Jenkins et al., 2003 combined ribotype patterns generated using *Eco*RI and *Pvu*II from Black Angus steers to determine that *E. coli* types within the animals changed throughout the season, based on 240 patterns from 451 isolates collected over 10 months. Only 20 ribotypes were shared between sampling times. Hartel et al., 2002 used a total of 298 *E. coli* isolates collected from both wild and captive deer to conclude that diet affected *E. coli* diversity in a host organism, as the wild deer produced more than 3 times (35 to 11) more unique patterns than those from the captive deer.

As a means to increase the discriminatory ability of the method the use of multiple restriction enzymes has been suggested (Samadpour, 2002). Samadpour et al., 2005 demonstrated success in generated 7 to 15 hybridizing bands using both *Eco*RI and *Pvu*II to create 27 unique patterns from 120 isolates from seven host sources.

Within the two U.S.-based MC studies, ribotyping identifications of blind isolates produced extremely variable results between labs using different organisms, restriction enzymes and protocols (Myoda, et al., 2003; Stoeckel et al., 2004). In the SCCWRP study two labs identifying *E. coli* (*Eco*RI and *Pvu*II) and one using enterococci (*Pst*I) correctly identified between 38 and 86% for 2- to 4-way classification splits. However, false positive identifications ranged from 14 to 57%. In the USGS study the identification of *E. coli* isolates using *Hin*dIII correctly identified from 13 to 49% of isolates for a 2-, 3-, and 8-way classification split. Using

*Eco*RI, 90 to 100% of isolates were correctly identified however classifications were attempted on only 5-6% of provided blind isolates. At present, the consensus of the MST community is that ribotyping is most useful as a means for separating human and non-human fecal sources (USEPA 2005).

Repetitive extragenic palindromic PCR (rep-PCR)

Repetitive extragenic palindromic PCR (rep-PCR) uses conserved, repeated sequences within the bacterial genome as a priming site for PCR amplification (Martin et al., 1992). Three families of these palindromic units (PUs) have been identified: repetitive extragenic palindromic (rep) sequences (35-40 bp), enterobacterial repetitive intergenic consensus (ERIC) sequences (125-127 bp), and the BOX sequence (154 bp, Louws et al., 1994). MST researchers applying rep-PCR have dominantly utilized either REP or BOX primers to amplify the DNA sequences located in between PUs. Known-source libraries are generated from the banding patterns of these DNA fragments when resolved in an agarose gel (see Appendix section C).

In a single laboratory study, rep-PCR was shown to be superior to ribotyping (Carson et al., 2003) and rRNA intergenic spacer region (ISR)-PCR (Seurnick et al., 2003) in both accuracy and reproducibility. Using 154 *E. coli* isolates from seven animal sources, Dombek et al. 2000 used REP and BOX primers to correctly classify 65.7% and 93% of isolates, respectively, using calculated Jaccard similarity coefficients. Results indicated that use of the BOX primers was superior to that of the REP primers for source identifications.

Using discriminant analysis, Carson et al., 2003 achieved an ARCC of 88.1% for an eight animal split (humans, cattle, pigs, horses, dogs, chickens, turkeys, and geese), and 96.7% for a human/non-human classification split using rep-PCR on 482 *E. coli* isolates (136 human, 346

non-human). In a less successful application, Leung et al., 2004 used ERIC-PCR to fingerprint 62 *E. coli* isolates, but only correctly classified 0.0%, 28.6% and 75.0% of cow, human and pig isolates, respectively using discriminant analysis.

Several field applications of rep-PCR have also been reported across the country. Baffaut et al., 2005 used a 1029-isolates rep-PCR library to identify geese and humans as the primary fecal polluters in a Missouri watershed. In a separate study in the Ozark National Scenic Riverways in Missouri MST researchers identified cattle, horse and sewage as the dominant fecal contributors across multiple sampling locations (Davis and Barr, 2006).

Within the SCCWRP MC study, three labs attempted to use rep-PCR to correctly identify isolates within a set of blind water samples (Myoda et al., 2003). All three labs attempted to use the BOXA1R primers and all produced similar results, correctly classifying 43-80% of isolates for 2- and 4-way classification splits. Although, false positive rates for the methods ranged from 17 to 52%. In the USGS MC study rep-PCR was performed by two laboratories using REP and BOX primers on a blind set of *E. coli* isolates (Stoeckel et al., 2004). Using the REP primers, researchers correctly identified from 22 to 63% of isolates for a 2-, 3-, and 8-way classification split. Using the BOX primers in a different lab, 22 to 77% of isolates were correctly identified however classifications were only attempted on 67-69% of provided blind isolates, with the remaining isolates producing unrecognized patterns.

The rep-PCR fingerprinting method has also been applied to one of the most significant questions in MST, genetic diversity in host organisms and environmental samples. Using REP and ERIC primers, McClellan et al., 2003 and McClellan et al., 2004 suggested that a large collection of known-source isolates are necessary to accurately represent the genetic diversity of *E. coli* from a given host animal, although no numerical recommendations were given. In

addition, a study using BOX-PCR identified *Enterococcus spp*. as a preferrable organism to *E. coli* for MST applications based on the correct identification of blind, non-library isolates at a rate of 98% and 70%, respectively (Hassan et al., 2007).

This method has been very popular among MST researchers due to its relative simplicity and rapid turnaround time. And, perhaps due to the relative ease with which libraries can be generated, this method has been a component in several studies of the effectiveness of commonly used statistical classification algorithms (Albert et al., 2003; Ritter et al., 2003; Zhong, 2003; Albert et al., 2004; Ritter and Robinson, 2004). Even with the wide variety of MST methods available to researchers, and in spite of the limitations in accuracy and reproducibility seen in the U.S.-based MC studies (Myoda et al., 2003; Stoeckel et al., 2004), rep-PCR continues to remains one of the most popular genotypic, library-dependent methods applied to field studies (Davis and Barr, 2006).

Library-Independent Source Tracking

Soon after the development of library-dependent MST, researchers began to look for organisms, or sequences within the genome of organisms, that were consistently exclusive to pollution from a particular fecal source. Known as library-independent methods, in addition to using source specific markers found in some indicator organisms (Scott et al., 2005), researchers have frequently expanded the search into non-indicator fecal organisms such as: *Bifidobacterium spp.* (Rhodes et al., 1999), *Bacteroides* (Bernhard et al., 2000), F-specific RNA coliphages (Hsu et al., 1995; Cole et al., 2003; Sundram et al., 2006), methanogens (Ufnar et al., 2006), and human or livestock specific enteric viruses such as enterovirus (Noble et al., 2001; Fong et al., 2005), adenovirus (Jiang et al., 2001; de Motes et al., 2004; Fong et al., 2005), and teschovirus

(Jimenez-Clavero et al., 2003). Library-independent methods do not require the collection and fingerprinting of known-source isolates, offering a distinct advantage in terms of time and overall expense to perform MST in a watershed. However, few if any current methods have proven to be consistently unique to a specific species, detectable in significant quantities in environmental waters, and/or geographical transferable in different regions of the United States or Europe.

Fecal coliform/Fecal streptococci ratio

Measuring the ratio of FC to FS was one of the first MST attempts to identify contributing fecal sources in a water body. Based on observations that human fecal samples contained higher FC counts, while animal samples contained higher FS counts, a ratio >4.0 FC/FS, was calculated to serve as an indicator of human pollution, while a ratio of ≤0.7 was to indicate animal pollution (Geldreich and Kenner, 1969). The major advantages of this method are the low cost, ease with which these organisms can be cultured and enumerated, and the ability to produce rapid results. However, application of this ratio has been largely abandoned due to evidence of differential survival rates of streptococci in both surface waters and in water treatment processes as well as frequent detections of ratios between 0.7 and 4.0 (Pourcher et al., 1991; Sinton et al., 1993; APHA, 1998).

Bacteriophages

Coliphages

Coliphages are bacteriophages which infect *E. coli*. Abundance in sewage, as well as the ability to survive water treatment processes and persist in the environment has led some to

suggest coliphages as an indicator of human pathogenic viruses in fecally polluted waters (Payment and Franco, 1993; Skraber et al., 2004). Coliphages are divided into two major groups: somatic and male-specific (F+) coliphages.

Somatic coliphages attach to viral receptors on the cell wall of *E. coli* (Muniesa et al., 2005). However, their replication potential in *E. coli* and other enterobacteria occurring naturally in environmental waters has limited their potential use as an indicator and an MST tool (Solo-Gabriella et al., 2000). F+ RNA coliphages, however, rarely multiply in the environment and are associated consistently and, thus far, exclusively with feces, allowing for potential use as both a fecal indicator (Kamiko and Ohgaki, 1993) and a MST tool. F+ RNA coliphages, which attach to and infect only *E. coli* containing an F pilus (carried on the F plasmid) can be divided into four groups using group-specific antisera (Beekwilder et al., 1996) or nucleic acid hybridization (Hsu et al., 1995). Generally, groups II and III have been found associated with sewage pollution, while groups I and IV are believe to be indicative of animal feces (Havelaar et al., 1986). Although these relationships have deemed statistically significant (Schaper et al., 2002a) exceptions have been noted, such as the occurrence of groups II and III in pigs (Hsu et al, 1995; Cole et al., 2003) and group III in wastewater from chickens (Sundram et al., 2006). In addition, the European-based MC study detected group III F+ RNA coliphages in 33% of animal fecal samples analyzed (Blanch et al., 2006)

Other potential drawbacks include the variable presence of F+RNA coliphages in animal feces and their occurrence in less than 10% of individual human fecal samples (Calci et al., 1998). In addition, differential survival rates have been noted, as groups III and IV coliphage have been found to be the least resistant, and group I the most resistant, to environmental stresses

(Brion et al. 2002; Schaper et al., 2002b). These variations could potentially undermine attempts of MST researchers to detect and quantify fecal sources in a water body.

Application of F+RNA coliphages in field studies in Florida (Griffin et al., 2000), Kentucky (Brion et al., 2002) and South Carolina (Stewart-Pullaro et al., 2006) found that majority of research sites contained group I coliphage as the dominant type, suggesting animal fecal sources. Performance of the F+ RNA coliphage typing in the SCCWRP MC study (Noble et al., 2003), was generally good if coliphages could be detected (5 of 12 freshwater samples) in the fecal/sewage spiked samples. Sewage was correctly identified in the NOAA lab in 3 of 4 samples, and falsely identified in the one remaining, non-sewage-containing sample. In samples typed at the UNC lab, again F+ RNA coliphages were only detected in 5 of 12 samples, and all 5 samples were correctly identified as human/sewage containing, however, non human source remained unidentified or falsely identified in 4 of the 5 samples. This inability to detect coliphages in spiked water samples suggests application would be most successful in detecting pollution sources containing large numbers of individuals such as animal feed lots and sewage leaks. Detection may be unsuccessful in areas where the number of individual sources is small such as a failing septic system.

Bacteroides phages

The potential for use of *Bacteroides*-infecting bacteriophages have been reported as successful for the detection of human feces in *Bacteroides fragilis* strain HSP40 (Tartera et al., 1989; Kator and Rhodes, 1992; Grabow et al., 1995) and RYC2056 (Puig et al., 1999). However, in European studies, only the bacteriophage infecting *B. thetaiotamicron* GA17 (Payan et al., 2005) displayed a consistently high specificity to human fecal pollution, although plaque

clarity appeared to be geographically specific (Blanch et al., 2004) and may require development of regionally specific strains of *Bacteroides* (Payan et al., 2005). An advantage in the continuing attempts to utilize bacteriophages is their failure to replicate in the environment (Yan and Sadowsky, 2006).

Host-specific microorganisms

Enterococcus spp.

In a study of 392 isolates identified as *Enterococcus spp*., Wheeler et al., 2002 found that *Enterococcus faecalis* was limited to a host range of humans, dogs, and chickens, based on isolates speciated from humans, dogs, pigs, cattle, deer, chickens and Canada geese. No field applications, or further testing of *Enterococcus faecalis* have been conducted to date, a likely result of the limited usefullness of detection given such a diverse host range. Kühn et al., 2003 typed 17,157 enterococci isolates collected from pigs, cattle, and humans (sewage, individuals, and hospital waste) in Denmark, Spain, Sweden, and the U.K. Different concentrations of *Enterooccus faecium*, *E. faecalis*, and *E. hirae* were reported for human vs. non-human sources, depending on their country of origin. However, detection of these species ratios was inconsistent in the European-based MC study (Blanch et al., 2006).

Bifidobacteria

Bifidobacteria are non-spore forming obligate anaerobic bacteria found consistently in the feces of humans and pigs, less frequently in cattle and sheep, and no in other animals (Mara and Oragui, 1983). The detection of sorbitol-fermenting bifidobacteria have been assessed as indicators of human fecal pollution (Mara and Oragui, 1983; Resnick and Levin, 1981; Rhodes

and Kator, 1999) based on their relative abundance in the intestines of humans and rarity in that of other animals.

Human Bifid Sorbitol agar (HBSA), developed by Mara and Oragui 1983, has allowed for the detection and enumeration *Bifidobacterium adolescentis* and *B. breve*, two sorbitol-fermenting human-specific stains which represent 58 to 90% of the bifidobacteria found in humans (Resnick and Levin, 1981). However, the survival of bifodobacteria in the environment can be highly variable, with die-off rates measured in hours in both fresh and marine sterile water (Resnick and Levin, 1981), and warm water temperatures significantly decreasing survival rates (Rhodes and Kator, 1999). Field evaluations have been conducted in areas containing large sewage inputs (Carillo et al., 1985, Jagals and Granbow, 1996) as well more diffuse, non-point sources (Rhodes and Kator, 1999). While detection was generally successful in areas of significant pollution, non-point human sources proved difficult to detect with current cultivation methods.

Recently, molecular methods have been developed to detect *B. dentium* through 16S RNA gene probes (Nebra et al., 2003) and *B. adolescentis* and *B. dentium* with a multiplex PCR using 16S rRNA primers (Bonjoch et al., 2004). Both *B. adolescentis* and *B. dentium* have been found in sewage but absent in the wastewaters of animals. However, European tests using the multiplex PCR assay found both human-specific bifidobacteria present in some of the animal samples and absent in some of the human samples tested (Blanch et al., 2006). Given the limited survivability of bifidobacteria in the environment, detection may only remain useful for episodes of very recent contamination.

Rhodococcus coprophilus

The use of *Rhodococcus coprophilus*, an actinomycete, was proposed by Mara and Oragui 1981 as an indicator of fecal contamination in domesticated herbivorous animals. Correlations with FS have implied a fecal origin, with transmission from animal to animal likely through the fecal-oral route. At present, detection is limited by a cultivation time of up to 21 days (Seurinick et al., 2005). Although, questions of abundance and survival time remain unknown, the recent development of a real-time PCR assay allowing for both detection and enumeration in 2-3 days (Savill et al., 2001) will allow for continued investigations into potential use of this organism as an MST tool.

Bacteroides thetaiotaomicron

Believed to be present in large quantities in humans and largely absent in animals, a primer set producing an amplicon specific to *Bacteroides thetaiotaomicron* was reported by Carson et al., 2003. Detection of this amplicon successfully identified 92% of human fecal samples, and was not detected in cow, pigs, horses, geese, turkeys, or chickens. However detection in 16% of dog fecal samples will likely limit application to a presence/absence test of human or human/canine fecal pollution.

Enteric viruses

Detection of viruses has received an increasing amount of attention in recent years as potential indicators of fecal pollution. In general, viruses have shown a prolonged persistence in natural waters as compared to the current FIB (Moce´-Llivinia et al., 2005), leading some to suggest they may serve as better indicators of health risk. As enteric viruses are believed to be

strongly host specific, application in MST has been explored in multiple studies as potential source-specific markers. A variety of viruses have been targeted for MST including enterovirus (Noble and Fuhrman, 2000; Fong et al., 2005), adenovirus (Jiang et al., 2001; de Motes et al., 2004; Fong et al., 2005), and teschovirus (Jimenez-Clavero et al., 2003). Although initial studies focused largely on viruses specific to humans, the application of animal-specific viruses to water quality problems has received increasing attention as well (de Motes et al., 2004; Fong et al., 2005).

Enteroviruses are a group of small, non-enveloped RNA viruses containing over 80 serotypes (Oberste et al., 2005). Able to survive in freshwater for prolonged periods of time (Rzezutka and Cook, 2004), enteroviruses have also been detected in marine waters (Noble and Fuhrman, 2000; Moce´-Llivina et al., 2005) and even Korean tap waters (Lee and Kim, 2002). Adenoviruses are the only human enteric viruses that contain DNA instead of RNA. These viruses cause respiratory infections in humans, and are considered more stable than many other enteric viruses such as poliovirus or hepatitis A virus in tap and seawater (Enriquez et al., 1995). In addition, human adenoviruses are believed to more abundant with a greater persistence in environmental waters than enteroviruses (Formiga-Cruz, et al., 2003). This has led some to suggest their potential for use as an indicator of human fecal pollution using a nested PCR assay (Pina et al., 1998).

As viruses are not easily isolated and cultivated from environmental samples, MST researchers have begun to employ variations of reverse-transcriptase PCR or quantitative PCR (q-PCR, see Appendix section H)) for detection. Similar to other viral MST methods, such as F+ RNA coliphages, the detection of human-specific viruses appears to be more useful in

identifying sewage pollution rather than individual fecal samples (Noble et al., 2003), as these viruses are not harbored in substantial percentage of the human population.

In field studies, a reverse transcription- and nested PCR procedure was applied to 30 water samples collected from the lower Altamaha River, Georgia (Fong et al., 2005). Researchers found that either human enterovirus or human adenovirus were present in 20 of the samples, with both present in eight. In addition, bovine enterovirus was detected in 11 of the 30 samples. Porcine or bovine adenoviruses were detected from 70% of the swine lagoons sampled and 3 of 4 cattle fecal samples, respectively, using two sets of oligonucleotide primers (de Motes et al., 2004). Gregory et al., 2006 used a quantitative reverse transcriptase PCR (qRT-PCR) assay to detect enteroviruses in environmental waters. The procedure is described as rapid (less than 5 hrs.) and highly sensitive (detects as few as 25 viral genomes in 91% of samples). Jiang et al., 2001 detected human adenovirus in southern Californian and Mexican coastal waters using nested-PCR. However, the collection of 20 to 40-liter water samples was required for adequate detection, an issue with all enteric virus methods.

Viruses have the potential to serve as highly specific indicators of fecal contamination. As detection procedures continue to improve, increases in speed and the ability to quantify results show promise not only as an MST tool, but as a supplement or potential replacement for the current FIB. At present, a major drawback of viral MST is difficulties detecting dilute concentrations in environmental waters. In addition, with the exception of enteroviruses, the presence of most viruses is variable in sewage effluents Formiga-Cruz et al., 2003).

Gene-specific targets

Escherichia coli and *Enterococcus faecium*

PCR-based methods targeting the STIb, STII and LTIIa heat stable enterotoxin genes in *E. coli* have been developed for humans (Oshiro and Olson, 1997), cattle (Khatib et al., 2002) and swine (Khatib et al., 2003) respectively. In the SCCRWP MC study the STIb gene was detected in all sewage spiked samples, but only successfully detected 2 of the 4 samples containing fecal material from individual humans, as well as falsely identifying human in a sample where it was not present (Field et al., 2003). The LTIIa gene was detected in all 5 samples containing feces from cattle, but only after samples were unblinded and analyses modified.

Source-specific markers in *E. coli* for geese and ducks we developed using subtractive hybridization by Hamilton et al., 2006. The marker DNAs identified successfully classified 76% of isolates tested from goose, and 73% of duck isolates, but proved to be geographically limited to the development area. Targeting of the 680 bp gene for an enterococcal surface protein (esp), a virulence factor in *Enterococcus faecium* believed to be present only in enterococci of human origin (Scott et al., 2005) has been applied in waters across the U.S. (McDonald et al., 2006; Soule et al., 2006; Francy et al., 2007). Unfortunately, only about 10% of humans are believed to harbor enterococci containing the esp gene. The advantages of these gene-specific MST methods are library-independence and seeming specificity to target genes. The disadvantages include the absence or limited presence in host animal of bacteria containing the toxin genes or surface protein, and a limited testing of host geographic range.

Methanobrevibacter smithii

The *nif*H gene in *Methanobrevibacter smithii* has been identified as a culture-independent means of detecting sewage contamination (Ufnar et al. 2006). *M. smithii*, a methanogen, is a common inhabitant of the human intestinal tract and remains undetected in any animal species (Lin and Miller, 1998). The designed PCR primers targeted the nitrogen fixation gene, *nif*H, based on observed sequence differences in methanogens (Ohkuma et al., 1999). The amplified 222-bp product was detected in sewage (93%), human fecal samples (29%) and human contaminated water (100%), and was not found in feces of cows, pigs, sheep, dogs, horses, goats, turkeys, chickens, deer, and geese. In the corresponding field test (Ufnar et al. 2006), *M. smithii* was detectable in waters for as long as 24 days. This method utilizes an anaerobe believed to be unlikely to replicate in environmental waters, and so far, not detected in sediments. To date, no further studies have been reported.

Bacteroides

Assays for the detection of host-specific *Bacteroides* are currently the most frequently used culture-independent MST methods (USEPA, 2005). The selection of *Bacteroides*, a strict anaerobe, as a target organism was based largely on abundance in human feces, with populations more prevalent than *E. coli* by as much as three orders of magnitude (Fiksdal et al., 1985).

A 16S rDNA marker was developed using length heterogeneity PCR (LH-PCR, see Appendix section E) and terminal restriction fragment length polymorphism (T-RFLP, see Appendix section E) for cattle (227 bp sequence) and humans (119 bp sequence) and tested using 16S primers and selected restriction endonucleases in Tillamook Bay, OR (Bernhard and Field 2000a). Based on these sequences, host-specific specific primers were designed and successfully

tested on fecal samples from cattle and humans (Bernhard and Field, 2000b). Estimates were made that 10% of *Bacteroides* cells present in the feces tested contained these host-specific markers. Additional applications of these markers include their use: at seven Lake Michigan beaches to detect the presence of human and ruminant fecal pollution at levels below the *E. coli* recreational standard (235 cfu/100 mL; Bower et al., 2005); at Catalina Island in California (Boehm et al., 2002); and with MAR and the human-specific esp gene (Scott et al., 2005) at Lake Erie beaches in Ohio (Francy et al., 2007).

Recently, Dick et al., 2005 designed primers from sequences believed to be unique to *Bacteroides* from pigs and horses. The primers allowed for the successful amplification of target sequences in all 19 pig samples, and 9 out of 10 horse samples tested. In addition, the target sequences were not detected in the fecal matter of any other host animals. Currently no field tests of these markers have been reported.

In an effort to obtain more rapid, and quantifiable results, a q-PCR assay was developed for all *Bacteroides* species (Dick and Field, 2004), as well as human-specific (Seurinick et al., 2005) and cow-specific *Bacteroides* (Layton et al., 2006). Although primarily still used as a presence/absence test for human or bovine fecal pollution, continued research into quantification procedures and further testing of the geographic stability of these markers may eventually allow for correlations of these *Bacteroides* populations with FIB and provide health managers with better ways to assess public health risks in fecally polluted recreational waters. In the SCWRRP MC study, *Bacteroides*-based identifications were plagued by difficulties recovering DNA from water samples, as well as problems with false-positives, which were attributed to contaminated samples.

Chemical-Based Source Tracking

In spite of the variety of MST methods available to researchers and watershed managers, numerous difficulties associated with tracing the origin of fecal pollution using microorganisms still remain. Prior to implication of one or several source tracking methods, project sponsors must take into account several factors comprising the current state of MST, including project cost and sample turnaround time, as well as sensitivity of a method, the geographic and temporal variability of the microorganisms used, questions concerning microbial host specificity, the ability of a method to quantify the relative proportions of fecal contributors, and for library-independent methods the sources for which markers are available. The detection of chemicals specific to human wastewater or animal sources has emerged as a viable alternative, or at minimum, supplemental line of evidence to support MST methods. The use of chemical indicators have several advantages over most microbial-based tests including, decreased sample preparation and analysis time (Rangdale et al., 2003), and reduced potential for geographic variations. The disadvantages, to date include a lack of correlations with recognized FIB, and the sensitivity of the methods, often requiring the detection of quantities ranging from μg to ng/L. In addition researchers employing chemical methods are interesting primarily in the identification of human-derived waste, through the detection of one or more chemicals unique to wastewater, as most methods have no means with which to identify other fecal contributors. Four major types of chemicals have been investigated as markers of anthropogenic or animal pollution: pharmaceuticals, including drugs, sanitation and hygienic products and cosmetics; fecal stanols, the metabolic by-products of cholesterols; optical brighteners (or fluorescent whitening agents), prevalent in household detergents and toilet papers; and caffeine, an alkaloid chemical commonly found in food and drinks.

Pharmaceuticals

Chemicals used in the pharmaceutical industry have been targeted as both indicators of human wastewater pollution (Leeming and Nichols, 1996) and in studies on the impacts on aquatic ecosystems, including acute and chronic toxicity (Brun et al., 2006; Fent et al., 2006), and effects on growth and reproduction (Binzcik et al., 2004). The successful detection of a variety pharmaceutical chemicals has been demonstrated in freshwaters (Hilton and Thomas, 2003; Roberts and Thomas, 2006), seawater (Weigel et al., 2004), estuaries (Thomas and Hilton, 2004), sediments (Bolz et al., 2001), sewage sludge (Temes et al., 2002) and wastewater effluents (Hilton and Thomas, 2003; Roberts and Thomas, 2006). Recently, the specific application of pharmaceutical chemicals to detect sewage-based human pollution has been investigated (Buser et al., 1999).

Polycyclic musks, chemicals used as fragrance ingredients in a variety of household products, including soaps, detergents, cosmetics, perfumes and air fresheners, have been tested as potential anthropogenic marker (Buerge et al., 2003a). While detection proved successful, abundance in both treated and untreated wastewaters appears problematic for real world applications (Buerge et al., 2003b).

Buser et al., 1999 detected concentrations of the pharmaceutical drug ibuprofen in wastewater influents at concentrations of as high as 3 μg/L, as well as the urinary metabolites, hydroxy-ibuprofen and carboxy-ibuprofin, at even greater concentrations. The effective degradation (>95%) of ibuprofen and its corresponding metabolites during waste treatment processes as determined for effluent samples and laboratory experiments suggests the potential for use as a human-specific chemical marker.

A study of 24 pharmaceuticals and personal care products in treated effluents and nearby wells and creeks was conducted in northwestern Washington. Of these products only 16 were consistently detected in effluents, while only caffeine, nicotine, and Metformin, a drug used to treat diabetes, were detected (at concentrations of 25 μg/L or less) in samples from nearby creeks and wells (Johnson et al., 2004). These results indicate additional testing of these chemicals as potential anthropogenic markers, may be warranted.

In the largest chemical detection test organized to date, researchers from the USEPA and USGS successfully detected 78 different compounds, including antibiotics, food additives and household chemicals, using liquid chromatography/mass spectrometry, in wastewater effluent samples collected across the continental U.S. (Glassmeyer et al., 2005). Results suggested that as many as 35 of these compounds may prove useful as indicators of anthropogenic pollution.

Other chemicals which have been investigated include carbamazepine, a drug used in the treatment of epilepsy, trigeminal neuralgia and bipolar disorder (Clara et al., 2004), propranolol, a commonly prescribed β-blocker (Fono and Sedlack, 2005), and triclosan (Weigel et al., 2004), an antimicrobial compound frequently found in hand soaps.

Advantages of the pharmaceutical approach include the specificity of many of these synthetic chemicals, which lack a potential natural source, to human wastewater, and the apparent success in initial detections. Problems however include the cost of equipment and trained personnel, sensitivity of detection in large bodies of water, as well as requirement of compound use within a locality. A problem which may limit application to large cities where the prevalence of compound available may be higher than in smaller communities resulting from the increased number of people contributing to the local sewage lines. In addition, correlation to

health risks, though FIB, and an understanding of environmental degradation rates remains uncertain for most of these compounds.

Fecal Sterols/Stanols

The detection and use of fecal sterols and stanols as a means of identifying human- or animal-derived fecal pollution has received attention on every continent, with attempted applications in: the United States (LeBlanc et al., 1992; Noblet et al., 2004); Brazil (da Costa and Carreira, 2005); the United Kingdom (Elhmmali et al., 1999); Greece (Bull et al., 2003); Spain (Maldonaldo et al., 1999); France (Tolosa et al., 2003); Japan and Vietnam (Isobe et al., 2004); Hong Kong (Chan et al. 1998); New Zealand (Gregor et al., 2002); Australia (Leeming et al., 1996; Leeming and Nichols, 1996); and even Antarctica (Edwards et al., 1998). Fecal stanols are formed in the guts of animals from the metabolism of sterols, with the metabolic end products often varying in concentration between animal groups based on diet and intestinal flora (Leeming et al., 1996). The sterol, cholesterol, is dominantly metabolized to the 5β-stanol, coprostanol within humans, representing approximately 60% of the total stanols within the human gut (MacDonald et al., 1983). In contrast, the relative quantity of coprostanol present in the feces of non-human animals is typically much lower (Bull et al., 2002). The metabolism of cholesterol by microorganisms within the environment largely generates cholestanol.

The fecal material from animals such as cattle, horses, and sheep contains a greater relative proportion of 5β-campestanol and 5β-stigmastanol, metabolic products of the sterols campesterol and sitosterol, abundant in the diet of these herbivorous animals (Leming et al., 1996). In contrast, cholesterol was found to dominant dog feces, while bird feces contained dominantly cholesterol and sitosterol (Leeming et al., 1997), likely due to the absence of specific

intestinal bifidobacteria necessary to utilize sterols. Leeming et al., 1996, determined that while coprostanol and 5β-campestanol/5β-stigmastanol production is not exclusively limited to humans and herbivores, respectively, the predominance of formation within these groups is significant enough to serve as potential source-specific markers.

At present, a standard method for the analysis of 5β-stanols has yet to be adopted, creating difficulties in correlating results from different studies (Bull et al., 2002). In most procedures a total lipid extract is created using organic solvents, fractionated by thin-layer chromatography or sold-phase extraction and analyzed by gas chromatography or gas chromatography/mass spectroscopy (Bull et al., 2002).

Several ratios based on a comparison of stanol concentrations have been proposed as a means to differentiate fecal sources. Quantification of the isomers 5α- and 5β-stanols led to the creation of a 5β/(5α + 5β) ratio by Grimalt et al., 1990; with a ratio greater than 0.7 believed to be indicative of human fecal pollution. Use of a ratio of coprostanol to 5β-stigmastanol was proposed by Evershed and Bethell 1996, to separate human and ruminant pollution with ratios above 1.5 considered positive for human fecal contamination. Leeming et al., 1997 proposed that if coprostanol / (coprostanol+5β-stigmastanol) was greater than 0.73, pollution may be as much as 100% human in origin, if the ratio is less than 0.28 than herbivores may be responsible for up to 100% of the fecal pollution.

In an application to a California river, Noblet et al., 2004 using the 5α/(5β +5α) ratio determined that sewage was unlikely to be the major fecal contributor. Gregor et al., 2002, determined that effluent from septic tanks contain higher concentrations of fecal stanols than treatment plant effluents. LeBlanc et al., 1992, used coprostanol detected in surface waters and sediments to trace sewage pollution to discharges into the Providence River, in Rhode Island.

Concentrations were found of 0.02-0.22 μg/L in waters and 0.22-0.33 μg/L in sediments, with values decreasing away from the mouth of the Providence River.

While some field applications seem to have been successful, questions concerning the presence of natural sterols in soils and sediments remain, as well as the expense involved in running samples, and concerns with the sensitivity of sterol/stanol detection in larger bodies of water (Pond et al., 2004). In addition, degradation, or lack there of, in deposited sediments remains an unknown (Bull et al., 2002). In the European MC, study β-stanols, coprostanol and 24-ethylcoprostanol (animal origin) were significantly different in human and non-human samples, however the number of incorrect classifications among individual water samples prevented the study from recommending use as an exclusive detection method (Blanch et al., 2006).

Although a direct relationship between pathogenic organisms and fecal sterols has yet to be established (Scott et al., 2002), a correlation between sterols and FIB has been attempted. Coprostanol concentrations of 60 and 400 ng/L were found to correlate with elevated FC concentrations of 150 and 1000 cfu/100 mL, respectively, and enterococci concentrations of 35 and 230 cfu/100 mL, respectively, in sewage-polluted Tasmanian waters (Leeming et al., 1997). In addition, Isobe et al., 2004 found a strong correlation between *E. coli* densities and coprostanol concentrations, although seasonal variations were observed. However, direct measurements of coprostanol concentrations are presently discouraged due to its apparent ubiquity, in trace amounts, in both soils and sediments (Bull et al., 2002), and its tendency to degrade under aerobic conditions, with a half life of about 10 days at 20ºC (Isobe et al., 2004).

Optical Brighteners/Fluorescent Whitening Agents

A potentially rapid and inexpensive method of identifying human pollution is the detection of optical brighteners (OBs) in environmental waters. OBs, also known as fluorescent whitening agents, are organic compounds used in household detergents to "whiten clothing" that absorb long-wave ultraviolet light (365 nm) and reemit it most strongly within the blue portion of the visible spectrum (415 nm to 435 nm). This reemission of blue light from the OBs that bind to clothing in a wash helps to balance the natural yellow color of cotton fabrics, making them appear bright white. Anywhere from 25-95% of the OBs used in a wash cycle are bound to clothing during the wash (Poiger et al., 1998), while the remaining portion of these OBs are discharged from the home within grey water. The mixing of wastewater and grey water in household plumbing systems, allows for the detection of OBs in both septic systems (Close et al., 1989; Boving et al., 2004) and untreated sewage (Poiger et al., 1998). In addition, the industrial use of OBs to whiten a variety of products, including toilet papers (Hagedorn et al., 2005a) allows for additional input into wastewaters. The detection of OBs, with no known non-anthropogenic source, is indicative of human wastewater, with primary entry into environmental waters from septic tank discharges or community sewage lines.

Three methods currently exist for detecting OBs in water, with advantages and disadvantages to each. The simplest and most cost effective approach is to place dye-free cotton pads in environmental waters (or suspended in storm drains) and for a period of time (two to three days) allowing for OBs to bind to the fibers in the pads. With a subsequent exposure to UV light, the pads will fluoresce if OBs were present in the water (Dixon et al., 2005). The major disadvantage of this method is both a low specificity and inability to quantify results. Because numerous other chemicals may be present in the environment that will fluoresce under UV light

(Dickerson et al., 2007a), false positives can be extremely high. The second available method is high performance liquid chromatography (Shu and Ding, 2005). This method possesses a high degree of sensitivity, but is greatly hindered by a high per sample cost resulting from the expense of both the equipment and necessity of employing a highly skilled technician. The final option is the use of a fluorometer, an instrument which varies in initial cost ($1000-$20,000), but is relatively easy to use, and has a reasonably high degree of sensitivity (Hagedorn et al., 2005a).

In the field, OBs have been successfully detected in numerous watersheds in the U.S. (Hagedorn et al., 2003b; McDonald et al., 2006; Dickerson et al., 2007a), Japan (Hayashi et al., 2002), and New Zealand (Close et al., 1989; Gilpin et al., 2002). Often high fluorometric readings correspond to high bacterial counts (Hagedorn et al., 2003b; McDonald et al., 2006; Dickerson et al., 2007a); however, this has not been the case for all studies (Close et al., 1989; Wolfe, 1995). Many of these high OB/ low bacteria situations are likely the result of unidentified chemicals which absorb and emit light at wavelengths similar to that of OBs, such as organic matter, or aromatic compounds (Hagedorn et al., 2005a). One successful fluorometer-based study separated these non-OB sources by exposing samples to UV light (Hartel et al., 2007). OBs are known to degrade under UV light (Kramer et al., 1996), and based on previous unpublished research at Virginia Tech (Dickerson, pers. comm.), prolonged exposure OBs were successfully degraded through UV exposure of the water sample for times varying from 30 minutes to 4 hours. The change in fluorometric readings, from before and after UV exposure represents the approximate amount of OBs present in the water sample.

Dickerson et al., 2007a found difficulties in detecting OBs in river and bay waters resulting from dilution, but did report success in storm drain outfalls. A fluorometer, possesses the additional advantage of the ability to convert to a continuous-flow mode (Hagedorn et al.,

2005b) allowing for the recoding of continuous readings along shorelines or within sections of a waterbody. The use of a fluorometer in continuous-flow mode has been used to successful detect point sources of contamination in the waters of eastern Virginia from faulty septic systems (Hagedorn et al., 2003b) and a broken sewer line (Dickerson et al., 2007a).

The detection of OBs is an emerging chemical method of detecting fecal pollution of human origin. The advantages currently include the relative simplicity of the method and ubiquity with which OBs are used in wash waters, globally. The current drawbacks include dilution of OBs below detection limits in large bodies of water, and the potential for interference from unknown compounds. At present, fluorometry is recommended as a means to trace point sources of human fecal contamination, or as a supplemental detection method with other MST methods (Dickerson et al., 2007a).

Caffeine

Another potential chemical indicator of human wastewater is the alkaloid, caffeine (1,3,7-trimethylxanthine). While present in over 60 plants, most caffeine-containing species are not native to the U.S.; with the exception of the yaupon holly (*Ilex vomitoria*), found in southeastern states (Peeler et al., 2006). As an ingredient in coffee, tea, soda, energy drinks, and chocolates, the average person (globally) consumes about 70 mg of caffeine per day (Buerge et al., 2003b). With even higher consumption levels in the United States (210 mg/day) and the United Kingdom (440 mg/day) the potential as an anthropogenic chemical marker has been noted (Seiler et al., 1999; Siegener and Chen, 2002). While only about 3% of caffeine ingested is not metabolized and excreted in the urine, disposal of unconsumed coffee, tea, soft drinks, or other caffeinated beverage into the sink may represent significant quantities in wastewater (Seiler et al., 1999).

Caffeine has been detected at levels of 37 µg/L and 4.0 µg/L in wastewater influent and effluent, respectively (Paxéus and Schröder, 1996) indicating the relative efficiency of removal via wastewater treatment processes, and lowering concerns of false positive detections from treated effluents. In addition, caffeine concentrations between 100-120 µg/L have been detected in effluent collected from septic tanks (Seiler et al., 1999). Caffeine has also been successfully isolated in freshwater rivers and lakes (Buerge et al., 2003b; Peeler et al., 2006), groundwater (Seiler et al 1999; Chen et al., 2002) and marine waters (Siegner and Chen 2002; Peeler et al., 2006). However, some studies have reported difficulties correlating high concentrations of fecal colifoms and enterococci with detectable levels of caffeine (Buerge et al., 2003b)

Caffeine is generally extracted using liquid-liquid extraction or solid-phase extraction (Rangdale et al., 2003), and detected by gas chromatography–mass spectrometry (GC–MS) or by high performance liquid chromatography–mass spectrometry (HPLC–MS). Results are typically expressed in units of µg/L or ng/L for natural waters.

Buerge et al., 2003b reported that caffeine has potential as a quantitative marker of human wastewater contamination in surface waters in area without high industrial or natural sources. The successful use of caffeine in tracing human pollution to specific tributaries in Boston Harbor has been noted (Siegener and Chen, 2000). However, caffeine has also been detected in presumably uncontaminated marine water samples collected in the Tromsø-Sound, Norway (Weigel et al., 2004). And, in spite of the high rates of removal from wastewater during treatment processes, caffeine has been found in all but the most remote Swiss lakes and rivers at concentrations from 6 to 250 ng/L (Buerge et al., 2003b).

Peeler et al., 2006 successfully detected caffeine levels in small Georgia creeks with known anthropogenic sources. However, as these creeks drained into larger rivers

concentrations were diluted below detectable levels. In addition, caffeine was detected in non-impacted Georgia wetlands containing populations of yaupon hollies, raising questions concerning the abundance of natural caffeine in surface waters. Although, failure to detect caffeine in rural creeks has suggested that the wetlands detection may be unique.

Although caffeine has been successfully linked to anthropogenic sources, use as an indicator of human pollution at best requires additional development. At present, relationships between caffeine concentrations and FIB have yet to be established, and with a half-life of approximately 12 days when exposed to natural light (Buerge et al., 2003b), questions of environmental persistence still remain. In addition, the expense and expertise in equipment and personnel required for detection may limit method application in most localities.

Method Comparison Studies

Three multi-laboratory, independently-evaluated method comparison (MC) studies have been conducted to date. The first, sponsored by the Southern California Coastal Water Research Project (SCCRWP) encompassed a wide variety of both library-dependent and –independent methods attempting to identify fecal sources from southern California (Griffith et al., 2003). The second study, sponsored by the United States Geological Survey (USGS), involved only library-dependent methods, and used *E. coli* isolates collected from West Virginia (Stoeckel, et al., 2004). The final MC study was conducted in Europe focusing solely in library-independent methods and attempted identification of human and farm wastewaters from Spain, France, Sweden, Cyprus, and the U.K. (Blanch et al., 2006).

Southern California Coastal Water Research Project

Goals

The primary goal of the SCCRWP MC study was the evaluation of numerous MST protocols for the detection, or identification of fecal sources present in laboratory-created, blind water samples. With 11 total sponsors, including public health and sanitation districts, a primary focus for all methods involved was the ability to correctly identify the presence, or absence, of human fecal pollution. The additional identification of the dominant fecal material within samples, and ability to detect all sources present in a sample, were applied to largely library-dependent methods, as well as selected library-independent methods with protocols capable of distinguishing between a greater variety of sources. Results of this study were published in a series of seven papers (Field et al., 2003, Griffith et al., 2003, Harwood et al, 2003, Myoda et al., 2003, Noble et al, 2003, Ritter et al., 2003, Stewart et al., 2003) which were unfortunately poorly coordinated, making comparisons across method types difficult.

Study Design

Of the three MC studies, the SCCRWP study tested the widest variety of protocols, selecting 22 different researchers to analyze blind water samples using 12 different MST methods, often with several variations on each. Source-tracking methods involved in the study included both phenotypic and genotypic library-dependent methods, as well a variety of library-independent methods. The SCCRWP study was the only U.S-based MC study to examine the ability of library-independent methods to identify the host origin of fecal pollution.

All methods tested would use water samples created in a laboratory to contain fecal material from a single, or combination of animal sources. These samples were constructed using

feces directly from humans, birds, cattle, dogs or sewage influent collected in Southern California. Only the methods detecting *Escherichia coli* toxin-genes, adenovirus, and the library-based PFGE were not tested by at least two separate laboratories. However within the study, the use of identically named methods could not be considered replicates, as the protocols and data analysis were not standardized, and varied between laboratories.

Library-dependent studies

Methods

Investigations of genotypic and phenotypic library-based methods were performed separately and published in separate papers (Harwood et al., 2003; Myoda et al., 2003). A negative aspect of this approach was the reporting of data in different forms and with different procedural variations applied to final data analyses. This oversight makes the direct comparison across these general method categories increasingly difficult.

The phenotypic methods investigated (Harwood et al., 2003) include four ARA protocols each selecting a different set or subset of indicator organisms, including enterococci, *E. coli*, fecal coliforms, and fecal streptococci. Similar in protocols, the concentrations and antibiotics selected varied for each ARA method, as did the researcher's criteria for considering bacterial growth as positive or negative. Two CSU methods using *E. coli* and fecal streptococci were also included, as were MAR and Kirby-Bauer antibiotic susceptibility tests both testing isolates of *E. coli.*

The genotypic portion of the study (Myoda et al., 2003) selected six researchers to employ one or more of three MST methods. The popular rep-PCR was performed by three researchers all using the fecal indicator *E. coli* and BOXA1R primers. Ribotyping was also

selected for use by three investigators; with two using the restriction endonuclease *Eco*RI on *E. coli* and the third using *Pst*I for enterococci. The final genotypic method selected was PFGE, in which the sole researcher elected to use the enzyme XbaI on isolates of *E. coli*. Although procedures were often similar, variations existed in the similarity thresholds (none were used in phenotypic study) required for a positive identification (e.g. 80% match) as well as both the selection and analyses of banding patterns produced. Unfortunately, the reporting of results differed significantly between the two library-dependent studies.

Results

The averaged sensitivity for all-source isolates for the phenotypic methods was high for the majority of the methods ranging from 100% to 78%, with only the Kirby-Bauer test falling below this range, correctly identifying less than 50% of sources present. However, for all methods, false-positives were high. A later attempt to improve results in the data analysis phase was the application of a minimum detectable percentage (MDP), or a minimum percentage of isolates which must be classified within a source category for that source to be considered relevant. A MDP of 5% showed a slight decrease in false positives, while the application of a 15% cut-off, decreased false positives substantial for most of the methods.

Within the genotypic study, in general, the false negatives were low, however, this was due large to the abundance of false positives reported. Although no water sample contained fecal material from all four potential sources, collectively, the genotypic methods indicated all four sources were present in 37% of the samples analyzed. Overall two of the ribotyping methods, and PFGE produced the lowest numbers of false positives and the highest number of true

negatives however all used 100% similarity thresholds, and the number of total number of unclassified isolates not reported.

In evaluations based strictly for the ability to correctly identify the dominant source in a water sample (unreported in the phenotypic study), most genotypic methods performed reasonably well, correctly identifying the dominant source in greater than two-thirds of the total water samples. When the identified secondary source was included, dominant source identification of five of the seven methods increased by 10% or more.

Library-independent studies

Methods

Researchers employing five different gene-specific library- and culture-independent MST methods were provided with blind, laboratory-created samples to identify all fecal sources possible for the method and present in the sample (Field et al., 2003). Unlike the library-dependent methods, protocols and/or primers did not exist for the genetic marker detection of all four of the possible sources in 4 of the 5 methods used. Selected for this study were: community terminal restriction length polymorphism (T-RFLP), which was the only method with the ability to identify fecal material from all four source categories; *Bacteriodetes* T-RFLP, able to detect human or cattle pollution ; *Bacteriodetes* PCR (OSU) with the ability to detect human, cattle, and dog markers; a second *Bacteriodetes* PCR (USC) method able to distinguish between human and cattle; and a PCR-based detection of *E. coli* toxin genes specific to either humans or cattle.

Although the blind waters samples were heavily spiked with fecal material, all methods were unable to obtain sufficient DNA from at least one of the water samples. In addition, none of the methods were able to quantify contributing sources. While, *Bacteroidetes* T-RFLP

produced no false positives and correctly identified all five water samples with cow manure present, human and sewage were only correctly identified in 25% of spiked samples. *Bacteroidetes* PCR (OSU) correctly identified all human and sewage contaminated water samples. Additionally dog fecal material was correctly identified in all four samples while incorrectly identified in three samples. Cattle were correctly identified in all five samples with one false positive. *Bacteroidetes* PCR (USC) correctly identified cattle in all five cattle spiked samples. However, the detection of human material was considerably lower where only two out of four sewage-, and one of four human-containing samples were correctly identified. *Bacteroidetes* PCR (USC), like the *Bacteroidetes* T-RFLP did not produce any false positives.

The *E. coli* toxin gene protocol correctly identified sewage in all four samples, while correctly identifying two of four human correctly with two false positives. No cattle were identified in any of the samples initially, however after samples sources were revealed, and additional analyses were conducted, all five cattle containing water samples were identified correctly.

Within these methods, a few instances did occur of false positives and false negatives which were attributed to method insensitivity, insufficient specificity, low fecal counts, or errors in method application. Often false negatives were the result of unavailability of primers for a specific source or failure to obtain sufficient DNA.

Human Viral Detection

The subtyping of F+ RNA coliphages, and identification of enteroviruses or adenoviruses were evaluated in the identification of blind water samples containing human and animal fecal waste (Noble et al., 2003). Results of the study indicated that all methods were sensitive to the

detection of sewage contaminated samples, with three methods correctly identifying the human source in 75% of samples, and F+ coliphage typing correctly identifying 100% of water samples containing sewage. False positives were absent for 3 of the 4 methods, with only the F+ coliphage incorrectly detecting a human signature in 1 of the 12 samples analyzed. Problems for all of the methods arose when attempting to identify samples spiked from a single human fecal source, as opposed to the composite type sample found in untreated sewage, with methods only identifying 0 to 25% of samples.

Problems in sensitivities for non-sewage human samples were attributed primarily to virus levels below the detection limit, or the absence of these viruses in the individual human samples used for this study. Often not found in the majority of the population, these viruses are more readily detectable from the population as a whole. However, using fecal material from portions of entire human population, as found in a sewage sample, allowed for the occurrence and subsequent detection of these viruses in the majority of the sewage contaminated samples. Application of this method in natural waters may be problematic resulting from high dilution, thus requiring collection and processing of large water samples (5-10 L).

Conclusions

Conclusions of the SCCWRP MC study include the need to: optimize existing MST methods, improve quality control techniques in laboratories, evaluated library size issues, and geographic variability. As a result SCCWRP recommended the use of multiple methods when attempting to perform MST in natural waters. In addition, issues of contamination in samples were raised with the library-independent studies. As each water sample provided to researchers

was a division of a larger sample created to keep blind samples as uniform as possible, if contamination was a factor, results from all methods in this study are potentially invalid.

United States Geological Survey

Goals

The goal of the USGS MC study was an evaluation of the ability of seven commonly used source-tracking protocols to correctly identify *E. coli* isolates of unknown origin using identical known-source libraries (Stoeckel et al., 2004). Method effectiveness was determined by the correct source classification of pre-selected individual isolates, as opposed to the general source category identification of randomly selected isolates from a laboratory inoculated water sample, as was the case in the SCCWRP study. The success of a given protocol was assessed through the ability to correctly classify blind isolates: already present within the library (REPLICATE), not present but from a source animal represented in the library (ACCURACY), and from a source animal not present in the library (RINGER).

Study Design

One prominent source-tracking laboratory was selected for each method to perform the necessary protocols for ARA, CUP, ribotyping (with *Hin*dIII or *Eco*RI), PFGE, BOX-PCR, and rep-PCR. Known-source libraries were constructed from fecal samples collected within Berkeley County, WV, where scat was collected and *E. coli* isolates were obtained centrally, for distribution to individual laboratories. Libraries of 900 identical isolates were constructed from nine sources (human, beef cattle, dairy cattle, chickens, swine, white-tailed deer, horses, dogs,

and Canada geese). However, researchers using CUP, Ribotyping-*Hin*dIII, and rep-PCR elected to use only a smaller subset of 630 isolates drawn from the 900.

Focusing strictly on the ability of a method to correctly classify blind isolates, three distinct collections of isolates (REPLICATE, ACCURACY, and RINGER), and three source-level splits (2-way, human and non-human; 3-way, human, livestock and wildlife; and 8-way, human cattle, chickens, swine, deer, horses, dogs, and geese) were used to assess method effectiveness. Identical sets of blind isolates, collected post-library construction were provided to researchers for source identification.

Results

The REPLICATE set should have allowed for the correct identification of the majority of isolates tested (although ARCCs were not reported), as all 26 isolates were present within the libraries of each method. However, the results were extremely varied and the percentage of correctly identified isolates using the 8-level split ranged from 13% (Ribotyping-*Hin*dIII) to 100% (PFGE).

A total of 150 isolates, composing the ACCURACY test, were provided to researchers to evaluate method classification of non-library isolates. Only the ARA, CUP, Ribotyping - *Hin*dIII, and BOX-PCR attempted to classify all isolates provided, with the highest classification rate of only 27% for the 8-way split. Of those methods employing thresholds, Ribotyping-EcoRI only attempted 5% of isolates provided, but correctly classified 90%.

Classification rates increased for all methods using the 3 source-level split. The highest averages of those attempting to classify all isolates was BOX-PCR at 48%, and of those not attempting all isolates, Ribotyping-EcoRI successful classifying 100% of the 6% of isolates

attempted. Classification based on 2 source-levels again increased classification rates. The highest average of those attempting to classify all isolates was REP-PCR at 77%, while of the methods not attempting the classification all isolates, Ribotyping-EcoRI again, correctly identified 100% of the 6% of isolates attempted.

The RINGER test was not applied to all protocols, as several of the methods force isolates into one of the predetermined categories, and did not possess a means to separate those which do not resemble, or closely fit a pattern present in the library. The most successful protocols for placing ringer isolates into an unknown category were Ribotyping-EcoRI and PFGE, which correctly failed to classify 24 and 16 of the 24 isolates respectively. The successful use of an unknown category, as pointed out by the authors, is of greatest importance when defining categories based on species.

Conclusions

Library sizes, although greatly increased in size from the SCCWRP study, were still deemed too small to accurately represent an entire county watershed, with scats samples from only 20 host organisms used per source category. As the diversity of *E. coli* genotypes and phenotypes present in the scat of warm-blooded animals currently remains poorly understood, the standardization of isolates used in the library and challenge sets of this study allowed for increased confidence in contrasting results.

Stresses in quality assurance/quality control (QA/QC) proposed in the conclusion of the SCCRWP study should have prompted the enlistment of at least one replicate lab for each protocol tested. The employment of replicate protocols would have provided information in method reproducibility beyond the results of the REPLICATE test. Temporal strain variability,

as acknowledged by researchers, may have also played a role in the low number isolates correctly identified from the ACCURACY set, which was collected nine months after samples used in library construction.

The single most significant result of this study revealed the lack of knowledge among the source-tracking community concerning known-source library construction. As pointed out in the author recommendations, better host species representation is needed in the libraries. An additional recommendation was the development of better methods, although in light of the library representativeness problems, a retesting of currently methods may be warranted once greater understanding of how to build a known-source library has been established.

European Union Project

Goals

A MC study was conducted in Europe to assess the abilities of 16 library-independent methods to identify wastewater samples isolated from sewage or farm sources (Blanch et al., 2006). All blind water samples contained considerably more fecal pollution that would be found in natural waters in order to assess the methods ability under ideal conditions, with the desire to first identify methods or a combination of methods able to successfully discriminate sources. The methods selected included: enumeration/measurements of FC, enterococci, clostridia, sorbitol-fermenting and total bifidobacteria, somatic and F-specific RNA coliphages; bacteriophages which infect *Bacteriodes fragilis* RYC2056 or *Bacteroides thetaiotaomicron;* identification of four fecal sterols; the detection of *Bifidobacterium adolescentis*, and *Bifidobacterium dentium*, genotyping of F-specific RNA phages, and CUP profiles of FC and enterococci.

The desires of the European Union sponsors were to:

- differentiated between human and non-human fecal sources;
- use highly polluted waters so, dilution was a minimum factor;
- study a wide geographic area to find the most discriminating and geographically consistent tracers;
- draw from the results of each method in an attempt to define successful ratios of multiple source indicators; and
- employ statistical or machine learning methods (discriminant analysis, k-nearest neighbor, and artificial neural networks) to develop applicable predictive models.

Study Design

Over the two year study, researchers each analyzed a portion of 230 blind water samples (114 exclusively human from WWTP influent, hospital wastewater, and military camp wastewater; and 116 of animal-origin from slaughterhouse effluent and farm slurries from cattle, sheep, pigs, horses and poultry) collected from Spain, France, Sweden, Cyprus, and the U.K. Collected water samples were divided and shipped to laboratories across Europe for analysis.

Results

FC, enterococci, clostridia, somatic coliphages, and bifidobacteria (total counts) were found in both humans and animals. While differences found in the abundance of these microbes generally favored animals, this was believed to be a result of animal samples containing overall higher concentrations of these enteric microbes. No geographical variations were detected.

Results indicated that phages infecting *Bacteroides fragilis* RYC2056 and *Bacteroides thetaiotaomicron,* and sorbitol-fermenting bifidobacteria where either not present in animal fecal samples, or present at significantly lower concentrations. No geographic variations were observed, although the *Bacteroides thetaiotaomicron* infecting phages from the UK produced considerably less visible plagues than those from other parts of Europe.

The genotyping of F-specific RNA coliphages indicated that genotypes I and IV were significantly more abundant, and genotype II significant less abundant, in animal than human fecal samples. However, all three genotypes were detected in multiple animal- and human-derived samples. Genotype III, believed to be indicative of human wastewater, was detected at a slightly higher rate in humans than animals, but this was not deemed statistically significant.

A higher diversity of phenotypic profiles of FC and enterococci were detecting in fecal samples from humans than animals. And, although this was geographical consistent, numerous overlaps in patterns were observed making differentiation between sources difficult.

When enterococci from the samples were speciated, *E. faecium* and *E. faecalis* were found to be the dominant species in human samples, and *E. hirae* was dominant in animal samples. This observation was geographically consistent, however species were generally not unique to humans or animals, and enough overlap in the relative percentages between sources were found to prevent meaningful use as an MST tool.

The detection of the potentially human-specific *Bifidobacterium adolescentis*, and *Bifidobacterium dentium* in samples found them absent in 45% and 6.3% of human samples, respectively, and present in 9.5% and 24.5% of animal samples, respectively. The study concluded that neither of these bacteria are useful tracers by themselves.

Concentrations of 24-ethylcoprostanol, epicoprostanol and cholestanol were deemed significantly different between human and animal sources. However, a number of overlaps occurred between sources, inhibiting the ability of study participant to calculate a source-specific reference concentration.

In an attempt to apply a combination of methods to improve source identifications, results from all 26 methods tested were combined. Using DA, 100% of water samples were correctly identified as human or non-human. However the study acknowledged that conducting this many tests is not feasible for a field application. Generation of a predictive model using the ratio of somatic coliphages to bacteriophages infecting *B. thetaiotaomicron* GA17 plus somatic coliphages was deemed the best of models attempted using machine learning/neural networks on a limited number of methods.

Conclusions

Conclusions of this study include that: none of the microbial or chemical methods tested were deemed optimal when used alone, however, use of multiple methods showed potential; and only the bacteriophage infecting *B. thetaiotaomicron* GA17 was consistently specific to humans. However, geographic variations are suspected, as seen in plaque clarity.

The major drawbacks of study include: the use of only highly concentrated fecal samples, which are unlikely to be present in environmental waters; and the failure to include fecal samples from wildlife (only livestock and human), which have the potential to significantly impact watersheds. However, future testing is anticipated by the European Union of many of the library-independent methods and predictive models developed in this study under the more dilute conditions prevalent in natural waters.

Lessons of the Method Comparison Studies

Lessons of the MC studies should be taken with an understanding that the source-tracking community did, and still does, not understand the diversity of bacteria or abundance of viruses and chemical markers likely to be present in host organisms and/or natural waters. Questions relating to MST that must be addressed in the future include issues of fecal abundance and source-specificity of tracers. For those engaged specifically in bacterial source tracking, issues of host and environmental strain prevalence, library-size, as well as temporal, seasonal, geographic, and within-host variability of fecal indicator communities must also be addressed.

Limitations and Unanswered Questions in Source Tracking

Strain Variability

After more than a decade of research into MST methods and applications, the most effective means to construct a known-source library, or collect isolates to test a suspected library-independent method, still remains largely unknown. Although certain aspects have received minor attention, library-based methods have relied on untested assumptions, and generally ignored research that does not support the current direction of source-tracking. The sole purpose of a known-source library is to provide a collection of 'fingerprints' in which to reference environmental isolates of unidentified origin. Isolates of known-origin are most commonly obtained through the collection of fecal scat, sewage, or septage samples; however questions remain concerning how, when, and where isolates should be obtained for library construction or marker testing.

Known-Source Collection

Only a few publications report the number of isolates obtained from each scat/sewage sample, although this should be reported in every source tracking paper, including those based on library independent methods. The studies that report these numbers typically used from 1 to 5 isolates per sample (Dombek et al., 2000; Albert et al., 2003; Hagedorn et al., 2003a; McLellan et al., 2003; Johnson et al., 2004; Moore et al., 2005; Dickerson et al., 2007a) although some have used numbers as high as 20 to 32 (Graves et al., 2002; Ahmed et al., 2005a). It has been suggested that a dominant *E. coli* strain exists in individuals (McLellan et al., 2003), however the results of the SCWWRP study seemed to indicate that five isolates from each fecal samples is probably inadequate to account for the variability within one host organism (Stewart et al., 2003). Classification rates were generally low among all methods when attempting to identify additional isolates collected from the library-building scat samples. The single exception was the success of classifying isolates, human in origin, an anomaly with multiple probable explanations. One possibility is the imbalance of the libraries, as isolates of human origin were represented by twice as many isolates as any other single category, and represented 40% of the total library, which can tend to bias classifications (Ritter and Robinson, 2004). Another plausible explanation is that the use of 60 isolates from a single untreated sewage sample was sufficient to account for the variability within that one sample.

Studies using PFGE have shown that only an average of 12 distinct patterns exists within a single sewage sample of less than 1 L (Simmons et al., 2002; O'Brien et al., 2005). A single sewage influent sample likely consists of fecal material from a relatively small number of individuals; however no studies have been conducted to determine the diversity of isolates in sewage samples collected over regular hourly, daily, or seasonal intervals from a single sewage

treatment plant. The USGS study, also using 5 isolates from each fecal sample, suffered from similar poor classification rates among all methods involved. Although challenge set isolates were obtained from different scat/fecal samples than those used for library construction, the considerably larger (than the SCCWRP library) 630- or 900-isolate library still proved insufficient.

Several studies have reported higher than expected diversity in *E. coli* isolated from farm animals. In a sample of 481 isolates, 240 different ribotype patterns were observed in two herds of cattle (Jenkins et al., 2003). Isolates collected from cattle, swine, chicken and horses in Georgia and Idaho, displayed 213 ribotypes in a collection of 568 *E. coli* isolates (Hartel et al., 2002). A sampling of *E. coli* diversity from pigs found an average of 2.4 phenotypes per 9.5 strains tested from an individual (Katouli et al., 1995). Using rep-PCR, Lu et al., 2005, predicted that approximately 60 different genotypes may be present in single herd of swine or dairy cattle on a given day. Analysis of two collections taken from manure holding tanks on these swine or cattle farms yielded estimates that between 70-158 genotypes were likely present (Lu et al., 2005). Using PCR-based methods, a required 32 isolates per individual was deemed sufficient for fecal characterization of sewage, horse, cow, gull, and dog fecal samples (Seurinck et al., 2003).

These studies indicate that a larger sampling of isolates per fecal sample may be necessary for account for the variability within a single individual, with size likely dependent on the MST method employed. The possibility of such variability within a single host raises questions concerning how much of the reported geographic variability is a result of actual differences in the individuals from different areas, and how much is by pure chance, as studies generally collected far less than 32 isolates per individual for analyses. Although largely uninvestigated,

the possibility of dominant strains within an individual may still prove to be geographically dependent as seen in one library-independent method (Hamilton et al., 2006). However, once outside of the host animals, some strains may come to dominate in an environmental setting such as a septic system (Gordon, 2001) or irrigation water (Lu et al., 2004), and the once dominant intra-host strain may display an increasing die-off rate relative to less prevalent strains under environmental conditions.

One suggestion to help increase library diversity is the collection of both fresh and dry manure from livestock (Weaver et al., 2005). Dry manure is generally more abundant on farms, and *E. coli* and fecal streptococcus populations showed a decrease in numbers over time in scat from horse, sheep and cattle populations. Thus, the collection of both dried and fresh scat may increase library diversity and yield information about the differential survival of strains outside the host organism. However, in a different study by Graves et al., 2007 used the Biolog™ system to identify enterococci at the species level in both dry and fresh livestock manure. No significant differences were detected in species survival rates within the environment. The conclusion was made that, the frequency of enterococci species likely to enter waterways would be consistent with those found in collected manure samples and used for the construction of known-source libraries. However, the possibility exists that differences in strain survival are present, but undetectable using the Biolog™ system when only looking at the species level.

Strain Overlap

Further compounding source identification is the possibility of strain overlap between host-sources. Using rep-PCR, shared banding patterns have been observed between gulls and sewage samples (McLellan et al., 2003) and overlapping phenotypic profiles were observed for human

and livestock in the European MC study (Blanch et al., 2006). The potential acquisition of new strains from external sources also requires the consideration of source-tracking researchers. Although a study from Kenya isolated distinct strains of *E. coli* from children living in close proximity to chickens (Kariuki et al., 1999), other publications point to the possibility of acquiring and incorporating environmentally persistent strains into an individual's intestinal flora. Gulls have been shown to contain an incredible diversity of strains (Fogarty et al., 2003), possibly resulting from the feeding habits of these birds that frequent areas of polluted rivers, water treatment facilities or areas containing human trash, such as landfills. New strains may also be obtained from ingested food and water (Neidhardt et al., 1987). Further supporting the idea of this acquired diversity is a study that found the range of genotypes present in a single mouse increased as the age increased (Gordon, 1997), suggesting the acquisition of long-term enteric residents over time. Additionally, *E. coli* from an isolated subset of a litter of pigs found more similarity among *E. coli* from the isolated litter than other members of the same litter that had mixed with other pigs (Katouli et al, 1995).

Geographic Variability

Issues of geographic variability in fecal indicator strains isolated from identical host species have been a major concern of library–dependent methods, and a largely untested aspect of library-independent methods. A study of 568 *E. coli* isolates from Idaho and Georgia (3 locations) from cattle, swine, chicken and horses found that similarity in strains decreased with increasing geographic distance for cattle and horses (although not for swine and chickens; Hartel et al., 2002). Additional difficulties have also been encountered in attempts to distinguish between farm animals in Florida over a large region (Scott et al., 2003). Using isolates collected

from mice inhabiting two geographically isolated farms, some *E. coli* genotypes that were dominant at one farm proved to be very rare at the other, thus demonstrating localized geographic variability (Gordon, 1997). It has been widely assumed that geographic variability could be circumvented with library-independent approaches (and field applications continue to be reported), although this assumption remains largely unproven. One recent study demonstrated that sets of *E. coli*-based library-independent markers found in waterfowl were geographically limited, although it is not known if the reported regional specificity will apply to other sources or other organisms (Hamilton et al., 2006).

Seasonal/Temporal Variability

Additional questions still remain concerning when to collect known-source samples for library construction. Issues of seasonal or temporal variability have received only minor attention, and application of these principles has generally not been reported within the literature. Inadvertently, the USGS MC study has raised the concern for seasonal variability of enteric communities within host organisms. The nine-month separation between the collection of library isolates and the collection of what would be the challenge set of isolates was suggested as a possibility for the poor performance of those methods tested,. A study of sewage-isolated *E. coli* saw significant changes in strain prevalence over four months (Pupo and Richardson, 1995), although sample sizes were generally small. Similar results were seen in populations of mice over a period of 6 months (Gordon, 1997). In addition, dietary changes have been shown to shift enteric populations in rats, pigs, cattle and humans (Silvi et al., 1999; Leser et al, 2000; Russell et al., 2000; McDonald et al., 2001). Katouli et al., 1995, tested isolates obtained from individual pigs through their life and concluded that *E. coli* populations changed over the lifetime of the

pigs. A dietary influence on *E. coli* populations was found in deer, with the nutritionally diverse wild deer displaying a greater variety of genotypes than those in captivity (Hartel et al., 2003). More diversity was seen in dairy cattle or swine manure holding tanks than from individuals on the farm, and may represent a temporal variability as enteric populations shift, and tanks receive additional input throughout the year (Lu et al., 2005). For known-source collections assuming that as the seasonal food supply of wildlife changes enteric populations undergo a shift as well, warrants the required simultaneous collections of water and scat samples.

Environmental Persistence

Further compounding the efforts of an MST researcher to collect isolates representative of the fecal pollution in an area may be bacterial survival in soils, sediments or on vegetation. Both *E. coli* and enterococci have been found to survive at least 19 weeks from cow manure in soil at temps from 9-21°C (Lau and Ingham, 2001). The pathogenic *E. coli* O157:H7, following land application, was recoverable from up to 6 weeks post-application on vegetation (Avery et al., 2004) and 15 weeks in soil (Ogden et al., 2002). Cow manure is believed to provide a substantial *E. coli* contribution for overland runoff for periods greater than 30 days (Muirhead et al., 2005). Such increased persistence in the soil and potential to enter natural waters in surface runoff may require MST researchers to begin collecting source samples in a watershed several months before collecting water samples.

Within streambed sediments, *E. coli* has been reported to remain viable for up to 6 weeks (Jamieson et al., 2005), lending further support to the idea of early known-source collections. Survival of *E. coli* in filtered irrigation water was found to exceed 8 weeks, however survival rates differed between strains, with one of the three strains tested increasing in relative

percentage over the 8 week period (Lu et al., 2004). The survival of *E. coli* for greater than 28 days in coastal sediments has been observed (Craig et al., 2004) noting that survival is improved at lower temperatures. Resuspension of sediments and release of bacteria into the water column is considered to be primarily the result of wave action (LeFevre and Lewis, 2003). This may require researchers to begin library collecting known-source samples even earlier in colder areas or in field projects spanning primarily winter months.

Differences have been observed between *E. coli* strain persistence in freshwater (Anderson et al., 2005). Failure to apply these differences in attempts at modeling water systems can misrepresent the true fecal inputs (Barnes and Gordon, 2004). However, as the relative risk of a water body is currently determined for fecal pollution strictly based on the prevalence of indicator organisms (USEPA, 1986), source tracking protocols employing the recommended fecal indicators should provide guidance to implement management practices aimed at lowering bacterial counts nevertheless.

Library Representativeness

The usage of average rate of correct classification (ARCC) and average frequency of misclassification (AFM) to assess predictive abilities has resulted in the construction and application of potentially undersized and non-representative known-source libraries. Although few studies have been conducted to date, results have shown that small libraries tend to have better internal classification rates than larger libraries (Dickerson et al., 2007b), but are generally less capable of correctly identifying non-library isolates than their larger counterparts (Wiggins et al., 2003). In the post-MC field of source-tracking, all testing or field applications of MST methods should now report an assessment of method/library competence using a validation set of

known-source isolates (Harwood et al., 2003; Stoeckel et al., 2004: Moore et al., 2005; Dickerson et al., 2007a). However, since the publication of the final U.S.-based MC study in 2004, the willingness to employ a validation set in library-based applications has been largely ignored. Application has been sporadic, ranging from collection of isolates specifically for challenging a known-source library (Moore et al., 2005), to use of a holdout set of library isolates (Wiggins et al., 2003; Carroll et al., 2005), to the failure to apply a validation set of any kind (e.g. Ahmed et al, 2005a; Edge and Hill, 2005; Shehane et al., 2005).

The optimal size of a known-source library for a given watershed is likely dependent on the MST method(s) selected. Any future attempts to determine library size should now be accompanied by a validation set. Moore et al., 2005 used a set of 97 *E. coli* and 99 *Enterococcus* isolates to assess the capabilities of ribotype library consisting of 997 *E. coli* isolates, as well as two ARA libraries of 3657 *Enterococcus* and 3477 *E. coli* isolates. Using ARCC as the desired level of proficiency, only the *Enterococcus* ARA library was able to correctly classify validation set isolates equally both internally and external, although ARCC was generally low for the library itself (44%). The recommendation of validation set application from this paper is consistent with the MC studies. Using a hold-out sample of isolates from an ARA library, Wiggins et al., 2003 reported that only the largest library, of 2931 isolates, was able to reach acceptable ARCC levels. Using a smaller, 717 isolate *E. coli* ARA library with a high correct classification rate (89.2%) and a human/non-human split, a hold-out, cross-validation ARCC of 93.2% was achieved (Carroll et al., 2005), however fecal samples were obtained from a potentially undersized sampling of fecal sources.

Based primarily on reasons discussed in previous sections, some general guidelines should be applied to validation set applications. The potential for environmental persistence indicates that

known-source fecal acquisition should begin prior to the collection of initial water samples. Efforts to avoid potential seasonal variability warrants the collection of all validation set isolates simultaneously with both the library and water sample isolates. None of the validation set isolates should be obtained from the same fecal scat sample used to construct the library. Multiple isolates from the same sample may possess identical genotypes or phenotypes, and a potential biasing of challenge set isolates may occur. Use of the internal classification ability (ARCC and AFM) as seen in Moore et al., 2005, should serve as the desired external classification standard for assessing library representativeness.

Available Source Tracking Options (and defense of ARA)

Early in the development of field applicable MST protocols, the idea was introduced that antibiotic resistance patterns provided less stability for library construction than those based on genotypic patterns (Parveen et al., 1999), and thus represented an inferior long-term source-tracking tool. However, since that time, evidence has not emerged supporting the temporal instability of antibiotic resistance patterns, nor has evidence been reported to advocate the stability over time of genotypic patterns in MST applications. The only MST publications to assess temporal stability concluded that the antibiotic resistance patterns used for known-source library construction were stable over periods of one and two years (Wiggins et al., 2003; Dickerson et al., 2007a). In the short-term, antibiotic resistance was found to persist in enteric bacteria isolated from swine manure lagoons (Graves, pers. comm.). Although debated within the MST community, the long-term stability of antibiotic resistant enteric bacteria has been clearly demonstrated by medical researchers.

One such study has shown that antibiotic resistance is prevalent within some (unexposed) wild animal populations, concluding that a discontinuity in the supply of antibiotics to animals harboring resistant strains is unlikely to decrease populations (Gilliver et al., 1999). Antibiotic resistant genes are considered stable even in the absence of obvious antibiotic selections pressures (Salyers and Amábile-Cuevas, 1997), making them difficult to eliminate from enteric populations. In a comparative study of oral streptococci obtained from healthy Mexican and Cuban volunteers, researchers found that in Mexico, where antibiotics are freely available (and used in agriculture), multi-drug resistance was high. Streptococci isolates from Cuba, where antibiotics have been tightly controlled since 1990 (having a very limited availability and no agricultural application), displayed similar resistance profiles to those found in Mexico, implying retention of resistance genes over more than a decade (Díaz-Mejía et al., 2002).

Genes encoding resistance to the antibiotic chloramphenicol were found in poultry broiler litter despite its discontinued use for greater than 15 years (Lu et al., 2003). Additional evidence has refuted claims that the retention of resistance genes reduces bacterial fitness (Björkman et al., 2000). In the environment antibiotic resistance genes have been found to provide a selective advantage in systems polluted with heavy metals (McArthur et al., 2000) or organic solvents (Li et al., 1998). Thus, the antibiotic resistance would seem to provide suitable stability (although possibly geographically restrictive) for MST employment over extended periods of time. Genotypic libraries have shown geographic limitations, however, strain variability and limited sample sizes may have unfairly skewed these results; larger libraries will be required to fully test this contention.

Reproducibility has proven to be problematic for both the genotypic and phenotypic methods. Poor method reproduction was seen in six of the seven protocols used in the USGS

study (Stoeckel et al., 2004). Of the limited studies addressing the duplication of results, problems were found in both ARA and ribotyping methods when performed in the same lab over multiple days (Moore et al., 2005). Only limited research, however, has been conducted to assess the variability of identical protocols in generating fingerprint patterns across independent laboratories (Lefrense et al., 2004). The relative complexity of genotypic methods, as compared to most phenotypic methods requires increasing quality control and expertise, and may limit application to only select laboratories. As widespread application to water quality issues is a major goal of source-tracking, reproducibility across labs is becoming increasingly important.

Using any MST method also involves dealing with the cost per isolate of performing analyses (factoring in the price of equipment, materials, and trained staff). These costs need to be considered for watershed analysis when evaluating appropriate library and sample size concerns. Pulsed-field gel electrophoresis, although considered by both MC studies to be one of the most accurate options, also features a high per isolate cost and a very slow turnaround time (Hager, 2001a), potentially limiting construction of a representative library and analyses of a sufficient number of isolates per sample. With varying degrees of severity, these fiscal and temporal problems exist for numerous genotypic methods. One source-tracking researcher using ribotyping has drawn criticism for using only two to three isolates per water sample (Hager, 2001b). Phenotypic methods, such as ARA, have received an increased application for field studies as seen at the ASM 2006 General Meeting with seven posters present, more than double that of any other single method. The requirements of only basic laboratory equipment, minimal personal training, and a turn-around time of about one week (Hager, 2001a) allows for the necessary testing of the large number of isolates required for simultaneous library building and water samples analyses. Additionally, the ability of phenotypic methods to produce rapid and

high-volume results can allow for the simultaneous construction of a known-source library with a field application; a limiting factor in the usage of many genotypic methods. While library-independent methods may offer a less expensive alternative to the analyses of thousands of isolates, these methods still have problems with rates of false-negatives, due to scarcity of the target bacteria or viruses in open waters, from host sources (Soule et al., 2006), or outside of a geographic range (Hamilton et al., 2006). Additionally, means of quantification (Field et al., 2003) and correlation to fecal indicators has yet to be established (Jiang et al., 2001). However, in their current forms these methods still make desirable secondary methods (when applicable) for validation of results.

The Future of Source Tracking

In the future, MST will likely move away from library-dependent methods in exchange for the less expensive, and more rapid library-independent means of assessing and identifying fecal pollution in waterways (if such can be found). Currently, however, most library-independent methods are restricted by the undeveloped protocols for multiple potential sources, inadequate sensitivity (leading to high false-negatives), inability to quantify source contributions (Field et al., 2003; Noble et al., 2003), and possible geographic limitations (Hamilton et al., 2006). Questions still remain as to how the detection of a bacterial species, viruses, chemicals, or gene-specific markers can be quantitatively related to the FIB used by state agencies. In addition, the major supporters of MST research, the European Union, USEPA, USGS, and SCCWRP are at present working independently. Failure of these agencies to collaboratively set goals and objectives in the future will continue to fragment the field of MST and hinder advances in the development of consistent, reproducible and widely accessible methods.

The possibility exists that as methods improve, public health standards for fecal indicators as set by USEPA, 2002 may be redefined in terms of percentage of animal contributions. However, the current TMDL projects continue to use quantifications of animal fecal inputs, a request which continues to remain beyond the scope of the available library-independent methods. Current options still involve the use multiple methods (both library-dependent and independent) as recommend by Stewart et al., 2003 and applied by McDonald et al., 2006 and Dickerson et al., 2007a for cross-validations and increased confidence in results.

References

Ahmed, W., R. Neller, and M. Katouli. 2005a. Host species-specific metabolic fingerprint database for enterococci and *Escherichia coli* and its application to identify sources of fecal contamination in surface waters. Appl. Environ. Microbiol. 71:4461-4468.

Ahmed, W., R. Neller, and M. Katouli. 2005b. Evidence of septic system failure determined by a bacterial biochemical fingerprinting method J Applied Microbiol. 98:910–920.

Albert, J.M., J. Munkata-Marr, L. Tenorio, and R.L. Siegrist. 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. Environ. Sci. Technol. 37:4554-4560.

Albert, J.M., J. Munkata-Marr, L. Tenorio, and R.L. Siegrist. 2004. Experimental and statistical evaluation of bacterial source tracking in a subalpine region by rep-PCR DNA fingerprinting On-Site Wastewater Treatment X, Conference Proceedings, 21-24 March 2004 (Sacramento, California USA). pp. 470-481.

Anderson, K.L., J.E., Whitlock, and V.J. Harwood. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl. Environ. Microbiol. 71:3041-3047.

APHA. 1998. American Public Health Association, American Water Works Association, Water Pollution Control Federation. Standard methods for the Examination of Water and Wastewater. 20th ed. American Public Health Association, Washington D.C.

Avery, L.M., K. Killham, and D.L. Jones. 2005. Survival of *E. coli* O157:H7 in organic wastes destined for land application. J. Appl. Microbiol. 98:814-822.

Baffaut, C., C.A. Carson, and W. Rogers. 2005. Little Sac River watershed: bacterial source tracking analysis. FAPRI-UMC Report #06-05 May 2005.

Barnes, B. and D.M. Gordon. 2004. Coliform dynamics and the implications for source tracking. Environ. Microbiol. 6:501-509.

Beekwilder, J., R. Nieuwenhuizen, A. H. Havelaar, and J. van Duin. 1996. An oligonucleotide hybridization assay for the identification and enumeration of F-specific RNA phages in surface water. J. Appl. Bacteriol. 80:179-186.

Bedendo, J. and A.C.C. Pignatari. 2000. Typing of *Enterococcus faecium* by polymerase chain reaction and pulsed field gel electrophoresis. Braz. J. Med. Biol. Res. 33:1269-1274.

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

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

Binzcik, G.A., J.L. Gates, L.E. Gray Jr., L.J. Guilette Jr., M.K. Horton, A.S. Kolok, C.S. Lambright, E.F. Orlando, A.M. Soto. 2004. Endocrine-disrupting effects of cattle feedlot effluent on an aquatic sentinel species, the fathead minnow. Environ. Health Perspect. 112: 353-358.

Björkman, J., I. Nagaev, O.G. Berg, D. Hughes, and D.I. Anderson. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science. 287: 479-1482.

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

Boehm, A. B., J. A. Fuhrman, R.D. Mrše, and S.B. Grant. 2002. Tiered approach for identification of a human fecal pollution source at a recreational beach: Case study at Avalon Bay, Catalina Island, California. *Environ. Sci. Technol*. 37:673-680.

Bolz, U., H. Hagenmaiser, and W. Komer. 2001. Phenolic xenoestrogens in surface water, sediments, and sewage sludge from Baden-Wurttemberg, south-west Germany. Environ Pollut. 115:291-301.

Bonjoch, X., E. Ballesté, and A.R. Blanch. 2004. Multiplex PCR with 16S rRNA gene-targeted primers of *Bifidobacterium spp*. to identify sources of fecal pollution. Appl. Environ. Microbiol. 70:3171-3175.

Booth, A.M., C. Hagedorn, A.K. Graves, S.C. Hagedorn, and K.H. Mentz. 2003. Sources of fecal pollution in Virginia's Blackwater River. J Environ. Eng. 6:547-552.

Boving, T.B., D.L. Meritt and J.C. Boothroyd. 2004. Fingerprinting sources of bacterial input into small residential watersheds: Fate of fluorescent whitening agents. Environ. Geol. 46:228-232.

Bower, P.A., C.O. Scopel, E.T. Jensen, M.M. Depas, and S.L. McLellan. 2005. Detection of genetic markers of fecal indicator bacteria in Lake Michigan and determination of their relationship to *Escherichia coli* densities using standard microbiological methods. Appl. Environ. Microbiol. 71:8305-8313.

Brion, G.M., J.S. Meschke, and M.D. Sobsey. 2002. F-specific RNA coliphages: occurrence, types, and survival in natural waters. Water Res. 36:2419–2425.

Brun, G.L., M. Bernier, R. Losier, K. Doe, P. Jackman, and H.B. Lee. 2006. Pharmaceutically active compounds in Atlantic Canadian sewage treatment plant effluents and receiving waters, and potential for environmental effects as measured by acute and chronic aquatic toxicity. Environ. Toxicol. Chem. 25:2163-76.

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

Buerge, I.J., H.R. Buser, M.D. Muller, and T. Poiger. 2003a. Behavior of the polycyclic musks HHCB and AHTN in lakes, two potential anthropogenic markers for domestic wastewater in surface waters. Environ. Sci. Technol. 37:5636-5644.

Buerge, I.J., T. Poiger, M.D. Muller, H.R. Buser. 2003b. Caffeine, an anthropogenic marker for wastewater contamination of surface waters. Environ. Sci. Technol. 37:691-700

Burnes, B.S. 2003. Antibiotic resistance analysis of fecal coliforms to determine fecal pollution sources in a mixed-use watershed. Environ. Mon. Assess. 85:87–98.

Buser, H.R., T. Poiger, and M.D. Muller. 1999. Occurrence and environmental behavior of the chiral pharmaceutical drug ibuprofen in surface waters and in wastewater. Environ. Sci. Technol. 33:2529-2535.

Butaye, P., L.A. Devriese, and F. Haesebrouck. 2001. Differences in antibiotic resistance patterns of *Enterococcus faecalis* and *Enterococcus faecium* strains isolated from farm and pet animals. Antimicrob Agents Chemother. 45:1374-1378

Bull, I.D., M.J. Lockheart, M.M. Elhmmali, D.J. Roberts, R.P. Evershed. 2002. The origin of faeces by means of biomarker detection. Environ. Int. 27:647–654.

Bull, I.D., M.M. Elhmmali, D.J. Roberts, and R.P. Evershed. 2003. The application of steroidal biomarkers to track the abandonment of a roman wastewater course at the Agora (Athens, Greece). Archaeometry 45:149-161.

Calci, K.R., W. Burkhardt III, W.D. Watkins, and S.R. Rippy. 1998. Occurence of male-specific bacteriophage in feral and domestic animal wastes, human feces, and human-associated wastewaters. Appl Environ Microbiol. 64: 5027-5029.

Carillo, M., E. Estrada, and T.C. Hazen. 1985. Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed. Appl. Environ. Microbiol. 50:468–476.

Carroll, S., M. Hargreaves, and A. Goonetilleke. 2005. Sourcing faecal pollution from onsite wastewater treatment systems in surface waters using antibiotic resistance analysis. J. Appl. Microbiol. 99:471-482.

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

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

Chan, K.H., M.H.W. Lam, K.F. Poon, H.Y. Yeung, and T.K.T. Chiu. 1998. Application of sedimentary fecal stanols and sterols in tracing sewage pollution in coastal water. Water Res. 32:225-235.

Chen, Z., P. Pavelic, P. Dillon, and R. Naidu. 2002. Determination of caffeine as a tracer of sewage effluent in natural waters by on-line solid-phase extraction and liquid chromatography with diode-array detection. Water Res. 36:4830–4838.

Choi K-H. and F.C. Dobbs. 1999. Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. J Microbiol Meth. 36:203-213.

Chun, J., A. Huq, and R.R. Colwell. 1999. Analysis of 16S-23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. Appl. Environ. Microbiol. 65:2202-2208.

Clara, M., B. Strenn, N. Kreuzinger. 2004. Carbamazepine as a possible anthropogenic marker in the aquatic environment: investigations on the behaviour of carbamazepine in wastewater treatment and during groundwater infiltration. Water Res. 38:947–954.

Close, M.E., L.R. Hodgson, and G. Tod. 1989. Field evaluation of fluorescent whitening agents and sodium tripolyphosphate as indicators of septic tank contamination in domestic wells. N. Z. J. Mar. and Fresh. Res., 23:563-568.

Cole, D., S.C. Long, and M.D. Sobsey. 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl. Environ. Microbiol. 69:6507-6514.

Condon, C., J. Philips, Z.Y. Fu, C. Squires, and C.L. Squires. 1992. Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. EMBO J. 11:4175–4185.

Craig, D.L., H.J. Fallowfield, and N.J. Cromar. 2004. Use of microcosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with *in situ* measurements. J. Appl. Microbiol. 96:922-930.

Crozier, J.B., B. Clark, and H. Weber. 2002. Identifying sources of fecal pollution in the Roanoke River, Roanoke County, Virginia. Va. J. Sci. 53:157-166.

da Costa, R.L. and R.S. Carreira. 2005. A comparison between faecal sterols and coliform counts in the investigation of sewage contamination in sediments. Braz. J. Oceanogr. 53:157-167.

Davis, J.V. and M.N. Barr. 2006. Assessment of possible sources of microbiological contamination in the water columm and streambed sediments of the Jacks Fork, Ozark National Scenic Riverways, Missouri—Phase III: U.S. Geological Survey Scientific Investigations Report 2006–5161, 32 p.

de Moissac, Y.R., S.L. Ronald, and M.S. Peppler. 1994. Use of pulsed-field gel electrophoresis for epidemiological study of *Bordetella pertussis* in a whooping cough outbreak. J. Clin. Microbiol. 32:398-402.

de Motes, C.M., P. Clemente-Casares, A.Hundesa, M. Martín, and R. Girones. 2004. Detection of bovine and porcine adenoviruses for tracing the source of fecal contamination. Appl. Environ. Microbiol. 70:1448-1454.

Díaz-Mejía, J.J., A. Carbajal-Saucedo, and C.F. Amábile-Cuevas. 2002. Antibiotic resistance in oral commensal streptococci from healthy Mexicans and Cubans: resistance prevalence does not mirror antibiotic usage. FEMS Microbiol. L. 217:173-176.

Dick, L. K. and K.G. Field. 2004. Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR Assay for 16S rRNA genes. Appl. Envir. Microbiol. 70: 5695-5697.

Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field. 2005. Host distributions of uncultivated fecal *Bacteroidales* bacteria

reveal genetic markers for fecal source identification. Appl. Environ. Microbiol. 71:3184-3191.

Dickerson, J. W. Jr., C. Hagedorn, and A. Hassall. 2007a. Detection and remediation of human-origin pollution at two public beaches in Virginia using multiple source tracking methods. Water Res. 41:3758-3770.

Dickerson, J.W. Jr, J. B. Crozier, C. Hagedorn, and A. Hassall. 2007b. Assessment of the 16S-23S rDNA intergenic spacer region in *Enterococcus spp.* for microbial source tracking. J. Environ. Qual. 36:1661-1669.

Dixon, L.K., H.M. Taylor, E. Staugler, and J.O. Scudera. 2005. Development of a fluorescence method to detect optical brighteners in the presence of varying concentrations of fluorescent humic substances: identifying regions in influenced by OSTDS in the estuarine waters of Charlotte Harbor. Mote Marine Laboratory Technical Report No. 1045.

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

Ebdon, J.E., J.L. Wallis and H.D. Taylor. 2004. A simplified low-cost approach to antibiotic resistance profiling for faecal source tracking. Wat. Sci. Tech. 1:185-191.

Edge, T.A. and S. Hill. 2005. Occurrence of antiobiotic resistance in *Escherichia coli* from surface waters and fecal pollution sources near Hamilton, Ontario. Can. J. Microbiol. 51:501-505.

Edwards, D.D., G.A. McFeters, and M.I. Venkatesan. 1998. Distribution of C*lostridium perfringens* and fecal sterols in a benthic coastal marine environment influenced by the sewage outfall from McMurdo Station, Antarctica. Appl. Environ. Microbiol. 64: 2596–2600.

Elhmmali M.M, D.J. Roberts, and R.P. Evershed. 2000. Combined analysis of bile acids and sterols/stanols from riverine particulates to assess sewage discharges and other fecal sources. Environ. Sci. Technol**.** 34:39-46.

Enriquez, C.E., C.J. Hurst, and C.P. Gerba. 1995. Survival of the enteric adenoviruses 40 and 41 in tap, sea, and wastewater. Wat. Res. 29:2548-2553.

Evenson, C.J. and K.A. Strevett. 2006. Discriminant analysis of fecal bacterial species composition for use as a phenotypic microbial source tracking method. Res. Microbiol. 157:437-444.

Evershed, R.P. and P.H. Bethell. 1996. Application of multimolecular biomarker techniques to the identification of faecal material in archaeological soils and sediments. ACS Symp. Ser. 625:157-172.

Farag, A.M., J.N. Goldstein, D.F. Woodward, M .Samadpour. 2001. Water quality in three creeks in the backcountry of Grand Teton National Park, USA. J. Freshwat. Ecol. 16:135-144.

Fent, K., A.A. Weston, and D. Caminada. 2006. Ecotoxicology of human pharmaceuticals. Aquat. Toxicol. 76:122–159.

Fiksdal, L., J.S. Maki, S.J. Lacroix, and J.T. Staley. 1985. Survival and detection of *Bacteroides* spp, prospective indicator bacteria. Appl. Environ. Microbiol. 49:148-150.

Field, K.G., E.C Chern, L.K. Dick, J. Fuhrman, J. Griffith, P.A. Holden, M.G. LaMontagne, J. Le, B. Olson, and M.T. Simonich. 2003. A comparative study of culture-independent genotypic methods of fecal source tracking. J. Wat. Health. 1: 181-194.

Fisher, P. J., O.H. Petrini, and M. Lappin-Scott. 1992. The distribution of some fungal and bacterial endophytes in m`aize (*Zea mays* L.) New Phytologist 122:299-305.

Fogarty, L.R., S.K. Haack, M J. Wolcott, and R.L. Whitman. 2003. Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull feces. J. App. Microbiol. 94:865-878.

Fono, L.J. and D.L. Sedlack. 2005. Use of the chiral pharmaceutical propranolol to identify sewage discharges into surface waters. Environ. Sci. Technol. 39:9244-9252.

Formiga-Cruz, M., A.K. Allard, A.C. Conden-Hansson, K. Hensilwood, B.E. Hernroth, J. Jofre, D.N. Lees, F. Lucena, M. Papapetropulou, R.E. Rangdale, A. Tsibouxi, A. Vantarakis, and R. Girones. 2003. Evaluation of potential indicators of viral contamination in shellfish with applicability to diverse geographical areas. Appl. Environ. Microbiol. 69:1556–1563.

Francy, D.S., E.E. Bertke, D.P. Finnegan, C.M. Kephart, R.A. Sheets, J. Rhoades, and L. Stumpe. 2006. Use of spatial sampling and microbial source-tracking tools for understanding fecal contamination at two Lake Erie beaches, U.S. Geological Survey Scientific Investigations Report 2006-5298, 29 p.

Geldreich, E.E. and B.A. Kenner. 1969. Concepts of fecal streptococci in stream pollution. J. Wat. Poll. Control Fed. 41:R336-R352.

Gilliver, M.A., M. Bennett, M. Begon, S.M. Hazel, and C.A. Hart. 1999. Antibiotic resistance found in wild rodents. Nature. 401:233-234.

Gilpin, B.J., J.E. Gregor, and M.G. Savill. 2002. Identification of the source of faecal pollution in contaminated rivers. Wat. Sci. Technol. 46:9-15.

Glassmeyer, S.T., E.T. Furlong, D.W. Kolpin, J.D. Cahill, S.D., Zaugg, S.L. Werner, M.T. Meyer, and D.D. Kryak. 2005. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. Environ. Sci. Technol. 39:5157-5169.

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

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.

Grabow, W.O.K., T.E. Neubrech, C.S. Holtzhausen, and J. Jofre. 1995. *Bacteroides fragilis* and *Escherichia coli* bacteriophages: excretion by humans and animals. Wat. Sci. Technol. 31:223–230.

Graham, T., E.J. Golsteyn-Thomas, V.P. Gannon, and J.E. Thomas. 1996. Genus- and species specific detection of *Listeria monocytogenes* using polymerase chain reaction assays targeting the 16S/23S intergenic spacer region of the rRNA operon. Can. J. Microbiol. 42:1155-62.

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

Graves, L.M., B. Swaminathan, M.W. Reeves, S.B. Hunter, R.E. Weaver, B.D. Plikaytis and A. Schuchat. 1994. Comparison of ribotyping and multilocus enzyme electrophoresis for subtyping of *Listeria monocytogenes* isolates. J Clin Microbiol. 32:2936-2943.

Graves, A.K., C. Hagedorn, A. Brooks, R.L. Hagedorn, and E. Martin. 2007. Microbial source tracking in a rural watershed dominated by cattle. Wat. Res. 41:3729-3739.

Gregor, J., N. Garrett, B. Gilpin, C. Randall, D. Saunders. 2002. Use of classification and regression tree (CART) analysis with chemical faecal indicators to determine sources of contamination. N. Z. J. Mar. and Freshwater Res. 36:387–398.

Gregory, J.B., R.W. Litaker, and R.T. Noble. 2006. Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. Appl. Environ. Microbiol. 72:3960-3967.

Griffin, D. W., R. Stokes, J. B. Rose, and J. H. Paul. 2000. Bacterial Indicator occurrence and the use of an F(+) specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida. Microb. Ecol. 39:56-64.

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. Wat. Health. 1:141-151.

Grimalt, J.O., P. Fernandez, J.M. Bayona, and J. Albaiges. 1990. Assessment of fecal sterols and ketones as indicators of urban sewage inputs to coastal waters. Environ. Sci. Technol. 24:357- 63.

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.

Guertler, V. and V.A. Stanisich. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. Microbiol. 145:3-16.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneneau. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522-5531.

Hagedorn, C., J.B. Crozier, K.A. Mentz, A.M. Booth, A.K. Graves, N.J. Nelson, and R.B. Reneau Jr. 2003a. Carbon source utilization profiles as a method to identify sources of fecal pollution in water. Appl. Environ. Microbiol. 94:792-799.

Hagedorn, C., R.B. Reneau Jr., M. Saluta, A. Chapman. 2003b. Impact of onsite wastewater systems on water quality in coastal regions. Virginia Coastal Resources Management Program Memorandum of Agreement 50312-01-13-PT. Virginia Dept. of Conservation and Recreation, Virginia Dep. of Health, Richmond.

Hagedorn, C., M. Saluta, A. Hassall, and J. Dickerson. 2005a. Fluorometric detection of optical brighteners as an indicator of human sources of water pollution. Part I. Description and detection of optical brighteners. Crop and Soil Environmental News, November 2005. [online] http://www.ext.vt.edu/news/periodicals/cses/2005-11/part1.html last accessed 27 November 2007.

Hagedorn, C., M. Saluta, A. Hassall, and J. Dickerson. 2005b. Fluorometric detection of optical brighteners as an indicator of human sources of water pollution. Part II. Development as a source tracking methodology in open waters. Crop and Soil Environmental News, November 2005. online http://www.ext.vt.edu/news/periodicals/cses/2005-11/part2.html.

Hager, M.C. 2001a. Detecting bacteria in coastal waters: Part 1. Stormwater: Buyers Guide 2002. Vol. 2. No. 3.

Hager, M.C. 2001b. Detecting bacteria in coastal waters: Part 2. Stormwater. Vol 2. No. 4.

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.

Hartel, P.G., C. Hagedorn, J.L. McDonald, J.A. Fisher, M.A. Saluta, J.W. Dickerson, Jr., L.C. Gentit, S.L. Smith, N.S. Mantripragada, K.J. Ritter, and C.N. Belcher. 2007. Exposing water samples to ultraviolet light improves fluorometry for detecting human fecal contamination. Wat. Res. 41:3629-3642.

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

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

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

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

Hassan, W.M., R.D. Ellender and S.Y. Wang. 2007. Fidelity of bacterial source tracking: *Escherichia coli* vs *Enterococcus spp* and minimizing assignment of isolates from nonlibrary sources. J. Appl.Microbiol. 102:591–598.

Havelaar, A.H., Furuse, K. and W.M. Hogeboom. 1986. Bacteriophages and indicator bacteria in human and animal faeces. J. Appl. Bacteriol. 60:255–262.

Hayashi, Y., S. Managaki, and H. Takada. 2002. Fluorescent whitening agents in Tokyo Bay and adjacent rivers: their application as anthropogenic molecular markers in coastal environments. Environ. Sci. Technol. 36:3556 -3563.

Hilton, H.J and K.V. Thomas. 2003. Determination of selected human pharmaceutical compounds in effluent and surface water samples by high-performance liquid chromatography-electrospray tandem mass spectrometry. J Chromatogr A. 1015:129-141.

Holmes B., M.M. Costa S.L.W. Ganner, and O.M. Stevens. 1994. Evaluation of Biolog system for identification of some Gram negative bacteria of clinical importance. J Clin. Microbiol. 32:1970-1975.

Hsu, F. C., Y.S. Shieh, J. van Duin, M.J. Beekwilder, and M.D. Sobsey. 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. Environ. Microbiol. 61:3960-3966.

Isobe, K.O., M. Tarao, N.H. Chiem, L.Y. Minh, and H. Takada. 2004. Effect of environmental factors on the relationship between concentrations of coprostanol and fecal indicator

bacteria in tropical (Mekong Delta) and temperate (Tokyo) freshwaters. Appl. Environ. Microbiol. 70:814–821.

Jagals, P. and W.O.K. Grabow. 1996. An evaluation of sorbitol-fermenting bifidobacteria as specific indicators of human faecal pollution of environmental water. Water S. Africa 22:235–238.

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

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.

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

Jimenez-Clavero, M.A., C. Fernandez, 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.

Johnson, A., B. Carey, and S. Golding. 2004. Results of a screening analysis for pharmaceuticals in wastewater treatment plant effluents, wells, and creeks in the Sequim-Dungeness area. Environmental Assessment Program, Olympia, Washington 98504-7710. Publication Number: 04-03-051.

Johnson, L.K., M.B. Brown, E.A. Carruthers, J.A. Ferguson, P.E. Dombeck, 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.

Kamiko, N. and S. Ohgaki. 1993. Multiplication characteristics of F-RNA phage and its utility as an indicator for pathogenic viruses. Wat. Sci. Technol. 27:133-136.

Kator, H. and M. Rhodes. 1992. Evaluation of *Bacteroides fragilis* bacteriophage, a candidate human-specific indicator of fecal contamination for shell-fish-growing waters. Final Report. Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Va.

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

Katouli, M., A. Lund, P. Wallgren, I. Kühn, O. Söderlind, and R. Möllby. 1995. Phenotypic characterization of intestinal *Escherichia coli* of pigs during suckling, postweaning, and fattening periods. Appl. Environ. Microbiol. 61:778–783.

Kelsey, R.H., G.I. Scott, D.E., Porter, B. Thompson, and L.Webster. 2003. Using multiple antibiotic characteristics to determine sources of fecal coliform bacterial pollution. Environ. Mon. Assess. 81: 337–348.

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. Biotech. 59:97-104.

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. Biotech. 63:231-238.

Kostman, J.R., T.D. Edlind, J.J. LiPuma and T.L. Stull. 1992. Molecular epidemiology of *Pseudomonas cepacia* determined by polymerase chain reaction ribotyping. J. Clin. Microbiol. 30:2084-2087.

Kramer, J.B., S. Canonica, and J. Hoigne. 1996. Degradation of fluorescent whitening agents in sunlit natural waters. Environ. Sci. Technol. 30:2227-2234.

Kühn, I., A. Iversen, L.G. Burman, B. Olsson-Liljequist, A. Franklin, M. Finn, F. Aarestrup, A.M. Seyfarth, A R. Blanch, X. Vilanova, H. Taylor, J.Caplin, M.A. Moreno, L. Dominguez, IA. Herrero, and R. Mollby. 2003. Comparison of enterococcal populations in animals, humans, and the environment— a European study. Int. J. Food Microbiol. 88:133–145.

Lasalde, C., R. Rodriguez, and G.A. Toranzos. 2005. Statistical analyses: possible reasons for unreliability of source tracking efforts. Appl. Environ. Microbiol. 71:4690-4695.

Lau, M.M. and S.C. Ingham. 2001. Survival of faecal indicator bacteria in bovine manure incorporated into soil. L. Appl. Microbiol. 33:131-136.

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.

LeBlanc, L.A., J.S. Latimer, J.T. Ellis and J.G. Quinn J. G. 1992. The geochemistry of coprostanol in waters and surface sediments from Narragansett Bay. Estuar., Coast. and Shelf. Sci. 34:439-458.

Lee, S.H. and S.J. Kim. 2002. Detection of infectious enteroviruses and adenoviruses in tap water in urban areas in Korea. Wat. Res. 36:248–256.

Leecaster M.K. and S.B. Weisberg. 2001. Effect of sampling frequency on shoreline microbiology assessments. Mar. Pollut. Bull. 42:1150-1154.

Leeming, R. and P.D. Nichols. 1996. Concentration of coprostanol that correspond to existing bacterial indicator guideline limits. Water Res. 30:2997-3006.

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

Leeming R., V. Latham, M. Rayner, and P. Nichols. 1997. Detecting and distinguishing sources of sewage pollution in Australian inland and coastal waters and sediments. ACS Symp. Ser. 671:306-319.

LeFevre, N.M. and G.D. Lewis. 2003. The role of resuspension in enterococci distribution in water at an urban beach. Wat. Sci. Technol. 47:205-210.

Lefresne, G., E. Latrille, F. Irlinger, and P.A.D. Grimont. 2004. Repeatability and reproducibility of ribotyping and its computer interpretation. Res. Microbiol. 155:154-161.

Leser, T.D., R.H. Lindecrona, T.K. Jensen, B.B. Jensen, and K. Moller. 2000. Changes in bacterial community structure in the colon of pigs fed different experimental diets and after infection with *Brachyspira hyodysenteriae*. Appl. Environ. Microbiol. 66:3290-3296.

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.

Li, X.Z., L. Zhang, and K. Poole. 1998. Role of the multidrug efflux systems of *Pseudomonas aeruginosa* in organic solvent tolerance. J. Bacteriol. 180:2987-2991.

Lin, C. and T.L. Miller. 1998. Phylogenetic analysis of *Methonobrevibacter* isolated from feces of humans and other animals. Arch. Microbiol. 169:397-403.

Louws, F.J., D.W. Fulbright, C.T. Stephens, and F.J. de Bruijn. 1994. Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. Appl. Environ. Microbiol. 60:2286-2295.

Lu, J., S. Sanchez, C. Hofacre, J.J. Maurer, B.G. Harmon, and M.D. Lee. 2003. Evaluation of broiler litter with reference to the microbial composition as assessed by using 16S rRNA and functional gene markers. Appl. Environ. Microbiol. 69:901-908.

Lu, L., M.E. Hume, K.L. Sternes, and S.D. Pillai. 2004. Genetic diversity of *Escherichia coli* isolates in irrigation water associated sediments: implications for source tracking. Wat. Res. 38:3899-3908.

Lu, Z., D. Lapen, A. Scott, A. Dang, and E. Topp. 2005. Identifying host sources of fecal pollution: diversity of *Escherichia coli* in confined dairy and swine production systems. Appl. Environ. Microbiol. 71:5992-5998.

MacDonald, I.A., V.D. Bokkenheuser, J. Winter, A.M. McLernon, and E.H. Mosbach. 1983. Degradation of fecal sterols in the human gut. J. Lipid Res. 24:675–694.

Maldonaldo C., J. Dachs, and J.M. Bayona. 1999. Trialkylamines and coprostanol as tracers of urban pollution in waters from enclosed seas: the Mediterranean and Black Sea. Environ. Sci. Technol. 33:3290-3296.

Manero, A., X. Vilanova, M. Cerda-Cuellar, and A.R. Blanch. 2002. Characterization of sewage waters by biochemical fingerprinting of enterococci. Wat. Res. 36:2831–2835.

Mara, D.D. and J.I. Oragui. 1981. Occurrence of *Rhodococcus coprophilus* and associated Actinomycetes in feces, sewage, and freshwater. Appl. Environ. Microbiol. 42:1037-1042.

Mara, D.D. and J.I. Oragui. 1983. Sorbitol-fermenting bifidobacteria as specific indicators of human faecal pollution. J. Appl. Bacteriol. 55:349–357.

Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D.A. Morrison, G.J. Boulnois, and J-P Claverys. 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae.* Nucleic Acids Res. 20:3479-3483.

McArthur, J.V. and R.C. Tuckfield. 2000. Spatial patterns in antibiotic resistance among stream bacteria: effects of industrial pollution. Appl. Environ. Microbiol. 66:3722-3726.

McDonald, D.E., D.W. Pethick, B.P. Mullan, and D.J. Hampson. 2001. Increasing viscosity of the intestinal contents alters small intestinal structure and intestinal growth, and stimulates proliferation of enterotoxigenic *Escherichia coli* in newly-weaned pigs. Br. J. Nutr. 86:487–498.

McDonald, J.L., P.G. Hartel, L.C. Gentit, C.N. Belcher, K.W. Gates, K. Rodgers, J. A.Fisher, K.A. Smith, and K.A. Payne. 2006. Identifying sources of fecal contamination inexpensively with targeted sampling and bacterial source tracking. J. Environ. Qual. 35:889-897.

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

Moce´-Llivina, L., F. Lucena, and J. Jofre. 2005. Enteroviruses and bacteriophages in bathing waters. Appl. Environ. Microbiol. 71:6838–6844.

Moore, D. F., V. J. Harwood, D.M. Ferguson, J. Lukasik, P. Hannah, M. Geitrich, and M. Brownell. 2005. Evaluation of antibiotic resistance analysis and ribotyping for

identification of fecal pollution sources in urban watershed. J. Appl. Microbiol. 99:618-628.

Muirhead, R.W., R.P. Collins, and P.J. Bremer. 2005. Erosion and subsequent transport state of *Escherichia coli* from cowpats. Appl. Environ. Microbiol. 71:2875-2879.

Muniesa M., F. Lucena and J. Jofre. 1999. Study of the potential relationship between the morphology of infectious somatic coliphages and their persistence in the environment. J. Appl. Microbiol. 87:402-409.

Myoda, S. P., C. A. Carson, J. J. Fuhrmann, B.K. Hahm, P.G. Hartel, H. Yampara-Iquise, L. Johnson, R.L. Kuntz, C.H. Nakatsu, M.J. Sadowsky, and M. Samadpour. 2003. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. J. Wat. Health. 1:167-180.

Nebra, Y., X. Bonjoch, and A.R. Blanch. 2003. Use of *Bifidobacterium dentium* as an indicator of the origin of fecal water pollution. Appl. Environ. Microbiol. 69:2651–2656.

Neidhardt, F.C. (editor). 1987. *Escherichia coli* and *Salmonella typhimurium* : cellular and molecular biology. Textbook. Washington, D.C.: American Society for Microbiology.

Noble, R.T. and J.A. Fuhrman. 2000. Enteroviruses detected by reverse transcriptase polymerase chain reaction from the coastal waters of Santa Monica Bay, California: low correlation to bacterial indicator levels. Hydrobiologia. 460: 175-184.

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

Noblet, J.A., D.L. Young, E.Y. Zeng, and S. Ensari. 2004. Use of fecal steroids to infer the sources of fecal indicator bacteria in the Lower Santa Ana River Watershed, California: sewage is unlikely a significant source. Environ. Sci. Technol. 38:6002-6008.

O'Brien, T.L., D. Bailey, A. Gill, and P. Station. 2005. Considerations for using sewage versus direct human samples to generate a bacterial source tracking database. ASM Poster Q-416.

Oberste, M.S., K. Maher, S. M. Michele, G. Belliot, M. Uddin, and M.A. Pallansch. 2005. Enteroviruses 76, 89, 90 and 91 represent a novel group within the species human enterovirus A. J. Gen. Vir. 86:445–451.

Ogden, I.D., N.F. Hepburn, M. MacRae, N.J.C. Strachan, D.R. Fenlon, S.M. Rusbridge, and T.H. Pennington. 2002. Long-term survival of *Escherichia coli* O157 on pasture following an outbreak associated with sheep at a scout camp. L. Appl. Microbiol. 38:100-104.

Ohkuma, M., S. Noda, S., and T. Kudo. 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. Appl. Environ. Microbiol. 65:4926-4934.

Oshiro, R.K. and B.H. Olson. 1997. Occurrence of *STh* toxin gene in wastewater. *In* Coliforms and *E. coli*: Problem or Solution? D. Kay and C. Fricker (ed.), pp. 255-259. Royal Society of Chemistry, Cam-bridge, UK.

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

Parveen, S., K.M. Portier, K. Robinson, L. Eminston, 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.

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

Paxéus, N., and H.F. Schröder. 1996. Screening for non-regulated organic compounds in municipal wastewater in Göteborg, Sweden. Wat. Sci. Technol. 33:9–15.

Payan, A., J. Ebdon, H. Taylor, C. Gantzer, J. Ottoson, G. T. Papageorgiou, A. R. Blanch, F. Lucena, J. Jofre, and M. Muniesa. 2005. Method for isolation of *Bacteroides* bacteriophage host strains suitable for tracking sources of fecal pollution in water. Appl. Environ. Microbiol. 71:5659–5662.

Payment, P., and E. Franco. 1993. *Clostridium perfringens* and somatic coliphages as indicators of the efficiency of drinking water treatment for viruses and protozoan cysts. Appl. Environ. Microbiol. 59:2418-2424.

Peeler, K.A., S.P. Opsahl, and J.P Chanton. 2006. Tracking anthropogenic inputs using caffeine, indicator bacteria, and nutrients in rural freshwater and urban marine systems. Environ. Sci. Technol. 40:7616-7622.

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.

Poiger, T., J.A. Field, T.M. Field, H. Siegrist, and W. Giger. 1998. Behavior of fluorescent whitening agents during sewage treatment. Wat. Res. 32:1939-1947.

Pond, K. R., R. Rangdale, W.G. Meijer, J. Brandao, L. Falcão, A. Rince, B. Masterson, J. Greaves, A. Gawler, E. McDonnell, and A. Cronin, S. Pedley. 2004. Workshop Report: Developing pollution source tracking for recreational and shellfish waters. Environ. For. 5:237-247.

Pourcher, A.M., L.A. Devriese, J.F. Hernandez, and J.M. Delattre. 1991. Enumeration by a miniaturized method of *E. coli*, *Streptococcus bovis*, and enterococci as indicators of the origin of faecal pollution in waters. J. Appl. Bacteriol. 70:525-530.

Puig, A., N. Queralt, J. Jofre, and R. Araujo. 1999. Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater. Appl. Environ. Microbiol. 65:1172–1176.

Pupo, G.M. and B.J. Richardson. 1995. Biochemical genetics of a natural population of *Escherichia coli*: season changes in alleles and haplotypes. Microbiol. 141:1037-1044.

Rabinovici, S.J.M., R.L. Bernknopf, A.M. Wein. 2003. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Tech. 38:2737-2745.

Rangdale, R E., W. Meijer, A. Rincé, J. Brandão, L. Falcão, J. Machado, C. Nogueira, M.L. Rosado, and C. Veríssimo. 2003. Development of methods for pollution source tracking- consolidated literature review. Completed for Environment Agency North West Region INTERREG IIIB Atlantic Area Programme Improved Coastal and Recreational Waters (ICReW).

Resnick, I.G. and Levin, M.A. 1981. Assessment of bifidobacteria as indicators of human fecal pollution. Appl. Environ. Microbiol. 42:433–438.

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.

Riffard, S., F. Lo Presti, P. Normand, P. Forey, M. Reyrolle, J. Etienne, and F. Vandenesch. 1998. Species identification of *Legionella* via intergenic 16S–23S ribosomal spacer PCR analysis. Int. J. Syst. Bacteriol. 48:723–730.

Ritter, K.J., C. 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 using in library-based approaches to microbial source tracking. J. Wat. Health. 1:209-223.

Ritter K. and B. Robinson. 2004. Statistical appraisal of disproportional versus proportional libraries. Presentation, presentation given during the Golf of Mexico Bacterial Source Tracking workshop in Biloxi, MS, Nov 11-13, 2004. Moderator: R.D. Ellender, University of Southern Mississippi, Hattisburg, MS.

Roberts, P.H. and K.V. Thomas. 2006. The occurrence of selected pharmaceuticals in wastewater effluent and surface waters of the lower Tyne catchment. Sci Total Environ. 56:143-53.

Russell, J. B., F. Diez-Gonzalez, and G.N. Jarvis. 2000. Symposium: Farm health and safety. Invited review: Effects of diet shifts on *Escherichia coli* in cattle. J. Dairy Sci. 83:863–873.

Rzezutka, A. and N. Cook. 2004. Survival of human enteric viruses in the environment and food. FEMS Microbiol. Rev. 28:441–453.

Salyers, A.A. and C.F. Amábile-Cuevas. 1997. Why are antibiotic resistance genes so resistant to elimination? Antimicro. Agents Chemother. 41:2321-2325.

Samadpour, M. and Chechowitz, N. 1995. Little Soos Creek microbial source tracking: a survey. Seattle, WA: University of Washington, Department of Environmental Health.

Samadpour, M. 2002. Microbial source tracking: principles and practice, US EPA Workshop on Microbial Source Tracking, 5–9 http://www.sccwrp.org/tools/workshops/source_tracking_agenda.html.

Samadpour, M., M.C. Roberts, C. Kitts, W. Mulugeta, D. Alfi. 2005. The use of ribotyping and antibiotic resistance patterns for identification of host sources of *Escherichia coli* strains. L. Appl. Microbiol. 40:63-68.

Savill, M. G., S.R. Murray, P. Scholes, E.W. Maas, R.E. McCormick, E.B. Moore, and B.J. Gilpin. 2001. Application of polymerase chain reaction (PCR) and TaqMan™ PCR techniques to the detection and identification of *Rhodococcus coprophilus* in faecal samples. J. Microbiol. Meth. 47:355-368.

Sayah, R.S., J.B. Kaneene, Y. Johnson, and R. Miller. 2005. Patterns of antimicrobial resistance observed in *Escherichia coli* isolates obtained from domestic- and wild-animal fecal samples, human septage, and surface water. Appl. Environ. Microbiol. 71:1394-1404.

Schaper, M., J. Jofre, M. Uys, and W.O. Grabow. 2002a. Distribution of genotypes of F-specific RNA bac-teriophages in human and non-human sources of faecal pollution in South Africa and Spain. J. Appl. Microbiol. 92:657-667.

Schaper, M., A.E. Duran, and J. Jofre. 2002b. Comparative resistance of phage isolates of four genotypes of F-specific RNA bacteriophages to various inactivation processes. Appl. Environ. Microbiol. 68:3702-3707.

Schoonmaker, D., T. Heimberger, and G. Birkhead. 1992. Comparison of ribotyping and restriction enzyme analysis using pulsed-field gel electrophoresis for distinguishing *Legionella pneumophila* isolates obtained during a nosocomial outbreak. J. Clin. Microbiol. 30:1491-1498.

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.

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 human, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. 69:1089-1092.

Scott, T.M. J. Caren, G.R. Nelson, T.M. Jenkins, and J. Lukasik. 2004. Tracking sources of fecal pollution in a South Carolina watershed by ribotyping *Escherichia coli*: a case study. Environ. For. 5:15-19.

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.

Sechi, L.A. and L. Daneo-Moore. 1993. Characterization of Intergenic Spacers in Two *rrn* of *Enterococcus hirae* ATCC 9790 operons. J. Bacteriol. 175:3213-3219.

Seiler, R.L., S.D. Zaugg, J.M. Thomas and D.L. Harcroft. 1999. Caffeine and pharmaceuticals as indicators of wastewaster contamination in wells. Ground Wat. 37:405-410.

Seurinck, S., W. Verstraete, and S.D. Siciliano. 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. 69:4942-4950.

Seurinck, S., W. Verstraete, and S.D. Siciliano. 2005. Microbial source tracking for identification of fecal pollution. Rev. Environ. Sci. Bio/Technol. 4:19-37.

Siegener R. and R.F. Chen. 2002. Caffeine in Boston Harbor seawater. Mar. Pollut. Bull. 44:383-387.

Shehane, S.D., V.J. Harwood, J.E. Whitlock, and J. B. Rose. 2005. The influence of rainfall on the incidence of microbial faecal indicators and the dominant sources of faecal pollution in a Florida river. J. Appl. Microbiol. 98:1127-1136.

Shu, W.C., and W.H. Ding. 2005. Determination of fluorescent whitening agents in laundry detergents and surface waters by solid-phase extraction and ion-pair high-performance liquid chromatography. J. Chrom. A. 1-2:218-223.

Silvi, S., C.J., Rumney, A. Cresci, and I.R. Rowland. 1999. Resistant starch modifies gut microflora and microbial metabolism in human flora-associated rats inoculated with faeces from Italian and UK donors. J. Appl. Microbiol. 86:521-530.

Simmons, G.M., Jr., S.A. Herbein, and C.M. James. 1995. Managing nonpoint fecal coliform sources to tidal inlets. Wat. Res. Update 100: 64-74.

Simmons, G.M., D.F. Waye, S. Herbein, S. Myers, S. and E. Walker. 2002. Estimating nonpoint fecal coliform sources in Northern Virginia's Four Mile Run watershed, p. 143-167. In Younos, T. (ed.), Advances in Water Monitoring Research. Water Resources Publications, LLC.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2002. Microbial source tracking: state of the science. Environ Sci. Technol. 36:5280-5288.

Sinton, L.W., Donnison, A.M., and Hastie, C.M. 1993. Faecal streptococci as faecal pollution indicators: a review. Part II: Sanitary significance, survival, and use. N. Z. J. Mar. and Freshwater Res. 27: 117-137

Skraber, S., B. Gassilloud, and C. Gantzer. 2004. Comparison of coliforms and coliphages as tools for assessment of viral contamination in river water. Appl. Environ. Microbiol. 70:3644–3649.

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

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.

Stevenson, A.H. 1953. Studies of bathing water quality and health. J. Amer. Public Health. Assoc. 43:529-538.

Stewart, J.R., R.D.Ellender, J.A.Gooch, S. Jiang, S.P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons learned from a methods comparison study. J. Wat. Health. 1:225-231.

Stewart-Pullaro, J., J.W. Daugomah, D.E. Chestnut, D.A. Graves, M.D. Sobsey, and G.I. Scott. 2006. F+ RNA coliphage typing for microbial source tracking in surface waters. J. Appl. Microbiol. 101: 1015-1026.

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

Stubbs, S.L.J., J.S. Brazier, G.L. O'Neill, and B.I. Duerden. 1999. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 Different PCR ribotypes. Appl. Environ. Microbiol. 37:461-463.

Sundram, A., N. Jumanlal, and M.M. Ehlers. 2006. Genotyping of F-RNA coliphages isolated for wastewater and river samples. Water SA. 32:65-70.

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

Temes, T.A., H. Andersen, D. Gilberg, and M. Bonerz. 2002. Determination of estrogens in sludge and sediments by liquid extraction and GC/MS/MS. Anal. Chem. 74:3498–3504.

Thomas K.V. and M.J. Hilton. 2004. The occurrence of selected human pharmaceutical compounds in UK estuaries. Mar. Pollut. Bull. 49:436-44.

Tolosa, I., N. LeBlond, C. Copin-Montégut, J.C. Marty, S. de Mora, and L. Prieur. 2003. Distribution of sterol and fatty alcohol biomarkers in particulate matter from the frontal structure of the Alboran Sea (SW Mediterranean Sea). Mar. Chem. 82:161-183.

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

USEPA. 1986. Ambient water quality criteria for bacteria - 1986. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA-440/5-84-002.

USEPA. 2002. Implementation guidance for ambient water quality criteria for bacteria. U.S. Environmental Protection Agency, Washington, D.C. May 2002 Draft.

USEPA. 2005. Microbial source tracking guide document. U.S. Environmental Protection Agency, Washington, D.C. EPA/600-R-05-064.

Vantarakis, A., D. Venieri, G. Komninou, and M. Papapetropoulou. 2005. Differentiation of faecal *Escherichia coli* from humans and animals by multiple antibiotic resistance analysis. L. Appl. Microbiol. 42:71-77.

Wallis, J.L. and H.D. Taylor. 2003. Phenotypic population characteristics of the enterococci in wastewater and animal faeces: implications for the new European directive on the quality of bathing waters. Wat. Sci. Tech. 47:23-32.

Weaver, R.W., J.A. Entry, and A. Graves. 2005. Numbers of fecal streptococci and *Escherichia coli* in fresh and dry cattle, horse, and sheep manure. Can. J. Microbiol. 51:847-851.

Webster, L.F., B.C. Thompson, M.H. Fulton, D.E. Chestnut, R.F. Van Dolah, A.K. Leight, G.I. Scott. 2004. Identification of sources of *Escherichia coli* in South Carolina estuaries using antibiotic resistance analysis. J. Exper. Mar. Biol. Ecol. 298:179-195.

Weigel, S., U. Berger, E. Jensen, R. Kallenborn, H. Thoresen, and H. Huhnerfuss. 2004. Determination of selected pharmaceuticals and caffeine in sewage and seawater from Tromsø/Norway with emphasis on ibuprofen and its metabolites. Chemosphere. 56:583–592

Wheeler, A.L., P.G. Hartel, D.G. Godfrey, J.L. Hill, and W.I. Segars. 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J. Environ. Qual. 31:1286-1293.

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. Wat. Res. 36:4273-4282.

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.

Wiggins, B.A., R.W. Andrews, R.A. Conway, C.L. Corr, E.J. Dobratz, D.P. Dougherty, J.R. Eppard, S.R. Knupp, M.C. Limjoco, J.M. Mettenburg, J.M. Rinehardt, J. Sonsino, R.L. Torrijos, and M.E. Zimmerman. 1999. Identification of sources of fecal pollution using discriminant analysis: Supporting evidence from large datasets. Appl. Environ. Microbiol. 65:3483-3486.

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.

Wolfe, T. M. 1995. A comparison of fecal coliform densities and fluorescent intensities in Murrells Inlet, a highly urbanized estuary and in North Inlet, a pristine forested estuary. M.S. thesis. University of South Carolina, Columbia.

Wünsche, L., L. Brüggemann and W. Babel. 1995. Determination of substrate utilization patterns of soil microbial communities: An approach to assess population changes after hydrocarbon pollution. FEMS Microbiol. Ecol. 17:295-305.

Yan, W., N. Chang, and D.E. Taylor. 1991. Pulsed-field gel electrophoresis of *Campylobacter jejuni* and *Campylobacter coli* genomic DNA and its epidemiologic application. J. Infect. Dis. 163:1068-1072.

Yan, T. and M.J. Sadowsky. 2006. Determining sources of fecal bacteria in waterways. Environ. Monit. Assess. [Online] http://www.springerlink.com/content/8356661161tp0764/ last accessed 28 November 2007.

Zhong, X. 2004. A study of several statistical methods for classification with application to microbial source tracking. Worchester Polytechnic Institute. Master's Thesis.

III. **Identification of Sources of Enterococci in Virginia's Public Beaches Using Antibiotic Resistance Analysis and Fluorometry**

Abstract

Numerous public beaches in Virginia were forced to post swimming advisories during the summer of 2004, as the result of high *Enterococcus spp.* concentrations detected in beach waters. The microbial source tracking (MST) methods of antibiotic resistance analysis (ARA) and fluorometry (to detect optical brighteners) were used in the summers of 2004 and 2005 to determine the animal origin of fecal pollution at beaches with a past history of, or the potential for, high enterococci counts and posted advisories. Although several beaches were suspected of contamination via sewage pollution, ARA results concluded that dominantly birds and/or wildlife were responsible for elevated enterococci levels at most beaches. The application of fluorometry proved difficult in opens waters due to high levels of dilution, and several instances of suspected false positives due to oil-based or unknown compounds prevalent in some storm drains (SDs).

Introduction

The detection of bacterial levels in recreational waters exceeding state standards requires the posting of advisories or closing of swimming areas by public health officials in the interest of protecting public health. Human exposure to fecally contaminated recreational waters has been linked to an increased risk in the contraction of gastroenteritis, as well as ear, eye, nose and skin infections (USEPA, 1986; Cheung et al., 1990). The greatest correlation between

gastrointestinal illness and bacterial densities in marine recreational waters was found with the detection of the fecal organism *Enterococcus spp*. (Cabelli et al., 1982). In addition to the risks to public health, extended swimming advisories/closures on public beaches can result in significant revenues losses for local economies (Rabinovici et al., 2004). As a result, beach officials have a considerable interest determining the origin of fecal pollution in beach waters, thereby allowing for the most efficient use of best management practices (BMPs) to reduce fecal counts and prevent sustained beach advisories/closures.

Within a given watershed, a large number of potential fecal pollution sources may be present, including: wastewater treatment plants, septic systems, animal farms (lagoons or field runoff), boat-waste discharges, urban storm water runoff, or wildlife. A visual inspection of land use patterns can often lead researchers to a suspected source however in many instances the fecal contributors are not obvious. The origin and development of microbial source tracking (MST) beginning in the mid-1990s has provided a means to discriminate between potential fecal sources in polluted waterbodies. The ability to identify the fecal sources which are present, or dominant, in polluted waters has provided watershed managers with an important tool for aiding in the reduction of fecal loads in surface waters. The potential success of MST is based on the assumption that strains of enteric bacteria are both adapted, and restricted to a single, or group of closely-related host organisms. A wide variety of both genotypic and phenotypic differences among enteric strains have been utilized in attempts to distinguish between host fecal sources (Scott et al., 2002; Simpson et al., 2002). As applied to U.S. recreational waters, MST researchers and project sponsors seek an accurate identification of the sources of fecal pollution within a given waterbody.

Two classes of MST methods have been reported in the literature (Pond et al., 2004; Blanch et al., 2006), those requiring the construction of database, or library, of isolates from known-sources, and those methods not requiring a database. MST methods requiring a database, or known-source library, require the collection of fecal samples from known animal sources to provide isolates and generate a collection of 'fingerprint' patterns used to compare and match the patterns produced by isolates of unknown origin (from environmental water samples). Although numerous assays have been developed, antibiotic resistance analysis (ARA) remains as one of the most widely tested and applied MST methods available to researchers (Wiggins, 1996; Hagedorn et al., 1999; Harwood et al., 2000; Carroll et al., 2005; Geary, et. al., 2006; Orosz-Coghlan et al., 2006). Researchers using ARA employ differences in the ability of bacterial strains to grow in the presence of varying concentrations of antibiotics (Wiggins, 1996, Hagedorn et al., 1999; Harwood et al., 2000; Graves et al., 2002;, Whitlock et al., 2002). The differential exposure of animal groups and their gastrointestinal flora to a variety of antibiotics or other xenobiotic compounds can lead to the selection of bacteria containing antibiotic resistant genes, and thus provide ARA researchers with a means to distinguish between host animals.

As an alternative to the use of microorganisms and/or construction of a known-source library, chemical-based MST has emerged as a means of detecting the chemical signature of specific fecal sources, most commonly those human in origin (Glassmeyer et al., 2005). Although the detection of a wide variety of chemicals has been tested, most require technical expertise and expensive equipment, which is beyond the budgets of many localities. However, the detection of chemicals known as fluorescent whitening agents, or optical brighteners (OBs) can be performed using a relatively inexpensive piece of equipment known as a fluorometer (Hagedorn et al., 2003; Picture 1). OBs are organic compounds which absorb ultraviolet light

(300-400nm) and re-emit it in the blue portion of the visible spectrum with maximum re-emittance between 415 and 435 nm. Used in laundry and dishwashing detergents, anywhere from 25-95% of the OBs are bound to clothing during the wash (Poiger et al., 1998). As a result a portion of these OBs are discharged with household grey water, which in most communities is combined with fecal material in community sewage lines. The detection of OBs, since there is no known non-anthropogenic source, is indicative of human wastewater, with primary entry into environmental waters from septic tank discharges or leaking sewage lines. OB detection has been successfully applied in numerous watersheds across the globe, including: the U.S. (Hagedorn et al., 2003; McDonald et al., 2006), Japan (Hayashi et al., 2002), and New Zealand (Gilpin et al., 2002).

Virginia's public beaches are monitored during the summer swimming season by local health departments. Virginia Beach and Norfolk have undergone monitoring since 1990 (NRDC, 2006), but in late 2003, Virginia's public beach monitoring program was greatly expanded to include numerous other coastal, bay and riverfront beaches within the state. In 2004, Virginia formally adopted the use of the fecal indicator bacteria *Enterococcus spp*. and the USEPA recommendations for monitoring water quality in marine waters. Advisories would be posted on beaches where enterococci densities were found to exceed a single sample maximum of 104 cfu/100 mL, or a geometric mean of 35 cfu/100 mL for 5 of more samples collected in a calendar month. Results of improved monitoring efforts led to increases in the number of reported exceedances of the state regulatory standard. In 2004, Virginia beaches spend a total of 141 days under a posted advisory resulting from elevated enterococci counts. With Hilton and Anderson beaches, located along the James River, accounting for 61 of these days. Monitoring results

from 2005 showed the total number of days dropped to 42 due partly to remediation efforts at Hilton and Anderson beaches in Newport News (Dickerson et al., 2007).

In this study, conducted over two summers, objectives were to located beaches with, or the potential for, persistently high enterococci counts, and to apply the MST methods, ARA and fluorometry to determine the origin of fecal enterococci detected in beach water samples. The application of multiple methods, as was used in this study, has been emphasized by all three method comparison studies (Stewart et al., 2003; Stoeckel et al., 2004; Blanch et al., 2006) in light of the deficiencies found in all MST methods examined. While the summer of 2004 served primarily as a time to assess the enterococci levels on beaches, and locate visible and non-visible storm drains (SDs), MST was performed on all collected water samples providing at least eight confirmed, enterococci isolates. SD outfalls were of particular interest at these beaches, as SDs have been indicted as a means of transporting fecal waste and potential human pathogens to beach waters. Epidemiological studies in the Santa Monica Bay, CA found that 373 of every 100,000 swimmers developed disease-related symptoms after swimming near SD outfalls (Haile et al., 1999); a rate considerably higher than that in individuals swimming at a greater distance.

All beaches examined in 2005 were those with a past history of (Table 1), or the potential for, high enterococci counts and posted advisories (typically beaches with one of more SDs). Water samples were collected by Virginia Tech (VT) researchers twice a month (April-September) from both beach waters and SDs (if applicable) and subjected to both ARA and fluorometry, except for the beaches of the Eastern Shore, for which only ARA, and not fluorometry, was employed per request of the sponsor.

Materials and Methods

Site Descriptions and Sampling Locations

All beaches in this study are located in the Tidewater region of Virginia. Figure 1 is a map of the eastern region of Virginia, showing the location of each beach along the Atlantic Ocean, Chesapeake Bay, or the James, York or Potomac Rivers. A brief description of specific sampling locations within these beaches is found in Table 2.

While the Peninsula Health District routinely monitors five beaches in the city of Newport News, only two have been included in this chapter. Yorktown beach consistently produced low enterococci counts, with no exceedances recorded in 2004 or 2005 (Table 1) and did not have any SDs present on the beach, and was thus excluded from MST analyses. Hilton and Anderson were the two most polluted in all of Virginia during the summer of 2004 and have been excluded from this chapter, due to successful location and remediation efforts in 2004 and 2005, the details of which have been chronicled in Chapter IV.

Huntington (Figure 1; Pictures 2-4) in Newport News is 100 m stretch of beach located on the northern banks of the James River at the War Memorial Museum, adjacent to the James River Bridge. One location from which samples were collected (Huntington B) is also the regular monitoring location of the Peninsula Health District along the southwestern end of the beach, in a 50 m wide swimming area enclosed by yellow buoys. In addition, collections were made from a single SD (Huntington SD) located at northeastern end of the beach, near the James River Bridge (Picture 4). In 2005 an additional site (Huntington A) was added centrally between the yellow buoyed swimming area and SD.

King-Lincoln Park (Figure 1; Pictures 5-6) is roughly 300 m in length, running southwest to northeast along the northern banks of the James River. The park is flanked on the west by the

Aqua Vista apartment complex and contains a fishing pier (rebuilt in 2005) on the northern end. Samples are collected by the VDH in water waist deep at the northeastern end of the beach just north the pier (King-Lincoln B). In addition, collections were made from a single SD (King-Lincoln SD) located directly behind King-Lincoln B off the beach and not directly visible due to extensive vegetative cover. In 2005, an additional site (King-Lincoln A) was added on the southern side of the fishing pier.

Norfolk beaches (Figures 1-2; Pictures 7-10) encompass an 8 km stretch along the southern side of the Chesapeake Bay between the Norfolk Naval Station and USN Little Creek Amphibious Base. The Norfolk Health Department monitored nine locations (Figure 2) weekly during the 2004 beach season and reported only minor problems with enterococci counts (Table 1). The Norfolk beaches however, contain numerous SD outfalls, most of which empty directly into swimming areas, and many of which tested above state standards for enterococci in samples collected by VT staff in 2004. The enterococci collected from the potentially problematic sites Norfolk 8 and 9 (selected for proximity to SD outfalls extending into the surf), along with Norfolk 4 and 12 (marking the eastern and western edges of the beachfront, respectively) and three SDs surrounding Norfolk 9 (Norfolk SD-E, SD-W, and SD2) were subjected to MST protocols in 2004 and 2005.

The Hampton City Department of Health monitors five beach sites along the western edge of the Chesapeake Bay, Buckroe (3), Salt Pond, and Grandview Pier (Figure 1). However, only samples collected from Buckroe North and South (Pictures 11-12), and a nearby SD produced any elevated enterococci concentrations in 2004 and were subjected to MST protocols in 2004 and 2005. Buckroe North and South are separated by about 200 m, and bisected by a fishing pier, rebuilt in 2005. Buckroe South is located 100 m south of the fishing pier and 5 m

north of a SD outfall (also sampled in 2004), while Buckroe North is located 100 m north of the pier and 20 m south of a rock breakwall.

Fairview beach (Figures 1 and 3) is located in King George County on the southern banks of the Potomac River, northeast of Fredericksburg. The Rappahannock Health District routinely monitors four sites Fairview #1-4 weekly across the length of the 400 m stretch of beach (Figure 3). In 2004, a sinkhole was discovered by VT researchers at 8th Street (Fairview Sinkhole; Picture 13), off the beach and behind Fairview #2. This sinkhole was filled with water from a leaking, exposed SD. The SD outfall was visible extending out into the swimming area, but was not observed flowing. Samples were collected from the sinkhole in 2004, however prior to the start of the 2005 monitoring season the sinkhole was filled in with concrete by an unknown party.

The Eastern Shore Health District monitors four different beaches on both the ocean-side (1 beach) and bayside (3 beaches) of the peninsula. In Accomack County, Assateague Island National Seashore (Figure 1; Picture 14) is a large ocean-side beach on the northern part of the peninsula with four sampling locations, evenly distributed down a 6 km shoreline. Also, in Accomack, Guard Shore (Figure 1; Picture 15) is a smaller bayside beach (100 m) with two sampling locations on opposite ends of the beach. Kiptopeke (Figure 1; Picture 17) is a 200 m beach on the Chesapeake Bay, located within a state park on the southern peninsula in Northampton County. Kiptopeke has two sampling locations, as monitored by the department of health, at opposite ends of the beach. Cape Charles harbor (Figure 1; Picture 16) is a bayside beach located north of Kiptopeke in Northampton County. Cape Charles contains four monitoring locations evenly spaced across a 200 m stretch of beach, and is bordered on the north by several groins and partially submerged SD outfalls.

Water sample collection

Samples used for enumerations and MST were collected in sterilized plastic bottles by local health departments or VT staff in accordance with the DRAFT - VDH Beach Monitoring Protocol July 2004. All samples were collected in the surf in waist to chest deep water no more than 20 meters from the shore. Samples collected specifically for this study and processes at VT were kept at 4ºC for no more than 28 hours before beginning analyses. This holding time is longer than the maximum allowed time of 6 hours (VDH, 2004) to use enumerations, however acceptable by TMDL (USEPA, 2000) protocols where comparable bacterial counts have been reported (Pope et al., 2003).

Known-source sample collection

ARA libraries were constructed using *Enterococcus spp.* isolates from the lower Chesapeake Bay area. Isolates were collected from a variety of locations, every two to three weeks during the summer swimming season (June-August) and four to six weeks over the remainder of the year, in areas along the lower portions of the James River, the York River, around Lynnhaven Bay, public parks and piers at Virginia Beach, and at parks and embayments adjacent to the Chesapeake Bay. The success of multi-watershed libraries for regional geographic areas was used as a guide in designing the sampling plan (Wiggins et al., 2003). The wildlife category was represented by enterococci isolated from deer (*Odocoileus virginianus*) and raccoons (*Procyon lotor*) and in 2005, house mice (*Mus musculus*). The pets category was represented by dogs (*Canis familiaris*) and cats (*Felis cattus*) and the bird category contained isolates from ring-billed (*Larus delawarensis*), herring (*Larus argentatus*) and laughing (*Larus atricllia*) gulls, Canada (*Branta canadensis*) and snow (*Chen caerulescens*) geese, as well as

mallard duck (*Anas platyrhynchos*). Sewage influent samples from all twelve treatment plants within the region were provided during this time period by the Hampton Roads Sanitation District. None of the wastewater treatment plants contained combined sewers (sewage and surface runoff) therefore material labeled sewage should have represented a composite of almost exclusively human enterococci. All fecal samples were collected in sterile plastic bags and kept on ice until processed (within 24 hr). Both fresh and dried fecal samples were collected from animals in each category, except sewage, based on opportunity at the time samples were collected.

Isolation of enterococci

For known-source samples, a portion of each fecal or untreated sewage sample was diluted into tubes of sterile distilled de-ionized (DDI) water and spread on m-Enterococcus (m-Ent) agar (Baltimore Biologics Laboratory) to obtain *Enterococcus spp*. isolates. Collected beach water samples were filtered through a 0.45 µm pore-size membrane filter and transferred to a 50-mm Petri dish containing approximately 5mL of m-Ent agar. All m-Ent plates were incubated for 48-hour at 35ºC. Randomly selected red to burgundy colonies were picked from each plate (16-24 per water sample if available) using sterile toothpicks and inoculated into Enterococcosel Broth (Baltimore Biologics Laboratory) in a 96-well microtiter plate for confirmation as enterococci (black color after 48-hour incubation at 35ºC). Only confirmed enterococci isolates were utilized in ARA.

Antibiotic resistance analysis

Approximately 1µL of Enterococcosel Broth containing the enterococci isolates, described above, were transferred onto 31 different Trypticase Soy Agar (TSA; Difco™) plates containing one of 30 concentrations of nine different antibiotics (Table 3) and a control (no antibiotics) using a 48-prong replica-plater (Sigma Chemical Company, St. Louis, MO). Antibiotic concentrations were selected based on successful field applications from Graves et al., 2002. Plates were made by the addition of a given concentration of liquid antibiotics to flasks containing warm, sterile, TSA agar, swirled gently to mix, and poured into 80 mm Petri plates. After inoculation, plates were incubated at 35ºC for 48 hours. Resistance to a given concentration of antibiotic was evaluated based on a comparison of the growth of an isolate on the antibiotic plates versus growth on the control plate. Data were converted to binary, with growth on a given antibiotic concentration represented by a 1, and no growth represented by a 0. Any isolates which failed to grow on the control plates were excluded from the analysis. Results for each isolate were entered into SAS-JMP 5.0.1. ARA was not performed on water samples unless a minimum of eight enterococci could be cultured and verified in enterococcosel broth. This decision was based on previous research directed at obtaining a representative sampling of enterococci from a water sample. (Graves et al., 2003)

Creation of known-source library (host-origin database) and validation set

An ARA library was constructed during the summer of 2004 [1,057 isolates + 50 validation set (VS) isolates] and increased in size during 2005 to create a large multi-watershed library of 2,326 isolates + 100 VS isolates. All libraries and validation sets were created using no more than four isolates per fecal source and eight isolates per sewage sample. The number of

fecal samples was larger (for non-sewage samples, approximately 50/source category for 2004, and approximately 120/source category for 2005), the number of isolates obtained per sample was smaller, and the period over which fecal samples were obtained was longer (20 months), than other reports where ARA was used for MST (Hagedorn et. al., 1999; Harwood et al., 2000; Graves et al., 2002). Isolates obtained for the VS were collected at the same time as beach water samples. None of the fecal/sewage samples used to generate VS isolates were used to construct the known-source library. The 2004 VS isolates were not used calculate a minimum detectable percentage (MDP) for the 2005 library. All 2005 VS isolates were collected during the summer of the 2005 simultaneously with the fecal samples used to increase the library size, and the beach water samples.

Statistical analyses

Isolates collected from known sources were used to create a library of fingerprint patterns for the four selected source categories (birds, pets, sewage, and wildlife) using the statistical algorithm, discriminant analysis (DA; Wiggins, 1996). Analyses of isolates of unknown origin and the validation set were conducted in SAS-JMP statistical software (version 5.0.1, SAS Institute Inc.). The classification table generated in SAS-JMP was used to calculate the rate of correct classification (RCC) of each group of isolates, the average rate of correct classification (ARCC), and average frequency of misclassification (AFM) for the library and VS of isolates, and the minimum detectable percentage (MDP) for the library.

Calculation of average rate of correct classification, average frequency of misclassification and minimum detectable percentage

The ARCC for the library and VS was calculated by adding the percentage of isolates correctly classified from each source category (before rounding), and dividing by the total number of source categories. The average AFM was calculated by adding the percentage of isolates incorrectly classified from each source category and dividing by the total number of source categories (Harwood et al., 2003; Ritter et al., 2003). A separate MDP was determined for both the 2004 and 2005 ARA libraries, by averaging the values of the calculated AFM of both the library and corresponding VS. Calculating the MDP based on the average of the AFM of both the library and VS gave equal weight to the large number of isolates present in the library and the small number of non-library isolates used for validation. When classifying isolates of unknown origin (water samples), source categories identified at percentages below the MDP were considered a negligible contributing source.

Fluorometry

A Turner Designs 10-AU Fluorometer (Sunnyvale, CA; Picture 1) was used to detect the presence of fluorescent agents such as optical brighteners (OBs, found in laundry and dishwashing detergents) in water samples. The fluorometer was calibrated with known standards of a commercially available OB, Fluorescent Brightener 28 (FB-28; Sigma Chemical Co.), at concentrations ranging from 0 to 500 µg/ L. Standards and blanks (control) were made in volumetric flasks using DDI water and FB-28. The basic sensitivity of the fluorometer was set between 25 and 35% of full-scale using 250 µg/ L FB-28. The fluorometer was blanked using 0 µg/ L FB-28 in DDI water from the same carboy used to make-up the standards. The

fluorometer was calibrated to subtract the value of the blank from all measurements, and the blanking percentage was set between 0 and 5%. The standard solution concentration was set at 125, and calibrated using the 125 µg/ L FB-28 solution. After calibration of the fluorometer, a standard curve was constructed using concentrations of 0, 15, 30, 60, 125, 250, and 500 µg/ L FB-28. The fluorometer was recalibrated if readings did not fall within + or - 10% of the constructed standard values. Due to the tendency of FB-28 to degrade over time, new standards were made every two weeks and the fluorometer was recalibrated using the new standards and a standards curve was reestablished. Because variations were found in the levels of fluorescence between two different OBs [with Tinopal CBS-X (Ciba Specialty Chemicals) displaying approximately 60 times more fluorescence at the same concentration than FB-28] all units are relative to the OB used for calibration.

The determination of OB concentrations constituting a high (sewage positive) fluorometric reading was based on data gathered from samples of untreated sewage, ambient water samples containing virtually no fecal pollution (anthropogenic or otherwise), and attempts to correlate readings with high levels of sewage isolates, as indicated by ARA (Dickerson et al., 2007). Using the FB-28 calibration method, 24 untreated sewage samples, collected from 12 different sewage treatment plants in the coastal region of Virginia, yielded values ranging from 164 to 558 µg/ L with an average value of 259.5 µg/ L. Open water samples from areas with no potential sources of OBs from human wastewater, still yielded "background" levels of fluorescence ranging from 10 to 45 µg/ L based on the FB-28 calibration method. Accounting for the expected dilution of sewage when mixed with open waters, and the results obtained from ARA, a value of 100 µg/ L was selected as the minimum detectable threshold for a water sample to be considered positive for the presence of optical brighteners (Dickerson et al., 2007).

Results

Validation of 2004 and 2005 known-source libraries and establishment of MDPs

Within the 2004 library, the ARCC for a four-way split was 81.8% for isolates within the library, and 83.8% using the validation set (VS) of isolates (Table 4). The AFM for the 2004 library and VS was 18.2% and 16.3%, respectively. Using the AFM values, a MDP of 17.3% [(18.2 + 16.3) / 2] was calculated and applied to all isolates collected from beach water samples in 2004. Thus, if the number of isolates from a 2004 water sample classified into a given category was represented by a percentage below 17.3%, that source was considered a negligible contributor.

The ARCC for a four-way classification split in the 2005 library was 72.1% for isolates within the library, and 68.0% for the VS (Table 5). Using the AFM of 27.9% for the library and 32.0% for the VS, a MDP of 30.0% was applied to 2005 water isolates run against the 2005 ARA library.

Huntington Beach

Huntington Storm Drain (SD) was only sampled a single time in 2004 (Table 6A). ARA results indicated birds were the only fecal contributors in the sample, at the only percentage (66.7%) above the MDP for the 2004 library (17.3%). Although a fluorometric reading of 196 µg/L was detected in the sample, which is well above the 100 µg/L threshold, an oily sheen was apparent within the sample as a possible contaminant known to produce false positive readings (Hagedorn et al., 2003). Samples collected from Huntington B in 2004 (Table 7A), including two exceeding the state standard, indicated the only significant fecal polluters were birds and wildlife, with all fluorometric readings negative for sewage. The potentially conflicting ARA

and fluorometric results for the SD, coupled with several exceedances in 2004 (Table 1) warranted a more intensive MST monitoring of Huntington Beach in 2005.

In 2005, an additional site was added (Huntington A) midway between the SD and site B. Of the samples collected from Huntington SD (Table 6B) half (6 of 12) exceeded the maximum enterococci concentration (104 cfu/100 mL), and the two highest counts were found with the only two samples exceeding the 100 ug/L threshold for OBs. However, as was the case with the 2004 samples, only birds were identified as fecal contributors, classified as 62.5% of enterococci isolates. At Huntington A and B (Tables 7B and 8), 79.3% and 69.4% of enterococci isolates were identified a bird in origin, with no other source category was identified over the MDP (30.0%). In addition, no high fluorometric readings were detected at either location.

King-Lincoln

King-Lincoln SD (Table 9A) was sampled only a single time in 2004, results indicated that the majority of isolates were of sewage origin (75%) however, OB-levels were well below 100ug/L. In 2005, due to the sporadic flows from the SD, samples were collected only six times (Table 9B). Applying an MDP of 30.0% to ARA results, only birds and pets were deemed significant fecal contributors, however three of the six samples generated high fluorometric readings. Although like the Huntington SD samples, an oily sheen was also apparent in all samples.

Samples collected from King-Lincoln B in 2004 (Table 10A) indicated that only birds and wildlife were fecal contributors using an MDP of 17.3%. In 2005 only birds were detected as a major contributor (Table 10B). High levels of OBs were not detected in any the samples collected 2004 or 2005 for King-Lincoln B.

Results from King-Lincoln A (added in 2005, Table 11) agreed with those from King-Lincoln B in 2005, with no high levels of OBs and only birds were identified as a fecal polluter when applying a MDP of 30.0%.

Norfolk

In 2004, Norfolk 4 and 8 samples were determined to contain only significant contributions of enterococci of wildlife-origin, and no elevated OB readings (Tables 12A and 13A). After more extensive sampling in 2005, only bird was identified from Norfolk 4, and only wildlife from Norfolk 8 as contributing to fecal enterococci concentrations, again with no elevated fluorometric readings detected (Tables 12B and 13B).

At Norfolk 9, pets and wildlife were identified from two water samples, both containing elevated enterococci counts, collected in 2004 (Table 14A). However, after nine total samples were analyzed in the summer of 2005, with no elevated enterococci levels, only isolates from birds and wildlife were identified (Table 14B). No elevated OB concentrations were detected in any of the Norfolk 9 samples in 2004 or 2005.

Norfolk 9 SD-E and SD-W both contained low concentration of OBs, and high enterococci counts in the single samples collected in 2004, with isolates of bird- and wildlife-origin identified above the MDP (Tables 15A and 16A). Samplings in 2005 found extremely elevated levels of OBs in samples collected from SD-E, with soap bubbles apparent in several of the samples. Using an MDP of 30.0% for the 2005 library, birds and pets were identified as fecal contributors (Table 15B). Only two samples were collected from SD-W, which was not seen flowing most of the summer. Low OBs and only isolates identified as bird in origin were

identified (Table 16B), although a high enterococci count was found in the sample from 6/15/2005.

Two samples collected from Norfolk SD2 in 2004 contained elevated enterococci counts, moderately high, but not positive OB readings, and identified wildlife, pets, and sewage as fecal sources (Table 17A). Based on the results of the six samples collected in 2005, only five of which contained enough isolates to conduct ARA, only birds were identified as significant fecal source (Table 17B), however, the sample collected on 6/15/2005 contained a high fluorometric reading (110µg/L).

The one sample collected from Norfolk 12 in 2004 contained low OBs and enterococci counts however pets, sewage, and wildlife were identified as fecal sources (Table 18A). Following the more extensive sampling in 2005, only wildlife could be considered a significant source using the MDP of 30.0%.

Buckroe

Buckroe SD, sampled twice in 2004, was found to contain high enterococci counts (295 and 405 cfu/100 mL), wildlife, bird, and sewage isolates, and fluorometric readings above background levels but not considered positive for sewage (Table 19). However, prior to the start of the 2005 swimming season, Buckroe SD was capped, and water flowing into the beach waters was diverted. Thus, no samples were collected from the SD in 2005.

In the single samples collected from Buckroe North and South in 2004, both identified birds and pets as significant fecal sources using an MDP of 17.3% (Tables 20A and 21A). Although the sample from Buckroe South yielded a fluorometric reading above determined background levels (53.2 µg/L), both sites contained OBs well below the 100 µg/L threshold.

Sampling in 2005 produced only two samples, both from Buckroe North with high enterococci concentrations (Tables 20B and 21B), no elevated fluorometric readings were found, and wildlife birds were found to be the only fecal source of significance.

Fairview

Two samples were collected from the Fairview Sinkhole in 2004, before it was filled-in with concrete, by an unknown party, prior to the start of the 2005 swimming season. Both samples contain elevated enterococci levels, and one sample produced a high fluorometric reading of 171µg/L (Table 22). In both samples, human was identified as the major fecal source, with wildlife also a prevalent contributor.

Samples collected from Fairview #1 in 2004 yielded high enterococci counts, and were identified as bird, wildlife, and pet in origin (Table 23A). No fluorometric readings were found to be positive for sewage contamination. In 2005, four of the five samples exceeded the state standard for enterococci, however the VDH practice of averaging counts across all four Fairview sites prevented the posting of advisories for several of these counts (Table 23B). Enterococci were identified as bird in origin.

Enterococci collected from Fairview #2, in 2004, were identified as originating from birds, pets, and wildlife (Table 24A). Samples collected a Fairview #3 and #4 (Tables 25A and 26A) were only classified as birds and wildlife using an MDP of 17.3%. In 2005, enterococci isolated from sites #2 and #3 (Tables 24B and 25B) were determined to be of bird-origin, while #4 (Table 26B) contained isolates from both birds and wildlife. None of the sites produced high fluorometric readings in 2004 or 2005.

Assateague Island

Enterococci concentrations in the beach waters of Assateague Island were consistently low, as measured by the VDH. No exceedances were recorded in 2004 or 2005 (Table 1) and counts were generally so low that isolates from the four locations along the beaches were combined for ARA, as was the case for the all beaches of the Eastern Shore (Guard Shore, Cape Charles, and Kiptopeke). In addition, water samples were not collected by, or shipped to Virginia Tech. Only m-Ent plates were received, with colonies present on membrane filters. Therefore, data on concentrations of *Enterococcus spp*. or fluorometric data are not available.

ARA results from Assateague Island concluded that only wildlife isolates were found from the two collections in 2004 (Table 27A). Only birds were identified as levels about the MDP (30.0%) from 2005 collections (Table 27B).

Guard Shore

Samples collected in 2004 from Guard Shore identified birds and wildlife at levels above the MDP of 17.3% for the 2004 library and were thus considered the only significant fecal contributors (Table 28A). Birds were sole fecal source identified in beach waters in 2005 (Table 28B).

Cape Charles

Results from Cape Charles sample collections indicated that enterococci from wildlife dominated in 2004 (Table 29A). However, in 2005, birds were the only source identified as a fecal contributor in the beach waters (Table 29B).

Kiptopeke

Samples collected from Kiptopeke in 2004 found only enterococci from birds and wildlife at levels above the MDP of 17.3% (Table 30A). Birds were found to be the sole fecal source identified on the beach in 2005 (Table 30B).

Discussion

Huntington Beach

Huntington beach posted three advisories in 2004 and a high fluorometric reading from the SD, warranting additional monitoring and MST testing in 2005. Results from all three locations in 2005 identified isolates collected from both the beach and SD as bird in origin. These results correlate very well with a visual inspection of the beach area. Posted signs prohibit people from walking pets on the beach (Picture 2) and every sampling trip, dozens of shorebirds and ducks were observed both on the beach and in the surrounding waters (Picture 3). While the SD did flow consistently (Picture 4), sewage leakage was not deemed as probable based on the ARA results and noticeable iridescent sheens, typical of automotive oils and known to interfere with fluorometry readings (Hagedorn et al., 2005), present in several of the samples.

King-Lincoln

King-Lincoln beach spent 21 days under a swimming advisory in 2004. Although ARA determined that only birds and wildlife enterococci were present in the water, the large number of advisory days, coupled with the detection of suspected sewage isolates in the SD prompted further monitoring in 2005. The King-Lincoln SD was found to be flowing (typically a very slow trickle) only six of the twelve sampling trips in 2005, which is not indicative the expected

high flow rate of an outfall cross-connected with a sewer line. In addition, the only high fluorometric readings were present in samples with oily-sheens, as was the case for Huntington SD, which most likely produced false positive readings. The ARA results for King-Lincoln SD in 2005 confirmed this by only finding birds and pets as fecal contributors. Birds were also identified as the major problem in open beach waters, a likely result of the rebuilt fishing pier located centrally on the beach (Pictures 5 and 6) which was observed frequently with 20 or more seabirds perched on the hand rails.

Norfolk

Norfolk received only minor attention in 2004 in terms of MST monitoring, due largely to a lack of cooperation from the Norfolk Health Department. The large number of SDs present on the beach, combined with the suspected detection of sewage isolates in Norfolk 9 SD2 (Picture 7; Table 17A), validated further MST applications in 2005. Norfolk 4 and 12 located at opposite ends of the Norfolk beaches (Figure 2) were both located next to SDs (Picture 8). It was thought these locations would allow for monitoring of enterococci swept along the length of the beach from the tides coming into and exiting the Chesapeake Bay. However, the massive inflow and outflow of water in the bay during tidal changes likely diluted samples keeping counts very low in 2005, particularly at site 4, where only six of the twelve samples contained enough verified *Enterococcus spp*. isolates to perform ARA (at least eight required).

Norfolk SD-W was rarely observed flowing and was probably on a minor contributor to any beach pollution. In addition isolates were identified primarily as bird in origin in both 2004 and 2005 (Picture 9). Enterococci collected from Norfolk SD-E (Picture 9) were identified primarily as birds and pets, however, extremely high levels of OBs were detected in seven of the

eight collected samples, with most of these containing what appeared to be soap bubbles. Observations of the surrounding area located a laundry mat, just off the beach, not more than 100 m from the SD (Picture 10). The high fluorometric readings within this SD are the probable result of an illegal connection from the laundry mat to the SD. However, we were never made aware of a follow-up study into this hypothesis by local officials. Only birds and pets were detected at Norfolk SD2 in 2005, and only one unexplained high fluorometric reading, with no obvious contaminants visible. It is likely that this single sample from SD2 represented an isolated incident of fluorescent material from surface runoff. In addition, the sewage detected in the single 2004 could have been the odd sample or inaccurate classifications by the 2004 library. However, evidence for SD2 from 2005 does not support a probable sewage leak.

Buckroe

Buckroe beach was under advisory for only six days in 2004 however, samples collected from Buckroe SD in 2004 indicate a minor sewage leak, as indicated by ARA, but not fluorometry, might be present. The diversion of the SD prior to 2005 prevented a follow-up study. However, the number of posted advisories decreased only slightly (from three to two; Table 1). A pet signature was found in 2004 and dogs, as well as a dog park (Picture 11), were observed on the beach. Although pets were the second highest source identified in 2005 (19.6%), levels still remaining below the MDP for the 2005 library. The reconstruction of the fishing pier, located centrally, between the two sampling locations (Picture 12) likely contributed to the bird enterococci detected.

Fairview

Fairview beach, spending 25 days under an advisory in 2004, warranted further application of MST in 2005. In addition, one of the two samples collected from the Fairview Sinkhole (Picture 13) in 2004 yielded high fluorometric readings (with the second well above background values), with both samples indicating a human signature was present as determined by ARA. A leaking sewage line was likely contributing to the water in the sinkhole, however once filled with concrete prior to 2005, additional samples were unavailable. While it is possible sewage contaminated water could have entered the swimming area via subsurface flow out of the sinkhole, significant human isolates were never detected in beach water samples. With several exceedances of the state entercocci standards detected in both 2004 and 2005 at one or more sites, Fairview beach may contain undetected SDs or be subject to the contamination of beach waters from upstream pollution on the Potomac River. However, with only bird and wildlife (and pets at site #1 in 2004) implemented as fecal contributors, remediation efforts will be difficult.

Eastern Shore Beaches (Assateague Island, Guard Shore, Cape Charles, Kiptopeke)

Based on sampling trips from 2004, all results are consistent with observations made at each beach. Enterococci from birds and/or wildlife were detected in water samples collected from 2004. Assateague Island (Picture 14) contains no human residences and no pets are allowed on the island. Guard Shore (Picture 15) was observed as a fairly remote beach, and although pets are allowed, none were observed during the 2004 visit. Both Cape Charles (Picture 16) and Kiptopeke (Picture 17) were observed with large bird populations. The major difference between ARA results in 2004 and 2005 was the absence of wildlife in 2005.

The heavy rainfall totals experienced during 2004 (Table 31) likely contributed to increases surface runoff and elevated the wildlife signature, while birds were visibly present on rocks, SDs, and groins (if applicable) and in the sand and water at all beaches. With rainfall considerably lower in 2005, runoff contributions from wildlife were probably significantly less, resulting in the detection of only isolates of bird-origin in 2005.

Fluorometry

No high readings were detected in open waters, and based on results from Dickerson et al., 2007 (see Chapter IV) the dilution of OB concentrations to background fluorescence levels in open bodies of water would have prevented detection. Fluorometry showed potential in SDs such as Norfolk SD-E, and interference from aromatic compounds is usually obvious in a sample. However, a few samples, such as Norfolk SD2 displayed high fluorometric readings and low sewage-origin isolates with no obvious signs of interference from other compounds. Several compounds commonly used by automobiles have yielded high fluorometric readings in previous tests (Hagedorn et al., 2005). A potential solution to this problem was worked out using UV light to photodegrade OBs within water samples and determining their presence based on the change in fluorometic readings. Based on initial tests by J.W. Dickerson Jr. at Virginia Tech (pers. comm.) performed in August 2005, a larger study was undertaken at the University of Georgia. Results indicated that a reduction in fluorometric reading of 30% or more after 30 minutes under a UV light tended to indicate a sample contained OBs (Hartel et al., 2007). Continued testing in smaller waterbodies, with less potential for dilution is warranted as is continued research into possible interference compounds. The potential for detection of OBs by

measuring their photodegradation under a UV light remains promising as a means to distinguish between other compounds fluorescing within wavelengths similar to OBs.

Conclusions

Differences between 2004 and 2005 ARA results with the beaches studied can be attributed to several factors.

1. Changes within the structure of the beaches. Several fishing piers were rebuilt before the summer of 2005, which appeared to attract shorebirds to the area that were not present during sampling trips in 2004.
2. Increasing the number sample collections – Results from 2004 were only based on a few samples collected from each beach, and only spanning the end of the swimming season. Increased frequency of sampling in 2005 allowed for a reduction in the significance of sampling anomalies, when results for the entire season were combined.
3. Rainfall – The summer of 2004 saw significantly more rain than 2005 (Table 31), particularly during the months in which ARA was performed in 2004 (end of June – August). The lower rainfall totals in 2005 restricted SD flow at many beaches (differences in flow rates were not measured), preventing fecal material from animals not normally present on the beach from entering the water.
4. Higher MDP between libraries – The MDP used for 2005 required the classification of isolates from a water sample into a given source category at 12.7 percentage points higher than 2004. This more stringent requirement in 2005 typically eliminated one category that would have been identified using the 2004 MDP. Larger libraries tend to have lower correct classification rates (Whitlock et al., 2002), but tend to perform better

in real-world situations (Wiggins et al., 1999). Therefore, the greater constraints put on the 2005 library are believed to provide more accurate source identifications of isolates of unknown-origin (Stewart et al., 2003).

References

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

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

Carroll, S., M. Hargreaves, and A. Goonetilleke. 2005. Sourcing faecal pollution from onsite wastewater treatment systems in surface waters using antibiotic resistance analysis. J. Appl. Microbiol. 99:471-482.

Cheung, W.H., K.C. Chang, R.P. Hung, and J.W. Kleevens. 1990. Health effects of beach water pollution in Hong Kong. Epidem. Infect. 105:139-162.

Dickerson, J. W. Jr., C. Hagedorn, and A. Hassall. 2007. Detection and remediation of human-origin pollution at two public beaches in Virginia using multiple source tracking methods. Water Res. 41:3758-3770.

Geary, P., V. Shah, H. Dunstan, P. Coombes, and T. Rothkirch. 2006. Tracing faecal contributions from on-site wastewater systems. Wat. 33:48-51.

Gilpin, B.J., J.E. Gregor, and M.G. Savill. 2002. Identification of the source of faecal pollution in contaminated rivers. Wat. Sci. Technol. 46:9-15.

Glassmeyer, S.T., E.T. Furlong, D.W. Kolpin, J.D. Cahill, S.D., Zaugg, S.L. Werner, M.T. Meyer, and D.D. Kryak. 2005. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. Environ. Sci. Technol. 39:5157-5169.

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

Graves, A.K. 2003. Identifying source of fecal pollution in water as a function of sampling frequency under low and high stream flow conditions. Virginia Polytechnic Institute and State University. Dissertation.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneneau. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522-5531.

Hagedorn, C., R.B. Reneau Jr., M. Saluta, A. Chapman. 2003. Impact of onsite wastewater systems on water quality in coastal regions. Virginia Coastal Resources Management Program Memorandum of Agreement 50312-01-13-PT. Virginia Dept. of Conservation and Recreation, Virginia Dep. of Health, Richmond.

Hagedorn, C., M. Saluta, A. Hassall, and J. Dickerson. 2005. Fluorometric detection of optical brighteners as an indicator of human sources of water pollution. Part I. Description and detection of optical brighteners. Crop and Soil Environmental News, November 2005. [online] http://www.ext.vt.edu/news/periodicals/cses/2005-11/part1.html 27 November 2007.

Haile, R.W., J.S. Witte, M. Gold, R. Cressey, C. McGee, R.C. Millikan, A. Glasser, N. Harawa, C. Ervin, P. Harmon, J. Harper, J. Dermand, J. Alamillo, K. Barrett, M. Nides, and G-Y. Wang. 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. Epidem. 10:355-363.

Hartel, P.G., C. Hagedorn, J.L. McDonald, J.A. Fisher, M.A. Saluta, J.W. Dickerson, Jr., L.C. Gentit, S.L. Smith, N.S. Mantripragada, K.J. Ritter, and C.N. Belcher. 2007. Exposing water samples to ultraviolet light improves fluorometry for detecting human fecal contamination. Wat. Res. 41:3629-3642.

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

Hayashi, Y., S. Managaki, and H. Takada. 2002. Fluorescent whitening agents in Tokyo Bay and adjacent rivers: their application as anthropogenic molecular markers in coastal environments. Environ. Sci. Technol. 36:3556 -3563.

McDonald, J.L., P.G. Hartel, L.C. Gentit, C.N. Belcher, K.W. Gates, K. Rodgers, J. A.Fisher, K.A. Smith, and K.A. Payne. 2006. Identifying sources of fecal contamination inexpensively with targeted sampling and bacterial source tracking. J. Environ. Qual. 35:889-897.

NRDC. 2006. Natural resources defense council. Testing the waters: a guide to water quality at vacation beaches - July 2006 [online] http://www.nrdc.org/water/oceans/ttw/sumvir.pdf last accessed 13 May 2007).

Orosz-Coghlan, P.A., P.A. Rusin, M.M. Karpiscak, and C.P. Gerba, C.P. 2006. Microbial source tracking of *Escherichia coli* in constructed wetlands. Wat. Environ. Res. 78:227-232.

Poiger, T., J.A. Field, T.M. Field, H. Siegrist, and W. Giger. 1998. Behavior of fluorescent whitening agents during sewage treatment. Wat. Res. 32:1939-1947.

Pond, K. R., R. Rangdale, W.G. Meijer, J. Brandao, L. Falcão, A. Rince, B. Masterson, J. Greaves, A. Gawler, E. McDonnell, and A. Cronin, S. Pedley. 2004. Workshop Report: Developing pollution source tracking for recreational and shellfish waters. Environ. For. 5:237-247.

Pope, M.L., M. Bussen, M.A. Feige, L. Shadix, S. Gonder, C. Rodgers, Y. Chambers, J. Pulz, K. Miller, K. Connell, and J. Standridge. 2003. Assessment of the effects of holding time and temperature on *Escherichia coli* densities in surface water samples. Appl. Environ. Microbiol. 69:6201-6207.

Portsmouth Weather Services Records. [online] http://members.cox.net/wxr/month.htm last accessed 13 May 2007).

Rabinovici, S.J.M., R.L. Bernknopf, A.M. Wein. 2003. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Tech. 38:2737-2745.

Ritter, K.J., C. 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 using in library-based approaches to microbial source tracking. J. Wat. Health. 1:209-223.

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.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2002. Microbial source tracking: state of the science. Environ Sci. Technol. 36:5280-5288.

Stewart, J.R., R.D.Ellender, J.A.Gooch, S. Jiang, S.P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons learned from a methods comparison study. J. Wat. Health. 1:225-231.

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

USEPA. 1986. Ambient Water Quality Criteria for Bacteria - 1986. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA-440/5-84-002.

USEPA. 2000. Protocols for Developing Pathogen TMDLs. U.S. Environmental Protection Agency, Washington, D.C. EPA-841-R-00-002.

VDH. 2004. DRAFT–Virginia Department of Health Beach Monitoring Protocol July 2004 (unpublished).

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. Wat. Res. 36:4273-4282.

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.

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.

Figure 1 – Map of beaches monitored by VDH in 2004 and 2005 (Festival Beach was added in 2006; image courtesy of VDH).

Figure 2 – Map of the VDH sampling locations at Norfolk beaches. Sites 4, 8, 9, and 12 were sampled by VT researchers (image courtesy of VDH).

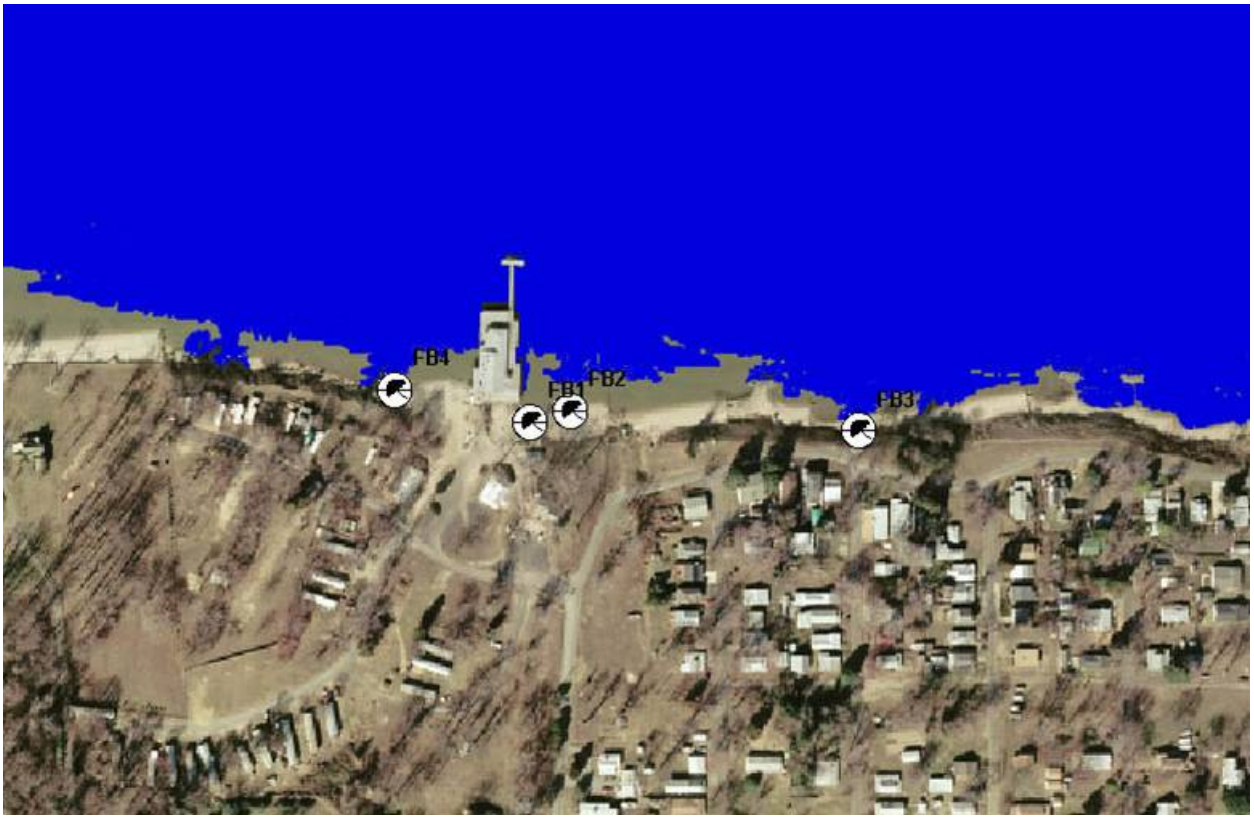


Figure 3 – Map of the VDH sampling locations at Fairview Beach. All sites were sampled by VDH staff, water samples and enterococci plates were shipped to VT (image courtesy of VDH).

Table 1 – City/county location, number of sampling sites, and number of advisories and total days spent under an advisory of beaches in Virginia.

Beach	City/County	# of sampling sites	# of advisories (total days) 2004	2005
Anderson^	Newport News	1	4(27)	1(2)
Hilton^	Newport News	1	4(34)	2(13)
Huntington*	Newport News	2	3(13)	0(0)
King-Lincoln*	Newport News	1	3(21)	4(7)
Yorktown	Newport News	1	0(0)	0(0)
Buckroe*	Hampton	3	3(6)	2(2)
Grandview Pier	Hampton	1	0(0)	0(0)
Salt Ponds	Hampton	1	0(0)	0(0)
Norfolk*	Norfolk	9	5(12)	0(0)
Assateague Island*	Accomack	4	0(0)	0(0)
Guard Shore*	Accomack	2	0(0)	2(8)
Cape Charles*	Northampton	4	0(0)	0(0)
Kiptopeke*	Northampton	2	0(0)	0(0)
Fairview*	King George	4	4(25)	2(8)
Gloucester Point	Gloucester	1	1(2)	0(0)
Virginia Beach	Virginia Beach	24	0(0)	2(2)
Total		61	**27(141)**	**15(41)**

^ results included in Chapter IV

* results included in this Chapter (Chapter III)

Table 2 – Beach sampling location and description.

Sample Site	Site Description
Huntington SD	Storm drain outfall on east end of the beach. Sample taken at the end of the pipe prior to entry into the surf
Huntington A	Sample collected at the midpoint between swimming area (designated by yellow buoys) and the storm drain.
Huntington B	Regular VDH monitoring site. in middle of swimming area (designated by yellow buoys).
King-Lincoln SD	Storm drain outfall that empties into the beach about 10 m from the shoreline. Drain is located 10 m north of the pier.
King-Lincoln A	Sample collected 20 m south of pier (located at the northern end of the beach)
King-Lincoln B	Regular VDH monitoring site, located 20 m north of pier (located at the northern end of the beach)
Buckroe S	Regular VDH monitoring site, located on northern end of the beach, 20 m north of the visible storm drain outfall.
Buckroe N	Sample collected on northern end of the beach, 20 m south of the visible storm drain outfall.
Norfolk 4	Regular VDH monitoring site, located behind the "Ship's Cabin' restaurant. Sample collected behind the 2nd westernmost break wall.
Norfolk 8	Regular VDH monitoring site, located at North Community Beach. Sample collected about 5m west of the double storm drain outfall.
Norfolk 9	Regular VDH monitoring site, located at Ocean View Park about 5 m west of the jetty.
Norfolk SD2	Storm drain 30m east of Norfolk 9. Sample collected at the end of the pipe.
Norfolk SD-E	Eastern pipe of double storm drain outfall 40m west of Norfolk 9. Sample collected at the end of the pipe.
Norfolk SD-W	Western pipe of double storm drain outfall 40m west of Norfolk 9. Sample collected at the end of the pipe.
Norfolk 12	Western end of the beach; regular Health Department monitoring site; 13th view about 5 meters to the west of the jetty, waist deep in the surf.
Assateague Island	Regular VDH monitoring sites. Four sampling locations evenly spaced along the beach 200m apart.
Guard Shore	Regular VDH monitoring sites. Two sampling locations at opposite ends of the beach separated by 40m.
Cape Charles	Regular VDH monitoring sites. Four sampling locations evenly spaced along the beach 50m apart.
Kiptopeke	Regular VDH monitoring sites. Two sampling locations at opposite ends of the beach separated by 100m.
Fairview Sinkhole	Sinkhole off the beach located immediately behind Fairview #2, 20m from shoreline.
Fairview #1	Regular VDH monitoring site, located 5m east of fishing pier/restaurant.
Fairview #2	Regular VDH monitoring site, located 20m east of fishing pier/restaurant.
Fairview #3	Regular VDH monitoring site, located 100m east of fishing pier/restaurant.
Fairview #4	Regular VDH monitoring site, located 40m west of fishing pier/restaurant.

Table 3 – Antibiotics and concentrations (given in ng/mL) used for antibiotic resistance analysis.

Antibiotic	**Concentrations (ng/mL)**
Amoxicillin	0.0625
Cephalothin	1, 1.5, 3, 5
Chloratetracycline	6, 8, 10
Erythromycin	1, 1.5, 3, 5
Oxytetracycline	2, 4, 6, 8, 10
Streptomycin	4, 6, 8, 10
Tetracycline	1, 1.5, 3, 5, 10
Neomycin	4, 6, 8
Vanomycin	0.25

Table 4 – Classification table of library and validation set isolates for the 2004 *Enterococcus* known-source library.

	Percentage (number) of isolates classified as:				
Source (number of isolates)	Bird	Dog	Sewage	Wildlife	
Bird (n = 248)	**80.24 (199)**	6.45 (16)	8.87 (22)	4.44 (11)	
Dog (n = 168)	6.55 (11)	**77.38 (130)**	10.12 (17)	5.95 (10)	ARCC = 81.8%
Sewage (n = 368)	8.70 (32)	7.07 (26)	**78.53 (289)**	5.71 (21)	AFM = 18.2%
Wildlife (n = 261)	2.30 (6)	3.83 (10)	2.68 (7)	**91.19 (238)**	
					MDP =17.3%
Validation Bird (n = 10)	**100.00 (10)**	0.00 (0)	0.00 (0)	0.00 (0)	
Validation Dog (n = 10)	0.00 (0)	**70.00 (7)**	20.00 (2)	10.00 (1)	ARCC = 83.8%
Validation Sewage (n = 20)	10.00 (2)	5.00 (1)	**85.00 (17)**	0.00 (0)	AFM = 16.3%
Validation Wildlife (n = 10)	0.00 (0)	10.00 (1)	10.00 (1)	**80.00 (8)**	

†Bold indicates rate of correct classification (RCC) for each source category
#Italics indicates calculated minimum detectable percentage (MDP)
Average rate of correct classification (ARCC); Average frequency of misclassification (AFM)

Table 5 - Classification table of library and validation set isolates for the 2005 *Enterococcus* known-source library.

	Percentage (number) of isolates classified as:				
Source (number of isolates)	Bird	Dog	Sewage	Wildlife	
Bird (n = 630)	**69.84 (440)**	10.95 (69)	7.46 (47)	11.75 (74)	
Dog (n = 463)	19.87 (92)	**69.98 (324)**	6.70 (31)	3.46 (16)	ARCC = 72.1%
Sewage (n = 675)	7.11 (48)	10.22 (69)	**75.26 (508)**	7.41 (50)	AFM = 27.9%
Wildlife (n = 558)	15.59 (87)	1.25 (7)	9.86 (55)	**73.30 (409)**	
					MDP = *30.0%*#
Validation Bird (n = 25)	**68.00 (17)**	20.00 (5)	4.00 (1)	8.00 (2)	
Validation Dog (n = 25)	24.00 (6)	**56.00 (14)**	12.00 (3)	8.00 (2)	ARCC = 68.0%
Validation Sewage (n = 25)	8.00 (2)	12.00 (3)	**76.00 (19)**	4.00 (1)	AFM = 32.0%
Validation Wildlife (n = 25)	4.00 (1)	20.00 (5)	4.00 (1)	**72.00 (18)**	

†Bold indicates rate of correct classification (RCC) for each source category
#Italics indicates calculated minimum detectable percentage (MDP)
Average rate of correct classification (ARCC); Average frequency of misclassification (AFM)

Table 6 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Huntington SD 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/24/2004	**1480**	16	3	2	3	24	**196**
Total	-	16	3	2	3	24	
Percentage	-	**66.7**	12.5	8.3	12.5		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	30	10	0	2	4	16	83.7
4/15/2005	50	12	0	1	3	16	43.5
5/1/2005	**13500**	7	6	0	3	16	**172**
5/15/2005	**190**	14	1	1	0	16	50
6/1/2005	45	12	0	0	4	16	68.4
6/15/2005	**6960**	5	7	1	3	16	**215**
7/15/2005	**540**	9	1	2	4	16	96.7
7/31/2005	**1600**	7	4	2	3	16	73.1
8/15/2005	**220**	12	0	1	3	16	92.9
8/30/2005	**1450**	9	3	1	3	16	89.9
9/15/2005	80	11	0	0	5	16	64.4
9/30/2005	40	12	0	1	3	16	27.5
Total	-	120	22	12	38	192	
Percentage	-	**62.5**	11.5	6.3	19.8		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 7 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Huntington B 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/24/2004	70	11	0	3	2	16	34.1
8/17/2004	**410**	2	1	1	10	14	51.7
8/18/2004	**180**	4	1	0	11	16	49.7
Total	-	17	2	4	23	46	
Percentage	-	**37**	4.3	8.7	**50**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	5						38.3
4/15/2005	35	12	2	0	2	16	27.2
5/1/2005	**1680**	12	0	0	4	16	42.2
5/15/2005	15	7	0	0	1	8	53.6
6/1/2005	68	3	3	2	8	16	34
6/15/2005	**462**	9	3	0	4	16	36.9
7/15/2005	94	12	0	0	4	16	31.8
7/31/2005	26	16	0	0	0	16	34.3
8/15/2005	24	11	2	0	3	16	37
8/30/2005	40	13	1	0	2	16	29.4
9/15/2005	6						27.3
9/30/2005	9	5	1	0	2	8	32.3
Total	-	100	12	2	30	144	
Percentage	-	**69.4**	8.3	1.4	20.8		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 8 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Huntington A 2005.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	0						36.8
4/15/2005	20	9	1	2	4	16	26.1
5/1/2005	**1890**	15	0	0	1	16	42.6
5/15/2005	8	6	1	0	1	8	26.1
6/1/2005	80	11	0	2	3	16	34.3
6/15/2005	**1470**	12	1	0	3	16	43.2
7/15/2005	90	14	0	0	2	16	30.2
7/31/2005	30	12	0	3	1	16	33.9
8/15/2005	16	13	0	0	3	16	40.3
8/30/2005	14	13	0	0	1	14	31
9/15/2005	10	8	0	0	0	8	25
9/30/2005	9	6	1	0	1	8	26.1
Total	-	119	4	7	20	150	
Percentage	-	**79.3**	2.7	4.7	13.3		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 9 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for King-Lincoln SD in 2004 (A) and 2005 (B)

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
9/14/2004	**480**	1	0	12	3	16	30.1
Total	-	1	0	12	3	16	
Percentage	-	6.3	0	**75**	**18.8**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005							
4/15/2005	50	0	12	4	0	16	23.1
5/1/2005	**680**	7	0	4	5	16	**230**
5/15/2005	**1200**	14	0	0	2	16	**162**
6/1/2005	25	6	4	3	3	16	34.1
6/15/2005	**460**	5	8	2	1	16	**232**
7/15/2005							
7/31/2005							
8/15/2005							
8/30/2005	100	4	6	2	4	16	26.1
9/15/2005							
9/30/2005							
Total	-	36	30	15	15	96	
Percentage	-	**37.5**	**31.3**	15.6	15.6		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 10 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for King-Lincoln B in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/15/2004	**560**	11	4	2	7	24	20.2
6/24/2004	**400**	3	9	2	9	24	32.3
6/29/2004	**1580**	20	1	0	3	48	37.8
8/17/2004	**830**	0	0	0	11	11	45.6
Total	-	34	14	4	30	107	
Percentage	-	**31.8**	13.1	3.7	**28**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	10	3	9	4	0	16	33.3
4/15/2005	20	3	3	8	2	16	28.3
5/1/2005	62	4	2	5	5	16	27.2
5/15/2005	15	9	1	0	6	16	17.2
6/1/2005	0						21.7
6/15/2005	26	5	6	0	5	16	22.9
7/15/2005	34	6	6	0	4	16	20.3
7/31/2005	27	6	3	0	7	16	25.1
8/15/2005	13	9	4	0	3	16	26.2
8/30/2005	102	6	5	0	5	16	28
9/15/2005	**335**	4	4	0	8	16	26.9
9/30/2005	9	7	3	0	6	16	21.3
Total	-	62	46	14	51	176	
Percentage	-	**35.2**	26.1	8	29		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 11 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for King-Lincoln A in 2005.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	10	2	1	3	2	8	36.5
4/15/2005	25	2	7	6	1	16	30
5/1/2005	24	12	1	1	2	16	26
5/15/2005	15	13	0	3	0	16	22.6
6/1/2005	4						22.9
6/15/2005	24	7	3	1	5	16	23.3
7/15/2005	40	6	5	0	5	16	20.1
7/31/2005	33	8	4	0	4	16	39.7
8/15/2005	4						28.3
8/30/2005	84	7	5	0	4	16	25.5
9/15/2005	10	3	4	0	9	16	27.6
9/30/2005	17	4	3	0	5	12	20.3
Total	-	64	33	14	37	148	
Percentage	-	**43.2**	22.3	9.5	25		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 ml), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 12 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 4 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	51	0	0	0	8	8	18
Total	-	0	0	0	8	8	
Percentage	-	0	0	0	100		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	0						26
4/15/2005	0						22.3
5/1/2005	6	12	0	2	2	16	20.2
5/15/2005	2						15.1
6/1/2005	3.3						22.9
6/15/2005	10	10	0	0	6	16	18
7/15/2005	14						16.6
7/31/2005	2	12	0	0	4	16	18.3
8/15/2005	13	13	0	0	3	16	29.9
8/30/2005	2						19.3
9/15/2005	55	10	1	0	5	16	21.7
9/30/2005	9	11	2	0	3	16	14.5
Total	-	68	3	2	23	96	
Percentage	-	**70.8**	3.1	2.1	24		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 13 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 8 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	215	0	0	1	7	8	19
Total	-	0	0	1	7	8	
Percentage	-	0	0	12.5	**87.5**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005							
4/15/2005							
5/1/2005	250	3	3	5	5	16	25.8
5/15/2005	16	5	2	2	7	16	15.4
6/1/2005	18.3	3	0	2	11	16	24.3
6/15/2005	60	5	0	11	0	16	18.5
7/15/2005	5	3	0	2	3	8	16
7/31/2005	31	5	4	2	5	16	18.4
8/15/2005	27	2	4	1	7	14	27.2
8/30/2005	1						26.8
9/15/2005	51	4	2	0	10	16	19.6
9/30/2005	20	7	2	1	5	15	16.6
Total	-	37	17	26	53	133	
Percentage	-	27.8	12.8	19.5	**39.8**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 14 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 9 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	185	0	5	0	3	8	18
8/2/2004	216	0	6	1	4	11	28.7
Total	-	0	11	1	7	19	
Percentage	-	0	**57.9**	5.3	**36.8**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	20	1	2	4	9	16	32.4
4/15/2005	5						22.7
5/1/2005	4						26.2
5/15/2005	100	13	0	1	2	16	19.6
6/1/2005	3.3	0	4	2	10	16	22.5
6/15/2005	40	5	2	3	6	16	25.8
7/15/2005	7						15.5
7/31/2005	23	1	4	5	6	16	18.1
8/15/2005	36	5	5	1	5	16	24.7
8/30/2005	16	9	0	2	3	14	19.7
9/15/2005	25	6	4	1	5	16	19.9
9/30/2005	37	4	3	0	8	15	18.9
Total	-	44	24	19	54	141	
Percentage	-	**31.2**	17	13.5	**38.3**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 15 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 9 SD-E in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	**470**	8	0	1	7	16	25
Total	-	8	0	1	7	16	
Percentage	-	**50**	0	6.3	**43.8**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005							
4/15/2005							
5/1/2005	1150	10	2	1	3	16	**801**
5/15/2005	990	5	10	0	1	16	**407**
6/1/2005	5960	2	13	0	1	16	**518**
6/15/2005	5						**233**
7/15/2005	1840	8	3	4	1	16	**493**
7/31/2005							
8/15/2005	15700	4	3	1	8	16	**>999**
8/30/2005	30	4	3	2	7	16	19.1
9/15/2005	550	7	0	2	7	16	**174**
9/30/2005							
Total	-	40	34	10	28	112	
Percentage	-	**35.7**	**30.4**	8.9	25		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 16 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 9 SD-W in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	**475**	12	1	0	3	16	28
Total	-	12	1	0	3	16	
Percentage	-	**75**	6.3	0	**18.8**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005							
4/15/2005							
5/1/2005							
5/15/2005							
6/1/2005							
6/15/2005	**2740**	13	2	0	1	16	30.3
7/15/2005							
7/31/2005							
8/15/2005							
8/30/2005	20	10	4	2	0	16	18.5
9/15/2005							
9/30/2005							
Total	-	23	6	2	1	32	
Percentage	-	**71.9**	18.8	6.3	3.1		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 17 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 9 SD2 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	830	0	5	0	3	8	71
8/2/2004	960	0	2	4	6	12	88
Total	-	0	7	4	9	20	
Percentage	-	0	**35**	**20**	**45**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005							
4/15/2005	40	3	7	5	1	16	22.8
5/1/2005							
5/15/2005							
6/1/2005	**185**	5	2	0	9	16	64.3
6/15/2005	**2460**	13	1	0	2	16	**110**
7/15/2005	100	1	9	1	5	16	95.7
7/31/2005	**700**	4	5	2	5	16	80.3
8/15/2005							
8/30/2005	0						21.2
9/15/2005							
9/30/2005							
Total	-	26	24	8	22	80	
Percentage	-	**32.5**	**30**	10	27.5		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 18 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Norfolk 12 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/19/2004	65	0	3	3	2	8	25
Total	-	0	3	3	2	8	
Percentage	-	0	**37.5**	**37.5**	**25**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	4	1	0	4	3	8	29.4
4/15/2005	5	1	9	4	2	16	23.8
5/1/2005	**240**	1	14	1	0	16	29.9
5/15/2005	62	10	0	6	0	16	18.9
6/1/2005	3						27.6
6/15/2005	37	4	0	9	2	15	18.6
7/15/2005	31	1	0	0	15	16	20.7
7/31/2005	14	4	0	2	10	16	18.6
8/15/2005	23	6	1	0	9	16	25.2
8/30/2005	6						19.6
9/15/2005	56	4	5	1	6	16	19
9/30/2005	5						18.6
Total	-	32	29	27	47	135	
Percentage	-	23.7	21.5	20	**34.8**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 19 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Buckroe SD in 2004 (drainage redirected away from the beach before 2005).

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/24/2004	**295**	8	5	5	6	24	77.3
7/14/2004	**405**	3	0	4	17	24	85.2
Total	-	11	5	9	23	48	
Percentage	-	**22.9**	10.4	**18.8**	**47.9**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 20 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Buckroe South in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/24/2004	65	11	9	3	1	24	53.2
Total	-	11	1	3	9	24	
Percentage	-	**45.8**	**37.5**	12.5	4.2		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	5						29
4/15/2005	5						23.5
5/1/2005	38	9	2	2	3	16	46
5/15/2005	8						16.6
6/1/2005	10						18.5
6/15/2005	31	11	3	0	2	16	15.2
7/15/2005	48	8	6	1	1	16	17
7/31/2005	19	16	0	0	0	16	22.1
8/15/2005	53	11	3	1	1	16	24.7
8/30/2005	27	11	4	0	1	16	21.1
9/15/2005	71	12	4	0	0	16	22.1
9/30/2005	9						19.7
Total	-	78	22	4	8	112	
Percentage	-	**69.6**	19.6	3.6	7.1		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 21 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Buckroe North in 2005.

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/24/2004	26	5	2	0	1	8	22.8
Total	-	5	2	0	1	8	
Percentage	-	**62.5**	**25**	0	12.5		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
3/31/2005	10						32.1
4/15/2005	20	12	2	0	2	16	24.7
5/1/2005	10						21.7
5/15/2005	28	10	2	1	3	16	16.5
6/1/2005	70	10	1	2	3	16	19.8
6/15/2005	8						18.1
7/15/2005	**206**	9	4	3	0	16	19.1
7/31/2005	25	13	2	0	1	16	21.8
8/15/2005	31	14	2	0	0	16	24.9
8/30/2005	24	9	0	0	7	16	22.2
9/15/2005	**208**	11	2	0	3	16	22.7
9/30/2005	6						22.3
Total	-	88	15	6	19	128	
Percentage	-	**68.8**	11.7	4.7	14.8		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 22 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Fairview Sinkhole in 2004.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/16/2004	**280**	1	3	8	12	24	78.7
6/23/2004	**485**	1	3	16	4	24	**171**
Total	-	2	6	24	16	48	
Percentage	-	4.2	12.5	**50**	**33.3**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 23 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Fairview #1 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/16/2004	**300**	12	5	0	7	24	42.6
6/23/2004	**560**	10	8	0	4	16	89.2
9/17/2004	**244**	1	0	9	10	20	68.1
Total	-	23	13	9	21	60	
Percentage	-	**38.3**	**21.7**	15	**35**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/22/2005	**177**	2	8	0	6	20	29
7/6/2005	**161**	9	6	1	0	16	61.7
7/13/2005	**340**	13	3	3	1	20	NA
7/20/2005	**128**	6	8	3	3	20	66.5
7/27/2005	46	3	2	1	17	23	52.8
Total	-	33	27	8	27	99	
Percentage	-	**33.3**	27.3	8.1	27.3		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 24 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Fairview #2 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/16/2004	45	14	4	0	6	24	24.6
6/23/2004	**400**	8	5	0	11	24	87.1
9/17/2004	**228**	0	5	8	7	20	57.6
Total	-	22	14	8	24	68	
Percentage	-	**32.4**	**20.6**	11.8	**35.3**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/22/2005	**181**	6	1	0	9	20	30.3
7/6/2005	**170**	17	2	0	0	19	59
7/13/2005	14	14	2	0	2	18	NA
7/20/2005	**276**	4	2	4	11	21	59
7/27/2005	84	10	0	5	8	23	46
Total	-	51	7	9	30	101	
Percentage	-	**50.5**	6.9	8.9	29.7		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 25 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Fairview #3 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/16/2004	25	9	7	0	8	24	40.2
6/23/2004	**265**	12	2	0	10	24	90.5
9/17/2004	85	0	0	7	13	20	64.8
Total	-	21	9	7	31	68	
Percentage	-	**30.9**	13.2	10.3	**45.6**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/15/2005	7	19	0	0	1	20	31
6/22/2005	22	8	3	2	1	14	73.8
7/6/2005	**170**	14	5	1	0	20	47.4
7/13/2005	6	2	2	9	1	14	45.8
7/20/2005	88	3	1	1	15	20	NA
7/27/2005	44	4	1	3	15	23	48.8
Total	-	50	12	16	33	111	
Percentage	-	**45**	10.8	14.4	29.7		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 26 – Enterococci concentrations (cfu/100 mL), source classifications (number of isolates), total source classifications for the summer (total number of isolates and percentage), and concentration of optical brighteners (OB) in µg/L for Fairview #4 in 2004 (A) and 2005 (B).

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/16/2004	25	11	6	0	7	24	43.4
6/23/2004	**385**	9	3	4	8	24	90.5
9/17/2004	**152**	0	1	1	10	20	57.9
Total	-	20	10	5	25	68	
Percentage	-	**29.4**	14.7	7.4	**36.8**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/15/2005	29	5	5	0	6	16	30.4
6/22/2005	14	7	2	3	3	15	54.8
7/6/2005	**182**	23	0	1	0	24	64.8
7/13/2005	36	2	0	15	0	17	73.5
7/20/2005	**148**	2	1	0	21	24	
7/27/2005	76	1	1	3	19	24	48.4
Total	-	40	9	22	49	120	
Percentage	-	**33.3**	7.5	18.3	**40.8**		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 27 – Source classifications (number of isolates) and total source classifications for the summer (total number of isolates and percentage) for Assateague Island National Seashore in 2004 (A) and 2005 (B). As samples were received as m-Ent plates (no water samples), and enterococci from different sampling locations along the same beach were combined, no data are available for OB or enterococci concentrations.

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/13/2004		5	0	0	5	10	
8/4/2004		7	0	1	8	16	
Total	-	12	0	1	13	26	
Percentage	-	**46.2**	0	3.8	**50**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
6/1/2005		3	6	2	5	16	
7/6/2005		7	3	4	2	16	
7/13/2005		6	0	0	2	8	
Total	-	16	9	6	9	40	
Percentage	-	**40**	22.5	15	22.5		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 28 – Source classifications (number of isolates) and total source classifications for the summer (total number of isolates and percentage) for Guard Shore in 2004 (A) and 2005 (B). As samples were received as m-Ent plates (no water samples), and enterococci from different sampling locations along the same beach were combined, no data are available for OB or enterococci concentrations.

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/13/2004		3	0	0	11	14	
7/15/2004		7	3	2	12	24	
9/2/2004		2	3	0	13	18	
Total	-	12	6	2	36	56	
Percentage	-	**21.4**	10.7	3.6	**64.3**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
5/25/2005		7	12	0	1	20	
6/1/2005		1	4	0	3	8	
6/8/2005		1	15	0	0	16	
6/15/2005		6	0	1	1	8	
7/6/2005		2	0	1	21	24	
7/13/2005		22	1	0	1	24	
7/14/2005		16	5	0	3	24	
Total	-	55	37	2	30	124	
Percentage	-	**44.4**	29.8	1.6	24.2		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 29 – Source classifications (number of isolates) and total source classifications for the summer (total number of isolates and percentage) for Cape Charles in 2004 (A) and 2005 (B). As samples were received as m-Ent plates (no water samples), and enterococci from different sampling locations along the same beach were combined, no data are available for OB or enterococci concentrations

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/13/2004		3	0	0	15	18	
7/19/2004		0	0	0	16	16	
7/26/2004		0	2	2	11	15	
8/2/2004		1	1	2	11	15	
8/9/2004		3	0	1	12	16	
Total	-	7	3	5	65	80	
Percentage	-	8.8	3.8	6.2	**81.2**		

B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
5/24/2005		16	4	0	0	20	
5/31/2005		10	5	0	1	16	
6/7/2005		3	4	0	1	8	
6/14/2005		5	10	0	1	16	
6/21/2005		13	4	6	1	24	
6/28/2005		14	0	0	8	22	
7/5/2005		9	5	4	6	24	
7/12/2005		21	1	0	2	24	
7/23/2005		18	2	1	0	21	
8/9/2005		9	3	1	11	24	
Total	-	118	38	12	31	199	
Percentage	-	**59.3**	19.1	6	15.6		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 30 – Source classifications (number of isolates) and total source classifications for the summer (total number of isolates and percentage) for Kiptopeke in 2004 (A) and 2005 (B). As samples were received as m-Ent plates (no water samples), and enterococci from different sampling locations along the same beach were combined, no data are available for OB or enterococci concentrations

A.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
7/13/2004		8	0	0	8	16	
9/2/2004		12	0	0	4	16	
Total	-	20	0	0	12	32	
Percentage	-	**62.5**	0	0	**37.5**		

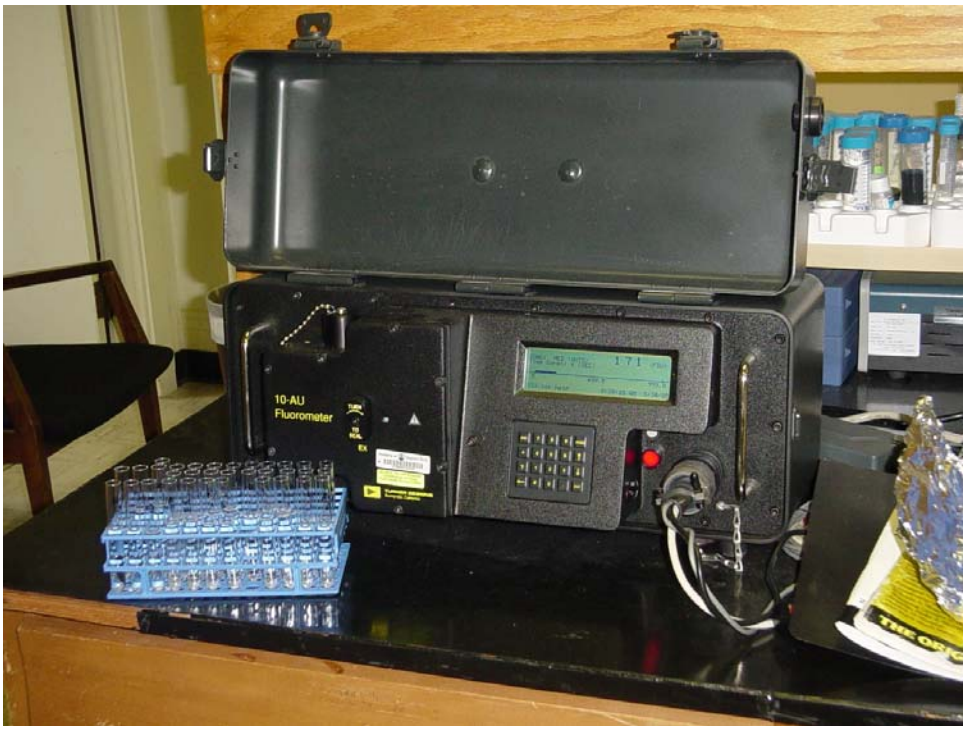
B.

Date collected	cfu/100 mL	Birds	Pets	Sewage	Wildlife	Total	OB (µg/L)
5/31/2005		4	1	1	2	8	
6/21/2005		5	1	0	2	8	
6/28/2005		7	4	4	1	16	
7/6/2005		13	2	0	1	16	
Total	-	29	8	5	6	48	
Percentage	-	**60.4**	16.7	10.4	12.5		

Gray boxes indicate that sufficient isolates were not available for ARA or the sample was not collected on that date.
Due to rounding, source category percentages may not add to 100%.
Bold indicates, enterococci counts exceeding the state standard (104 cfu/100 mL), source percentage above the MDP (17.3% for 2004, or 30.0% for 2005), or OB concentrations over 100 µg/L.

Table 31 – Monthly precipitation total (in inches) for the summers of 2004 and 2005.

Month/Year	Precipitation (in inches)
Apr-04	2.76
May-04	3.96
Jun-04	6.98
Jul-04	6.91
Aug-04	10.23
Sep-04	5.14
Total	35.98
Apr-05	2.94
May-05	4.55
Jun-05	4.89
Jul-05	6.34
Aug-05	3.98
Sep-05	2.36
Total	25.06



Picture 1 – Turner Designs 10AU Fluorometer.

Picture 2- Huntington Beach – 'No Pets Allowed' sign

Picture 3- Huntington beach, with yellow buoy present in foreground, blue-green storm drain visible in the background just as the rocks begin, and birds, which maintained a constant presence both on the beach and in the water.

Picture 4 – Sampling the Huntington SD, looking west down the length of the entire beach.

Picture 5 – King-Lincoln pier in 2004, with shore birds prevalent.

Picture 6 – Collecting samples at King-Lincoln A in 2005 with rebuilt pier visible in the background.

Picture 7 – Collection of a water sample from a flowing Norfolk 9 SD2.

Picture 8 – SD at Norfolk 12 almost completely covered, even at low tide.

Picture 9 – Norfolk SD-E and SD-W (double storm drain) visible looking westward down the beach at Norfolk 9. Both storm drains, if flowing, empty directly into the surf.

Picture 10 – Laundry mat immediately behind Norfolk SD-E and SD-W just off the beach. This facility is suspected of being illegally connected to SD-E and the source of extremely high fluorometric readings and visible soap bubbles in collected samples.

Picture 11 – Dog park at Buckroe beach was one of many locations where pet enterococci were collected to construct the ARA known-source library. Pets were frequently observed outside the park, with owners seldom cleaning up after their pets.


Picture 12 – Buckroe beach pier located central between the two sampling locations.

Picture 13 – Fairview Sinkhole, identified as a source of human-origin enterococci in 2004. Filled in with concrete by an unknown party before further testing could be conducted in 2005.

Picture 14 – Assateague Island, a frequently used beach with no visible storm drains, and no pets allowed on the island.

Picture 15 – Looking southward down the length of Guard Shore beach

Picture 16 – Numerous SDs and groins visible on the northern end of Cape Charles beach. Numerous birds were observed on these structures, but are not visible in the picture.

Picture 17 – Birds prevalent on Kiptopeke beach.

IV. Detection and Remediation of Human-Origin Pollution at Two Public Beaches in Virginia Using Multiple Source Tracking Methods*

*From Water Research, with permission (http://www.elsevier.com/wps/find/supportfaq.cws_home/rightsasanauthor). Water Res.
Dickerson, J.W. Jr., C. Hagedorn, and A. Hassall. 2007. 41:3758-3770.

Abstract

Two public beaches (Anderson and Hilton) in Newport News, Virginia, were frequently closed to swimming in 2004 due to high *Enterococcus spp.* counts that exceeded the regulatory standard. The microbial source tracking (MST) methods of antibiotic resistance analysis (ARA) and fluorometry (to detect optical brighteners) were used in the summer of 2004 to determine the origins of fecal pollution at the two beaches. Both MST methods detected substantial human-origin pollution at the two beaches, in locations producing consistently high levels of *Enterococcus spp.* Investigations by municipal officials led to the fluorometric detection and subsequent repair of sewage infrastructure problems at both beaches. The success of the mitigation efforts was confirmed during the summer of 2005 using ARA and fluorometry, with the results cross-validated by pulsed-field gel electrophoresis (PFGE).

Introduction

When bacterial levels in recreational waters exceed state standards, local authorities are required to post advisories or close swimming areas in the interest of protecting public health. Exposure to fecal-contaminated recreational waters has been shown to increase the risk of contracting gastroenteritis, as well as ear, eye, nose and skin infections (Cheung et al., 1990).

The greatest correlation between gastrointestinal illness and bacterial densities in marine recreational waters was found with the detection of the fecal organism *Enterococcus spp*. (Cabelli et al., 1982). In addition to the risks to public health, extended swimming advisories/closures on public beaches can result in significant revenue losses for local economies (Rabinovici et al., 2004). As a result, beach managers have a significant interest in the identification and quantification of the contributing sources of fecal pollution in beach waters. Specific knowledge of fecal sources allows for the most efficient use of best management practices (BMPs) in reducing fecal counts and preventing sustained beach advisories/closures.

A large number of potential fecal sources may exist within a watershed including: wastewater treatment plants, individual septic systems, livestock and poultry farms (lagoons or field runoff), boat-waste discharges, urban storm water runoff (dogs and birds), and wildlife. Knowledge of land use patterns can often lead researchers to a suspected source; however, in many instances the fecal contributors are not obvious. The origin and development of microbial source tracking (MST) over the last decade has provided a means to discriminate between fecal sources in polluted waters. When used properly, MST can assist watershed managers in the construction and implementation of BMPs to reduce fecal inputs and minimize risks to water users. The premise of MST is that enteric bacterial strains are uniquely adapted to, and thus reside exclusively in, the gastrointestinal tract of a single or group of closely-related (e.g. livestock, wildlife, domestic pets) host organisms. Utilizing differences in the phenotypes or genotypes of fecal organisms has been the basis of all MST methods developed to date to identify and quantify fecal contributions to polluted waters from different sources (Scott et al., 2002; Blanch et al., 2006).

Antibiotic resistance analysis (ARA) was one of the earliest MST methods to be developed and tested in a variety of different watershed environments (Wiggins, 1996; Hagedorn et al., 1999; Harwood et al., 2000) and is still widely used (Carroll et al., 2005; Geary, et al., 2006; Orosz-Coghlan et al., 2006). Pulsed-field gel electrophoresis, (PFGE) despite being both expensive and time consuming, has also found success as an MST tool, but has been employed less frequently than ARA (Simmons et al., 2002; Samadpour et al., 2005). Both ARA and PFGE were included in the two U.S.-based MST method comparison (MC) studies (Harwood et al., 2003; Myoda et al., 2003; Stoeckel et al., 2004). Although, problems were encountered with all the MST methods, PFGE emerged as one of the most accurate and reproducible methods evaluated. Substantial issues with accuracy were reported for ARA (as well as most other methods), but both studies described approaches, such as including a blind challenge (or validation) set of known-source isolates, using multiple MST methods, and including an alternative tracer, that could compensate for the methodology problems (Stewart et al., 2003; Stoeckel et al., 2004).

Chemical-based methods have recently emerged as a means of detecting the presence of compounds associated with fecal pollution, most commonly those that are human in origin, as an alternative to geographically variable microorganisms (McDonald et al., 2006). Although the detection of a wide variety of chemicals has been examined (Glassmeyer et al., 2005), most detection assays require both technical expertise and expensive equipment, which may be beyond the abilities and budgets of many state monitoring agencies. The use of optical brighteners (OBs) that can be detected using a relatively simple piece of equipment (a fluorometer) has been applied in numerous watersheds in the U.S. (Hagedorn et al., 2003; McDonald et al., 2006), Japan (Hayashi et al., 2002), and New Zealand (Gilpin et al., 2002).

Optical brighteners, also known as fluorescent whitening agents, are organic compounds used in household detergents that absorb long-wave ultraviolet light (365 nm) and reemit it most strongly within the blue portion of the visible spectrum (415nm to 435nm). Anywhere from 25-95% of the OBs used in a wash cycle are bound to clothing during the wash (Poiger et al., 1998), while the remaining portion of these OBs are discharged from the home within wastewater. The detection of OBs, with no known non-anthropogenic source, is indicative of human wastewater, with primary entry into environmental waters from septic tank discharges or community sewage lines.

The public beaches in Virginia are monitored weekly by local health departments, with beaches in both Virginia Beach and Norfolk having undergone regular bacterial monitoring since 1991 (NRDC, 2006). In 2003, Virginia's public beach monitoring program was greatly expanded to include numerous other coastal, bay and riverfront beaches within the state. In 2004, Virginia formally adopted the use of fecal indicator *Enterococcus* spp. and the requirements of the federal BEACH (Beaches Environmental Assessment and Coastal Health) Act for monitoring water quality in public beach waters. Advisories would be posted on beaches where enterococci densities were found to exceed a single sample maximum of 104 colony forming units (cfu)/100 mL, or a geometric mean of 35 cfu/100 mL for 5 of more samples collected in a calendar month. In 2004, Virginia beaches spend 141 total days under a posted advisory resulting from elevated *Enterococcus spp*. counts (NRDC, 2006). Hilton and Anderson beaches proved to be especially problematic in 2004 and accounted for 91 (64.5%) of the posted advisory days.

The goal of this study, beginning in 2004, was to use monitoring and MST to evaluate Virginia's public beaches as they were rebuilt following destruction from Hurricane Isabel in

September 2003. Specific objectives were to: i) identify beaches with persistent water quality issues, ii) determine the sources of fecal contamination at these problematic beaches, iii) develop and assess the utility of fluorometric detection of OBs as an indicator of human fecal pollution, iv) examine the reliability of ARA with a blind validation set and cross-validation with PFGE; and v) evaluate the success of mitigation efforts at the two beaches following detection and repair of infrastructure problems.

Materials and Methods

Research site description

Hilton and Anderson beaches are located in Newport News County in southeastern Virginia. Both beaches are adjacent to the James River, a 547 km, east-flowing tidal river, as wide as 8 km at the mouth, emptying into the Chesapeake Bay. The Virginia Department of Health (VDH) – Newport News District routinely collects a single weekly grab sample at both Hilton and Anderson beaches during the summer swimming season (May-September).

Hilton is a small beach located on the northern banks of the James River approximately 20 km upstream of the Chesapeake Bay. Only covering a linear distance of 75 m, the sandy section of the beach ranges from 1 to 10 m in width, depending on location and tidal level. Hilton beach contains a single storm drain outfall (SD) at the easternmost end of the beach that empties into the sand approximately 10 m above the high tide line, and drains into the surf 5 m from a metal fence extending into the water, marking the end of the public beach. The Hilton SD was observed in 2004 as a steadily flowing drain regardless of precipitation conditions over the preceding days. In 2005, less frequent and lower flow conditions were observed during sample collections.

Anderson beach is a located on the northwestern banks of the James River as the river turns to the northeast and enters the Chesapeake Bay. Anderson beach is divided into two sections separated by a boat landing and two jetties that allow access into the river. The northern section of the beach extends about 25 m before sand fades into large breakwater rocks. The southern and major recreational section of beach is 200 m long and 20 to 30 m wide. Anderson beach also contains a single visible SD centrally located along the southern stretch of beach. Emptying directly into the surf, and often partially covered at high tide, the Anderson SD flowed intermittently during both 2004 and 2005, with heaviest flows observed following precipitation events.

Water sample collection

All water samples were collected at Anderson and Hilton beaches in Newport News, VA, from storm drain outfalls or in open waters surrounding these outfalls. On Hilton beach, sample SD was collected directly from the storm drain outfall prior to entry into the surf zone. Hilton sample B (regular VDH monitoring location) was collected in waist deep water 10 m west of the storm drain. Hilton sample A, collected only during the summer of 2005, was obtained in waist deep water, 4 m east of the storm drain outfall. At Anderson beach, sample SD was collected directly from the storm drain outfall prior to entry into the surf zone (if not covered by a high tide). Anderson sample B (regular VDH monitoring location) was collected 40 m southwest of the SD, in waist deep water. Anderson sample A, was collected 10 m southwest of the SD, in waist deep water, only during summer 2005. No water samples were collected to the northeast of Anderson SD because of limited accessibility due to large breakwater rocks lining the edge of the beach and turbulent currents.

Water samples were collected in sterile plastic bottles in accordance with DRAFT - VDH Beach Monitoring Protocol (VDH, 2004). During the summers of 2004 and 2005, all weekly water samples found to exceed state enterococci standards (104 cfu/100 mL), by VDH were shipped on ice overnight to Virginia Tech to perform MST protocols. During 2005 samples were collected by Virginia Tech researchers twice a month, April 2005-September 2005, from storm drains, regular VDH monitoring sites, and selected locations of interest as described above. All samples were kept on ice during transport, and assayed within 18 hours of collection. Enterococci densities in water samples were determined in the VDH and Virginia Tech labs using the Enterolert system (IDEXX Laboratories, Westbrook, ME). Additionally, per request from VDH in 2004, several collections were made from the single storm drain outfall present on each beach.

Known-source sample collection

Both the ARA and PFGE libraries were constructed using *Enterococcus spp.* isolates from the lower Chesapeake Bay area. Isolates were collected from a variety of locations, every two to three weeks during the summer swimming season (June-August) and four to six weeks over the remainder of the year, in areas along the lower portions of the James River, the York River, around Lynnhaven Bay, public parks and piers at Virginia Beach, and at parks and embayments adjacent to the Chesapeake Bay. The success of multi-watershed libraries for regional geographic areas was used as a guide in designing the sampling plan (Wiggins et al., 2003). The wildlife category was represented by enterococci isolated from deer (*Odocoileus virginianus*) and raccoons (*Procyon lotor*) and in 2005, house mice (*Mus musculus*). The pets category was represented by dogs (*Canis familiaris*) and cats (*Felis cattus*) and the bird category

contained isolates from ring-billed (*Larus delawarensis*), herring (*Larus argentatus*) and laughing (*Larus atricllia*) gulls, Canada (*Branta canadensis*) and snow (*Chen caerulescens*) geese, as well as mallard duck (*Anas platyrhynchos*). Sewage influent samples from all twelve treatment plants within the region were provided during this time period by the Hampton Roads Sanitation District. None of the wastewater treatment plants contained combined sewers, therefore material labeled sewage should have represent a composite of almost exclusively human enterococci. All fecal samples were collected in sterile plastic bags and kept on ice until processed (within 24 hr). Both fresh and dried fecal samples were collected from animals in each category, except sewage, based on opportunity at the time samples were collected.

Isolation of enterococci

For known-source samples, a portion of each fecal or untreated sewage sample was diluted into tubes of sterile distilled de-ionized (DDI) water and spread on m-Enterococcus agar (Baltimore Biologics Laboratory) to obtain *Enterococcus spp*. isolates. Collected beach water samples were filtered through a 0.45 μm pore-size membrane filter and grown on m-Enterococcus agar. All m-Enterococcus plates were incubated for 48-hour at 35ºC. Randomly selected red to burgundy colonies were picked from each plate (24 per water sample if available) using sterile toothpicks and inoculated into Enterococcosel Broth (Baltimore Biologics Laboratory) in a 96-well microtiter plate for confirmation as enterococci (black color after 48-hour incubation at 35ºC). Only confirmed enterococci isolates were utilized in ARA and PFGE.

Antibiotic resistance analysis

Antibiotic Resistance Analysis was performed as described in Hagedorn et al., 1999 and Wiggins, 1996 on isolates collected in both 2004 and 2005 using a total of 30 concentrations of 9 different antibiotics (one to five concentrations selected for each antibiotic). Results for each isolate were analyzed (using SAS-JMP 5.0.1), against a library of isolates collected from known-sources, using discriminant analysis (DA; Wiggins, 1996) for four-way classifications (bird, sewage, pets, and wildlife) and logistic regression (LR; Dickerson et al., 2007) for two-way classifications (sewage/non-sewage). Antibiotic concentrations were selected based on successful field applications from Graves et al., 2002.

Resistance to a given concentration of antibiotic was evaluated based on a comparison of the growth of an isolate on the antibiotic plates versus growth on the control plate. Data were converted to binary, with growth on a given antibiotic concentration represented by a 1 and no growth represented by a 0. Any isolates which failed to grow on the control plates were excluded from the analysis.

Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis was performed as described in Simmons et al., 2002, on approximately 30% of 2005 water sample isolates (for cross-validations) and known-source isolates (to generate a known-source library). Enterococci isolates confirmed in Enterococcosel Broth (Baltimore Biologics Laboratory) were streaked from the microtiter plate onto individual plates of Tryptic Soy Agar (TSA, Difco™) and incubated overnight at 35 ºC. Cultures from TSA plates, were added to cell suspension buffer (1M NaCl, 10mM Tris-HCl, pH 7.6) treated with lysozyme (20 mg/ml), and embedded in 1.2% InCert™ agarose to form a cell suspension

plug. Cell plugs were added to a cell lysis buffer (50mM Tris-HCl, 50mM EDTA, pH 8) to complete lysing of the cells and incubated for 1 hr. at 37 ºC. Plugs were then rinsed with wash buffer (1mM Tris-HCl, 0.01 mM EDTA, ph 7.5), and incubated with proteinase K (20 mg/ml, Promega) at 55 ºC for 2 hrs. at 100 rpm in a Lab-Line® Orbit Environ-Shaker to digest proteins. The plugs were washed four times with wash buffer and stored at 4 ºC. Total DNA from each isolate (plug) was digested using the *Sma*I (Promega, 5'-CCC^GGG-3', 3'-GGG^CCC-5') restriction enzyme for 16 hrs. at 25 ºC. DNA fragments, and a Bio-Rad λ ladder (48.5-970 kbp), were separated on a 1% agarose gel using a Bio-Rad GenePath® system in 0.5X TBE at 14ºC for 20 hrs. The gels were stained with ethidium bromide (1µg/ml), and de-stained in distilled water. Photographs of the gels were taken using a Polaroid camera (aperture setting 7.5, shutter speed 1) on a long-wave UV transilluminator. Images were scanned and saved as Microsoft Windows Bitmap files. The molecular band sizes for each digest profile were determined using Jandel Scientific SigmaGel™ gel analysis software. Band lengths were entered into SAS-JMP statistical software (version 5.0. 1, SAS Institute Inc.) for each *Enterococcus* isolate, and classification models were generated using LR for a two-way source (sewage/non-sewage) split.

Creation of known-source library (host-origin database) and validation set

An ARA library was constructed during the summer of 2004 [1,057 isolates + 50 validation set (VS) isolates] and increased in size during 2005 to create a large multi-watershed library of 2,326 isolates + 100 VS isolates. All libraries and validation sets were created using no more than four isolates per fecal source and eight isolates per sewage sample. The number of fecal samples was larger (for non-sewage samples, approximately 50/source category for 2004, and approximately 120/source category for 2005), the number of isolates obtained per sample

was smaller, and the period over which fecal samples were obtained was longer (20 months), than other reports where ARA was used for MST (Hagedorn et. al., 1999, Harwood et al., 2000, Graves et al., 2002). Isolates obtained for the VS were collected at the same time as beach water samples. None of the fecal/sewage samples used to generate VS isolates were used to construct the known-source library. The 2004 VS isolates were not used calculate a minimum detectable percentage (MDP) for the 2005 library. However, the 2004 VS isolates were also run against the library to further assess the temporal stability of the library over the two summers. All 2005 VS isolates were collected during the summer of the 2005 simultaneously with the fecal samples used to increase the library size and the beach water samples. A smaller PFGE library of 420 isolates + 40 VS isolates was constructed for cross-validation of approximately 30% (selected at random) of the 2005 water sample isolates.

Statistical analyses

Analyses for ARA and PFGE data were conducted using two parametric classification methods, discriminant analysis (DA) and logistic regression (LR), in SAS-JMP statistical software (version 5.0.1, SAS Institute Inc.); comparing isolates from within the library and the VS against the model constructed using patterns from isolates in the known-source library. The classification table generated in SAS-JMP was used to calculate the rate of correct classification (RCC) of each group of isolates, the average rate of correct classification (ARCC), and average frequency of misclassification (AFM) for the library and VS of isolates, and the minimum detectable percentage (MDP) for the library.

Calculation of average rate of correct classification, average frequency of misclassification and minimum detectable percentage

The ARCC for the ARA and PFGE libraries and corresponding VS was calculated by adding the percentage of isolates correctly classified from each source category (before rounding), and dividing by the total number of source categories. The average AFM was calculated by adding the percentage of isolates incorrectly classified from each source category and dividing by the total number of source categories (Harwood et al., 2003; Ritter et al., 2003). A MDP was determined for both versions (two-way and four-way split) of the 2004 and 2005 ARA libraries, as well as the 2005 PFGE library, by averaging the values of the calculated AFM of both the library and corresponding VS. Calculating the MDP based on the average of the AFM of both the library and VS gave equal weight to the large number of isolates present in the library and the small number of non-library isolates used for validation. When classifying isolates of unknown origin (water samples), source categories identified at percentages below the MDP were considered a negligible contributing source.

Fluorometry

A Turner Designs 10-AU Fluorometer (Sunnyvale, CA) was used to detect the presence of fluorescent agents such as optical brighteners (OBs, found in laundry and dishwashing detergents) in water samples. The fluorometer was calibrated with known standards of a commercially available OB, Fluorescent Brightener 28 (FB-28, Sigma Chemical Co.), at concentrations ranging from 0 to 500.0 µg/ L. Standards and blanks (control) were made using DDI water and FB-28. The basic sensitivity of the fluorometer was set between 25 and 35% of full-scale using 250.0 µg/ L FB-28. The fluorometer was blanked using 0.0 µg/ L FB-28 in DDI

water from the same carbide used to make-up the standards. The fluorometer was calibrated to subtract the value of the blank from all measurements, and the blanking percentage was set between 0 and 5%. The standard solution concentration was set at 125, and calibrated using the 125.0 µg/ L FB-28 solution. After calibration of the fluorometer, a standard curve was constructed using concentrations of 0, 15, 30, 60, 125, 250, and 500 µg/ L FB-28. The fluorometer was recalibrated if readings did not fall within + or - 10% of the constructed standard values. Due to the tendency of FB-28 to degrade over time, new standards were made every two weeks and the fluorometer was recalibrated using the new standards and a standards curve was reestablished. Because variations were found in the levels of fluorescence between two different OBs [with Tinopal CBS-X (Ciba Specialty Chemicals) displaying approximately 60 times more fluorescence at the same concentration than FB-28] all units are relative to the OB used for calibration. The fluorometer, during the fall of 2004, was converted to continuous flow mode (Hagedorn et al., 2003), and used to detect plumes of fluorescence along the shoreline in open beach waters at Anderson beach.

The determination of OB concentrations constituting a high (sewage positive) fluorometric reading was based on data gathered from samples of untreated sewage, ambient water samples containing virtually no fecal pollution (anthropogenic or otherwise), and attempts to correlate readings with high levels of sewage isolates, as indicated by ARA. Using the FB-28 calibration method, 24 untreated sewage samples, collected from 12 different sewage treatment plants in the coastal region of Virginia, yielded values ranging from 164.0 to 558.0 µg/ L with an average value of 259.5 µg/ L. Open water samples from areas with no potential sources of OBs from human wastewater, still yielded “background” levels of fluorescence ranging from 10.0 to 45.0 µg/ L based on the FB-28 calibration method. Accounting for the

expected dilution of sewage when mixed with open waters, and the results obtained from ARA, a value of 100.0 µg/ L was selected as the minimum detectable threshold for a water sample to be considered positive for the presence of optical brighteners.

Results

2004 ARA database validation and establishment of MDP

The statistical algorithms DA and LR were used on the 2004 library for both a four-way (bird, pets, sewage, and wildlife) and two-way (sewage and non-sewage) split. Based on both internal (library isolates) and external validations, for the four-way split, slightly better classification rates (less than 5%) were achieved using DA (data not shown), conversely for a two-way split, classification rates were slightly better (less than 5%) using LR (data not shown).

The ARCC for the four-way split using DA was 81.8% for isolates within the library, and 83.8% using the validation set (VS) of isolates (Table 342A). The AFM for the 2004 library and VS was 18.2% and 16.3%, respectively. Using the AFM values, a MDP of 17.3% [(18.2 + 16.3) / 2] was calculated and applied to all isolates collected from beach water samples. Thus, if the number of isolates from a water sample classified into a given category was represented by a percentage below 17.3%, that source was considered a negligible contributor. The ARCC for the two-way classification library (sewage/non-sewage) using LR (Table 342B) was 87.2% for library isolates and 84.2% for isolates from the VS. The AFM of the library (12.8%) and the VS (15.8%) were averaged to obtain a MDP of 14.3% for the 2004 ARA sewage/non-sewage library.

2004 ARA results

Results from site Anderson B, the regular VDH monitoring location, indicated that water samples contained 22.1% birds, 12.8% pets, 33.7% sewage, and 31.4% wildlife (Table 343). Sewage represented the major contributor, with pets falling below the MDP for the library (17.3%). Dominated by wildlife, Anderson SD results indicated that water samples contained 7.5% birds, 2.5% pets, 7.5% sewage, and 82.5% wildlife. Application of the two-way classification library indicated isolates collected from Anderson B were 32.6% sewage and 67.4% non-sewage in origin (Table 34). Anderson SD contained 7.5% sewage and 92.5% non-sewage. As was the case with the four-way split, both sewage and non-sewage were significant contributors in Anderson B, while only non-sewage was deemed a fecal contributor in water samples collected from the storm drain.

Application of the MDP to Hilton beach sites indicated that only pets, sewage, and wildlife were significant fecal contributors at both Hilton B and SD (Table 36); with sewage representing the largest percentage of total isolates collected from each site. Using the two-way classification library (Table 34), both sewage and non-sewage isolates were classified above the MDP at both Hilton B and SD; with sewage isolates representing the majority (57.3%) of those collected from the SD.

2004 Fluorometric and Enterococcus MPN results

Samples collected in the summer of 2004 would serve as both an application and determination of detectable OB levels, and values positive for sewage traces in water. Based on the ARA results of 2004, although enterococci of sewage origin were detected in the beach waters at both Anderson and Hilton beaches, concentrations of OBs were consistently detected

within established background levels (Table 35). High counts did not correspond to high OB levels at Anderson B (6/15 and 7/20) or Hilton B (7/6 and 7/14). The Anderson beach storm drain (SD), while only sampled twice, yielded a value within the background range (40.2 μg/ L) and one slightly above (47.7 μg/ L), but still below the determined threshold level of 100.0 μg/ L. Only from grab samples collected at Hilton SD were OB levels of consistently over 100.0 μg/ L. In addition, all Hilton SD samples contained enterococci densities over 1500 MPN/100 mL (Table 35).

2004 Detection of sewage leaks

With the identification of enterococci isolates of human origin in the waters of Anderson and Hilton beaches, as well as the Hilton storm drain outfall, municipal engineers were dispatched to locate the point of sewage entry into public beach waters. The high levels of OBs detected during the summer of 2004 in the storm drain at Hilton beach prompted engineers to send a remote camera and sample collector (for fluorometric analyses) up the storm drain. Using the fluorometry protocol developed at Virginia Tech, officials were able to follow the signal of OBs past a fork in the storm drain and to a nearby mobile home community. Upon visual inspection of the area, workers discovered a drainage ditch containing sewage waste from the trailer park leaking into the storm drain. Prior to the start of the 2005 swimming season, the leak was repaired.

Since no OBs were detected in the Anderson beach storm drain, municipal engineers converted the fluorometer to continuous flow mode, consistent with Hagedorn et al., 2003, to search for plumes of OBs along the edge of Anderson beach. A plume of high fluorescence was detected approximately 50 m northeast of the storm drain against a stretch of rock breakwalls

lining the shoreline, and just below a sewage pump station which once served a local apartment complex. Although the line had been closed and capped for approximately 20 years, closer inspection of the pipe system unearthed a substantial sewage leak. A metal cap which had once sealed off the pipes from the pump station had corroded from the brackish water, failing, and allowing for the discharge of substantial amounts of untreated sewage into the beach waters. Prior to the start of the 2005 swimming season, the pipe was replaced and recapped to prevent further sewage leakage.

2005 ARA database validation and establishment of MDP

The calculated ARCC for a four-way classification split using DA on the 2005 ARA library was 72.1% for isolates within the library, and 68.0% for the VS (Table 36A). Using the AFM of 27.9% for the library and 32.0% for the VS, a MDP of 30.0% was applied to water isolates run against the 2005 four-category ARA library. The ARCC for the two-way classification split on the 2005 ARA library using LR was 83.5% for library isolates, and 86.7% for VS isolates (Table 36B). The AFM for the library was 16.5% and 13.4% for the VS, which yielded a MDP of 15.0%. In addition, the 2004 VS isolates were run against the 2005 ARA library to assess the stability of resistance patterns over two summers. Using the four-category library, the RCCs for the 2004 VS isolates from the bird and sewage categories were the same as those from the 2005 VS, the pet category increased (from 70% to 80%), while the wildlife category decreased (from 80% to 65%). In the two-category split, both the sewage and non-sewage categories decreased in RCCs, by 5% and 10%, respectively.

2005 PFGE database validation and establishment of MDP

Perhaps due to the small size of the PFGE library (n = 420 isolates), the two-way classification split performed considerably better than the four-way split in both library and VS classifications (data not shown). And as was the case in the ARA libraries, LR outperformed DA for two-way classifications (data not shown). The ARCC for the 2005 PFGE library using LR was 79.8% for isolates within the library, and 80.0% for VS isolates (Table 37). An AFM for the library of 20.2% and 20.0% for the VS was averaged to calculate a MDP of 20.1% for isolates run against the 2005 PFGE library.

2005 ARA Results

Applying a MDP 30.0% for the 2005 four-category ARA library, only birds and wildlife were determined to be major fecal contributors at Anderson A (a new site for 2005) and B, with isolates from only pets and wildlife present in Anderson SD. Using the two-way classification, library isolates from Anderson A, B, and SD classified as 88.1%, 96.5%, and 95.8% non-sewage, respectively (Table 34). Enterococci isolates from the Anderson sites identified as sewage all fell below the 15.0% MDP for the 2005 two-way ARA library.

Based on the MDP for the four-way classification library (30.0%), only birds were identified as a significant fecal contributors at Hilton A and SD (Table 343); with only birds and wildlife detected at Hilton B. Using the two-way ARA classification library, none of the Hilton sites yielded a significant proportion of isolates of sewage origin in 2005 (Table 34). Non-sewage classification percentages of 91.8%, 86.9%, and 91.4% were found at Hilton A, B, and SD, respectively.

2005 PFGE Results

Application of PFGE to a randomly selected subset of approximately 30% of the total isolates analyzed from each site yielded results very similar to those using ARA (Table 37). At Anderson beach, site A isolates were identified as 87.5% non-sewage, while sites B and SD contained 89.6% and 83.9% non-sewage isolates, respectively. Non-sewage isolates at Hilton A, B, and SD were classified as 86.2%, 82.8%, and 85.2%, respectively, of the total isolates analyzed. Using a MDP of 20.1%, none of the Anderson or Hilton water samples were identified as containing a significant proportion of sewage isolates in 2005.

2005 Fluorometric Results

Eleven to twelve water samples from each of the six 2005 sampling locations were analyzed with the fluorometer to detect OBs (Tables 38 and 39). All but two fluorometric measurements taken at Anderson and Hilton beaches in 2005 fell in the range considered either negative or inconclusive for the presence of OBs, and thus human wastewater. A single sample from Hilton SD (6/27, Table 39), and a single sample from Anderson SD (6/27, Table 38), although not identified as containing sewage based on the results of ARA and PFGE, produced the high fluorometric readings of 237.0 µg/ L and 149.0 µg/ L, respectively. Nine of the water samples fell into the inconclusive range (46-99 µg/ L), with the majority (six) from Hilton SD. However, neither ARA nor PFGE concluded sewage was present in any of the water samples. For the two regular VDH monitoring sites, only one sample at Anderson B (5/16, Table 38) and three samples at Hilton B (5/16, 7/17, and 8/2, Table 39) exceeded the 104 MPN/100 mL standard, a substantial improvement over 2004.

Discussion

The goal of this study, beginning in 2004, was the use of MST to evaluate public beaches in Virginia identified as problematic based on persistently high *Enterococcus spp.* counts. The development and application of fluorometry in beach waters, was used with ARA to identify contributing fecal sources at two beaches. The detection and repair of sewage leaks at both beaches allowed for the 2005 assessment of remediation efforts through the continued use of fluorometry and ARA, as well as the addition of PFGE for cross-validation of ARA, and to use with open waters where grab-sample fluorometry was deemed less successful.

The results from Anderson beach in 2004 found that birds, wildlife, and sewage were contributing to the high fecal counts at the regular VDH monitoring site (Anderson B, Tables 33 and 34). Although sewage isolates were detected using both a four-way and two-way classification, measured OB levels, perhaps due to dilution in the river, fell within background values for all collected water samples (Table 35). Analysis of samples collected from the single storm drain present at Anderson also provided no high fluorometric readings and indicated only wildlife as a fecal contributor. Conclusions following completion of the 2004 swimming season, based on ARA results, indicated that sewage was likely entering the swimming area on the beach however the point of entry remained unidentified. The storm drain outfall, having tested negative for both OBs and isolates of human origin, was considered an unlikely candidate as the source of sewage transport to the beach. The other possibility was that the ARA results (regarding sewage-origin isolates) were incorrect in spite of the successful identification of VS isolates (Tables 32A and 32B).

Sewage also appeared to be a major component of the isolates collected from the beach water and storm drain outfall at Hilton beach, regardless of the category divisions or statistical

algorithms used (Tables 33 and 34). In the beach waters, OBs again proved to be too dilute for fluorometric detection however, not subject to the dilution of the river, a strong reading was detected in all three storm drain water samples (Table 35). Following the completion of the 2004 swimming season, results indicated that not only was sewage present within waters of Hilton beach, but the storm drain outfall was the probable mode of transport into the swimming area.

Lessons of the 2004 results focused on the difficulties of detecting OBs in open waters and stressed the usefulness of deploying multiple MST methods for increased confidence in results. Although the detection of OBs in storm drains was of a sufficient concentration to yield numbers well above background levels, in open, ambient waters, dilution proved to be too great to reasonably distinguish between background levels and potentially sewage-polluted waters in grab samples. Based on the average fluorometric value obtained from raw sewage samples (259.5 μg/ L), sewage effluent entering surface waters would require as little as a 5 to 1 water to effluent dilution to reduce the concentration of OBs to background levels. When applying grab-sample fluorometry to large bodies of water, such as a river or lake, it would be necessary (but feasible) for researchers to collect samples very close to potential point sources in efforts to avoid dilution problems.

Failure to detect OBs in the Anderson SD allowed for city engineers to ignore the storm drain as a potential sewage source and to focus along the shoreline where a plume of high fluorescent readings was discovered. This resulted in officials locating and repairing a failed cap on an old sewer line adjacent to the beach that was no longer being used. At Hilton beach, there was much less water available for dilution in the storm drain, particular during low flow conditions, and optical brighteners (presumably from sewage) were successfully detected (Table

35). Based on the fluorometry and ARA results, city officials investigated the storm drain in the fall of 2004 and discovered and repaired a leaking sewer drain from a mobile home park.

The usefulness of the fluorometer in storm drains has been demonstrated, and the potential still remains for fluorometer application in smaller bodies of water such as creeks and streams. In larger open waters, continuous flow (Hagedorn et al., 2003) or an intensive, systematic collection of shoreline grab samples may be the most efficient way of detecting sewage leaks or failing on-site systems. The possible use of an alternative detection method with greater specificity for OBs, such as high-performance liquid chromatography (Poiger et al., 1996, Hayashi et al., 2002), may increase sensitivity allowing for the improved detection in larger bodies of water where dilution is problematic, but at a more time-consuming scale and greater per sample cost.

Following completion of the off-season remediation efforts at both Anderson and Hilton beaches, the focus of monitoring and MST in the summer of 2005 shifted to assessing the results of these efforts. Enterococci isolated from Hilton and Anderson water samples during 2005 were not identified as sewage in origin, as determined by both ARA and PFGE libraries (Tables 33, 34, and 38). Adding another sampling site at each beach, closer to the repaired sewer at Anderson beach (Anderson A) and the storm drain outfall at Hilton beach (Hilton A), provided isolates and water samples to better assess the repair efforts. Only birds and wildlife were identified as fecal sources in the open waters of Anderson beach (sites A and B) in 2005 (Tables 33 and 34). The storm drain contained fecal isolates from wildlife, as well as pets, but no isolates of sewage origin. Birds were not considered a significant contributor in the storm drain, and likely entered the beach directly from the gulls and occasionally Canada geese that were prevalent on the jetties and rock breakwalls adjacent to the water. All OB values for Anderson

beach in 2005, except for one sample (possibly due to unidentified chemical interference) from the storm drain, were negative (Table 38).

In 2005 sewage was not found to be a fecal source in the open water sites at Hilton beach (A and B, Tables 34 and 35). However, as opposed to 2004 where birds were not identified as a potential source of fecal pollution, the 2005 ARA library identified birds as the major contributing source and cause of the limited number of elevated enterococci counts (Table 34). A possible explanation for this discrepancy is the construction of a fishing pier at Hilton beach between 2004 and 2005. This pier attracted large numbers of shore birds to the area (mainly gulls), that were not present at the beach during 2004. This increase in bird populations was the likely cause of the increase in bird isolates obtained both in the beach waters and from the storm drain. Isolates from the storm drain, classified as bird and wildlife, were no longer identified as sewage in origin, and only one of the eleven samples collected in 2005 contained elevated OBs (Table 39).

The successful remediation efforts at both Hilton and Anderson beaches from 2004 to 2005 are most easily seen as a decrease in the number of posted advisory days at each beach. Anderson beach advisories decreased from 26 to 2 days, and Hilton beach advisories decreased from 65 to 13 days in 2004 and 2005, respectively (VDH 2006). However, the persistence of storm drain outfalls emptying directly into swimming areas may continue to cause sporadic advisories, as outfalls are typically associated with increased fecal indicator densities resulting from the concentrated transport of surface runoff into beach waters (Haile et. al., 1999). However, animal fecal pollution is believed to be less likely to cause swimmer illnesses than human waste at the same enterococci densities (Sinton et al., 1998).

The application of multiple methods as recommended by Southern California Coastal Water Research Project (SCCWRP) method comparison (MC) study (Stewart et al., 2003) increased confidence in results for both the initial detection of human-origin pollution and the validation of remediation efforts. The agreement of both ARA and PFGE in identifying fecal sources coupled with confirmation from fluorometric storm drain results added a high degree of validity to the source identifications. The sponsor of our project, VDH, was familiar with both MC studies and had requested multiple lines of evidence (ARA and PFGE) plus an alternative tracer (OBs).

Both the SCCWRP (Stewart et al., 2003) and USGS MC studies (Stoeckel et al., 2004) stressed the usefulness of a VS of non-library isolates as an additional means (with RCCs from library isolates) to assess the effectiveness of a known-source library in a field setting. However, few studies to date have employed a VS (in any form) prior to declaration of method success or field deployment (Moore et al., 2005). Within this project, VS isolates were collected concurrently with water samples and from different fecal samples than those used to construct the known-source library. This procedure was followed to attempt to correct now-recognized deficiencies in the MC studies. The SCCRWP study created a VS of water matrices from the same fecal samples used to generate the known-source library. The VS set in the USGS study was collected from fecal samples nine months after those used for library construction were collected, lending additional support to the possibility of seasonal variations within fecal indicator strain populations. Although further research is desirable, seasonal variations have been implicated as a potential cause for classification inaccuracies in library-based MST methods (Pupo et al., 1995; Jenkins et al., 2003; Stoeckel et al.; 2004, and Lu et al., 2005), and our year-round collection of library isolates may have contributed to the successful identification of the

isolates in the VS. Methods employing antibiotic resistance patterns averaged RCCs of 25% to 40% for the various VS sources in the MC studies (Harwood et al., 2003; Stoeckel et al., 2004), and our results (Tables 33 and 34) show definite improvement over those RCCs.

A MDP was calculated and applied to all isolates based on the classification abilities of the library as determined by the AFM of both the library and VS. Within the SCCRWP MC study, the application of a MDP to the phenotypic methods substantially increased the RCC among the blind water samples (Harwood et al., 2003). The MDP allowed for an increased confidence in the ability to identify the dominant sources by eliminating sources identified at low percentages, thus accounting for many of the limitations of the known-source library. To date there is no recognized method for calculating a MDP. Taking the average of the AFM for both the library and the VS was selected as a means to give equal weight to the large collection of fecal isolates used to construct the library and the small number of known-source isolates in the VS used to test the library. Regardless of how the MDP was calculated, using some basis for ignoring low percent source allocations, a recommendation of both MC studies, provided field results that agreed perfectly with the sewage detection at both beaches.

The statistical analyses in the SCCWRP study suggested that uneven source categories might result in isolates being "forced," and incorrectly classified, into the category with the largest number of isolates (Ritter et al., 2003). The USGS study was originally designed so that every source category possessed the same number of isolates (Stoeckel et al., 2004). Although this approach still led to an imbalance once several categories were combined (e.g., human vs. non-human). In such a case, there is no proven procedure for eliminating isolates to retain balanced source-categories. The four categories in our 2004 library were not balanced and ranged from 168 for pets to 368 for sewage (Table 32). This should have pulled isolates into the

sewage category if concerns about uneven categories are warranted. However, the two-category 2004 library, which was heavily weighted towards non-sewage isolates (677 vs. 368 for sewage, Table 32), and should have conversely favored identifications as non-sewage. There was no evidence that these libraries pulled isolates either way, as demonstrated by the high RCCs of the validation set in 2004 (Table 32) and the close agreement on the water sample results (e.g., 38.4% and 35.5% sewage for the four-category and two-category libraries, respectively, for Hilton B, 2004, Tables 33 and 34). Other water sample results were similarly close. Adding isolates in 2005 resulted in a library where the four categories were more even (Table 36), while the unbalanced two-category library continued to provide water sample results that agreed very closely with the four-category results (e.g., 14.0% and 13.1% sewage for the four-category and two-category libraries, respectively, for Hilton B, 2005, Tables 33 and 34).

Assembling a known-source library using fecal samples collected on a frequent basis, over the entire two year study resulted in a library that was stable for at least that period. The relatively consistent RCC values for the 2004 ARA VS isolates, when run against the 2005 ARA library, is encouraging for MST researchers attempting to conduct multi-year studies within a watershed.

Conclusions

This is the first MST report where fecal sources were identified, traced to a point source, physically repaired, and validated by subsequent MST results. The major conclusions of this project were:

1. Considering the limitations of all current MST methods (Stewart et al., 2003, Stoeckel et al., 2004), the validation of MST results using multiple methods is recommended.

Consensus between methods in identified sources increases confidence in conclusions, and is particularly important for localities willing to allocate funds to support remediation efforts.

2. Fluorometric results were encouraging for future use in storm drains and near-shore applications, as well as use in continuous-flow mode for OB detection in larger bodies of water. Although, continued testing is warranted, as the possibility of interference still exists from other chemicals, such as aromatics (Hagedorn et al., 2005). In its current form fluorometry can be considered most effective as a supplemental means for verifying a human signature or as a way to locate the point of sewage entry into a water body remains.

3. Similar results were obtained for the library-based methods of this study using both LR and DA. While within these particular libraries superior results were obtained for a two-way classification split using LR and a four-way split applying DA, the differences were generally less than 5% with regards to the ARCCs of library and VS isolates. Additional testing is still needed into the best classification algorithms for a given MST method, or even library, as the algorithms published to date can differ across methods or even across libraries employing the same method.

4. The ARA library constructed and applied over the two summers of this project showed a good degree of stability, as indicated by the RCCs from the 2004 and 2005 VS, and is encouraging for future MST projects spanning multiple years.

Acknowledgements

This research was funded by the Virginia Department of Health, Division of Zoonotic and Environmental Epidemiology. We thank the Hampton Roads Sanitation District and the Virginia Department of Conservation and Recreation for assistance in the collection of sewage and fecal samples.

References

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

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

Carroll, S., M. Hargreaves, and A. Goonetilleke. 2005. Sourcing faecal pollution from onsite wastewater treatment systems in surface waters using antibiotic resistance analysis. J. Appl. Microbiol. 99:471-482.

Cheung, W.H., K.C. Chang, R.P. Hung, and J.W. Kleevens. 1990. Health effects of beach water pollution in Hong Kong. Epidem. Infect. 105:139-162.

Dickerson, J.W. Jr, J. B. Crozier, C. Hagedorn, and A. Hassall. 2007. Assessment of the 16S-23S rDNA intergenic spacer region in *Enterococcus spp.* for microbial source tracking. J. Environ. Qual. 36:1661-1669.

Geary, P., V. Shah, H. Dunstan, P. Coombes, and T. Rothkirch. 2006. Tracing faecal contributions from on-site wastewater systems. Wat. 33:48-51.

Gilpin, B.J., J.E. Gregor, and M.G. Savill. 2002. Identification of the source of faecal pollution in contaminated rivers. Wat. Sci. Technol. 46:9-15.

Glassmeyer, S.T., E.T. Furlong, D.W. Kolpin, J.D. Cahill, S.D., Zaugg, S.L. Werner, M.T. Meyer, and D.D. Kryak. 2005. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. Environ. Sci. Technol. 39:5157-5169.

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

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneneau. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522-5531.

Hagedorn, C., R.B. Reneau Jr., M. Saluta, A. Chapman. 2003. Impact of onsite wastewater systems on water quality in coastal regions. Virginia Coastal Resources Management Program Memorandum of Agreement 50312-01-13-PT. Virginia Dept. of Conservation and Recreation, Virginia Dep. of Health, Richmond.

Hagedorn, C., M. Saluta, A. Hassall, and J. Dickerson. 2005. Fluorometric detection of optical brighteners as an indicator of human sources of water pollution. Part I. Description and detection of optical brighteners. Crop and Soil Environmental News, November 2005. online http://www.ext.vt.edu/news/periodicals/cses/2005-11/part1.html.

Haile, R.W., J.S. Witte, M. Gold, R. Cressey, C. McGee, R.C. Millikan, A. Glasser, N. Harawa, C. Ervin, P. Harmon, J. Harper, J. Dermand, J. Alamillo, K. Barrett, M. Nides, and G-Y. Wang. 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. Epidem. 10:355-363.

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

Harwood, V.J., B. Wiggins, C. Hageodorn, R.D. Ellender, J.Gooch, J. Kern, M. Samadpour, A.C.H. Chapman, and B.J. Robinson. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J. Wat. Health. 1:153-166.

Hayashi, Y., S. Managaki, and H. Takada. 2002. Fluorescent whitening agents in Tokyo Bay and adjacent rivers: their application as anthropogenic molecular markers in coastal environments. Environ. Sci. Technol. 36:3556 -3563.

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.

Lu, Z., D. Lapen, A. Scott, A. Dang, and E. Topp. 2005. Identifying host sources of fecal pollution: diversity of *Escherichia coli* in confined dairy and swine production systems. Appl. Environ. Microbiol. 71:5992-5998.

McDonald, J.L., P.G. Hartel, L.C. Gentit, C.N. Belcher, K.W. Gates, K. Rodgers, J. A.Fisher, K.A. Smith, and K.A. Payne. 2006. Identifying sources of fecal contamination

inexpensively with targeted sampling and bacterial source tracking. J. Environ. Qual. 35:889-897.

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

NRDC. 2006. Natural resources defense council. Testing the waters: a guide to water quality at vacation beaches - July 2006 Available at: http://www.nrdc.org/water/oceans/ttw/sumvir.pdf (verified 15 Jan 2007).

Orosz-Coghlan, P.A., P.A. Rusin, M.M. Karpiscak, and C.P. Gerba, C.P. 2006. Microbial source tracking of *Escherichia coli* in constructed wetlands. Wat. Environ. Res. 78:227-232.

Poiger, T., J.A. Field, T.M. Field, and W. Giger. 1996. Occurrence of fluorescent whitening agents in sewage and river water determined by solid–phase extraction and high-performance liquid chromatography. Environ. Sci. Technol. 30:2220-2226.

Poiger, T., J.A. Field, T.M. Field, H. Siegrist, and W. Giger. 1998. Behavior of fluorescent whitening agents during sewage treatment. Wat. Res. 32:1939-1947.

Pupo, G.M. and B.J. Richardson. 1995. Biochemical genetics of a natural population of *Escherichia coli*: season changes in alleles and haplotypes. Microbiol. 141:1037-1044.

Rabinovici, S.J.M., R.L. Bernknopf, A.M. Wein. 2003. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Tech. 38:2737-2745.

Ritter, K.J., C. 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 using in library-based approaches to microbial source tracking. J. Wat. Health. 1:209-223.

Samadpour, M., M.C. Roberts, C. Kitts, W. Mulugeta, D. Alfi. 2005. The use of ribotyping and antibiotic resistance patterns for identification of host sources of *Escherichia coli* strains. L. Appl. Microbiol. 40:63-68.

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.

Simmons, G.M., D.F. Waye, S. Herbein, S. Myers, S. and E. Walker. 2002. Estimating nonpoint fecal coliform sources in Northern Virginia's Four Mile Run watershed, p. 143-167. In Younos, T. (ed.), Advances in Water Monitoring Research. Water Resources Publications, LLC.

Sinton L.W., R.K. Finlay, and D.J. Hannah. 1998. Distinguishing human from animal faecal contamination in water: a review. N. Z. J. Mar. Freshwater Res. 32:323-348.

Stewart, J.R., R.D.Ellender, J.A.Gooch, S. Jiang, S.P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons learned from a methods comparison study. J. Wat. Health. 1:225-231.

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

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.

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.

VDH. 2004. DRAFT–Virginia Department of Health Beach Monitoring Protocol July 2004 (unpublished).

VDH. 2006. Virginia Department of Health Beach Monitoring Website. Available at http://www.vdh.virginia.gov/epi/dzee/beachmonitoring/index.asp (verified 15 Jan 2007).

Table 32 - Four-category library in 2004 using discriminant analysis and two-category library using logistic regression.

A – ARA library 2004 using discriminant analysis

	Percentage (number) of isolates classified as:				
Source (number of isolates)	Bird	Dog	Sewage	Wildlife	
Bird (n = 248)	**80.2(199)†**	6.5 (16)	8.9 (22)	4.4 (11)	
Dog (n = 168)	6.6 (11)	**77.4 (130)**	10.2 (17)	6.0 (10)	ARCC = 81.8%
Sewage (n = 368)	8.7 (32)	7.1 (26)	**78.5 (289)**	5.7 (21)	AFM = 18.2%
Wildlife (n = 261)	2.3 (6)	3.8 (10)	2.7 (7)	**91.2 (238)**	
					MDP =17.3%#
Validation Bird (n = 10)	**100.0 (10)**	0.0 (0)	0.0 (0)	0.0 (0)	
Validation Dog (n = 10)	0.0 (0)	**70.0 (7)**	20.0 (2)	10.0 (1)	ARCC = 83.8%
Validation Sewage (n = 20)	10.0 (2)	5.0 (1)	**85.0 (17)**	0.0 (0)	AFM = 16.3%
Validation Wildlife (n = 10)	0.0 (0)	10.0 (1)	10.0 (1)	**80.0 (8)**	

B – ARA Library 2004 using logistic regression

	Percentage (number) of isolates classified as:			
Source (number of isolates)	Sewage	Non-Sewage		
Sewage (n = 368)	**84.2 (310)**	15.8 (58)	ARCC = 87.2%	
Non-Sewage (n = 677)	9.9 (67)	**90.1 (610)**	AFM = 12.8%	
				MDP = 14.3%
Validation Sewage (n = 20)	**75.0 (15)**	25.0 (5)	ARCC = 84.2%	
Validation Non-Sewage (n = 30)	6.7 (2)	**93.3 (28)**	AFM = 15.8%	

†Bold indicates rate of correct classification (RCC) for each source category
#Italics indicates calculated minimum detectable percentage (MDP)
Average rate of correct classification (ARCC); Average frequency of misclassification (AFM)

Table 33 - MST results for 2004 and 2005, four-category library with discriminant analysis.

		Percentage (number) of isolates classified as:				
	Location	Bird	Pets	Sewage	Wildlife	
2004	Anderson A	NA	NA	NA	NA	*MDP = 17.3%#*
	Anderson B*	**22.2 (19)†**	12.8 (11)	**33.7 (29)**	**31.4 (27)**	
	Anderson SD	7.5 (3)	2.5 (1)	7.5 (3)	**82.5 (33)**	
	Hilton A	NA	NA	NA	NA	
	Hilton B*	8.0 (11)	**20.3 (28)**	**38.4 (53)**	**33.3 (46)**	
	Hilton SD	4.0 (3)	**24.0 (18)**	**53.3 (40)**	**18.7 (14)**	
2005						*MDP = 30.0%*
	Anderson A	**32.7 (52)**	13.2 (21)	13.8 (22)	**40.3 (64)**	
	Anderson B*	**37.0 (64)**	17.3 (30)	8.1 (14)	**37.6 (65)**	
	Anderson SD	28.3 (68)	**34.2 (82)**	4.6 (11)	**32.9 (79)**	
	Hilton A	**45.7 (106)**	23.3 (54)	7.8 (18)	23.3 (54)	
	Hilton B*	**41.5 (98)**	12.7 (30)	14.0 (33)	**31.8 (75)**	
	Hilton SD	**53.3 (130)**	9.8 (24)	14.3 (35)	22.5 (55)	

* Virginia Department of Health weekly monitoring site.
†Bold indicates a source percentage above the minimum detectable percentage (MDP)
#Italics indicates minimum detectable percentage (MDP) applied to isolates of unknown origin

Table 34 - MST results for 2004 and 2005, two-category library with logistic regression.

		Percentage (number) of isolates classified as:		
	Location	Sewage	Non-Sewage	
2004	Anderson A	NA	NA	*MDP = 14.3%#*
	Anderson B*	**32.6 (28)†**	**67.4 (58)**	
	Anderson SD	7.5 (3)	**92.5 (37)**	
	Hilton A	NA	NA	
	Hilton B*	**35.5 (49)**	**64.5 (89)**	
	Hilton SD	**57.3 (43)**	**42.7 (32)**	
2005				*MDP = 15.0%*
	Anderson A	12.0 (19)	**88.1 (140)**	
	Anderson B*	3.5 (6)	**96.5 (167)**	
	Anderson SD	4.2 (10)	**95.8 (230)**	
	Hilton A	8.2 (19)	**91.8 (213)**	
	Hilton B*	13.1 (31)	**86.9 (205)**	
	Hilton SD	8.6 (21)	**91.4 (223)**	

* Virginia Department of Health weekly monitoring site.
†Bold indicates a source category percentage above the minimum detectable percentage (MDP)
#Italics indicates minimum detectable percentage (MDP) applied to isolates of unknown origin

Table 35 - Enterococci densities and optical brighteners from 2004 Hilton and Anderson beach water samples.

Location	Sample Collection Date	Enterococci Densities (mpn/100 mL)	Optical Brighteners (µg/ L)
Anderson B*	6/15/2004	1440	22.3
Anderson B*	6/24/2004	540	31.6
Anderson B*	7/20/2004	2040	42.1
Anderson B*	8/17/2004	210	42.6
Anderson SD	8/16/2004	570	40.2
Anderson SD	9/14/2004	375	47.7
Hilton B*	6/15/2004	530	31.7
Hilton B*	6/24/2004	235	36.9
Hilton B*	7/6/2004	2380	48.1
Hilton B*	7/14/2004	3140	42.2
Hilton B*	8/2/2004	390	41.2
Hilton B*	8/17/2004	400	33.4
Hilton SD	8/2/2004	2230	**109**†
Hilton SD	8/16/2004	3840	**209**
Hilton SD	6/24/2004	1520	**170**

* Virginia Department of Health weekly monitoring site.
†Bold indicates an elevated optical brightener concentration (above 100µg/ L)

Table 36 - Four-category library in 2005 using discriminant analysis and two-category library using logistic regression.

A – ARA library 2005 using DA

	Percentage (number) of isolates classified as:				
Source (number of isolates)	Bird	Dog	Sewage	Wildlife	
Bird (n = 630)	**69.8(440)†**	11.0 (69)	7.5 (47)	11.8 (74)	
Dog (n = 463)	19.9 (92)	**70.0 (324)**	6.7 (31)	3.5 (16)	ARCC = 72.1%
Sewage (n = 675)	7.1 (48)	10.2 (69)	**75.3 (508)**	7.4 (50)	AFM = 27.9%
Wildlife (n = 558)	15.6 (87)	1.3 (7)	9.9 (55)	**73.3 (409)**	
					MDP = 30.0%#
Validation Bird (n = 25)	**68.0 (17)**	20.0 (5)	4.0 (1)	8.0 (2)	
Validation Dog (n = 25)	24.0 (6)	**56.0 (14)**	12.0 (3)	8.0 (2)	ARCC = 68.0%
Validation Sewage (n = 25)	8.0 (2)	12.0 (3)	**76.0 (19)**	4.0 (1)	AFM = 32.0%
Validation Wildlife (n = 25)	4.0 (1)	20.0 (5)	4.0 (1)	**72.0 (18)**	

B – ARA Library 2005 using LR

	Percentage (number) of isolates classified as:			
Source (number of isolates)	Sewage	Non-Sewage		
Sewage (n = 675)	**74.1 (500)**	25.9 (175)	ARCC = 83.5%	
Non-Sewage (n = 1651)	7.1 (117)	**92.9 (1,534)**	AFM = 16.5%	
				MDP = 15.0%
Validation Sewage (n = 25)	**80.0 (20)**	20.0 (5)	ARCC = 86.7%	
Validation Non-Sewage (n = 75)	6.7 (5)	**93.3 (70)**	AFM = 13.4%	

†Bold indicates rate of correct classification (RCC) for each source category
#Italics indicates calculated minimum detectable percentage (MDP)
Average rate of correct classification (ARCC); Average frequency of misclassification (AFM)

Table 37 - PFGE 2005 library with logistic regression and results at beach locations.

A – PFGE Library 2005 using LR

	Percentage (number) of isolates classified as:			
Source (number of isolates)	Sewage	Non-Sewage		
Sewage (n = 153)	**71.9 (110)†**	28.1 (43)	ACRR = 79.8%	
Non-Sewage (n = 267)	12.4 (33)	**87.6 (234)**	AFM = 20.2%	
				MDP = 20.1%#
Validation Sewage (n = 10)	**80.0 (8)**	20.0 (2)	ARCC = 80.0%	
Validation Non-Sewage (n = 30)	20.0 (6)	**80.0 (24)**	AFM = 20.0%	

B – PFGE Results 2005 using LR

	Percentage (and number) classified as:		
Location	Sewage	Non-Sewage	
Anderson A	12.5 (6)	**87.5 (42)@**	
Anderson B*	10.4 (5)	**89.6 (43)**	
Anderson SD	16.1(9)	**83.9 (47)**	
			MDP = 20.1%
Hilton A	13.8 (9)	**86.2 (56)**	
Hilton B*	17.2 (11)	**82.8 (53)**	
Hilton SD	14.8 (12)	**85.2 (69)**	

†Bold indicates rate of correct classification (RCC) for each source category
@Bold indicates a source category percentage above the minimum detectable percentage (MDP)
#Italics indicates calculated minimum detectable percentage (MDP)
Average rate of correct classification (ARCC); Average frequency of misclassification (AFM)

Table 38 - Enterococci densities and optical brighteners from 2005 Anderson beach water samples.

Location	Sample Collection Date	Enterococci Densities (mpn/100 mL)	Optical Brighteners (µg/ L)
Anderson A	3/31/2005	82	36
Anderson A	4/14/2005	5	27.9
Anderson A	5/16/2005	48	27.6
Anderson A	5/28/2005	<10	21
Anderson A	6/14/2005	<10	23
Anderson A	6/27/2005	34	20.9
Anderson A	7/17/2005	31	24.5
Anderson A	8/2/2005	20	27
Anderson A	8/15/2005	<10	27.7
Anderson A	8/24/2005	75	29.5
Anderson A	9/11/2005	42	27.5
Anderson A	10/3/2005	<10	20.9
Anderson B*	3/31/2005	20	33.3
Anderson B*	4/14/2005	62	28.9
Anderson B*	5/16/2005	114	30.6
Anderson B*	5/28/2005	13	18.1
Anderson B*	6/14/2005	<10	25.8
Anderson B*	6/27/2005	74	22.4
Anderson B*	7/17/2005	<10	22.1
Anderson B*	8/2/2005	10	31.4
Anderson B*	8/15/2005	10	28.2
Anderson B*	8/24/2005	75	24.2
Anderson B*	9/11/2005	20	27.3
Anderson B*	10/3/2005	<10	19.6
Anderson SD	3/31/2005	960	50.5
Anderson SD	4/14/2005	<10	16.2
Anderson SD	5/16/2005	900	40.3
Anderson SD	5/28/2005	157	33.3
Anderson SD	6/14/2005	44	21.2
Anderson SD	6/27/2005	396	**149**†
Anderson SD	7/17/2005	97	33.8
Anderson SD	8/2/2005	640	28
Anderson SD	8/15/2005	20	31.7
Anderson SD	8/24/2005	30	23.8
Anderson SD	9/11/2005	200	26.4
Anderson SD	10/3/2005	NA	NA

* Virginia Department of Health weekly monitoring site.

†Bold indicates an elevated optical brightener concentration (above 100.0µg/ L).

Table 39 - Enterococci densities and optical brighteners from 2005 Hilton beach water samples.

Location	Sample Collection Date	Enterococci Densities (mpn/100 mL)	Optical Brighteners (µg/ L)
Hilton A	3/31/2005	15	36.4
Hilton A	4/14/2005	30	27.1
Hilton A	5/16/2005	1810	41.5
Hilton A	5/28/2005	10	27.3
Hilton A	6/14/2005	75	33.9
Hilton A	6/27/2005	58	29.1
Hilton A	7/17/2005	216	33.8
Hilton A	8/2/2005	231	34.9
Hilton A	8/15/2005	10	40
Hilton A	8/24/2005	40	29
Hilton A	9/11/2005	<10	28
Hilton A	10/3/2005	10	26.2
Hilton B*	3/31/2005	15	38.9
Hilton B*	4/14/2005	15	27.5
Hilton B*	5/16/2005	1460	48.5
Hilton B*	5/28/2005	75	24.5
Hilton B*	6/14/2005	90	37.2
Hilton B*	6/27/2005	81	28.4
Hilton B*	7/17/2005	158	35
Hilton B*	8/2/2005	134	34
Hilton B*	8/15/2005	10	38.8
Hilton B*	8/24/2005	56	31.5
Hilton B*	9/11/2005	<10	64.4
Hilton B*	10/3/2005	<10	26
Hilton SD	3/31/2005	272	47.4
Hilton SD	4/14/2005	760	43.5
Hilton SD	5/16/2005	3435	46
Hilton SD	5/28/2005	680	61.9
Hilton SD	6/14/2005	269	40.4
Hilton SD	6/27/2005	574	**237**†
Hilton SD	7/17/2005	2595	72
Hilton SD	8/2/2005	1081	58.9
Hilton SD	8/15/2005	146	77.1
Hilton SD	8/24/2005	NA	NA
Hilton SD	9/11/2005	269	38.1
Hilton SD	10/3/2005	173	30.2

* Virginia Department of Health weekly monitoring site.
†Bold indicates an elevated optical brightener concentration (above 100.0µg/ L).

V. Assessment of the 16S-23S rDNA Intergenic Spacer Region in *Enterococcus spp.* for Microbial Source Tracking*

* From Journal of Environmental Quality, with permission. J. Environ. Qual. Dickerson, J.W. Jr., J.B. Crozier, C. Hagedorn, and A. Hassall. 2007. 36:1661-1669.

Abstract

This study developed and tested a new library-based microbial source tracking (MST) approach intended for initial application in the marine and coastal waters of Virginia. Host-origin isolates of *Enterococcus spp.* were collected from beaches and surrounding areas in the tidewater region of Virginia and used to construct a library based on the pattern of DNA band lengths produced by the amplification of the 16S-23S rDNA intergenic spacer (IGS) region, and subsequent digestion using the restriction endonuclease MboI. Initial results from small host-origin libraries (64 and 200 total isolates) with discriminant analysis (DA) and logistic regression (LR) yielded high average rates of correct classification (ARCC) for a 4-source classification split (birds, dogs, sewage, and wildlife), with ARCCs ranging from 83% to 100%. However, the poor results obtained when classification was attempted on a non-library validation set (VS, ARCCs of 47% and 48%, respectively, using DA and LR) demonstrated that a library of 200 isolates was insufficient to adequately represent the diversity of the enterococci in the sampled region. An increase in the library size to 1,029 total isolates was accompanied by a reduction in the ARCC of the library to 42.7% with DA and 45.7% with LR, plus similarly poor results obtained from the VS. The low correct classification rates generated by the larger known-source library were unsuitable for field application and the use of the rDNA IGS region in *Enterococcus spp.* with MboI was deemed unsuccessful as a potential MST method. The success of many reported MST methods have been based on results obtained using a small host-origin library

without the external verification of a VS. Our results indicate that such an approach can be very misleading, and that larger libraries and VSs are essential for the confirmation of preliminary results obtained with small libraries.

Introduction

The advent of MST over the last decade has provided watershed managers with a means to discriminate between fecal sources polluting surface waters. The underlying premise of MST is that certain enteric bacterial strains are uniquely adapted to, and thus reside exclusively in the gastrointestinal tract of a single, or group of closely-related host organisms. Differences in the phenotypes or genotypes of these strains can be used to determine the relative contributions of animal sources to the fecal pollution in a waterbody. Two major classes of MST methods are currently being developed and utilized in surface waters across the world (Sinton et al., 1998; Scott et al., 2002; Simpson et al., 2002; Pond et al., 2004; Blanch et al., 2006).

The earliest and most commonly applied genotypic and phenotypic methods involve the construction of a host-origin database, or library, of isolates from known fecal sources providing a collection of possible 'fingerprint' patterns allowing for a direct comparison with the fingerprints of isolates of unknown origin. The most commonly used phenotypic methods have employed differences in antibiotic resistance patterns (Wiggins, 1996; Hagedorn et al., 1999; Harwood et al., 2000; Harwood et al., 2003; Graves et al., 2002; Whitlock et al., 2002) or the ability to utilize varying nutrient-sources (Hagedorn et al., 2003; Harwood et al., 2003; Ahmed et al., 2005) of indicator organisms to determine fecal origins. An even greater variety of genotypic methods have been reported in the MST literature including: ribotyping (Parveen et al., 1999; Carson et al., 2001; Hartel et al., 2002; Carson et al., 2003; Scott et al., 2003), pulsed-field gel

electrophoresis (PFGE) (Simmons et al., 2002; Samadpour et al., 2005), microarrays (Indest et al., 2005), and repetitive sequence PCR (rep-PCR) (Dombek et al., 2000; Carson et al., 2003; Seurinck et al., 2003; Johnson et al., 2004).

Soon after the development of library-based MST methods, researchers began looking for organisms, or sequences within the genome of organisms, that were consistently exclusive to pollution from a particular fecal source. Known as library-independent methods, in addition to using source specific markers found in some recognized indicator organisms (Scott et al., 2005; USEPA, 2005), researchers have frequently expanded the search into non-indicator fecal organisms such as: *Bifidobacterium spp.* (Rhodes and Kator, 1999), *Bacteroides* (Bernhard and Field, 2000; Field et al., 2003; Simpson et al., 2003), F-specific DNA and RNA coliphages (Hsu et al., 1995; Cole et al., 2003; Long et al., 2005; Sundram et al., 2006), methanogens (Ufnar et al., 2006), and human or livestock specific enteric viruses such as enterovirus (Noble and Fuhrman, 2001; Fong et al., 2005), adenovirus (Jiang et al., 2001; Maluquer de Motes et al., 2004Fong et al., 2005), and teschovirus (Jimenez-Clavero et al., 2003).

A major drawback of library-based methods to date has been observable geographical limitations (Hartel et al., 2002; Scott et al., 2003), although similar spatial restrictions have been seen in at least one library-independent method as well (Hamilton et al., 2006). An additional disadvantage of the current status of library-independent methods is that only a limited number of methods are capable of consistently quantifying the contributing proportions of fecal inputs in polluted waters (Noble et al., 2003; Field et al., 2003); as most of these methods presently serve primarily as a presence/absence test of human, and a limited number of non-human, sources. Although human fecal contamination presents the greatest risk to public health (Sinton et al.,

1993), additional information on other potential sources is often useful in attempts to lower indicator bacteria concentrations to within USEPA levels of acceptable risk (USEPA, 1986).

A few recent studies have reported success using *E. coli* 16S-23S ribosomal DNA (rDNA) intergenic spacer (IGS) regions to discriminate between humans, cows, and chickens (Buchan et al., 2001), and to a lesser extent *E. coli* from sewage, horses, cows, gulls and dogs (Seurinck et al., 2003). The absence of selection pressures in the IGS region, as opposed to the highly conserved nature of the bordering rDNA, has proven useful as a target site for the molecular subtyping of a variety of pathogenic bacteria (Guertler and Stanisich, 1996; Stubbs et al., 1999; Graham et al., 1996; Chun et al., 1999; Riffard et al., 1998), providing a simpler (in both equipment needed and level of training required) and more cost-effective assay than more traditional MST methods such as PFGE (Bedendo and Pignatari, 2000) or ribotyping (Carson et al., 2003). Frequently present in multiple copies, the arrangement of the rDNA operon is almost always 16S-IGS-23S-IGS-5S in bacteria. The amplification of 16S-23S rDNA IGS regions within a bacterial genus such as *Enterococcus spp*. can be performed using primers that recognize the highly conserved sequences found in the flanking regions of 16S and 23S rDNA. The *Enterococcus spp*. genome contains as many as six rDNA operons (Sechi and Daneo-Moore, 1993) allowing for the amplification and digestion of multiple amplicons with the potential in MST to increase the diversity of banding patterns produced among strains in the search for banding or fingerprint patterns unique to strains from a specific host organism.

The objective of this study was to develop a method of detecting and quantifying source-specific enterococci from birds (ducks, geese, and gulls), dogs, sewage (presumed human), and wildlife (deer and raccoons). The completion of successful laboratory testing would allow for the application of a new library-based MST method in the coastal regions of Virginia.

Enterococci were selected as the target organisms due to their abundance in the fecal matter of warm-blooded animals (Devriese et al., 1987) and usage as fecal indicators in weekly monitoring procedures in the marine and coastal waters of Virginia (VDH, 2004). As library-size requirements likely vary between methods and watersheds, the use of a non-library collection of known-source isolates, or validation set (VS) was employed for external validation in order to better assess the number of isolates required to represent enterococcal diversity in the target watershed. This study addresses the method-specific sampling and performance criteria described by Stoeckel and Harwood, 2007.

Materials and Methods

Collection of Fecal Samples

Fecal samples from known animal sources were collected at public beaches, dog parks, and nature parks within the Tidewater region of southeastern Virginia from February to September of 2005, as described in Dickerson et al., 2007. Sewage influent samples from the twelve treatment plants within the region were provided on four separate occasions during this time period by the Hampton Roads Sanitation District. None of the wastewater treatment plants contained combined sewers, so sewage samples should have contained composite samples of enterococci of almost exclusively human origin. Both fresh and dried fecal samples were collected from animals in each category (birds, dogs, and wildlife), except sewage, based on opportunity at the time samples were collected. Gulls identified were dominantly Ring-billed (*Larus delawarensis*) and Herring (*Larus argentatus*) as well as an occasional Laughing gull (*Larus atricllia*). The geese and ducks from which scat was obtained were identified as: Snow Goose (*Chen caerulescens*), Canada goose (*Branta canadensis*), and Mallard (*Anas*

platyrhynchos). Dog (*Canis familiaris*) fecal samples were collected from local beaches and from several dog parks in the area. Wildlife (deer [*Odocoileus virginianus*] and raccoon [*Procyon lotor*]) scat was collected in Chickahominy Wildlife Management Area, Waller Mill Park, Newport News Park, and Pocahontas and York River State Parks in Eastern Virginia.

Isolation of Enterococci

A portion of each fecal or untreated sewage sample was diluted into tubes of sterile distilled deionized (DDI) water and spread on m-Enterococcus agar (Baltimore Biologics Laboratory). After 48-hour incubation at 35ºC (APHA, 1998), no more than 4 randomly selected red to burgundy colonies from each non-sewage source, and not more than 12 from each sewage source were picked from each plate using sterile toothpicks. All isolates were inoculated into Enterococcosel Broth (Baltimore Biologics Laboratory) in a 96-well microtiter plate for confirmation as enterococci (black color after incubation). All confirmed enterococcal isolates were regrown on TSA agar for use in PCR. The numbers of fecal samples collected and isolates selected from each sample, and the period over which fecal samples were obtained, was similar to other reports where PCR was used for MST (Carson et al., 2003; Hamilton et al., 2006;).

Polymerase Chain Reaction

PCR was used to amplify *Enterococcus* IGS regions located between the 16S and 23S rDNA regions. Based on sequences in the GenBank database, primers were designed manually that would anneal to highly conserved downstream 16S rDNA and upstream 23S rDNA sequences in virtually all enterococci such that entire IGS regions could be amplified from each isolate (primers produced by Invitrogen Corporation). Approximately 1.0 µL of a pure culture of

cells was diluted into 300.0 µL of sterile DDI water to serve as a template for PCR. PCR was carried out using PuReTaq Ready-To-Go PCR beads (Amersham Biosciences), in 22.0 µL of sterile DDI water, 1.0 µL of 16S primer (5'- GCCTAAGGTGGGATAGATGA-3', novel to this study), 1.0 µL of 23S primer (5'-CCCGTCCTTCATCGGCTCCTA–3', novel to this study), and 1.0 µL of diluted cell culture. Primers were used at a final dilution of approximately 0.2 µM. The PCR was initiated by incubating the reaction mixture at 95ºC for 6 minutes to lyse the cells, followed by 35 one minute cycles of 94ºC, 57ºC, and 72ºC. The final elongation step was completed at 72ºC for 7 minutes, followed by a 4ºC hold of all reaction mixtures. All PCR experiments contained a positive control (*E. faecalis*) to assess method reproducibility and stability (numbers of bands and length of each).

Restriction Digests

Restriction enzyme digests consisted of 10ul of restriction digest mix (34µl of 10X Buffer C, 17µl BSA, 17µl spermidine (100mM), 93.5 sterile DDI, 8.5µl MboI restriction enzyme (Promega, 5'-^GATC-3', 3'-CTAG^-5') combined with 10µl of PCR product into a centrifuge tube, centrifuged briefly and incubated at 37ºC for 3.5 hours.

Gel Electrophoresis

Restriction enzyme digests were mixed with loading dye and loaded on a 3% horizontal agarose gel (Agarose Low Melting, Fisher Scientific), with several 100 bp ladders and the positive control, in order to detect polymorphisms among isolates. All gels were run in 1X TAE (10mM Tris, 5mM acetate, 0.1 mM EDTA, pH 7.4 (Promega)) for 80 minutes at 100V with standard gels (10 by 15 cm). Gels were stained for 3 hours in a solution of 2X SYBR Green I

(Cambrex Bio Science Rockland, Inc), and photographed on a UV mini-transilluminator with a Polaroid DS34 camera. All photographs were digitally scanned in Gel-Pro 3.1 using a HP Scanjet 6300C.

Statistical Analysis of PCR Profiles for Source Prediction

Each digest, when visualized, exhibited between 4 and 14 total bands. Band lengths were quantified using Gel-Pro Software and converted to binary data based on 100 base-pair length categories ranging from <100 to >1000 bp in length. Analyses were conducted using both discriminant analysis (DA) and logistic regression (LR) in SAS-JMP statistical software (version 5.0.1, SAS Institute Inc.); comparing isolates from within the library (all isolates were left in the library) and the VS against the model constructed using patterns from isolates in the host-origin library. The classification table generated in SAS-JMP was used to calculate the rate of correct classification (RCC) of each group of isolates, with the ARCC for the library or VS of isolates calculated by dividing the sum of the number of isolates correctly classified across all four categories by the total number of isolates classified, similar to the estimate of correct classification (ECC) used in Albert el al., 2003.

Creation of Known-Source Library (Host-Origin Database)

An initial library of 64 isolates was created using 16 isolates from each source category, obtained from 2 sewage, 4 dog, 5 bird, and 5 wildlife fecal samples collected in February 2005. The library was increased in size to 200 isolates (50 per category) using isolates collected (May 2005) from 8 sewage, 14 dog, 15 bird, and 15 wildlife fecal samples. The final library contained a total of 1029 isolates (201 birds, 353 sewage, 266 dogs, 209 wildlife), from 52 bird, 42 sewage,

70 dog, and 54 wildlife fecal/sewage samples, with additions made from June-September 2005 collections. Clonal isolates within the library were not removed for initial tests. In addition to isolates collected for library construction, 100 additional isolates (12 birds, 48 sewage, and 20 each of dogs and wildlife), collected in May 2005 from 4 birds, 6 sewage, 6 dogs, and 7 wildlife fecal/sewage samples, were held out of the known-source library for use as a VS. No isolates obtained for the VS were collected from the same fecal/sewage samples used for library construction.

Following completion of the 1029-isolate library, a one-time random sampling of 201 isolates was selected (with the JMP software) from each of the three largest source categories (dogs, sewage, and wildlife) to generate a balanced known-source library of 804 isolates (201 isolates per source category). An additional library (323 isolates) and VS (62 isolates) was later generated from the larger 1029-isolate library and 100-isolate VS containing only unique banding patterns through the detection and removal all clonal isolates within and across source categories.

Results

PCR and subsequent restriction digest product was obtained for almost all isolates tested, yielding bands ranging from approximately 32 to 1055 bp in size. An initial test library was constructed using 64 total isolates (16 per source category) to select for the restriction endonuclease providing the most discriminating source-specific fingerprint patterns. This library produced an average rate of correct classification (ARCC) of 100% using MboI for the desired four-source split, with both DA (Table 41) and LR (Table 43), suggesting there was potential for this approach. A VS was not constructed for this early stage of library development.

Based on the success of the 64-isolate library (Tables 40 and 43), additional fecal samples were collected and enterococci isolates fingerprinted, increasing the library size to 200 isolates (50 per source category). With the increased library size, the ARCC decreased from 100% to 83% with DA (Table 41), with the highest RCC of 88% in both the dogs and wildlife categories. Using LR (Table 44) a higher ARCC of 88% was achieved, with both dogs and sewage yielding RCCs of 94%. Although these classification rates would be considered acceptable and promising for most library-based MST methods based on previous studies (Harwood et al., 2000, Whitlock et al., 2002, Choi et al., 2003, VanOmmeren et al., 2006), the use of a VS of isolates was implemented to provide an additional means of assessing library capabilities. The VS isolates were correctly classified at considerably lower rates than those composing the library with DA only identifying only 47 out of 100 isolates correctly (ARCC = 47%, Table 41) and LR only placing only 48 of the 100 isolates into the correct source category (ARCC = 48%, Table 44). The ineffectiveness of the known-source library in classifying non-library isolates suggested the library was of insufficient size and not representative of the strain diversity of enterococci in the coastal region of Virginia.

A final increase in the size of the library brought the total number of isolates to 1029 (201 birds, 266 dogs, 353 sewage, and 209 wildlife isolates). With DA (Table 42), the final ARCC for the classification of isolates composing the library (internal ARCC) was 42.7%, well below the classification rates of any libraries applied in field studies (Hagedorn et al., 1999; Harwood et al., 2000; Graves et al. 2002; Choi et al., 2003; Carroll et al., 2005). No source category in the 1029-isolate library produced a RCC greater than 70% with DA. The internal ARCC using LR (Table 45) was 45.7%, only slightly higher than the results generated by DA, with the wildlife category producing the highest RCC of 64.1%. Only 47 of the 100 VS isolates were correctly

classified by the 1029-isolate library using DA (ARCC = 47%, Table 42). And only 53 of 100 VS isolates (ARCC = 53%, Table 45) were correctly classified into respective host categories with LR.

Only a slight overall improvement was seen in classification rates when library source categories were balanced (201 isolates per category) by randomly selecting a subset from each of the three largest source categories (Tables 46 and 47). With DA (Table 40), the internal ARCC for the library increased slightly, from 42.7% to 43.3%, by balancing the source categories. However, the VS of isolates showed a slight decrease, as the ARCC declined from 45.7% to 44%. Library isolates from individual source categories showed minor changes, with only the smallest original category (birds) displaying an RCC increase (+5.5%). In the VS, changes within individual categories were also minor, with the only major change (>5.0 percentage points) being a 25 percentage point decrease in the RCC of bird isolates.

When a classification model was generated using LR for the balanced source-categories (Table 47), the ARCC for isolates within the library displayed a slight decrease, dropping from 45.7% to 45.1%. The VS isolates showed a larger decrease, as ten fewer isolates were correctly classified, decreasing the ARCC for the VS from 53% to 43%. Within the individual categories birds and wildlife, the smallest categories in the original unbalanced library yielded an increase in the RCC of 24.8 and 4.6 percentage points respectively once the source categories were equal in size. Conversely, the RCC of the dogs and sewage categories both decreased by 9.2 and 14.1 percentage points, respectively. The decrease in the ARCC for the VS of isolates was the largest average change seen. VS isolates from the two largest categories in the original library decreased by 10.4 (sewage) and 50 (dogs) percentage points.

The removal of clones from the 1029-isolate library reduced the library size to 323 unique isolates, a reduction of 68.6% (Tables 48 and 49). Although the ARCCs increased in some source categories and declined in others with both statistical algorithims for the library with clones removed (compare Tables 42 and 48 for DA, Tables 45 and 49 for LR), there was little effect on the overall ARCCs and the values remained considerably lower than those reported for other non-clonal libraries of comparable size (Dombek et al. 2000; Guan et al., 2002; Seurnick et al., 2003; Lasalde et al., 2005; Duran et al., 2006; Vantarakis et al., 2006). For example, the ARCCs (with DA) for the clonal library (Table 42) and the non-clonal library (Table 48) were 43.9% and 53.6%, respectively. The VS ARCCs (with DA) for the clonal library (Table 42) and the non-clonal library (Table 48) were 49.9% and 52.5%, respectively.

Discussion

A successful MST method requires the testing and/or fingerprinting of large numbers of fecal isolates within a geographic region to assess and account for host-strain diversity (Dickerson et al., 2007). The results of this study indicate that the validation of an MST method using isolates from a small number of sources tends to falsely inflate method effectiveness, which can result in improper and premature applications in field trials. The rDNA IGS method, both simpler to perform and less expensive than most other molecular MST methods, worked very well during the initial stages of testing (Tables 40 and 43) producing correct classification rates of 100% with both DA and LR. This initial success prompted the continued addition of known-source isolates working towards the construction of a host-origin library of adequate size to undergo field evaluations in the coastal region of Virginia. For the library of 200 isolates (Tables 41 and 44), the ARCC remained reasonably high and provided sufficient discrimination

between source categories using both statistical methods. However, the inability of the library to correctly classify isolates from the VS at a level comparable to those within the library indicated that the diversity of fecal isolates was inadequately represented. As the library continued to increase in size (to 1029 isolates), the method began to fail (Tables 42 and 45), falling to levels unsuitable for source discriminations and was thus deemed unsuccessful.

For this study, classification models were generated by both DA and LR, two parametric methods, for the identification of isolates which were both part of, and not part of the known-source library. While DA has been widely applied in library-based source tracking methods (Wiggins, 1996; Hagedorn et al., 1999; Harwood et al., 2000; Graves et al., 2002), LR has, to date, remained unutilized in the field of MST. The most frequently implemented classification method for biomedical applications (Dreiseitl and Ohno-Machado; 2002); LR requires fewer assumptions than DA such as a normal distribution and equal variances within groups among independent variables. The use of LR has been shown to more effectively classify unknowns, as compared to DA, under conditions of nonnormality, such as those using binary explanatory variables (Press and Wilson, 1978).

While methods of internal classification are commonly used in library-based MST, the practicality of a method lies exclusively in the ability to correctly classify isolates of unknown origin, such as those from water samples. One of the major recommendations to emerge from the Southern California Coastal Water Research Project (SCCRWP) and United States Geological Survey (USGS) sponsored method comparison (MC) studies (Stewart et al., 2003; Stoeckel et al., 2004) was the usefulness of the VS of non-library isolates, as opposed to internal validations, to assess the effectiveness of a known-source library. Commonplace in the medical (Terrin et al., 2003) or statistical (Press and Wilson, 1978) fields, the external verification of

library effectiveness has been used in very few MST studies (Moore et al., 2005). However modifications of the VSs used in each of the MC studies are necessary to correct potential flaws that were present in each. In the SCCRWP study, isolates were obtained from the same set of fecal samples used to construct the known-source library. The underperformance of most MST methods in this study, even with the use of a seemingly favorable VS, may have been the result of a greater level of fecal diversity than previously expected in a single fecal sample, or inadequate reproducibility of many MST methods (Stewart et al., 2003). The USGS sponsored MC study (Stoeckel et al., 2004) used a VS of isolates collected nine months after those used for library construction. The overall inadequate performance of the methods involved resurrected previous concerns of temporal instability of strains within host organisms (Jenkins 2003). Therefore in this study, once the library size was increased to 200 isolates, comparable to numbers frequently reported in several publications (Dombek et al. 2000; Guan et al., 2002; Seurnick et al., 2003; Lasalde et al., 2005; Duran et al., 2006; Vantarakis et al., 2006), a VS of isolates (collected simultaneously, but from different fecal samples) were fingerprinted to serve as an additional means of predicting the classification ability of the known-source library. In library-based MST methods, small libraries typically produce high ARCCs solely due to the random placement of isolates into defined categories (Whitlock et al., 2002). The poor classification rates and marked differences between library isolates and those not a part of the library is indicative of a library that is too small, and does not contain enough isolates to represent strain diversity in a watershed. Once the library-size was increased to over 1000 isolates classification rates decreased sharply, yielding correct classification rates below 50% and unsuitable for field applications.

No attempt was made during method development to distinguish between species of *Enterococcus*. Although the potential existed for greater source discrimination if analyses were limited to one or a few *Enterococcus* species, the requirements of speciating isolates would serve to increase both time-constraints and method costs, making any finding less desirable for application in the field, but nevertheless may have resulted in a workable method. In addition, as the population and proportions of specific enterococci species vary between organisms (Lauková et al., 1997; Wheeler et al., 2002), limiting analyses to specific species could unfairly bias relative fecal contributions or potentially eliminate the ability to detect animal sources not carrying the species selected.

The differences seen in the ARCC between libraries containing balanced and unbalanced source categories were small regardless of whether DA or LR was used to generate classification models (Tables 42, 45-47). The ARCC for the VS decreased using both algorithms, possibly due to a loss in overall library representativeness from a decrease in the total number of isolates. The major difference resulting from a balanced library could only be seen within individual source categories. The effects of balancing a library on the RCC for a category were minor for both the library and VS isolates when using DA. However, using LR, larger changes were observed in both library and VS isolates when the library was balanced; as an inverse relationship was seen between the number of isolates lost from the original unbalanced library and the change in RCC for a given source category. Thus decreases in the RCC were seen for the two categories (dogs and sewage) losing a significant portion of the total isolates, while the two smallest categories, forfeiting only seven (wildlife) or zero (birds) isolates, increased (or failed to decrease) in the category RCC. For the libraries generated in this study, there was also little to be gained by

removing clones. This would indicate that an inadequate clonal library cannot be substantially improved by removing clones; the non-clonal library will still be inadequate.

Conclusions of this study stress the dangers in using a small number of isolates/fecal sources in assessing the effectiveness of both library-based and, possibly, library-independent MST methods. Levels of strain diversity are undoubtedly method-dependent, however host strain variability is probably greater than previously assumed, especially in a non-conserved DNA region such as IGS, and small libraries are likely not capable of reporting results with a high level of confidence. The SCCWRP MC study concluded that libraries of ~300 were generally not successful at identifying fecal pollution in blind water samples, even when libraries were generated from the same fecal material used to construct the blinds (Griffith et al., 2003; Harwood et al., 2003; Myoda et al., 2003). Further research is needed into the diversity of genotypes and phenotypes of fecal bacteria both within a single host organism and a host population (Stoeckel and Harwood, 2007). Based on the changes observed in source category RCCs in the modified libraries, additional research is needed into the robustness of commonly used classification algorithms when library source categories are unequally represented, as well as the effects of clone removal on the classification ability of a known-source library. Results suggest that the use of a VS of isolates is a necessary tool for assessing the size requirements of a known-source library. This method was unsuccessful using the restriction enzyme MboI, however one or several other potential restriction enzymes may provide greater source discrimination in future tests, even as applied to the same amplicon.

Acknowledgements

This research was funded by the Virginia Department of Health, Division of Zoonotic and Environmental Epidemiology and an internal Roanoke College grant. We thank the Hampton Roads Sanitation District and the Virginia Department of Conservation and Recreation for assistance in the collection of sewage and fecal samples. Special thanks to Roanoke College students Tiffany Simpson, Margaret Mauney, and Marina Salama for the screening of multiple restriction enzymes and the initial testing of the method.

References

Ahmed, W., R. Neller, and M. Katouli. 2005. Host species-specific metabolic fingerprint database for enterococci and *Escherichia coli* and its application to identify sources of fecal contamination in surface waters. Appl. Environ. Microbiol. 71:4461-4468.

Albert, J.M., J. Munakata-Marr, L. Tenorio, and R.L. Siegrist. 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. Environ. Sci. Technol. 37:4554-4560.

APHA. 1998. American Public Health Association, American Water Works Association, Water Pollution Control Federation. Standard methods for the Examination of Water and Wastewater. 20th ed. American Public Health Association, Washington D.C.

Bedendo, J. and A.C.C. Pignatari. 2000. Typing of *Enterococcus faecium* by polymerase chain reaction and pulsed field gel electrophoresis. Braz. J. Med. Biol. Res. 33:1269-1274.

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

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

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

Carroll, S., M. Hargreaves, and A.Goonetilleke. 2005. Sourcing faecal pollution from onsite wastewater treatment systems in surface waters using antibiotic resistance analysis. J. Appl. Microbiol. 99: 471-482.

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

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

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

Chun, J., A. Huq, and R.R. Colwell. 1999. Analysis of 16S-23S rRNA Intergenic Spacer Regions of *Vibrio cholerae* and *Vibrio mimicus*. Appl. Environ. Microbiol. 65:2202-2208.

Cole, D., S.C. Long, and M.D. Sobsey. 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl. Environ. Microbiol. 69:6507-6514.

Devriese, L.A., A.van de Kerckhove, R. Kilpper-Baelz, and K.H. Schleifer. 1987. Characterization and identification of *Enterococcus* species isolated from the intestines of animals. Int. J. Syst. Bacteriol. 37:257-259.

Dickerson, J. W. Jr., C. Hagedorn, and A. Hassall. 2007. Detection and remediation of human-origin pollution at two public beaches in Virginia using multiple source tracking methods. Water Res. 41:3758-3770.

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.

Dreiseitl, S. and L. Ohno-Machado. 2002. Logistic regression and artificial neural network classification models: A methodology review. J.Biomed. Informatics. 35:352-359.

Duran, M., B.Z. Haznedaroglu, and D.H. Zitomer. 2006. Microbial source tracking using host specific FAME profiles of fecal coliforms. Water Res. 40:67-74.

Field, K.G., A.E. Bernhard, and T.J. Brodeur. 2003. Molecular approaches to microbiological monitoring: fecal source detection. Environ. Mon. Assess. 81:313-326.

Fong, 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.

Graham, T., E.J. Golsteyn-Thomas, V.P. Gannon, and J.E. Thomas. 1996. Genus- and species specific detection of *Listeria monocytogenes* using polymerase chain reaction assays targeting the 16S/23S intergenic spacer region of the rRNA operon. Can. J. Microbiol. 42:1155-62.

Graves, A.K., C. Hagedorn, A. Teetor, M. Mahal, A.M. Booth, and R.B. Reneau Jr. 2002. Determining sources of fecal pollution in water for a rural Virginia watershed. J. Environ. Qual. 31:1300-1308.

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.

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.

Guertler, V. and V.A. Stanisich. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. Microbiol. 145:3-16.

Hagedorn, C., J.B. Crozier, K.A. Mentz, A.M. Booth, A.K. Graves, N.J. Nelson, and R.B., Reneau Jr. 2003. Carbon source utilization profiles as a method to identify sources of fecal pollution in water. J. Appl. Microbiol. 94:792-799.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneau Jr. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522-5531.

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.

Hartel, P.G., J.D. Summer, J.L. Hill, V. 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.

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.

Harwood, V.J., B. Wiggins, C. Hagedorn, R.D. Ellender, J. Gooch, J. Kern, M. Samadpour, A.C.H. Chapman, and B.J. Robinson. 2003. Phenotypic library-based microbial source

tracking methods: efficacy in the California collaborative study. J. Wat. Health. 1:153-166.

Indest, K.J., K. Betts, and J.S. Furey. 2005. Application of oligonucleotide microarrays for bacterial source tracking of environmental *Enterococcus sp.* isolates. Int. J. Environ. Public Health. 2:175-185.

Hsu, F., Y.S.C. Shieh, J. van Duin, M.J. Beekwilder, and M.D. Sobsey. 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. Environ. Microbiol. 61:3960-3966.

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.

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

Jimenez-Clavero, M.A., C. Fernandez, 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.

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.

Lasalde, C., R. Rodriguez, and G.A. Toranzos. 2005. Statistical Analyses: Possible reasons for unreliability of source tracking efforts. Appl. Environ. Microbiol. 71:4690-4695.

Lauková, A and Juris, P. 1997. Distribution and characterization of *Enterococcus* species in municipal sewages. Microbios. 89:73-80.

Long, S.C., S.S. El-Khoury, S.J.G. Oudejans, M.D. Sobsey, and J. Vinje. 2005. Assessment of sources of diversity of male-specific coliphages for source tracking. Environ. Eng. Sci. 22:367-377.

Maluquer de Motes, C., P. Clemente-Casares, A. Hundesa, M. Martin, and R. Girones. 2004. Detection of bovine and porcine adenovirsuses for tracing the source of fecal contamination. Appl. Environ. Microbiol. 70:1448-1454.

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

Myoda, S.P., C.A. Carson, J.J. Fuhrmann, B. Hahn, P.G. Hartel, R.L. Kuntz, C.H. Nakatsu, M.J. Sadowsky, M. Samadpour, and H. Yampara-Isquire. 2003. Comparing genotypic bacterial source tracking methods that require a host origin database. J. Wat. Health 1: 167-180.

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

Noble, R.T., and J.A. Fuhrman. 2001. Enteroviruses detected in the coastal waters of Santa Monica Bay, California: Low correlation to bacterial indicators. Hydrobiologia 460:175-184.

Parveen, S., K.M. Portier, K. Robinson, L. Eminston, 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.

Pond, K. R., R. Rangdale, W.G. Meijer, J. Brandao, L. Falcão, A. Rince, B. Masterson, J. Greaves, A. Gawler, E. McDonnell, and A. Cronin, S. Pedley. 2004. Workshop Report: Developing pollution source tracking for recreational and shellfish waters. Environ. Forensics 5:237-247.

Press, S. J. and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. J. Amer. Stat. Assoc. 73:699-705.

Riffard, S., F. Lo Presti, P. Normand, P. Forey, M. Reyrolle, J. Etienne, and F. Vandenesch. 1998. Species identification of *Legionella* via intergenic 16S–23S ribosomal spacer PCR analysis. Int. J. Syst. Bacteriol. 48:723–730.

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.

Samadpour, M., M.C. Roberts, C. Kitts, W. Mulugeta, and D. Alfi. 2005. The use of ribotyping and antibiotic resistance patterns for identification of host sources of *Escherichia coli* strains. L. Appl. Microbiol. 40:63-68.

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.

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* isolated from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. 69:1089-1092.

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.

Sechi, L.A. and L. Daneo-Moore. 1993. Characterization of intergenic spacers in two *rrm* of *Enterococcus hirae* ATCC 9790 operons. J. Bacteriol. 175: 3213-3219.

Seurinck, S., W. Verstraete, and S.D. Siciliano. 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. 69:4942-4950.

Simmons Jr., G. E., D.F. Waye, S. Herbein, S. Myers, and E. Walker. 2002. Estimating nonpoint source fecal coliform sources using DNA profile analysis. p. 143-168. *In* T. Younos (ed.) Advances in Water Monitoring Research. Water Resources Publications, Denver, CO.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2002. Microbial source tracking: State of the science. Environ. Sci. Technol. 36:5279-5288.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2003. Assessment of equine fecal contamination: The search for alternative bacterial source-tracking targets. FEMS Microbiol. Ecol. 47:65-75.

Sinton, L.W., A.M. Donnison, and C.M. Hastie. 1993. Faecal streptococci as faecal pollution indicators: a review. Part II: Sanitary significance, survival, and use. N. Z. J. Mar. Freshwater Res. 27:117-137.

Sinton L.W., R.K. Finlay, and D.J. Hannah. 1998. Distinguishing human from animal faecal contamination in water: a review. N. Z. J. Mar. Freshwater Res. 32:323-348.

Stewart, J.R., R.D. Ellender, J.A. Gooch, S. Jiang, S.P. Myoda, and S.B. Weisberg. 2003. Recommendations for microbial source tracking: Lessons learned from a methods comparison study. J. Water Health 1:225-231.

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

Stoeckel, D.M., and J. Harwood. 2007. Performance, design, and analysis in microbial source tracking studies. Appl. Environ. Microbiol.73: 2405-2415.

Stubbs, S.L.J., J.S. Brazier, G.L. O'Neill, and B.I. Duerden. 1999. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 Different PCR ribotypes. Appl. Environ. Microbiol. 37:461-463.

Sundram, A., N. Jumanlal, and M.M. Ehlers. 2006. Genotyping of F-RNA coliphages isolated for wastewater and river samples. Water SA 32:65-70.

Terrin, N., C.H. Schmid, J.L. Griffith, R.B. D'Agostino Sr., H.P. Selker. 2003. External validity of predictive models: A comparison of logistic regression, classification trees, and neural networks. J. Clin. Epidemiol. 56:721-729.

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

USEPA. 1986. Ambient Water Quality Criteria for Bacteria – 1986. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA-440/5-84-002.

USEPA. 2005. Microbial Source Tracking Guide Document. U. S. Environmental Protection Agency, Office of Research and Development: Washington, D.C.

VanOmmeren, L. and E.W. Alm. 2006. Development and application of rapid antibiotic resistance analysis for microbial source tracking in the Black River watershed, Michigan. Lake and Reservoir Management. 22:240-244.

Vantarakis, A., D. Venieri, Komninou, and M. Papapetropoulou. 2006. Differentiation of faecal *Escherichia coli* from humans and animals by multiple antibiotic resistance analysis. L. Appl. Microbiol. 42:71-77.

VDH. 2004. DRAFT–Virginia Department of Health Beach Monitoring Protocol. July 2004 (unpublished).

Wheeler, A.L., P.G. Hartel, D.G. Godfrey, J.L. Hill, and W.I. Segars. 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J. Environ. Qual. 31:1286-1293.

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.

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.

Table 40 – Classification table displaying the percentages (and number) of isolates classified using discriminant analysis for a 64-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=64	Birds (16)	**100.0(16)**†	0.0 (0)	0.0 (0)	0.0 (0)
	Dogs (16)	0.0 (0)	**100.0 (16)**	0.0 (0)	0.0 (0)
	Sewage (16)	0.0 (0)	0.0 (0)	**100.0 (16)**	0.0 (0)
	Wildlife (16)	0.0 (0)	0.0 (0)	0.0 (0)	**100.0 (16)**

†Bold indicates rate of correct classification (RCC) for each source category. Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 41 – Classification table displaying the percentages (and number) of isolates classified using discriminant analysis for a 200-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=200	Birds (50)	**76.0 (38)**†	14.0 (7)	0.0 (0)	10.0 (5)
	Dogs (50)	6.0 (3)	**88.0 (44)**	6.0 (3)	0.0 (0)
	Sewage (50)	4.0 (2)	10.0 (5)	**80.0 (40)**	6.0 (3)
	Wildlife (50)	2.0 (1)	8.0 (4)	2.0 (1)	**88.0 (44)**
	Validation Birds (12)	**25.0 (3)**	25.0 (3)	25.0 (3)	25.0 (3)
	Validation Dogs (20)	45.0 (9)	**50.0 (10)**	5.0 (1)	0.0 (0)
	Validation Sewage (48)	12.5 (6)	31.3 (15)	**45.3 (22)**	10.4 (5)
	Validation Wildlife (20)	15.0 (3)	15.0 (3)	10.0 (2)	**60.0 (12)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 42 – Classification table displaying the percentages (and number) of isolates classified using discriminant analysis for a 1029-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=1029	Birds (201)	**36.8 (74)**†	25.9 (52)	21.4 (43)	15.9 (32)
	Dogs (266)	25.9 (69)	**31.6 (84)**	26.3 (70)	16.17 (43)
	Sewage (353)	13.3 (47)	33.1 (117)	**39.7 (140)**	13.9 (49)
	Wildlife (209)	7.7 (16)	14.4 (30)	10.5 (22)	**67.5 (141)**
	Validation Birds (12)	**50.0 (6)**	0.0 (0)	25.0 (3)	25.0 (3)
	Validation Dogs (20)	55.0 (11)	**35.0 (7)**	5.0 (1)	5.0 (1)
	Validation Sewage (48)	20.8 (10)	18.8 (9)	**39.6 (19)**	20.8 (10)
	Validation Wildlife (20)	5.0 (1)	10.0 (2)	10.0 (2)	**75.0 (15)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 43 - Classification table displaying the percentages (and number) of isolates classified using logistic regression for a 64-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=64	Birds (16)	**100.0 (16)**†	0.0 (0)	0.0 (0)	0.0 (0)
	Dogs (16)	0.0 (0)	**100.0 (16)**	0.0 (0)	0.0 (0)
	Sewage (16)	0.0 (0)	0.0 (0)	**100.0 (16)**	0.0 (0)
	Wildlife (16)	0.0 (0)	0.0 (0)	0.0 (0)	**100.0 (16)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 44 - Classification table displaying the percentages (and number) of isolates classified using logistic regression for a 200-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=200	Birds (50)	**84.0 (42)**†	10.0 (5)	0.0 (0)	6.0 (3)
	Dogs (50)	2.0 (1)	**94.0 (47)**	4.0 (2)	0.0 (0)
	Sewage (50)	0.0 (0)	6.0 (3)	**94.0 (47)**	0.0 (0)
	Wildlife (50)	2.0 (1)	4.0 (2)	14.0 (7)	**80.0 (40)**
	Validation Birds (12)	**33.3 (4)**	16.7 (2)	25.0 (3)	25.0 (3)
	Validation Dogs (20)	55.0 (11)	**40.0 (8)**	0.0 (0)	5.0 (1)
	Validation Sewage (48)	8.3 (4)	27.1 (13)	**52.1 (25)**	12.5 (6)
	Validation Wildlife (20)	0.0 (0)	10.0 (2)	35.0 (7)	**55.0 (11)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 45 - Classification table displaying the percentages (and number) of isolates classified using logistic regression for a 1029-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=1029	Birds (201)	**10.9 (22)**†	45.8 (92)	29.4 (59)	13.9 (28)
	Dogs (266)	5.6 (15)	**42.5 (113)**	37.2 (99)	14.7 (39)
	Sewage (353)	2.3 (8)	28.9 (102)	**56.9 (201)**	11.9 (42)
	Wildlife (209)	1.4 (3)	15.3 (32)	19.1 (40)	**64.1 (134)**
	Validation Birds (12)	**0.0 (0)**	33.3 (4)	41.7 (5)	25.0 (3)
	Validation Dogs (20)	5.0 (1)	**75 (15)**	15.0 (3)	5.0 (1)
	Validation Sewage (48)	4.2 (2)	31.3 (15)	**47.9 (23)**	16.7 (8)
	Validation Wildlife (20)	5.0 (1)	5.0 (1)	15.0 (3)	**75.0 (15)**

†Bold indicates rate of correct classification (RCC) for each source category
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 46 - Classification table displaying the percentages (and number) of isolates classified with discriminant analysis for a source-category balanced 804-isolate library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=804	Birds (201)	**31.3 (63)**†	32.8 (66)	22.4 (45)	13.4 (27)
	Dogs (201)	25.9 (52)	**32.3 (65)**	25.4 (51)	16.4 (33)
	Sewage (201)	12.9 (26)	27.4 (55)	**43.3 (87)**	16.42 (33)
	Wildlife (201)	7.5 (15)	11.9 (24)	14.4 (29)	**66.2 (133)**
	Validation Birds (12)	**25.0 (3)**	25.0 (3)	25.0 (3)	25.0 (3)
	Validation Dogs (20)	55.0 (11)	**30.0 (6)**	10.0 (2)	5.0 (1)
	Validation Sewage (48)	14.6 (7)	25.0 (12)	**41.7 (20)**	18.8 (9)
	Validation Wildlife (20)	5.0 (1)	5.0 (1)	15.0 (3)	**75.0 (15)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 47 - Classification table displaying the percentages (and number) of isolates classified with logistic regression for a source-category balanced 804-isolate (201 per category) library.

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=804	Birds (201)	**35.8 (72)**†	31.3 (63)	18.9 (38)	13.9 (28)
	Dogs (201)	25.9 (52)	**33.3 (67)**	24.9 (50)	15.9 (32)
	Sewage (201)	11.0 (22)	26.9 (54)	**42.8 (86)**	19.4 (39)
	Wildlife (201)	5.0 (10)	12.9 (26)	13.4 (27)	**68.7 (138)**
	Validation Birds (12)	**41.7 (5)**	8.3 (1)	25.0 (3)	25.0 (3)
	Validation Dogs (20)	55.0 (11)	**25.0 (5)**	15.0 (3)	5.0 (1)
	Validation Sewage (48)	14.6 (7)	29.2 (14)	**37.5 (18)**	18.8 (9)
	Validation Wildlife (20)	5.0 (1)	5.0 (1)	15.0 (3)	**75.0 (15)**

†Bold indicates rate of correct classification (RCC) for each source category.
Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 48 – Classification table displaying the percentages (and number) of isolates classified with discriminant analysis for a 323-isolate library (clones removed from the 1,029-isolate library, Table 42).

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=323	Birds (61)	**32.8 (20)**†	32.8 (20)	24.6 (15)	9.8 (6)
	Dogs (60)	6.7 (4)	**68.3 (41)**	18.3 (11)	6.7 (4)
	Sewage (150)	17.3 (26)	22.0 (33)	**52.0 (78)**	8.7 (13)
	Wildlife (52)	1.9 (1)	25.0 (13)	11.5 (6)	**61.5 (32)**
	Validation Birds (7)	**0.0 (0)**	42.9 (3)	42.9 (3)	14.3 (1)
	Validation Dogs (14)	0.0 (0)	**92.9 (13)**	7.1 (1)	0.0 (0)
	Validation Sewage (33)	21.2 (7)	15.2 (5)	**54.6 (18)**	9.1 (3)
	Validation Wildlife (8)	0.0 (0)	25.0 (2)	12.5 (1)	**62.5 (5)**

†Bold indicates rate of correct classification (RCC) for each source category. Classification rates have been rounded, and as a result may not add to 100.0% across rows.

Table 49 – Classification table displaying the percentages (and number) of isolates classified with logistic regression for a 323-isolate library (clones removed from the 1,029-isolate library, Table 45).

Library Size	Source (number of isolates)	Percentages (and number) of isolates classified as: Birds	Dogs	Sewage	Wildlife
n=323	Birds (61)	**26.2 (16)**†	13.1 (8)	52.5 (32)	8.2 (5)
	Dogs (60)	3.3 (2)	**46.7 (28)**	43.3 (26)	6.7 (4)
	Sewage (150)	4.7 (7)	8.0 (12)	**85.3 (128)**	2.0 (3)
	Wildlife (52)	1.9 (1)	11.5 (6)	28.9 (15)	**57.7 (30)**
	Validation Birds (7)	**0.0 (0)**	14.29 (1)	71.4 (5)	14.3 (1)
	Validation Dogs (14)	0.0 (0)	**71.4 (10)**	28.6 (4)	0.0 (0)
	Validation Sewage (33)	6.1(2)	12.1 (4)	**78.8 (26)**	3.0 (1)
	Validation Wildlife (8)	0.0 (0)	12.5 (1)	37.5 (3)	**50.0 (4)**

†Bold indicates rate of correct classification (RCC) for each source category. Classification rates have been rounded, and as a result may not add to 100.0% across rows.

VI. An Evaluation of Source Category Imbalance in Microbial Source Tracking Libraries Using Multiple Statistical Algorithms

Abstract

While much research in the field of microbial source tracking (MST) has moved into methods not involving the construction of known-source libraries, the ongoing development of total maximum daily loads (TMDLs) in the United States continues to rely on several library-based MST methods for the quantification of fecal pollution sources. One question facing MST field application involves source classification bias when one source category contains a greater number of total isolates. In this study, discriminant analysis (DA), logistic regression (LR), and k-nearest neighbor (k-NN) statistical classification algorithms were tested on antibiotic resistance analysis (ARA) libraries (and an unchanged validation set (VS)) generated where one source category was made disproportionately large. Findings indicated the DA was the most robust algorithm for use with source category imbalance when measuring both correct and incorrect classification rates. Conversely k-NN was identified as the most sensitive algorithm to imbalances with the greatest levels of distortion obtained from the highest k values. Results from the VS were similar to those of the small constructed libraries but at lower rates of correct classification. Thus, when the desires of a watershed manager extend beyond the scope of available library-independent methods a library-based approach can still prove useful given careful selection of size, sources and analyses within a constructed library.

Introduction

Numerous surface waters throughout the United States currently fail to meet the minimum water quality standards required by the Clean Water Act (USEPA, 2007). Applying federal recommendations, water pollution standards are established by individual states, territories, and tribes based on intended water uses, such as drinking or recreation. Those water bodies identified as impaired require the development of a Total Maximum Daily Load (TMDL) to quantify the amount of a pollutant from point and nonpoint sources, both natural and anthropogenic in origin. The completion of a TMDL specifies the amount of pollution that may be legally discharged into a body of water while still maintaining water quality standards. As frequent source of violations, elevated concentrations of fecal indicator bacteria (FIB; *Escherichia coli*, *Enterococcus spp*. or fecal coliforms) require a knowledge of both fecal sources and modes of entry into waterways to create useful and accurate TMDLs (Kern et al., 2002). While, visual inspection of land use patterns can often lead researchers to a suspected source, in many instances the fecal contributors are not obvious. The development of microbial source tracking (MST) over the last decade has provided watershed managers with a means to identify the fecal sources contributing to elevated bacterial concentrations in surface waters.

The idea behind MST is that certain enteric bacterial strains or viruses are uniquely adapted to, and thus reside exclusively in the gastrointestinal tract of a single, or group of closely-related host organisms. Differences in the phenotypes or genotypes of these strains can be used to identify animal source contributions to the fecal pollution in a waterbody. Two major classes of MST methods (library-dependent and -independent) are currently being utilized in surface waters across the world (Scott et al., 2002; Simpson et al., 2002; Blanch et al., 2006; Field and Samadpour, 2007).

Library-dependent methods rely on the construction of a database, or library of isolates collected from known fecal sources. Bacterial isolates obtained from these known fecal sources are assayed to provide a collection of possible 'fingerprint' patterns allowing for a direct comparison with the fingerprints of isolates of unknown origin using statistical classification algorithms. Although a variety of statistical methods have been employed in library-based MST, discriminant analysis (DA) has seen the most frequent application (Wiggins, 1996; Hagedorn et al., 1999; Dickerson et al., 2007a; Graves et al., 2007). An alternative parametric algorithm, logistic regression (LR), has seen more limited applications (Harwood et al., 2003; Dickerson et al., 2007a), while non-parametric, k-nearest neighbor (k-NN) and maximum similarity [considered homologous to a nearest neighbor value of k=1 (Wilbur and Whitlock, 2007)] have largely received attention in studies comparing classification proficiencies (Albert et al., 2003; Ritter et al., 2003; Zhong, 2003; Albert et al., 2004; Ritter and Robinson, 2004; Robinson *et al*, 2007).

A second class of MST methods, known as library-independent, has become increasingly popular in recent years. The use of specific microbes or DNA sequences within these microbes, believed to be unique to a given host can eliminate the need to collect and fingerprint large numbers of isolates to generate a known-source library. However, most library-independent MST methods do not use the currently recognized FIB, lack the ability to detect multiple fecal sources (Noble et al., 2003), or have demonstrated a limited geographic host marker range (Hamilton et al., 2006) similar to library-dependent methods and thus remain largely limited for TMDL applications. As a result, library-dependent methods continue to be utilized, fulfilling the desire of many watershed managers, to quantify the percentage of fecal material present in a

water body from given animal sources This data is used to generate TMDL load allocations (Stoeckel and Harwood, 2007).

Numerous library-dependent MST methods have been employed as part of TMDL development, including applications in Texas of repetitive extragenic palendromic polymerase chain reaction (rep-PCR), ribotyping, carbon source utilization, antibiotic resistance analysis (ARA), and pulsed-field gel electrophoresis (PFGE; Bacterial TMDL Report, 2006). Additional examples in TMDLs include rep-PCR in Mississippi (Baffaut et al., 2005), ARA in South Dakota (Troelstrup et al., 2005), and ARA and PFGE in Virginia (Culver et al., 2002; VDEQ, 2003; VDEQ, 2004).

Conclusions presented in the Identifying Sources of Fecal Pollution edition of Water Research, published in August 2007, indicated that combining multiple library-based MST methods was successful for correctly identifying fecal sources (Ahmed et al., 2007; Dickerson et al., 2007a; Edge and Hill, 2007). In addition, the construction of a library of sufficient size (Dickerson et al, 2007b) and tested with a challenge, or validation set (VS) of non-library isolates (Dickerson et al, 2007a; Graves et al., 2007; Stoeckel and Harwood, 2007) increased confidence in the correct classification of isolates of unknown origin. One of the problems with the construction of known-source libraries has been the relative ease with which isolates from some fecal sources can be obtained, such as sewage, and the difficulties associated with the collection of other fecal samples, such as those from birds or wildlife. A concern exists that the construction of libraries containing one or more source categories with significantly more isolates than others in the library may have the tendency to bias the classification of isolates of unknown origin into those larger categories due primarily to the larger diversity of fingerprint

patterns. The question of classification bias has also been compounded by the variety of algorithms and MST methods available to researchers. (Ritter et al., 2003; Robinson et al., 2007)

The use of imbalanced source categories in libraries generated for the United States Geological Survey (USGS) sponsored method comparison (MC) study was cited as a possible reason for the lack of success amongst all methods involved (Stoeckel et al., 2004). However, to date only one study has examined the effects of uneven source categories within a library. Results generated by Robinson et al., 2007 using a rep-PCR library indicated that the non-parametric algorithm, maximum similarity was the most sensitive to disproportionately large source categories, while the two parametric algorithms, average similarity and DA were the most immune to classification bias in imbalanced libraries. However k-NN, for the rep-PCR library, was determined to possess a lower bias similar to average similarity (a parametric algorithm) and DA with a high rate of correct classification (RCC) among isolates within the library. Our study looked to expand upon these results by 1) testing an ARA generated library using the statistical classification algorithms DA, LR, and k-NN (with k = 1, 5, 10, and 100); 2) increase the levels of imbalance of a single source category beyond that of previous studies through a stepwise addition of isolates in increments of 100, generating categories with as much as four times the number of isolates as any other single category; and 3) examined the effect of imbalance on the misclassification of isolates from the small (non-additive) categories into the disproportionately larger (additive) category.

Materials and Methods

Construction of the original known-source library

The library from which isolates were drawn for this study contained a total of 2502 enterococci and 100 validation set (VS) isolates, collected primarily during the summers of 2004 and 2005, and used in a previous MST project (Dickerson et al., 2007a). Isolates, which were not part of the previous MST project (226 total), were collected in January 2006, and used to replace half (50) of the VS isolates as well as to further increase the size of the known-source library. The VS isolates were collected at the same time as the majority of those used to construct the library. However, none of the fecal/sewage samples used to obtain VS isolates were used to construct the known-source library.

Classification results for the original 2502-isolate library when applying a four-way classification split (birds, pets, sewage, and wildlife) for both library and VS isolates and all three statistical algorithms are seen in Tables 1-3.

Antibiotic Resistance Analysis fingerprints were generated for all isolates as previous described in Dickerson et al., 2007a on both library and VS isolates using a total of 30 concentrations of 9 different antibiotics (one to five concentrations selected for each antibiotic). Antibiotic concentrations were selected for the previous project based on successful field applications from Graves et al., 2002 and Graves et al., 2007.

Construction of the imbalanced libraries

An initial 400-isolate library was generated using 100 randomly-selected isolates from each of the four source categories (bird, pets, sewage, and wildlife) present in the 2502-isolate ARA library, to create a source-category balanced library (100 isolates per source).

Classification models were developed for this source-category balanced library using both DA and LR. Isolates composing the library were then classified, and the rates of correct classification (RCCs) for each of the four source categories were recorded. Isolates from the VS were then run against the library using DA, LR, and k-NN, and RCCs recorded. As k-NN does not generate a model, only the VS isolates were run against the library.

Following the establishment of these baseline values, an additional 100 isolates from a single, selected source (randomly chosen from the remaining isolates) were added to the library creating an "additive category," generating a library of 500 total isolates. The total isolates remaining in the "non-additive categories" remained at 100 throughout the process. This newly generated library now contained a category with twice as many isolates as any other single category. Classification models were re-run for this new library, and both library and VS isolates were classified using DA, LR, and k-NN as described above. This process was repeated, adding isolates only to the additive category, in groups of 100, until the library contained a single category with four times as many isolates as any of the non-additive categories. Throughout the experiment, a total of 20 balanced, 400-isolate libraries were generated, with each source category (bird, pets, sewage, and wildlife) serving as the additive category five times.

Calculation of the percentage of correctly classified isolates and the percentage of isolates incorrectly classified into the additive category

All RCCs were based on a four-source classification split (bird, pets, sewage, and wildlife). Therefore data reported from the non-additive categories is an average of the isolates correctly classified from the three sources categories which remained at 100 isolates throughout the trial. While the 20 different baseline (evenly-balanced) libraries were generated, and

modified by the addition of isolates to the additive category, all VS isolates were held constant through out the experiment As an additional means of assessing the impact of source category imbalance, the percentage of incorrectly classified isolates were calculated as the total number of isolates from the non-additive categories identified as a member of the additive category, divided by the total number of non-additive isolates. Therefore, if 45 non-additive isolates were incorrectly identified as members of the additive source, then dividing 45 by the total number of additive isolates, 300, would yield an incorrect classification rate of 15.0%.

Statistical analyses

Analyses were conducted using two parametric classification methods, DA and LR, in SAS-JMP version 5.0.1 (SAS Institute Inc.), as described in Hagedorn et al., 1999, Graves, et al., 2002, and Dickerson et al., 2007a; and k-NN, a non-parametric method, using R version 2.4.1, as described in Zhong 2003. Discriminant analysis measured similarity using Mahalanobis distance; k-NN used Euclidean distance, while LR used maximum-likelihood estimation.

Four different k-values were applied to this experiment for k-NN. A value of k=1 was selected to assess the ability of isolates within the library to correctly classify VS isolates by matching a single isolate within the known-source library displaying the greatest degree of similarity. In addition, a value of k=1 is identical in principal to the maximum similarity algorithm reported in other MST publications. Values of k=5 and k=10 were selected because both fell within the range of optimal k values (4.5 to 11.7) based on a data set of ‘n’ isolates, with ‘n’ ranging from 400 to 700 isolates and applying the formula that the optimal k value will fall within $k=n^{2/8}$ and $k=n^{3/8}$ (Wilbur and Whitlock, 2007). The final value of k=100 was selected

to observe the results of library imbalance at its most extreme, as the non-additive categories each contained only 100 isolates.

Results

Original library

Classification results for both library and VS isolates applying all three statistical algorithms and using the entire 2502-isolate library and are seen in Tables 1-3. The highest average rate of correct classification (ARCC; average of RCCs) for the library was generated using LR (72.0%), however the highest ARCC for the VS was produced using k=10 (which was the only k value tested which fell within the calculated optimum range k=7 to k=19 for this size library).

Discriminant analysis

Applying DA to the 400-isolate, evenly-balanced and randomly-generated libraries, with 100 isolates/category (Figure 4), yielded classification rates for isolates composing the library similar to the larger library seen in Table 50. However, VS isolates were correctly classified by the 400-isolate library at appreciably lower rates than by the 2502-isolate library. Isolates composing the additive category in the 400-isolate library initially produced a RCC of 73.9%. An improvement of only 1.5 percentage points was seen in the classification of library isolates from the additive category, as the category increased to 400 isolates, creating a single category with more isolates than the remaining three categories combined (300 total isolates). A greater, but still modest, increase was seen in the RCC of VS isolates representing the additive category (62.8% to 68.2%) as it grew to four times its original size. Application of DA to non-additive

categories displayed only minor changes, as RCCs decreased by 3.2 and 1.2 percentage points for the library and VS isolates, respectively, as the additive category increased to 400 isolates.

Logistic regression

The application of LR to the evenly-balanced libraries (Figure 5) also produced higher RCCs for library isolates, and lower RCCs for isolates from the VS, than results using the complete library (Table 51). Library isolates within the additive categories began at a RCC of 77.5%, increasing to 84.5% as the additive category was doubled, and to 89.8% as the additive category increased to four times its original size. VS isolates of the additive category began at an RCC of 62.0% and increased by over 20 percentage points reaching 82.4% with a four-fold category increase. Library isolates from the non-additive categories began at an RCC of 75.9%, and steadily decreased to 62.0% as the additive category reached 400 isolates. Isolates from the non-additive categories of the VS were correctly classified at an average rate of 56.6% in the balanced library, and decreased to a RCC of 44.0% when the additive category was increased four-fold.

k-nearest neighbor

As no model is generated using k-NN, only the VS isolates were used to evaluate the effects of unequal source category distributions (Figure 6). The application of the smaller, evenly-distributed libraries to the validation set produced lower initial RCCs than the original known-source library (Table 52) for all k values. Validation set isolates from the additive category using k=1 and k=5 began at the highest RCCs of the four k values applied (Figure 6). A four-fold increase in the number of isolates present in the additive category produced an increase

in additive category RCC values of 22.0 (k=1) and 22.4 (k=5) percentage points. Even greater increases were seen in the additive categories with the larger k values as classification rates increased from 52.5% to 74.8% to 88.4% for k=10, and from 56.2% to 87.6% to 96.8% using k=100, as the additive category doubled and quadrupled in size, respectively. Isolates from the non-additive categories followed similar but inverse trends to those seen with additive isolates and corresponding k values. Application of k=1 displayed the most modest decline in correct classification rates while the most significant decrease was seen with k =100. The RCC values of k=5, and k=10 yielded declining rates in between the extremes of k=1 and k=100.

Misclassification of non-additive isolates into additive category

The percentage of isolates from non-additive categories misclassified into the additive category was calculated as an additional means to evaluate bias in unbalanced libraries. Results calculated using all three statistical algorithms are seen in Figures 7 and 8. Isolates from the library incorrectly identified as a member of the additive category was initially low (Figure 7) using both DA (6.5%) and LR (6.2%). However, as the isolates in the additive category doubled, a slight increase to 7.2% was seen with DA, while application of LR yielded a larger increase to 11.6%. Increasing the additive category to four times the original size produced another small increase to 8.5% using DA and a considerably larger increase to 20.5% with the application of LR. Isolates from the VS showed similar results, with the percentage of isolates misclassified into the additive category increasing more rapidly with LR than with DA. Applying k-NN (Figure 8), the greatest increases were produced by the higher k values, with k=1 not quite doubling in value from 16.5% to 31.0% as the additive category quadrupled. The calculated percentages using k=100, however more than doubled once the number of isolates in the additive

category increased two-fold, and more than tripled once the number of additive isolates was increased to 400.

Discussion

This initial baseline library generated in these experiments consisted of only 400 isolates. Although smaller libraries have been reported (Dombek et al. 2000; Guan et al., 2002; Seurinck et al., 2003; Lasalde et al., 2005; Duran et al., 2006; Vantarakis et al., 2006), deployment in field studies of such libraries has not been recommended in even the smallest of watersheds (Stewart et al., 2003). For the purposes of this experiment, the initial small libraries that were generated allowed for a greater distortion of the additive category relative to the other source categories.

Although LR did produce the highest ARCC (72%) of isolates within the original library, between the two parametric classification methods (Tables 1 and 2), the range between the library ARCC and that of the VS isolates was 8 percentage points, considerably greater than the 3.3 percentage point difference seen using DA. Although no publications have cited an acceptable range to achieve between VS and library isolates, DA, while producing a lower ARCC (69.3%) for the library, could be deemed the only library of adequate size if relatively strict requirements (e.g. $\leq$5%) are applied.

As a classification algorithm, DA proved to be consistently less sensitive to library imbalances caused by the inflation of one source category, as seen by the slower increase and decline in the RCC among additive and non-additive categories, respectively. In addition, the percentage of isolates from the non-additive category misclassified into the additive category was relatively small as compared to LR and k-NN. While the classification of VS isolates from

the additive category did display a modest increase as the additive category increased in size, again this change was the smallest among the algorithms tested.

Both library and VS isolates classified using LR and k-NN displayed a steady increase in the RCC of isolates from the additive categories and a steady decrease in the RCC of isolates from the non-additive categories. In general, k-NN was the least accurate algorithm for the classification of VS isolates in the randomly-generated, evenly-balanced libraries, and was subject to greatest changes as one source category became increasingly larger. The sensitivity to isolate imbalance between categories indicates that researchers employing either of these methods should take increasing care to keep known-source categories roughly equal in size.

Both of the U.S. based MC studies (Stewart et al., 2003; Stoeckel et al., 2004) emphasized the use of a VS of isolates to assess the classification ability of a known-source library from environmental unknowns. However, only a few studies to date have utilized such a set prior to declaration of the success of a method or field deployment (Moore et al., 2005; Dickerson et al. 2007a). The correct classification rates of VS isolates using DA and LR and the generated 400-isolate libraries were consistently lower (>10% when both the additive and non-additive categories were averaged) than the RCC for the isolates within the library, indicating a library which was of inadequate size for field applications. The application of a VS has shown to be useful in the determination of minimum library size requirements and in preventing a report of method success based on a modest library size (Dickerson et al., 2007b).

Library-dependent methods have found increased levels of success applying either a minimum detectable percentage (MDP, Harwood et al., 2003; Dickerson et al., 2007a) or similarity thresholds (Myoda et al., 2003) to aid in the identification of isolates of unknown origin. Application of a MDP eliminates sources which are not identified at or above a certain

percentage of the total isolates from a single, or collection, of water samples. Thus if the number of isolates are not identified above a certain percentage (e.g. 25%) that source is considered negligible. Application of a MDP in a MC study consistently increased the ability of many known-source libraries to correctly identify sources in blind water samples (Harwood et al., 2003). The advantage of using a MDP is the elimination of sources which are identified at levels typically below the classification limits of the library, thereby increasing confidence in reported results. The major disadvantage however is inability to report the relative percentages of isolates present from a given source, particularly when using a large MDP.

An alternative to the MDP is the use of similarity thresholds. Thresholds require that the fingerprint, produced by an isolate of unknown origin, is similar in pattern to an isolate or group of isolates present in the known source library by a designated percentage (e.g. 80%, 100%) using a given classification algorithm. This method has been shown to be successful in both U.S-based MC studies (Stewart et al., 2003; Stoeckel et al., 2004), although a large number of isolates (up of 95%) from a given water sample were excluded from analyses. Thus, when applied in field studies the number of excluded isolates should be reported. Failure to report these isolates identified as 'unknown' can lead to significant distortions in the contributing fecal sources reported.

The identification of k-NN as the most sensitive algorithm to imbalances in source category isolates is the opposite of the results reported by Robinson et al., 2007. However, their results were based on a library of isolates constructed using rep-PCR. This may indicate that these results may be limited ARA, or even an individual library, and further testing is warranted. In addition, continued research is needed into the effectiveness and optimization of both similarity thresholds and MDPs for source identifications. The hope of MST researchers is that

the continued development of library-independent methods will eliminate the need for the construction of known-source libraries. However, the current TMDL projects continue to use the quantifications of animal fecal inputs, a request which at present remains beyond the scope of the available library-independent methods.

Acknowledgements

This research was funded by the Virginia Department of Health, Division of Zoonotic and Environmental Epidemiology. We thank the Hampton Roads Sanitation District and the Virginia Department of Conservation and Recreation for assistance in the collection of sewage and fecal samples.

References

Ahmed, W., J. Stewart, T. Gardner, D. Powell, P. Brooks, D. Sullivan, and N. Tindale. 2007. Sourcing faecal pollution: A combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments **41**, 3771-3779.

Albert, J.M., J. Munkata-Marr, L. Tenorio, and R.L. Siegrist. 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. Environ. Sci. Technol. 37:4554-4560.

Albert, J.M., J. Munkata-Marr, L. Tenorio, and R.L. Siegrist. 2004. Experimental and statistical evaluation of bacterial source tracking in a subalpine region by rep-PCR DNA fingerprinting On-Site Wastewater Treatment X, Conference Proceedings, 21-24 March 2004 (Sacramento, California USA). pp. 470-481.

Bacteria TMDL Report. 2006. Bacteria total maximum daily load task force report. Prepared for: Texas commission on environmental quality and Texas state soil and water conservation board. [online] http://twri.tamu.edu/bacteriatmdl/TSSWCB%20Comments%20on%20Second%20Draft.doc. last accessed 12 October 2007.

Baffaut, C., C.A. Carson, and W. Rogers. 2005. Little Sac river watershed: bacterial source tracking analysis. FAPRI-UMC Report #06-05.

Blanch, A.R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iverson, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G.T. Papageorgiou, H. Taylor, J. Wallis, and J. Jofre. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl. Environ. Microbiol. 72:5915-5926.

Culver, T.B., Y. Jia, R. Tikoo, and J. Simsic. 2002. Development of the total maximum daily load (TMDL) for fecal coliform bacteria in Moore's Creek, Albemarle County, Virginia. Virginia Department of Environmental Quality and Virginia Department of Conservation and Recreation.

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

Duran, M., B.Z. Haznedaroglu, and D.H. Zitomer. 2006. Microbial source tracking using host specific FAME profiles of fecal coliforms. Water Res. 40:67-74.

Dickerson, J. W. Jr., C. Hagedorn, and A. Hassall. 2007a. Detection and remediation of human-origin pollution at two public beaches in Virginia using multiple source tracking methods. Water Res. 41:3758-3770.

Dickerson, J.W. Jr, J. B. Crozier, C. Hagedorn, and A. Hassall. 2007b. Assessment of the 16S-23S rDNA intergenic spacer region in *Enterococcus spp.* for microbial source tracking. J. Environ. Qual. 36:1661-1669.

Edge, T.A. and S. Hill. 2005. Occurrence of antiobiotic resistance in *Escherichia coli* from surface waters and fecal pollution sources near Hamilton, Ontario. Can. J. Microbiol. 51:501-505.

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

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

Graves, A.K., C. Hagedorn, A. Brooks, R.L. Hagedorn, and E. Martin. 2007. Microbial source tracking in a rural watershed dominated by cattle. Wat. Res. 41:3729-3739.

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. Applied and Environmental Microbiology **68**, 2690-2698.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneneau. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522-5531.

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.

Harwood, V.J., B. Wiggins, C. Hageodorn, R.D. Ellender, J.Gooch, J. Kern, M. Samadpour, A.C.H. Chapman, and B.J. Robinson. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J. Wat. Health. 1:153-166.

Kern, J., B. Petrauskas, P. McClellan, V.O. Shanholtz, and C. Hagedorn. 2002. Bacterial source tracking: A tool for total maximum daily load development. In T. Younos (Ed.). Advances in Water Monitoring Research. (pp. 125-142). Denver, Co.: Water Resources Publications.

Lasalde, C., R. Rodriguez, and G.A. Toranzos. 2005. Statistical analyses: possible reasons for unreliability of source tracking efforts. Appl. Environ. Microbiol. 71:4690-4695.

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

Myoda, S. P., C. A. Carson, J. J. Fuhrmann, B.K. Hahm, P.G. Hartel, H. Yampara-Iquise, L. Johnson, R.L. Kuntz, C.H. Nakatsu, M.J. Sadowsky, and M. Samadpour. 2003. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. J. Wat. Health. 1:167-180.

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

Ritter, K.J., C. 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 using in library-based approaches to microbial source tracking. J. Wat. Health. 1:209-223.

Ritter K. and B. Robinson. 2004. Statistical appraisal of disproportional versus proportional libraries. Presentation, presentation given during the Golf of Mexico Bacterial Source Tracking workshop in Biloxi, MS, Nov 11-13, 2004. Moderator: R.D. Ellender, University of Southern Mississippi, Hattisburg, MS.

Robinson, B.J., K.J. Ritter, and R. D. Ellender. 2007. A statistical appraisal of disproportional versus proportional microbial source tracking libraries. J. Wat. Health 5**:**503–509.

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 human, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. 69:1089-1092.

Seurinck, S., W. Verstraete, and S.D. Siciliano. 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. 69:4942-4950.

Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. 2002. Microbial source tracking: state of the science. Environ Sci. Technol. 36:5280-5288.

Stewart, J.R., R.D.Ellender, J.A.Gooch, S. Jiang, S.P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons learned from a methods comparison study. J. Wat. Health. 1:225-231.

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

Stoeckel, D.M., and J. Harwood. 2007. Performance, design, and analysis in microbial source tracking studies. Appl. Environ. Microbiol.73: 2405-2415.

Troelstrup, Jr., N.H, B. Bleakley, and E. Jorgenson. 2005. Fecal coliform strain identification project: final project completion report. South Dakota Department of Environment and Natural Resources.

USEPA. 2007. Total Maximum Daily Loads. [online] http://www.epa.gov/owow/tmdl/overviewfs.html last accessed 26 February 2008.

Vantarakis, A., D. Venieri, G. Komninou, and M. Papapetropoulou. 2005. Differentiation of faecal *Escherichia coli* from humans and animals by multiple antibiotic resistance analysis. L. Appl. Microbiol. 42:71-77.

VDEQ. 2003. Coan River watershed total maximum daily load (TMDL) report for six shellfish areas listed due to bacteria contamination. Virginia Department of Environmental Quality.

VDEQ. 2004. Implementation plan for the fecal coliform TMDL (total maximum daily load) for Four Mile Run, Virginia. Prepared for: Virginia Department of Environmental Quality and Virginia Department of Conservation and Recreation.

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.

Zhong, X. 2004. A study of several statistical methods for classification with application to microbial source tracking. Worchester Polytechnic Institute. Master's Thesis.

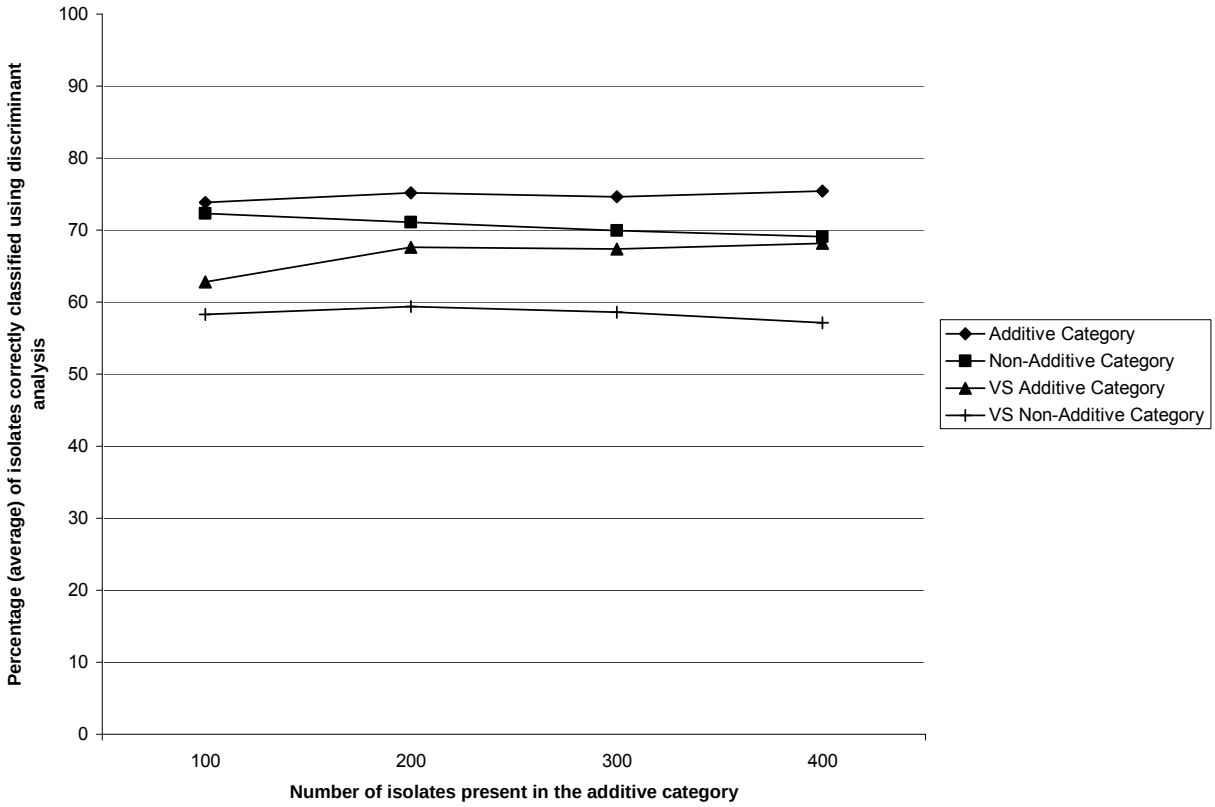


Figure 4 – Average of the rates of correct classification of library and validation set (VS) isolates from both the additive and non-additive source categories using all 20 constructed subsets of the ARA library and discriminant analysis as the additive category is increased in size.

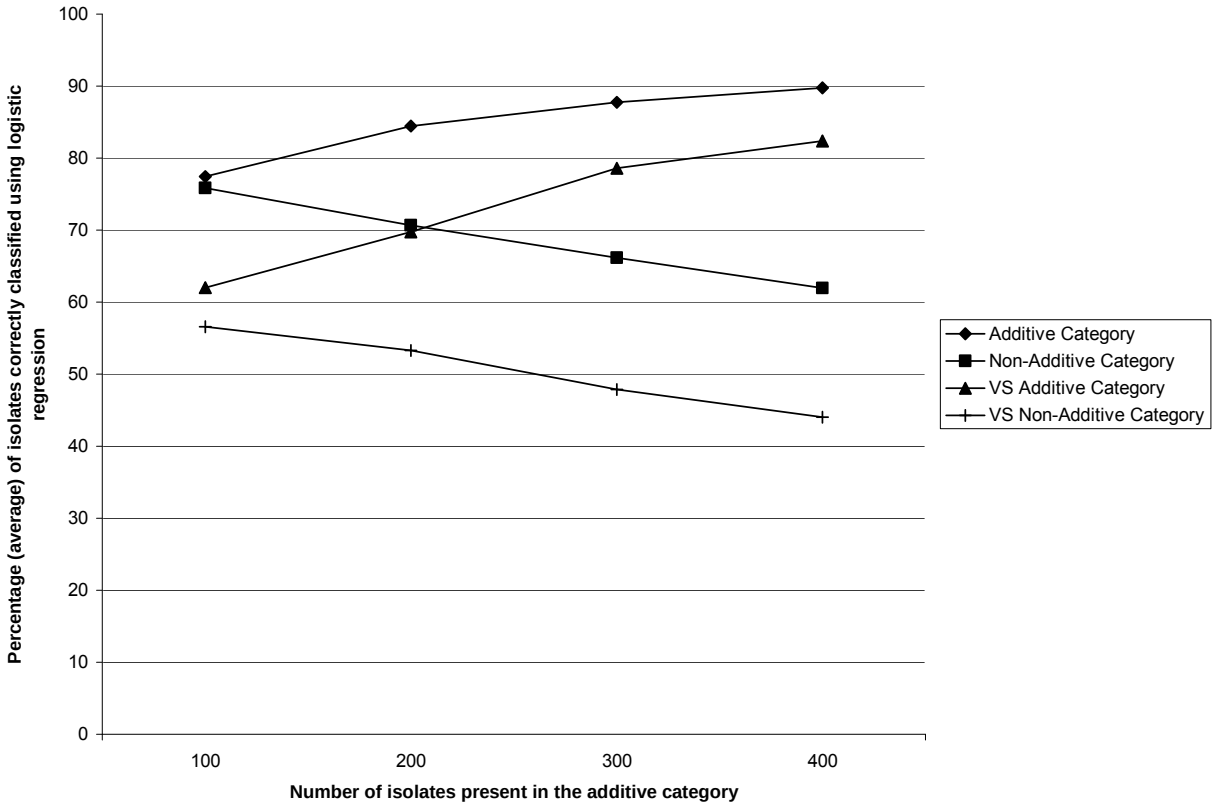


Figure 5 – Average of the rates of correct classification of library and validation set (VS) isolates from both the additive and non-additive source categories using all 20 constructed subsets of the ARA library and logistic regression as the additive category is increased in size.

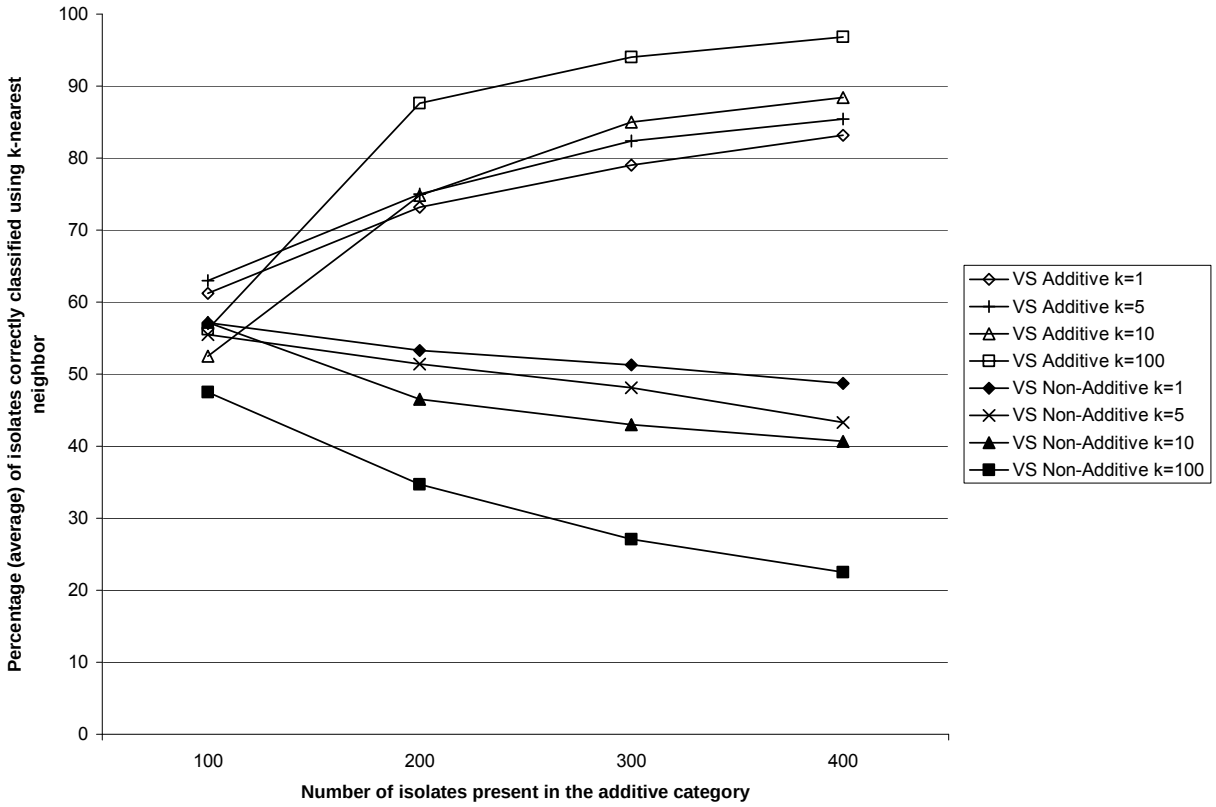


Figure 6 – Average of the rates of correct classification of validation set (VS) isolates from both the additive and non-additive source categories using all 20 constructed subsets of the ARA library and k-nearest neighbor (k values of 1, 5, 10, and 100) as the additive category is increased in size.

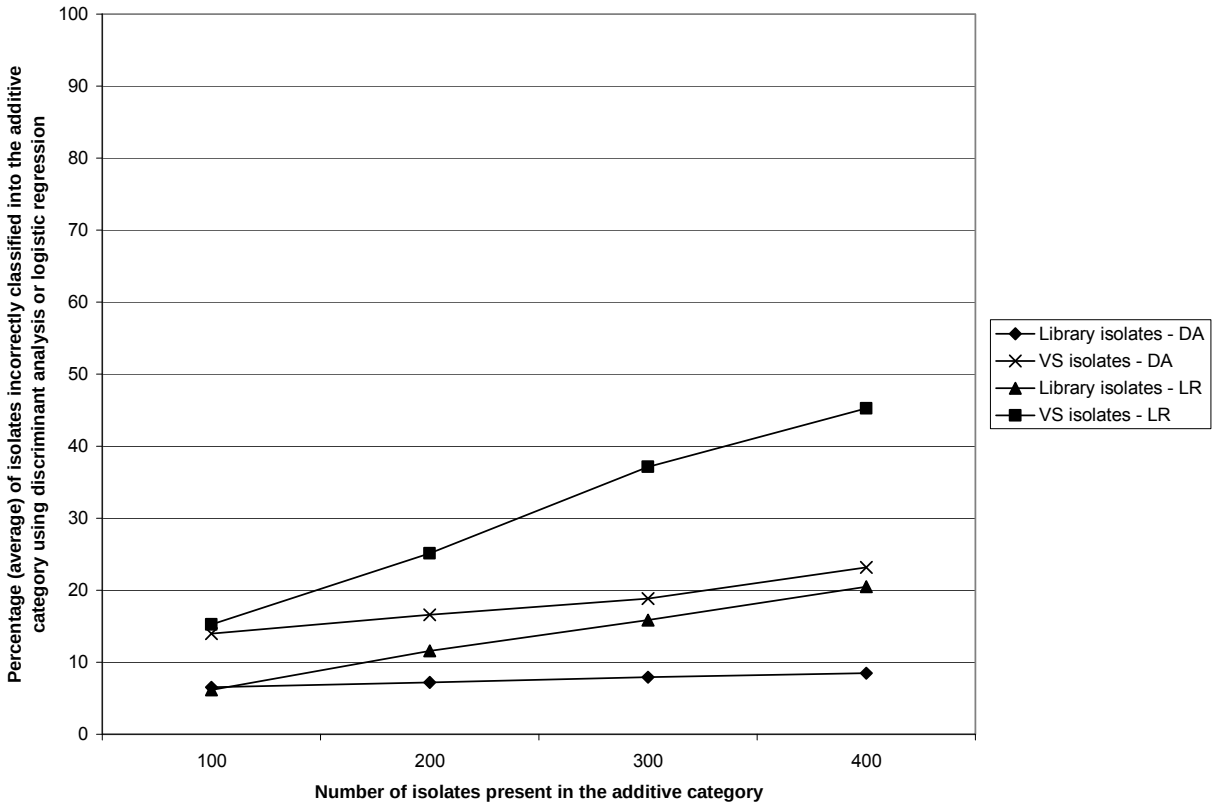


Figure 7 - Average of the percentage of non-additive category isolates from the library and validation set (VS) incorrectly classified into the additive category using discriminant analysis and logistic regression for all 20 constructed ARA library subsets as the additive category is increased in size.

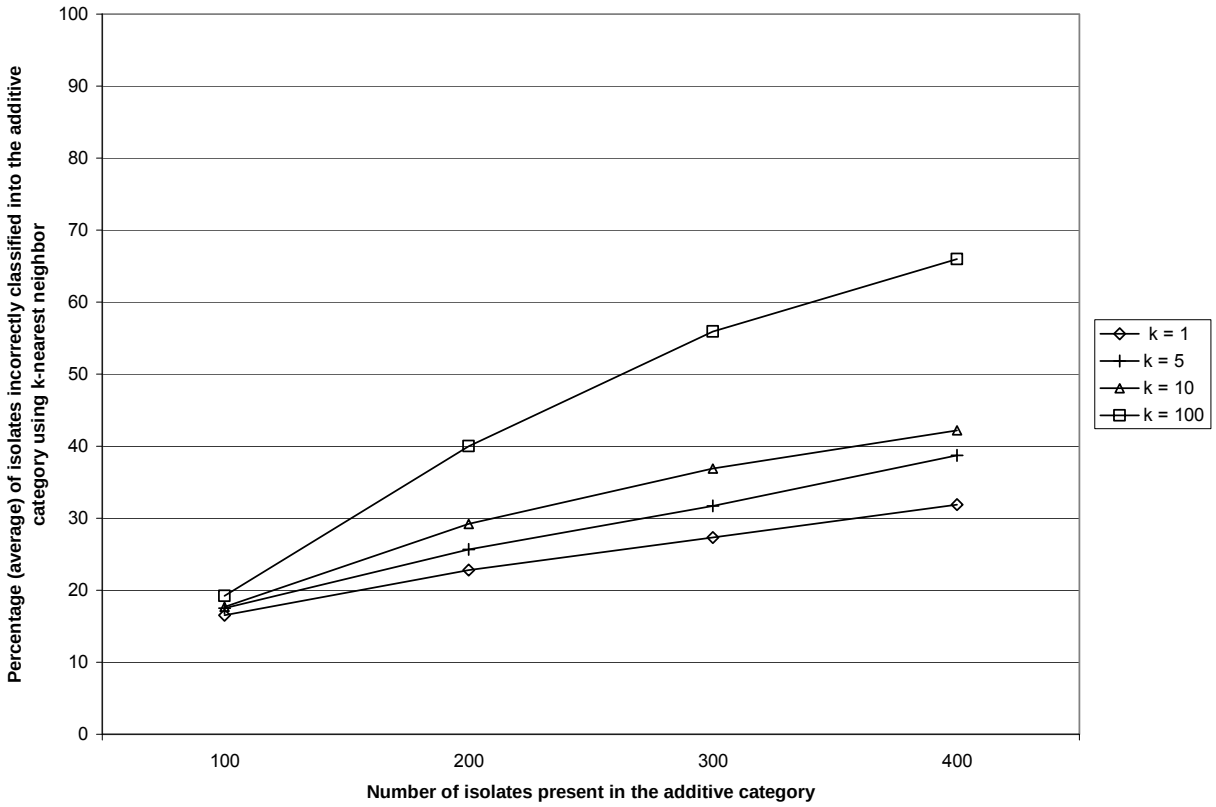


Figure 8 - Average of the percentage of non-additive category isolates from the library and validation set (VS) incorrectly classified into the additive category using k-nearest neighbor (k values of 1, 5, 10, and 100) for all 20 constructed ARA library subsets as the additive category is increased in size.

Table 50 - Classification rates of antibiotic resistance analysis library and validation set isolates using discriminant analysis.

	Percentage (number) of isolates classified as:			
Source (number of isolates)	Bird	Pets	Sewage	Wildlife
Bird (n = 690)	**68.0 (469)**	11.2 (77)	8.0 (55)	12.9 (89)
Pets (n = 491)	20.8 (102)	**68.4 (336)**	7.1 (35)	3.7 (18)
Sewage (n = 759)	8.7 (66)	13.0 (99)	**69.8 (530)**	8.4 (64)
Wildlife (n = 562)	17.4 (98)	2.1 (12)	9.4 (53)	**71.0 (399)**
Validation Bird (n = 25)	**64.0 (16)**	20.0 (5)	8.0 (2)	8.0 (2)
Validation Pets (n = 25)	24.0 (6)	**52.0 (13)**	16.0 (4)	8.0 (2)
Validation Sewage (n = 25)	4.0 (1)	16.0 (4)	**76.0 (19)**	4.0 (1)
Validation Wildlife (n = 25)	4.0 (1)	20.0 (5)	4.0 (1)	**72.0 (18)**

Table 51 - Classification rates of antibiotic resistance analysis library and validation set isolates using logistic regression.

	Percentage (number) of isolates classified as:			
Source (number of isolates)	Bird	Pets	Sewage	Wildlife
Bird (n = 690)	**69.9 (482)**	8.8 (61)	10.0 (69)	11.3 (78)
Pets (n = 491)	20.4 (100)	**64.4 (316)**	12.0 (59)	3.3 (16)
Sewage (n = 759)	10.7 (81)	7.5 (57)	**74.8 (568)**	7.0 (53)
Wildlife (n = 562)	11.2 (63)	1.1 (6)	9.1 (51)	**78.7 (442)**
Validation Bird (n = 25)	**64.0 (16)**	12.0 (3)	20.0 (5)	4.0 (1)
Validation Pets (n = 25)	28.0 (7)	**48.0 (12)**	24.0 (6)	0.0 (0)
Validation Sewage (n = 25)	4.0 (1)	8.0 (2)	**84.0 (21)**	4.0 (1)
Validation Wildlife (n = 25)	4.0 (1)	28.0 (7)	8.0 (2)	**60.0 (15)**

Table 52 - Classification rates of validation set isolates using antibiotic resistance analysis library and k-nearest neighbor (k values of 1, 5, 10 and 100).

		Percentage (number) of isolates classified as:			
		Bird	Pets	Sewage	Wildlife
k = 1	Validation Bird (n = 25)	**76.0 (19)**	8.0 (2)	12.0 (3)	4.0 (1)
	Validation Pets (n = 25)	20.0 (5)	**40.0 (1)**	24.0 (6)	12.0 (3)
	Validation Sewage (n = 25)	28.0 (7)	4.0 (1)	**64.0 (16)**	4.0 (1)
	Validation Wildlife (n = 25)	4.0 (1)	16.0 (4)	8.0 (2)	**72.0 (18)**
k = 5	Validation Bird (n = 25)	**80.0 (20)**	4.0 (1)	0.0 (0)	16.0 (4)
	Validation Pets (n = 25)	32.0 (8)	**48.0 (12)**	8.0 (2)	12.0 (3)
	Validation Sewage (n = 25)	20.0 (5)	8.0 (2)	**64.0 (16)**	8.0 (2)
	Validation Wildlife (n = 25)	8.0 (2)	16.0 (4)	8.0 (2)	**68.0 (17)**
k = 10	Validation Bird (n = 25)	**76.0 (19)**	4.0 (1)	4.0 (1)	16.0 (4)
	Validation Pets (n = 25)	32 (8)	**68.0 (17)**	0.0 (0)	0.0 (0)
	Validation Sewage (n = 25)	16.0 (4)	12.0 (3)	**60.0 (15)**	12.0 (3)
	Validation Wildlife (n = 25)	4.0 (1)	24.0 (6)	4.0 (1)	**68.0 (17)**
k = 100	Validation Bird (n = 25)	**48.0 (12)**	16.0 (4)	8.0 (2)	28.0 (7)
	Validation Pets (n = 25)	32.0 (8)	**36.0 (9)**	20.0 (5)	12.0 (3)
	Validation Sewage (n = 25)	16.0 (4)	8.0 (2)	**60.0 (15)**	16.0 (4)
	Validation Wildlife (n = 25)	0.0 (0)	24.0 (6)	4.0 (1)	**72.0 (18)**

VII. Conclusions and Advancement of the Field

This project represents one of the first attempts to apply several recommendations of the collective MST community. Library-based source tracking methods needs to undergo a dramatic shift in the way libraries are created and applied in the future to remedy numerous inconsistencies and inaccuracies. The application of a validation set, as well as a minimum detectable percentage, to known-source libraries aided significantly in accurately assessing the classification power of the library and reducing the false positive identification of contributing fecal sources. In addition, the validation of MST results using multiple methods is strongly recommended for field applications. Consensus between methods identifying fecal sources increases confidence in conclusions, and is particularly important for localities willing to allocate funds to support remediation efforts.

The use of fluorometry to detect OBs as a means of identifying human wastewater proved difficult in open bodies of water where dilution levels were high. Fluorometry did display potential for detecting sewage leaks in storm drains such as Norfolk SD-E and Hilton SD. However, interference from aromatic compounds, while usually obvious from an oily, iridescent sheen in the sample, provided false positive readings in several storm drain samples. The separation of these interfering compounds is believed to be possible by measuring the differences in fluorometeric readings of a water sample before and after the photodegradation of the OBs contained within the sample under a UV light. Results of initial tests indicated that a reduction in fluorometric readings of 30% or more after 30 minutes under a UV light tended to indicate a sample contained OBs. Continued research into UV exposure as well as application of fluorometry in smaller waterbodies with less potential for dilution should be continued in future projects.

The use of a small number of fecal isolates during the development and/or application of MST methods has been shown to be insufficient, allowing for little confidence in results obtained. Further research is needed into the diversity of genotypes and phenotypes of fecal bacteria both within a single host organism and a host population. Levels of strain diversity are undoubtedly method-dependent, however host strain variability is probably greater than previously assumed.

The application of statistical algorithms to MST libraries must be done with a cautious understanding of the varied results that are obtained from each, and the effect of source category imbalance on classification bias. A greater understanding of how algorithms for a given method/library should be selected is warranted for further studies.

In the future, MST will likely move away from library-dependent methods in exchange for the less expensive, and more rapid library-independent means of assessing and identifying fecal pollution in waterways (if such can be found). Currently, however, most library-independent methods are restricted by the undeveloped protocols for multiple potential sources, inadequate sensitivity (leading to high false-negatives), possible geographic limitations, and the inability to directly correlate with the fecal indicator bacteria used by state agencies. As testing is currently underway to select new, or verify old, organisms for use as indicators of fecal pollution, the field of MST will likely shift in a similar direction. Further advancement of the field will allow for more direct correlations with health risks through the development of methods utilizing the new fecal indicators, or focusing more strictly on the human waste believed to be the greatest risk to swimmers health.

VIII. Molecular Methods Appendix

A. DNA Purification

The purification of DNA, or isolation from other cell components, is useful for a variety of molecular biological procedures including PCR, Southern blotting, and DNA sequencing. Although numerous DNA purification kits currently exist, simplifying the purification process in both time and complexity, I will describe a more basic approach, so as to emphasis the steps involved in the process.

The first step in purification is to break open the cell, releasing the cell contents. This can be accomplished using either physical or chemical methods. Physical methods, while once popular, have largely been replaced by, or combined with, chemical methods in recently years. The act of physically breaking cells often involves large and/or expensive equipment, and can be difficult to consistently replicate. Examples of physical cell lysis methods include: use of a French press, sonication, and repeated freeze/thaw. The French press is a piston used to force a sample through tiny holes in under high pressure. This intense pressure mechanically breaks open the bacterial cells. Use of high frequency sound waves (sonication) utilizes the mechanical energy from a vibrating probe to form and burst tiny bubbles in a liquid sample, generating shock waves which rupture cells. Sonication is performed in short bursts while the sample is immersed in an ice bath to prevent excess heating. Repeated cycles of freezing and thawing causes the cells to expand and contract, ultimately breaking as a result of ice crystal formation. While simplistic in its procedure, the length of time required for freeze/thaw lysis can be significantly longer than other options. Cells solutions can also be treated with lysozyme, an enzyme which digests the polysaccharides present in the cell wall of bacteria.

Cell lysis or membrane disruption using detergents can be easier than physical alternatives. Detergents, like the lipids found in cell membranes, are both hydrophobic and hydrophilic at opposite ends of the molecule. The addition of detergents to the cell solution solubilizes membrane proteins and disrupts lipid-lipid interactions. The binding of detergents to the hydrophobic tail of the lipids destroys the phospholipid membrane structure. Non-ionic detergents (such as Triton-X) can be used to prevent the disruption of protein function, while ionic detergents (such as sodium dodecyl sulfate [SDS]) tend to denature proteins. With the addition of a detergent, a proteinase such as Proteinase K can be added to inactivate enzymes such as deoxyribonuclease (DNase) and ribonuclease (RNase) preventing the degradation of DNA and RNA, respectively.

The remaining proteins in solution can be denatured using organic solvents (causing precipitation) such as phenol or chloroform, or digested with proteases. These solvents also inactivate the proteinase, and leave the DNA and RNA in solution. The RNA is typically digested through the addition of RNase, which acts as a catalyst to hydrolyse RNA into smaller constituents.

Following removal of the proteins and RNA, a high concentration of salt may be added to allow DNA to aggregate, and the addition of an alcohol (ethanol) allows the DNA to precipitate while any remaining smaller molecules remain in solution. Most of the ethanol is removed physically, while the remainder is allowed to evaporate, leaving a pellet of DNA. The DNA pellet is resuspended in a buffer to prolonged preservation for future use.

B. Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) is an innovative method that allows for the *in vitro* production of an almost unlimited number of copies of a specific DNA sequence. The process consists of the repetitive synthesis of a specific (target) region of DNA using primer extension. Considered to be a very selective, PCR can allow for the amplification of a target sequence within a heterogeneous mixture of DNA, such as a complete bacterial genome.

This technology was first developed in 1983 by Kary Mullis at the Cetus Corporation in California. Mullis was awarded a $10,000 bonus from Cetus, and would later be award the Nobel Prize for Chemistry in 1993.

The process of PCR consists of multiple thermal cycles used to replicate DNA through denaturation (high temperature), primer binding (low temperature), and nucleotide addition by DNA polymerase (intermediate temperature). The desired region of DNA to be amplified is defined by primers which flank either side of the sequence and provide a site for DNA polymerase to begin adding complementary nucleotides. When the primer anneals the target sequence, the DNA has already been denatured at high temperatures, and is single-stranded. The process of denaturation, annealing, and elongation is done repetitively so that the amplified DNA product of one round of reaction is used as a template in the next round, thus working as a chain reaction. Since both strands of separated DNA are copied during a cycle there is an exponential increase in the number of copies of the target sequence. During the few hours required to run a complete reaction, a single DNA fragment can be amplified to a million copies. PCR generally works well with target sequences up 3000 base pairs (bp); however steps can be taken to amplify segments up to 50,000 bp in length.

The components used in a PCR reaction include:

- A sample of the target DNA sequence to be amplified. Theoretically only one molecule is necessary to begin.
- Two oligonucleotides (primers) complementary to opposite strands of DNA on either side of the target sequence
- The four deoxynucleotide triphosphates (dNTPs), dATP, dCTP, dGTP, dTTP in excess
- The DNA polymerase enzyme
- A variety of buffers and cofactors necessary for proper DNA polymerase functions (see Contents of a PCR Reaction section).

Major Steps in PCR

1. **Denaturation** (91- 97°C)

A DNA strand is held together in a double helix by the combined effects of hydrogen bonding between base pairs and van der Waals forces associated with base stacking interactions along the helix. These non-covalent forces can be disturbed and/or broken by the addition of sufficient heat.

Denaturation refers to the melting of a double-stranded DNA to generate two single strands. This occurs when the molecule of DNA is heated until the hydrogen bonds between the bases, which normally hold the double helix together, are broken. The temperature at which this occurs is known as the melting temperature or T_m (see Melting Temperature section). Within the denaturation step the T_m is generally 91-97°C for one to several minutes. The specific T_m of a piece of DNA will vary based on the length of the sequence, the composition of the bases, the geometric configuration, and the buffer composition.

The shorter the piece of DNA, the quicker and more easily the sequence will separate as there are less total bonds to be broken. Sequences rich in the A-T pairs which are connected by 2 hydrogen bonds are not as strong as sequences dominated by C-G pairs connected by 3-hydrogen bonds and thus require less energy (lower T_m) to pull the two strands apart. In addition, DNA regions containing alternating pyrimidine (cytosine and thymine) and purine (adenine and guanine) contain less stacking forces between neighboring base pairs and thus will melt more easily. The geometric structure of the DNA will also effect T_m, as closed circle or positively supercoiled DNA (more than 10.4 bases per helix) will require more energy to separate than will DNA that is linear, or heavily nicked (a nick is a break in the phosphodiester linkages along one of the strands). Salt or ion concentrations can alter T_m by decreasing the electrostatic repulsion present on a DNA molecule between negatively charged phosphates on the backbone. If the composition of the buffer contains a low concentration of salt or ions, the DNA molecule remains more negatively charged, which increases repulsion between strands requiring a lower T_m to denature the strands. When higher concentrations are present the charges tend to be neutralized and thus the energy required to denature the strands is increased. In general, at ionic strength lower than 150mM NaCl, denaturation can occur at temperatures below 100°C. Once the DNA has been melted, the temperature of the reaction is lowered to allow for the annealing of primers. If the mixture is left to cool too long it is possible for the DNA to renature as the complementary base pairs rejoin restoring the DNA to its double helix form.

Following each round of synthesis, the reaction contents must be heated again to denature the double-stranded DNA product.

2. **Annealing** (55-72°C)

When synthesizing a complementary strand of DNA, polymerase can only extend from an existing fragment. It cannot begin adding complementary base pairs to a fragment without the 3' end of piece of DNA on which to attach. This 3' end is provided by short pieces of DNA known as primers.

Primers

In the annealing step, the primers bind to the single-stranded DNA created in the denaturing step. Primers are short pieces of single stranded DNA (oligonucleotides) that are complementary to the starting point on the template of DNA for which amplification is desired. Prior knowledge of a target sequence is required to design the appropriate primers, typically approximately 18-26 nucleotides in length. The complementary nature of the primer to the template allows for the creation of a duplex due to the hydrogen bond formation between the nucleotide bases of the target DNA and the primers. Within the annealing step, bonds are continuously formed between the template DNA and the complementary primers; however greater stability is established between complementary sequences which more closely fit, with the most stable bonds forming between the exactly matching primer and DNA template strand. How specific amplification is depends on the specificity of the primers for the target DNA sequences relative to other sequences that may be present. In general, primers ranging from about 18-26 nucleotides long and are considered to be specific enough that the chance of another identical match elsewhere within the genome is considered very low. However, multiple amplification products can occur if primers contain part of a repetitive DNA sequence, as is the case with the popular rep-PCR microbial source tracking (MST) method.

Primer Design

Numerous computer programs exist to aid in the development of working, species-specific primers, many of which are available at no cost online. The first step to designing a primer is a determination of which organism and region of the genome one wishes to amplify. With the completed genome sequences of many organisms (and often multiple strains) available through GenBank, potential primer sequences can be located online. The selection of primers needs to account for several features/potential problems:

1) The PCR product is typically selected between 100-1000 base pairs in length.
2) Primers are typically 18-26 base pairs in length.
3) The C+G content of the primers are generally 40-60%.
4) Melting temperature (T_m) differences between primers should not be greater than 5ºC (can be calculated using the formula on page 9 or using numerous available online programs).
5) Use of T_m will allow for the calculation of an appropriate annealing temperature (T_a) for the primers.
6) The 3' end of the primer sequence should end in a C or G to increase stability (due to the stronger bonding of C-G bases).
7) Secondary structures such as hairpin loops, and primer dimer formation should be avoided (again, online programs are available to determine the potential for formation).
8) Species-specificity can be evaluated using the BLAST program available in GenBank to determine the homology of the primers with sequences in other species and sequences within the genome of the selected species.

With all of these factors taken into consideration, primers can be ordered from numerous biological supply companies found online.

Potential Primer Problems

Secondary Structure

Secondary structures are formed by intramolecular bonding, and can decrease PCR efficiency leading to little or no PCR product. Secondary structures such as hairpins are thus formed by the interaction of the primer with itself. A section of the primer curls back onto itself, creating hydrogen bonds between the base pairs and a partial DNA molecule that resembles a hairpin. This hairpin structure prevents or limits primer annealing to the DNA template.

Primer Dimers

Primer dimers occur because of the hybridization between two primers when part (or all) of one primer is complementary to the other (intermolecular bonding). When the primers anneal to one another rather than the DNA template they can significantly reduce the amount of primer available for PCR and therefore reduce product yield. In addition, the hybrid can form a substrate for DNA polymerase, resulting in the amplification of the primer plus additional nucleotides. If one primer is elongated by additional nucleotides, a new primer is created, which may increase the amplification of nonspecific, undesired sections of DNA.

Primer Annealing Conditions

A correct annealing temperature (T_a) is extremely important for correct PCR amplification, and varies based on the length, concentration, and base composition of the primers. The optimum primer annealing temperature is approximately 5°C lower than the T_m of

the primers ($T_a = T_m - 5°C$), and varies from 55-72°C depending on the set of primers. The primer's annealing temperatures can be calculated based on the T_m of the primers which uses the formula:

$$T_m = 4(G+C) + 2(A+T)\ °C$$

The T_m for a primer depends on the length of the primers and base pair composition where greater influence is given to the G-C content than the T-A due the increased strength of the triple bonded G-C pairs.

Using a lower T_a decreases the specificity of the primer for the desired sequence which can anneal to incorrect sequences and decreases the production of non-target DNA fragments caused by incorrect annealing. However, use of a higher T_a reduces the chance of the primers annealing resulting in the decreased production of target fragments to point at which appreciable yield are never obtained. The selection of primers should also be within 5°C of one another, or results may yield an asymmetric amplification where the most efficient primer at the selected T_a dominates the reaction.

The annealing time can range from about 30-45 seconds in a given PCR cycle unless T_a is too high, or primers are usually long (which increases required time). In general, longer annealing times will not significantly increase the amplification of a PCR, and may increase the chance of obtaining nonspecific amplification products as DNA polymerase may have a reduced activity at these temperatures.

Melting temperature

The strength of the primer-template DNA duplex in a PCR is measured by the melting temperature (T_m). The T_m is the temperatures at which one half of the DNA duplexes will dissociate and become single-stranded. Therefore, the annealing temperature (T_a) in the reaction must be set below the T_m to maximize primer annealing to the DNA template. For this reason T_m can be used to calculate a T_a for a set of primers. The two primers of a primer pair should have closely matched melting temperatures for maximized PCR product yield. The melting temperature of a set of primers should be closely matched to assure maximum template (amplicon) production. If the difference in T_m between the two primers is too large, less efficient or no amplification may result. The melting temperatures of primers are most accurately calculated using the nearest neighbor calculation, available on numerous websites.

3. **Extension** (72°C)

Once the primers have annealed to sections flanking the target sequence, the temperature is raised and DNA polymerase will attach to the 3'-hydroxyl end of the primer sequence and begin to add complementary nucleotides along the target strand.

The DNA polymerase from the original reaction, as created by Kary Mullis, used a DNA polymerase from *Escherichia coli*. When the reaction was heated in the denaturing step the DNA polymerase was permanently inactivated, therefore a fresh aliquot of these enzymes had to be added by manually at the beginning of each cycle. Fortunately, a thermostable polymerase enzyme was later isolated in the lab of Henry Erlich, from a microbe isolated from a hot spring. This DNA polymerase would remain stable at the high temperatures (~95°C) required to denature DNA and thus did not have to be added at the start of each cycle. This DNA

polymerase was isolated from the organism *Thermus aquaticus* which lived and reproduced in the extremely high temperatures of a hot spring. Currently the most widely used polymerase in PCR methods; the enzyme was named *Taq* polymerase after *Thermus aquaticus*. *Taq* polymerase functions at an optimal temperature of ~72°C, which was higher than that of the *E. coli* polymerase, so the reaction could proceed at a higher temperature allowing for the increased stringency of primer binding and a more accurate amplification, containing fewer non-specific products than had been possible in the past.

Similar to DNA replication within a cell, the primers provide a 3'-hydroxyl end for DNA polymerase to begin adding nucleotides in the 5'→ 3' direction. However, unlike DNA replication within the cell, the *Taq* polymerase does not possess a 3' → 5' exonuclease to allow polymerase to remove mismatched bases during replication. In the cell (although fairly rare), DNA polymerases can attach an incorrect base to the template DNA during replication. Usually these mismatched bases are recognized by the polymerase immediately due to distortion in the helix from incorrect base pairing. The 3' → 5' exonuclease allows the DNA polymerase to back up, removing the mismatched base, and as polymerization begins again the correct base is inserted. The absence of this exonuclease means that the synthesized amplicon has an error rate that is higher than *in vivo* copying and the final products are usually very similar, but not necessarily identical sequences of DNA. Currently, proofreading or high fidelity DNA polymerases are available at increased cost if higher accuracy is required. These enzymes work with a very low rate of nucleotide misincorporation. The first of these enzymes discovered was *Pfu* polymerase, from the thermophile *Pyrococcus furiosus.* Errors only occur about every 767,000 nucleotides using *Pfu* polymerase as compared to every 125,000 nucleotides with *Taq* polymerase. Heat-stable, like *Taq* polymerase, *Pfu* polymerase can tolerate temperatures

exceeding 95°C, making it particularly useful for the higher denaturing temperature necessary to amplify GC-rich targets.

The primers are extended by DNA polymerase through the addition of the complementary nucleotides to the single-stranded DNA, converting it into a double-strand. The newly synthesized DNA strands serve as templates for additional DNA synthesis in subsequent cycles. As the first cycle of synthesis begins each half of the denatured DNA has a single primer attached at the beginning or end of the target section. The DNA polymerase begins to add complementary nucleotides at the 3' end of the primer on each strand, with the addition continuing through the target sequence and beyond the area identical to the other primer, producing a long strain of DNA. The same process occurs on the other half of the original strand as well. As the second cycle begins the DNA are denatured again, and primers bind to the complementary regions in now separate original and long strands. More long strands are produced from the original templates. As the primers hybridize with the long strands, DNA polymerase adds nucleotide in the opposite direction in which the long strand was produced (still 5' → 3') running out of nucleotides at the end of the sequence complementary to the other primer. This produces a short strand containing only the DNA of the target sequence and the two primers. This is repeated over subsequent cycles, during which the short sequences accumulate at the much greater rate becoming about a million times more abundant by the 30^{th} cycles. Thus only the target sequence will be seen when visualizing the PCR product. Typically only about 30 amplification cycles are run per PCR reaction because *Taq* polymerase has a half life at 95 °C of only about 30 minutes. After all cycles have been completed, the amplicon(s) are often identified using electrical charges to separate products based on size differences.

Contents of a PCR Reaction

The contents for a PCR reaction generally begin with:

- ✓ **< 1 µg of template DNA** - DNA template concentration depends on the source
- ✓ **0.2-1.0µM each primer** –Primer concentrations which are too high increase the chance of nonspecific products due to mispriming. Low primer concentrations result in inefficient product yield.
- ✓ **50 - 200µM each dNTP -** Equal concentrations of each dNTP is important, as uneven concentrations may drastically increase the misincorporation level. Concentrations of 50uM are usually sufficient, however especially long templates may require more.
- ✓ **1-1.5µl of *Taq* DNA polymerase per 50µl of reaction mix** - In general, higher *Taq* polymerase concentrations may yield nonspecific products. However, higher amounts of *Taq* Polymerase (2-3µl) may be necessary to obtain a better yield if inhibitors, such as unpurified template, are present.

Polymerases are often supplied with a buffer for optimizing the reaction. A general polymerase buffer contains:

- ✓ **10-50mM Tris-HCl** – A pH buffer, which ranges from pH 7.5-9.0.
- ✓ **1-5mM $MgCl_2$** – Mg^{++} forms complexes with dNTPs, primers, and DNA templates, thus the appropriate concentration of $MgCl_2$ will vary for each experiment. An Mg^{++} concentration that is too low will result in a low product yield, whereas too high a concentration can increase non-specific products. In reactions containing chelators, such as EDTA, $MgCl_2$ concentration should be increases proportionally. The $MgCl_2$ concentration should be selected empirically, beginning with 1mM concentrations should be increasing in 0.1mM steps, until a sufficient product yield is obtained.

- ✓ **< 50mM KCl or NaCl –** Although necessary to facilitate primer annealing, higher than 50mM KCl or NaCl inhibits *Taq* polymerase.

C. Restriction Fragment Length Polymorphisms

Restriction fragment length polymorphisms (RFLPs) are variations in the size of DNA fragments seen between individuals when DNA is digested (cut) using one of more restriction enzymes. This cut is made in relation to a specific DNA sequence known as a recognition site. Often the presence or absence of a fragment allows for distinguishing between individuals of the same or closely-related species. Although extremely similar, individuals of the same species do not possess identical genomes. Variation in a few nucleotides will always be present due to minor base additions, deletions, or substitutions. If these differences produce new or destroy old recognition sites, the cuts made by the restriction enzymes will change, and different fragments will be produced.

A restriction enzyme is an endonuclease that recognizes and breaks a specific target nucleotide sequence in a DNA molecule. This break is caused by the cleaving (hydrolyzing) of the phosphodiester bonds that make up the phosphate sugar backbone of DNA. Discovered in 1968 at Johns Hopkins University, restriction enzymes were recognized for their ability to cut DNA into fragments outside of the bacteria in which they normally reside. Within bacteria, these enzymes serve a protective function by chopping up foreign DNA molecules, particularly bacteriophages (rendering them inactive), which enter the cell. Approximately 400 different restriction enzymes are commercially available today, covering roughly 100 recognition sites (although several thousand enzymes have been described in the literature).

Most restriction enzymes bind as dimers, meaning they sit on both strands of DNA simultaneously. These enzymes recognize a particular sequence within a DNA molecule and cut

on both sides of the DNA within or close to that sequence. Both specific (recognition site) and non-specific (at any position) binding may occur between the DNA molecule and the enzyme. Some restriction endonucleases bind non-specifically to the DNA and diffuse along the linear segment at rates on the order of 10^6 bp/s towards their recognition sites. Water fills the spaces between the enzyme and the DNA during this process known as linear diffusion. Once the recognition target is found, an enzyme-DNA complex is formed. This complex results in a conformational change and removal of the water at the enzyme-DNA interface to establish closer contact. About 15-20 hydrogen bonds between the enzyme and DNA hold the enzyme in a position close to the site of cleavage.

The recognition sequence is usually 4-6 base pairs in length (although a few are 8) and is palindromic, meaning the nucleotides within the site are identical on both side of the strand when read the 5' → 3' direction. However, numerous non-palindromic enzymes are known as well. Enzymes which share the same recognition sequence as known as isoschizomers. Differences often exist between isoschizomers, and some cut at different locations on a recognition site. The isoschizomers *Mbo*I and *Sau*3A differ in their ability to cut methylated DNA strands, as the former cannot while the later can. Some endonucleases also recognize discontinuous palindromes, that is, sites interrupted by stretch of additional DNA containing a specified length but no specified sequence.

The location of the cuts within the recognition sequence depends on the restriction enzyme. These can by symmetrical or asymmetrical within the recognition sequence, leaving 3', 5', or no overhanging base pairs. Cuts made symmetrically on exactly opposite sides of the DNA leave no unpaired bases in what are known as blunt ends. If the restriction enzyme cuts at staggered sites along the length of the DNA, this will leave an unpaired length of bases known as

sticky ends. Asymmetrical cuts such as these will leave fragments only held together by the weak hydrogen bonds between complementary base pairs. These hydrogen bonds are not strong enough to hold the DNA together and will quickly separate into two fragments. The sticky ends from one fragment have the potential to bind to a complementary sequence of bases on another fragment or at the opposite end of at same fragment if DNA ligase (see description below) is added to the mixture to join the backbones of the molecules. After the cuts are made, the enzyme will release from the complex, leaving the cleaved DNA molecule.

The processes shared by most restriction enzymes are: the binding site recognition, the DNA binding of the enzyme, and the DNA cleavage and the enzyme release. Several types of restriction enzymes exist, differing in how and where they cleave a DNA molecule in relation to its recognition site. Type II restriction endonucleases have found the most application in molecular biology for there ability to cleave DNA close to or within their specific recognition sequences. Type I and III restriction enzymes have less commercial application because they cut outside of their recognition sequence at sites a dozen or several thousand base pairs away. Type II restriction enzymes are small in size ranging from 200-350 amino acids in length with molecular weights between 20,000 and 100,000 daltons and require Mg^{++} as a cofactor for DNA cleavage.

Restriction enzymes, like all enzymes, only function properly under a specific set of conditions. Most work best at 37°C however, some require an incubation temperature close to 25°C due to a short half-life at 37°C. When enzymes are used that have been isolated from thermophilic bacteria, optimal function is typically between 50 and 65°C. Restriction enzymes are degraded easily and are typically stored in glycerol at -20°C when not in use.

The mixture used in a digest generally contains: the enzyme, the DNA to be cut, and a specific buffer mix. Within the buffer mix, a pH buffering agent is often accompanied by a salt, and $MgCl_2$. The buffering agent, usually Tris, is present to maintain a proper pH, usually around 8.0. Poor cutting rates can result from incorrect pH, and some enzymes are more sensitive to pH changes than others. Salts are also necessary to establish the correct ionic strength in a mixture, and usually present as NaCl or KCl although variations in optimal ionic strength may be required for different enzymes. Finally, $MgCl_2$ is present in the mixture to provide the Mg^{++} that is required as a cofactor for the activity of the enzyme. Sometimes the addition of bovine serum albumin (BSA) into the reaction mix can have a positive influence on enzyme activity due to its ability to stabilize the enzyme, bind impurities, prevent the adsorption of the enzyme onto surface of the tube, and protect the enzyme from proteases. Establishment of an improper mixture can result in a loss of fidelity or increase in cleavage at sites similar to the recognition site, in a phenomenon known as star activity.

The endonuclease used in pulsed-field gel electrophoresis by some source-tracking researchers is *Not*I. This enzyme recognizes, and cuts within the following sequence at the ^ marks.

5'-G C^G G C C G C-3'

3'-C G C C G G^C G-5'

An optimum temperature of 37°C is used, in a buffer containing the following: 50mM Tris-HCl (pH 7.5), 10mM MgCl2, 100mM NaCl, 0.1mg/ml BSA. A four to six hour incubation period is required for proper digestion.

After DNA has been digested, electrophoresis is commonly used for separation and visualization of the DNA fragments.

DNA ligase

DNA ligase is an enzyme that catalyzes the repair of a broken phosphodiester bond such as a nick in single-stranded, or a break in double-stranded DNA. When complementary sticky ends are present, DNA ligase can create covalent (phosphodiester) bonds between the 3'-hydroxyl and the 5'-phosphate ends of the two fragments. DNA ligase, in prokaryotes, uses NAD+ as a cofactor (ATP eukaryotes, viruses and archaebacteria), where in the absence of NAD+ bound to the active site of the ligase, recognition of the DNA nick does not occur.

D. Gel Electrophoresis

Once DNA has been amplified and/or cut into fragments using one or several restriction enzymes, fragments may be separated via gel electrophoresis. Electrophoresis separates DNA fragments based on size differences when run thorough a gel inside an electrophoresis chamber. The principle behind separation is based on the use of an electric current, and a gel (or molecule sieve). As an electric current is applied, DNA moves through the gel towards the positive electrode (due to the negatively charge phosphate backbone that makes up DNA molecules). The gel acts as a sieve, allowing small molecules to move more quickly towards the positive electrode and the larger molecules, restricted by the friction of the gel itself, move more slowly.

Two types of gels are commonly used for electrophoresis: agarose and polyacrylamide. Agarose gels are generally used for larger-sized molecules and offer poorer resolution than polyacrylamide gels. Although higher resolution is an important

feature of polyacrylamide gels, acrylamide is a neurotoxin and requires special care be taken by laboratory personnel. In general, polyacrylamide gels are used in protein analyses, but are not suitable for nucleic acids larger than 200,000 daltons due to the smaller pore sizes within the gel. Although pore size can be adjusted in a gel by decreasing the concentration of agarose or acrylamide, for polyacrylamide to be able to separate large molecules the concentration would need to be reduced to such a level that a gel would be unable to form (remain as a liquid). Regardless of the gel type selected, gel solutions (gel powder and buffer solution) are boiled until a clear liquid is obtained. While still hot, the liquid is poured into a thin, horizontal mold. A gel comb is placed at one end of the gel mold to create small wells, which partially penetrate the gel and allow for the DNA to be loaded into the gel. Once the gel has cooled and become rigid, the comb is removed.

Once rigid, the gel is placed into an electrophoresis chamber, and immersed in a buffer solution (such as TAE). The buffer solution conducts electricity between the two oppositely charged electrodes at a selected voltage. Selected DNA samples are loaded into the wells at the end of the gel closed to the negative electrode, with a loading dye. The dye increases the density of the DNA and prevents it from floating out of the top of the well, and allows the progress of DNA movement thorough the gel to be visually monitored. The loading dye however, does not allow for the visualization of individual bands. In addition, DNA containing fragments of known bp length (usually in round number such as 100 bp lengths), known as a ladder, are added to several of the wells to allow for interpretation of band lengths of the "unknown" DNA fragments. Once all DNA is loaded into the wells, an electric current is generated and the DNA molecules move through the gel. When the DNA and loading dye have

moved the desired amount through the gel the electric current is removed and the gel is taken to stain.

Staining of the gel allows visualization of the individual bands, which is used to determine fragment lengths. Most of the dyes used to stain DNA, such as ethidium bromide or SYBR green intercalate (binds to the DNA in between adjacent base pairs) the DNA sequence and the bands become visible under ultraviolet (UV) light. Some stains, such as methylene blue do not require UV light for visualization and are believed to form ionic bonds with the phosphate groups on the outside of the DNA. However, these dyes tend to not be as sensitive, will often heavily color the gels, require extensive destaining (ethidium bromide requires destaining, SYBR green does not), and the coloring can fade rapidly. Finally, the sizes of DNA fragments are determined by a comparison to the bands in the lane or lanes containing the ladder sequence, using gel imaging software such as Gel-Pro™. Some of the dyes not requiring UV light for visualization are believed to form ionic bonds with the phosphate groups on the outside of the DNA.

E. Terminal restriction fragment length polymorphism/ Length heterogeneity polymerase chain reaction

Both terminal restriction fragment length polymorphism (T-RFLP) analysis and length heterogeneity polymerase chain reaction (LH-PCR) are methods frequently utilized to assess structure of a microbial community based on the patterns of DNA amplified from a sample (soil, sewage sludge, fecal material, aquatic environments, etc.). While T-RFLP identifies variations in PCR fragment lengths based on variability in restriction sites, LH-PCR utilizes variability in the length of 16S rDNA sequences.

T-RFLP allows for the analysis of polymorphisms within a certain gene, or amplified marker, in a process involving PCR amplification, restriction enzyme digestion, and gel electrophoresis. Once extracted from a sample, DNA is amplified using primers targeting conserved sequences within a target gene. One of the primers is labeled with a fluorescent molecule on the 5'-end (selection of which primer depends on the sequence amplified, and research objectives). The amplicons are digested with one or multiple restriction enzymes, typically ones with a four base recognition sequence. The DNA fragments are separated on a gel, and a laser reader detects the fragments generating a profile of the community based on the band lengths. Use of fluorescently labeled primers limits analysis to the fragments representing the terminal portions of the amplicon. T-RFLP analysis has the advantages of high throughput, highly reproducibility, and being partially quantitative in examining the diversity of a particular gene

Length heterogeneity polymerase chain reaction (LH-PCR) amplifies regions within the 16S rDNA genes, considered variable. DNA purified from a sample is amplified by PCR using a fluorescently labeled primer (usually the forward primer). PCR products are separated on a polyacrylamide gel and measurements of fluorescent bands are converted to electropherograms (peaks representing band lengths, in base pairs). The intensities of the peaks from each size class are considered proportional to the original concentrations of template DNA. LH-PCR has the advantage of being relative simple, rapid and is considered very reproducible. However, application is limited to the bacterial species for which sequences are available.

F. Pulsed-Field Gel Electrophoresis

Pulsed-Field Gel Electrophoresis (PFGE) is a molecular technique used to separate very large fragments of DNA in a flat, horizontal gel using alternating electrical field angles. In normal gel electrophoresis DNA fragments up to about 50kb in size are separated using a continuous, unidirectional electrical field. Fragments larger than 50kb will not separate, and are seen as a large, single, diffuse band. However, by alternating the direction of an electrical current, the DNA is forced to reorient and change direction as it migrates through the gel matrix allowing for the separation and resolution of fragments over 10,000 kb. When the direction of the electric field is altered the DNA are forced to move in a different direction (towards the positive electrode). When this occurs, the DNA of smaller size will reorient and move faster than the larger DNA, thus separating the larger and smaller DNA. In most applications, the current is altered between 96 and 120° (reorientation angles) forcing the DNA to move through the gel in a zigzag pattern, but all the lanes and the bands will appear straight when complete.

Sample Preparation

Preparation for PFGE begins with the growth of a pure cell culture. A loopful of culture is collected and suspended in a buffer. The cells and buffer are embedded in agarose plugs and mixed with a lysozyme and protease to disrupt the cell membrane releasing the chromosomal DNA from the cells and digest the proteins. Due to the fact that long DNA strains can physically break if solutions are pipetted, and the only desirable breakage is controlled with restriction enzymes, immobilization in agarose plugs is necessary. After the plugs are washed to remove cell debris, they are cut to the necessary size, and treated with a selected rare-cutting restriction

enzyme (tends to recognize 8 bp or larger sequences) and incubated for a specified amount of time. An agarose gel is poured around the plugs, and the gel is run using varying combinations of field strength, pulse time, reorientation angles, etc. specific to the size of bands the researcher is attempting to resolve. The gel can be run in times varying from hours to days, and is typically stained using ethidium bromide to visualize the bands.

Variables of PFGE protocols

Pulse (or Switch) Time

Pulse time is the amount of time in which a continuous electric current is applied to the gel, before it is altered to a different angle. Pulse time primarily alters the separation of DNA based on size. Due to the fact that larger molecules take longer to reorient and change direction once the field angle has been switched, larger molecules migrate more slowly through the gel. Thus, extended pulse times are used to separate larger DNA molecules.

Reorientation Angle

As a general rule, wider reorientation angles produces better resolution of bands, while smaller angles cause the DNA to move faster through the gel. However, while angles of 90° provide little to no separation, angles between 105° and 165° have not been found to significantly effect separations.

Temperature

In PFGE, gels are run at temperatures between 4°C and 15°C, in contrast to the room temperature conditions usually used for conventional electrophoresis. While gels can be run at

room temperature, the elevated temperature increases the mobility of the DNA through the gel diminishing resolution. As a result of the heat generated by voltage gradients, the buffer is circulated through a heat exchanger to keep the temperature low.

Agarose Concentration

The concentration of agarose in a gel affects the speed at which DNA molecules migrate. The tendency of higher agarose concentrations to slow DNA movement is more pronounced in larger than smaller molecules. Resolution tends to be higher with greater agarose concentrations, but higher concentrations limit usage to smaller molecules and tend to decrease the width of individual bands. Typical agarose concentrations of 0.8-1.2% represent a compromise between resolution and a reasonable separation time.

Voltage/Field Strength

The voltage used in PFGE is varied based on the size of DNA molecule to be separated, and selection often represents a compromise between the required separation time, and desired resolution. Voltage gradients of 6-10 V/cm are effective for molecules as large as 1000 kb, however, at these voltages larger sized molecules can become immobile. Thus larger DNA molecules require lower voltage (2 V/cm or less) gradients and longer switch intervals. Lower voltages tend to produce better separation of bands, but are only useful for a limited range of sizes.

Buffers

Buffers are usually used at a low ionic strength to reduce the heat generated, and increase DNA mobility. Using TAE or TBE at 0.25 or 0.5x is preferred, as dilutions below 0.25x can decrease mobility.

PFGE Variations

The equipment required for PFGE includes a gel box, a high voltage power supply, a switch unit, computer program (to control the switch interval), and a cooler (to control buffer temperature). Several variations of PFGE exist, distinguishable primary by the equipment necessary, and the manner in which the electrical fields are alternated. The simplest method and the one requiring the least specialized equipment is field inversion gel electrophoresis (FIGE).

In FIGE, the polarity is simply reversed 180° at regular intervals. This reversal forces the DNA to alternate between forward and backward movements (no zigzagging), with forward intervals longer than the reverse (about 3 to 1), generating a net forward movement. Field inversion allows for resolution of DNA molecules between 10-2000 kb in size, and is popular for separation of smaller-sized fragments. FIGE can be performed in a conventional electrophoresis chamber with the addition of a device allowing for periodic electrical field reversal, and offers researchers the advantage of straight lanes with relatively simple equipment.

Another PFGE variation is rotating gel electrophoresis (RGE) in which the electric field remains constant, and the gel is rotated within the chamber to achieve the desired reorientation angles. Rotating the gel in a uniform electric field allows for the production of straight bands, as only a single set of electrodes are employed. This method allows for researchers to easily

experiment with varying reorientation angles, by varying the rotation angles of the gel. Molecules up to 6000 kb can be resolved using this method.

In transverse-alternation field gel electrophoresis (TAFE) the gel is vertically oriented with four electrodes placed outside the plane of the gel. This arrangement forces the DNA to zigzag across the thickness of the gel. In this set up all bands remain straight as they are subjected to a net force pulling them down through the gel. However, the movement further down the gel increases variations in the strength of the field and the angles of reorientation, although this is consistent for all lanes. To date, TAFE has been used for the separation of DNA fragments as large as 1600 kb in size.

Contour-clamped homogeneous electric fields (CHEF) uses 24 electrodes spaced evenly around the edges of a hexagonal set-up, to create reorientation angles of 120°. In the CHEF system, all electrodes are active simultaneously. The voltages at each of these 24 points are set by the power supply generating uniform, alternating electric fields. The uniform nature of the field allows for consistently and straight running gels for separating molecules up to 7000 kb in size.

G. Ribotyping

Ribotyping uses the ribosomal RNA (rRNA), composed of three units (5S, 16S, and 23S), and present in all bacteria to distinguish between bacterial strains. The DNA present in bacteria coding for the rRNA is highly conserved, and present in multiple copies. Isolated genomic DNA is cut using a restriction enzyme and separated on an agarose gel using electrophoresis. Bands will usually appear as a smear, as opposed to the clearly defined bands of other methods, due to the large number of fragments present. Prior to transfer to a nylon

membrane the DNA is denatured in the gel using NaOH separating bands into single stranded DNA.

The DNA is then transferred from the gel onto a nylon membrane. The membrane is placed either on top of, or underneath the gel and an even pressure is applied to the gel to facilitate transfer of the DNA onto the membrane. The pressure is generated by either vacuum suction or placing a weight on a stack of papers towels on top of the gel and membrane. Transfer of buffer and DNA from areas of high water potential (the gel) to areas of low water potential (the membrane or paper towels) via capillary action or suction pressure moves the DNA from the gel to the membrane in a process taking one (vacuum) to several hours. The negatively charged DNA binds to the positive charge present on the membrane. The DNA fragments retain the same separation patterns on the membrane that they had on the gel.

After DNA transfer, the membrane is then exposed UV light or a high temperature to permanently bind the DNA to the membrane through the formation of covalent bonds. The membrane is treated with chemiluminescent hybridization probes complementary to 5S, 16S and 23S portions of the rDNA. These probes will hybridize to the single-stranded DNA (ssDNA) present on the membrane with complementary to the strands forming double-stranded DNA (dsDNA). The excess probe is washed from the membrane following hybridization leaving only the probes which have annealed to complementary sequences present on the membrane. Visualization of the attached probes, by chemiluminescent methods or X-ray film, indicates that a specific DNA fragment, complementary to the probe, is present on the membrane. The banding patterns produced are believed to be highly specific for strain or species differentiations.

H. Real-time Polymerase Chain Reaction

Real-time polymerase chain reaction, also known as quantitative PCR (q-PCR) provides researchers with a means to both amplify and quantify a DNA sequence within a single reaction. q-PCR is based on the same overall concept as PCR, requiring: forward and reverse primers, *Taq* polymerase, free nucleotides, buffer, and ability to regulate increases and decreases in temperature. However, unlike traditional PCR, q-PCR does not require the additional step of electrophoresis to visually PCR products. Researchers have employed a variety of q-PCR variations to detect the presence of a specific DNA or RNA sequence (typically about 100 bp in size, but may be as much as 600 bp), or to quantify the number of sequence copies present. While the procedure follows the same principle as a general PCR, q-PCR allows for the quantification of DNA sequences after each round of amplification. As the reaction is run in a thermocycler, q-PCR employs fluorescent dyes to allow for the detection of amplified DNA sequences. After each round of amplification, the fluorescence levels increase proportionally to the number of amplicons produced. This fluorescence is measured following each cycle using a camera to detect and/or quantify PCR products.

In addition to use for DNA sequences, q-PCR protocols are available to measure messenger RNA (mRNA). In a procedure using reverse transcriptase, mRNA is transcribed to form a double-stranded molecule known as complementary DNA (cDNA) prior to PCR initiation. Although all q-PCR methods measure amplified DNA using fluorescence, a variety of fluorescent dyes/probes exist, each with procedural advantages and disadvantages.

One of the most commonly used q-PCR methods is TaqMan® real-time PCR. The TaqMan® probe is an oligonucleotide, containing a sequence complementary to a portion of the DNA to be amplified (usually 18-22 bp), containing fluorophores at opposite ends. A

fluorophore is a functional group which absorbs light at one specific wavelength and emits it at another. The TaqMan® probe contains both a reporter and a quencher fluorophore. The reporter fluorophore, typically attached to the 3' end, is used to provide the measurable fluorescent signal in the reaction, usually at a short-wavelength, in the green range. At the 5' end of the probe is the quencher fluorophore, which usually emits a long-wavelength in the red range. When the reporter is held in close proximity to the quencher, as is the case with both attached to the probe, the fluorescence of the reporter is significantly reduced. The quencher acts by absorbing the energy emitted from the reporter molecule, thereby not allowing for the emission of fluorescence from the probe itself in what is known as Förster, or fluorescence, resonance energy transfer, or FRET. Multiple reporters are available to researchers, such as fluorescein (FAM), rhodamine x (ROX), tetrachloro-6-carboxyfluorescein (TET), and tetramethylrhodamine (TAMRA), which are all frequently used with the quencher 4-4-dimethylaminophenylazobenzoic acid (DABCYL).

In a q-PCR, the TaqMan® probe binds the complementary template DNA on the sequence to be amplified, along with the primers, following the denaturating of DNA. As the temperature is raised (as in PCR) and as polymerization of the complementary strand reaches the probe, the 5' exonuclease activity of the *Taq* polymerase individually cleaves the nucleotides of the probe. With the reporter and quencher now separated, fluorescence from the reporter is now detectable. Each round of amplification increases the amount of fluorescent generated from the liberation of reporter fluorophores allowing for the quantification of DNA sequences present before and after PCR cycles.

Another variation of q-PCR is the molecular beacon method. This method uses a probe which, unless bound to a target, is wrapped into a hairpin structure. A molecular beacon is composed of a loop, a stem, and two fluorophores. Like the TaqMan® probe the fluorophores

are located at the 3'and 5' ends of the probe. The hairpin conformation is important because, when hybridized, the hairpin structure holds the reporter and quencher close together, suppressing fluorescence. The loop is typically 20-22 bp in length (increases in the loop length can leading to increased mismatches), and composed of the complementary target sequence. The stem is 3-7 nucleotides (about 5 C/G pairs and 1 A/T pair) of complementary sequences added to both the 3' and 5' ends of the loop. These stem sequences on the arms of the hairpin keep the probe stable, and allows the probe to rebind to target sequences for multiple cycles. The high C/G to A/T ratio increases the melting temperature above that of the primer annealing temperature (usually 7-10°C) which allows only exactly matching complementary sequences to hybridize, and ensures the probes which do not hybridize remain in hairpin conformation preventing fluorescence. Upon hybridization to the target sequence, the fluorophores are separated by sufficient distance to allow for detectable fluorescence. This fluorescence is measure during the PCR cycle, as the beacon only fluoresces when bound to the target site. Unlike the TaqMan® probe, molecular beacons remain intact after amplification has been completed.

An alternative to using specific probes is SYBR green, a fluorescent dye, which selectively binds to dsDNA but not ssDNA. A cheaper alternative, SYBR green is an intercalating agent which fluoresces very little when unbound in solution, however it will emit a strong signal when bound to dsDNA. As the SYBR green binds to the PCR products, application offers a simple way to measure the amount of DNA amplified from one cycle to the next. However, several disadvantages exist, including the high levels of optimization needed to reduce the possibilities of non-specific amplifications (which the SYBR green will bind to as well), and the emission of stronger signals by longer amplicons (may cause camera saturation). Upon

completion of PCR reactions, a melting-point assay is usually required. Generation of a melting curve is used to check for primer-dimers, mispriming, or contamination. If the reaction was successful, all PCR products should have the same melting temperature.

References

Birren, B.W., Lai, E., Clark, S.M., Hood, L., and Simon, M.I. 1988. Optimized conditions for pulsed field gel electrophoretic separations of DNA. Nucl. Acids Res. 16:7563-7582.

Dorak, M. T. 2006. Real-Time PCR. In Dorak M.T. (Ed): Real-Time PCR (Advanced Methods Series). Oxford: Taylor & Francis.

Gerhardt, P., R.G.E. Murray, W.A. Wood, and N.R. Krieg. 1994. Methods for General and Molecular Bacteriology. American Society for Microbiology. Washington, D.C.

Glick, B.R. and J.J. Pasternak. 1994. Molecular Research Procedures. In Molecular Biotechnology: Principles and Applications of Recombinant DNA. American Society for Microbiology. Washington, D.C. pp. 55-81.

Hunt, M. Real-time PCR. Microbiology and Immunology: University of South Carolina School of Medicine [online] http://pathmicro.med.sc.edu/pcr/realtime-home.htm last accessed 4 April 2007.

Joppa, B., S. Li, S. Cole, and S. Gallagher. Pulsed field electrophoresis for separation of large DNA. HSI Laboratories, Hoefer Scientific Instruments, San Francisco, CA. [online] http://www.nal.usda.gov/pgdic/Probe/v2n3/puls.html last accessed 24 March 2007.

Kitts, C.L. 2001. Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr. Issues Intest. Microbiol. 2:17-25.

Marsh, T.L. 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. Curr. Opin.Microbiol. 2:323-327.

Meays, C.L., K. Broersma, R. Nordin, and A. Mazumde. 2004. Source tracking fecal bacteria in water: a critical review of current methods. J.Environ. Manag. 73:71–79.

Pierce: The Protein People – Part of Thermo Fischer Scientific [online] http://www.piercenet.com/Proteomics/browse.cfm?fldID=72F377CD-2581-438C-9B27-5360226EA128&Format=Print#intro last accessed 21 April 2007.

Ritchie, N.J., M.E. Schutter, R.P. Dick, and D.D. Myrold. 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. Appl. Environ. Microbiol. 66:1668-1675.

Seurinck, S., W. Verstraete, and S.D. Siciliano. 2005. Microbial source tracking for identification of fecal pollution. Rev. Environ. Sci. Bio/Technol. 4:19-37.

USEPA. 2005. Microbial Source Tracking Guide Document. U.S. Environmental Protection Agency, Washington, D.C. EPA/600-R-05-064.