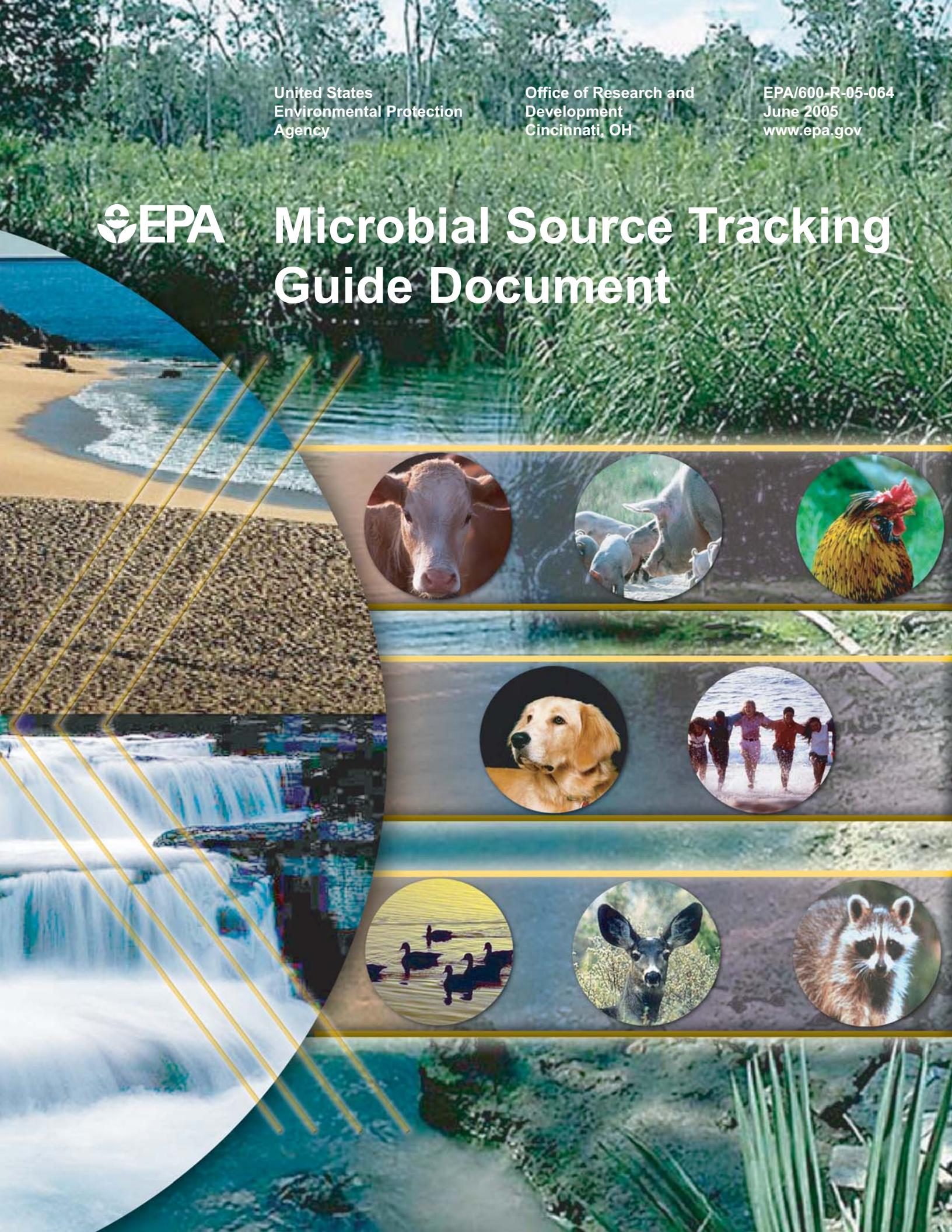
United States
Environmental Protection
Agency

Office of Research and
Development
Cincinnati, OH

EPA/600-R-05-064
June 2005
www.epa.gov

EPA

Microbial Source Tracking Guide Document



Microbial Source Tracking Guide Document

U. S. Environmental Protection Agency
Office of Research and Development

June 2005

Disclaimer

The intent of this guide document is to provide the reader with insight into various tools and approaches used to track sources of fecal contamination impacting water quality in streams, rivers, lakes, and marine beaches. Descriptions of research and several case studies gathered through workshops, literature searches, and phone interviews are also provided. An effort was made to showcase programs, activities, and analyses that incorporated diverse Microbial Source Tracking (MST) approaches and tools. EPA does not support or condone any of the uses of the MST data presented here; nor does it endorse any of the organizations discussed in the case studies. An extensive interpretive review of the scientific literature is included for those interested in learning more about the field. This document does not impose legally binding requirements on states, authorized tribes, or the regulated community and does not substitute for Clean Water Act (CWA) or Safe Drinking Water Act (SDWA) requirements, EPA's regulations, or the obligations imposed by consent decrees or enforcement orders.

Acknowledgements

Prepared by the U.S. Environmental Protection Agency
Office of Research and Development

in cooperation with

Environment Canada

Environment Canada Environnement Canada

Nat'l Oceanic and Atmospheric Admin.

U.S. Department of Agriculture

U.S. Geological Survey

Contributing authors

Thomas Edge, Environment Canada
John Griffith, Southern California Coastal Waters Research Program
Joel Hansel, USEPA Region 4
Valerie J. (Jody) Harwood, University of South Florida
Michael Jenkins, U.S. Department of Agriculture
Alice Layton, University of Tennessee
Marirosa Molina, USEPA Ecosystems Research Division
Cindy Nakatsu, Purdue University
Robin Oshiro, USEPA Office of Water
Michael Sadowsky, University of Minnesota
Jorge Santo Domingo, USEPA Office of Research and Development
Orin Shanks, USEPA Office of Research and Development
Gerald Stelma, USEPA Office of Research and Development
Jill Stewart, National Oceanic and Atmospheric Administration
Donald Stoeckel, U.S. Geological Survey
Bruce Wiggins, James Madison University
Jayson Wilbur, Worcester Polytechnic Institute

Working meeting participants

Steve Harmon, USEPA Office of Research and Development
Bonita Johnson, USEPA Region 4
Stephanie Harris, USEPA Region 10
Samuel Myoda, Delaware Department of Natural Resources and Environmental Control
Donald Reasoner, USEPA Office of Research and Development
Jutta Schnieder, Virginia Department of Environmental Quality
Mano Sivagenesan, USEPA Office of Research and Development

Other contributors

Charles Hagedorn, Virginia Polytechnic Institute and State University
Peter Hartel, University of Georgia

Other people consulted

Sally Gutierrez, USEPA Office of Research and Development
Ronald Landy, USEPA Region 3

Lead Editor

Jorge Santo Domingo, USEPA Office of Research and Development

Technical Editor

Jean Dye, USEPA Office of Research and Development

The lead editor gratefully acknowledges the help of Sue Schock (USEPA) for setting up the contract for the working meetings and Science Applications International Corporation (SAIC) for organizing the meetings. We also want to recognize the leadership of Ronald Landy (USEPA Region 3), Bobbey Smith (USEPA Region 9), Donald Reasoner (USEPA-ORD/NRMRL), and Sally Gutierrez (USEPA-ORD/NRMRL), who in many ways were responsible for several of the USEPA-ORD and regional offices sponsored activities associated with this document. The working meetings were supported by funds from the Water Supply and Water Resources Division and by Regional Applied Research Effort (RARE) Program of the Office Science Policy.

The cover illustration and graphic support was provided by Teresa Ruby (CSI Inc.). We also recognize the help provided by present and former members of the Microbial Contaminant Control Branch (Catherine Kelty, Randy Revetta, Jingrand Lu, Donald Reasoner, Margaret Williams, and Joyce Simpson) and the participants of the Microbial Source Tracking workshop held in August 2003 in Cincinnati, OH. Special thanks to the reviewers of this document: Sheridan Haack (USGS), Bane Schill (USGS), Donald Stoeckel (USGS), Mike Jenkins (USDA), Brian Robinson (NOAA), Laura Webster (NOAA), Jan Gooch (NOAA), Sally Gutierrez (EPA), Tom Edge (Environment Canada), James Goodrich (EPA), and Robin Oshiro (EPA).

FOREWORD

Water is vital to all biological systems, thus safeguarding our Nation's waters is a top priority to the U.S. EPA and the many federal and state agencies that are concerned with human health and environmental sustainability. While the majority of surface and ground waters in the U.S. meet regulatory standards, a significant portion of monitored surface waters contains fecal bacterial densities that exceed the levels established by state surface water quality standards. Reducing fecal pollution levels in natural water systems is particularly challenging as in most cases non-point sources of pollution are the primary contributors to high fecal bacterial levels. As the demand for agricultural activities increases, and urban expansion decreases wildlife's natural habitats, identifying the primary sources of fecal pollution will become even more important in the years to come. Currently, the identification of fecal pollution sources is determined using a variety of Microbial Source Tracking (MST) methods. While some existing reviews address the pros and cons of MST, researchers from the EPA Office of Research and Development (ORD) and EPA regional offices, USGS, NOAA, USDA, and Environment Canada recognized the need of a document that describes in greater detail many issues regarding source identification. Between the spring and fall of 2004, the National Risk Management Research Laboratory organized expert workshops to produce a guide document for stakeholders and environmental professionals interested in MST. This document is a product of these meetings and follow-up communications between federal, regional, state, and academic experts in the field of source identification and environmental monitoring. This document includes a comprehensive review of the literature and in some cases provides a critical view of the state of the science and current research gaps in MST. As advances in the fields of molecular biology and genomics continue to push the frontiers of science forward, it is very possible that new tools and approaches in MST will emerge. It is our intent to follow ~~such~~ these developments and update this guide document as advances are made.

Sally Gutierrez,
Acting Director
National Risk Management Research Laboratory
Cincinnati, OH

Table of Contents

Executive Summary

Approximately 13% of surface waters in the United States do not meet designated use criteria as determined by high densities of fecal indicator bacteria. Although some of the contamination is attributed to point sources such as confined animal feeding operation (CAFO) and wastewater treatment plant effluents, nonpoint sources are believed to contribute substantially to water pollution. Microbial source tracking (MST) methods have recently been used to help identify nonpoint sources responsible for the fecal pollution of water systems. Moreover, MST tools are now being applied in the development of Total Maximum Daily Loads (TMDL) as part of Clean Water Act requirements and in the evaluation of the effectiveness of best management practices. It is evident that MST is transitioning from the realm of research to that of application.

This is not a regulatory document; rather, this document was designed to be used as a reference guide by those considering MST tools for water quality evaluations and TMDL-related activities. However, in a broader sense, water quality managers addressing public health issues, beach/shellfish closures, microbial risk management, and ecosystem restoration should also benefit from the extensive materials contained in this document. Since some of the tools discussed are used in other areas of microbial water quality, environmental scientists and engineers in general would benefit from several of the Chapters of this document.

The guide is divided into seven Chapters. None of the Chapters is intended to stand alone; thus, the reader is encouraged to consult as many Chapters as possible to put into context the comments and suggestions made in various sections. A brief introduction to MST and the goals of this guide document are provided in Chapter 1. Many of the criteria used to decide which method to use in source identification are discussed in Chapter 2. Details relevant to each of the most current approaches used in MST and ways data are collected and analyzed are explained in the next two Chapters. Performance standards for MST studies are discussed in Chapter 5, followed by a critical evaluation of the general assumptions behind and limitations to application inherent in the various approaches. Examples of MST application are presented in Chapter 7.

Most MST studies have relied on matching "fingerprints" from bacterial strains isolated from a water system to those isolated from various hosts (e.g., humans, cows, pigs, raccoons, deer, geese, chickens, etc.) or known environmental sources (e.g., municipal wastewater). In essence, fingerprints are based on phenotypic traits (e.g., antibiotic resistance analysis) or genotypic profiles (e.g., rep-PCR, ribotyping) of individual microbial strains. Typically, hundreds of fingerprints of pure culture bacteria isolated from different sources (or known-source library) are generated in MST studies (Chapter 3). Although results from several studies support the use of the library-dependent approaches for MST, accuracy of these approaches in field application has been questioned because of various problems associated with the target organisms. Some of these problems relate to the level of complexity introduced by spatial and temporal vectors, the stability of the markers used, and issues of sampling design (Chapter 6). More recently, library-independent approaches have been proposed based on the amplification of host-specific markers. Reports are beginning to surface

summarizing results of studies that evaluated library-independent methods against real-world samples (Chapter 7). Much less is known about the library-independent approaches than the library-dependent approaches; therefore, it is not possible at this time to recommend one approach over the other except in specific circumstances (outlined in Chapter 2). It should be noted that in some cases, more than one approach could be utilized for the purpose of identifying fecal pollution sources. Furthermore, in some circumstances it might be necessary to use more than one approach to validate preliminary results obtained with a particular approach.

The complexity of environmental samples and the different variables affecting microbial survival and host specificity have an indirect impact of the efficacy of all MST tools. Moreover, selection of MST tools and approaches are dependent in large part on the goals of an individual effort. In all cases, accomplishing project goals will be impacted by the availability of technical and financial support. As a consequence, various MST methods might be deemed appropriate at sites with similar characteristics (Chapter 2). Regardless of project-specific criteria, the ultimate MST goal can generally be summarized as identification of the major sources of fecal contamination impacting the water system in question. In some cases, this goal has been achieved, while in others, the lack of strong experimental design and poor understanding of the limitations of MST have resulted in insufficient data analysis and poor decision making. Hence, environmental managers must consult the scientific literature and, whenever possible, consult experienced practitioners prior to embarking on source identification studies.

This document builds upon a history of cooperative work among federal, state, and academic partners. To aid our understanding of the reliability that can be expected from various MST tools and approaches, the EPA Office of Research and Development organized several multiagency meetings with the purpose of receiving input from scientists in specialized areas such as population genetics, population biology, host-microbe interactions, microbial physiology, and microbial ecology. MST researchers from academia, USEPA regions, states, federal government (USDA, USGS, NOAA, and USEPA), and Environment Canada participated in these meetings. Most of contributing authors of this guide document participated in the aforementioned meetings and in similar meetings in the U. S., Canada, and Europe; in many ways, this document captures the most relevant elements and suggestions that were discussed in prior meetings.

Chapter 1. Introduction to Fecal Source Identification

The Clean Water Act establishes that the states must adopt water quality standards that are compatible with pollution control programs to reduce pollutant discharges into waterways. In many cases the standards have been met by the significant reduction of loads from point sources under the National Pollutant Discharge Elimination System (NPDES). Point sources are defined as "any discernable, confined and discrete conveyance, including but not limited to any pipe, ditch or concentrated animal feeding operation from which pollutants are or may be discharged". However, more than 30 years after the Clean Water Act was implemented, a significant fraction of the U.S. rivers, lakes, and estuaries continue to be classified as failing to meet their designated uses due to the high levels of fecal bacteria (USEPA 2000b). As a consequence, protection from fecal microbial contamination is one of the most important and difficult challenges facing environmental scientists trying to safeguard waters used for recreation (primary and secondary contact), public water supplies, and propagation of fish and shellfish.

Microbiological impairment of water is assessed by monitoring concentrations of fecal-indicator bacteria such as fecal coliforms and enterococci (USEPA 2000a). These microorganisms are associated with fecal material from humans and other warm-blooded animals and their presence in water is used to indicate potential presence of enteric pathogens that could cause illness in exposed persons (Dufour, 1984). Fecally contaminated waters not only harbor pathogens and pose potential high risks to human health, but they also result in significant economic loss due to closure of shellfish harvesting areas and recreational beaches (Rabinovici et al., 2004). For effective management of fecal contamination to water systems, the sources must be identified prior to implementing remediation practices. Millions of dollars are spent each year on monitoring fecal-indicator bacteria in water and attempting to develop reliable methods for fecal source tracking. Reliable and accurate fecal source identification methods are imperative for developing best management practices (BMPs) to control fecal contamination from relevant animal sources, to protect recreational-water users from water-borne pathogens, and to preserve the integrity of drinking source water supplies.

The immediate demand for methods in MST has been stimulated by the current total maximum daily load (TMDL) requirements that states, territories, and tribes must comply with in the next five to ten years. A TMDL specifies the theoretical amount of a pollutant that a waterbody can receive and still meet water quality standards. Strict waste load allocations from point sources like sewage treatment plants or industrial discharge pipe have already been established with the purpose of meeting regulatory standards. For this reason it is believed that nonpoint-pollutant sources are mostly responsible for many water system impairments, especially after storm events. Most nonpoint sources are associated with agricultural operations, although urban associated pollution is also an important contributor due to the increase in residential, commercial, and industrial development, use of manure as fertilizers, persistence of combined sewer overflows, and malfunctioning septic systems. Wildlife is often assumed to be a relevant source of pollution in cases where no obvious contribution could be assigned to human activity and livestock farming. Due to the variety of potential fecal sources impacting watersheds, fecal source identification is a challenging task that often requires multidisciplinary teams to effectively implement.

Various approaches have been used to identify fecal sources in water samples (Sinton 1996, Jagals 1995, Simpson et al., 2002). For example, chemical analyses have been used to detect human-associated markers like caffeine, fragrances, and detergents (search Ed Furlong). Fecal constituents (e.g., fecal sterols, fecal stanols, and secretory immunoglobulins) have also been considered as source identifiers, since different congeners are preferentially present in different animal species. Some of the chemical-based approaches for fecal source identification are gaining acceptance within the environmental community; however, issues that relate to specificity, sensitivity, microbial biodegradation, and adsorption must be further investigated in order to validate their use as reliable source identification tools. While this document will focus on source tracking tools and approaches that use microorganisms as the source identifiers, it should be noted that chemical approaches could also be used in fecal source identification studies.

Early attempts to classify fecal sources based on microbial source identifiers focused on discriminating contamination sources in a broad fashion (i.e., human vs. nonhuman categories) based on the fecal coliforms to fecal streptococci (FC-FS) ratio. It is now widely accepted that this approach cannot accurately differentiate between human and animal sources because differences in die-off rates between fecal coliforms to fecal streptococci could affect FC-FS ratios in aged the ratios used to classify the sources are not consistently valid for different animals (Reference standard methods). Variability in bacterial survival rates between fecal coliforms and fecal streptococci affects this ratio, particularly when a temporal component is added to the equation. While the FC-FS ratio is seldom used in contemporary source identification studies, it should be recognized that the work of Geldreich and colleagues (Geldreich and Kenner, 1969; Geldreich et al., 1968; Geldreich and Clarke, 1966; Geldreich et al., 1964) is in large part responsible for encouraging other scientists to develop and evaluate new tools to discriminate between the different sources of fecal pollution.

More recently, a number of microbial source tracking (MST) approaches have been developed to associate various animals with fecal pollution of natural waters. MST is based on the assumption that, given the appropriate method and source identifier, the source of pollution can be detected. In general terms, MST methods could be grouped as library dependent methods (LDMs) and library independent methods (LIMs). LDMs require the development databases of genotypic or phenotypic fingerprints for bacterial strains isolated from suspected fecal sources. Fingerprints of isolates from contaminated water are compared with these libraries for classification. Bacterial indicators of fecal contamination (e.g., *E. coli* and enterococci) are commonly used for LDM development. LIMs do not depend on the isolation of targeted source identifier as detection is performed via the amplification of a genetic marker by a Polymerase Chain Reaction (PCR) step, although some methods often require a pre-enrichment to increase the sensitivity of the approach. Some LIMs target the16S rDNA (which is vital for protein synthesis and therefore present in all bacteria), while others target function-specific genes (which are present in a particular bacterial group) for PCR primer development. The advantages, limitations, and applications of a majority of these methods will be discussed in the following Chapters.

Several MST tools are now being applied in the development of TMDL plans and in the evaluation of best management practices. However, due the relatively recent development of MST, most environmental managers and scientists have little training and experience in the application of MST methods to TMDL plans. To date there is no single method that could be applied to all types of fecally contaminated water systems. This is due, in part, to the fact that several factors can control

the level of complexity of a particular water system, which has a direct impact in choosing the best method for the identification of primary sources of pollution. Moreover, there is a lack of consistency among the various laboratories performing some of the MST techniques that keeps them from sharing data. The Office of Research and Development recognizes the importance of effective pollution management measures and the need to develop, evaluate, validate, and standardize methods that could help stakeholders address current fecal pollution issues.

The purpose of this guide is to provide scientists, engineers, and environmental managers with a comprehensive, interpretive analysis of the current and relevant information (based on both lab and field data) related to MST. Descriptions of the various MST approaches, data collection tools, data analysis procedures, method application, performance standards, and assumptions and limitations associated with the field of MST will be provided in different Chapters. The Chapters were written by a diverse group of professionals from academia and government agencies. Regional and state environmental professionals were also consulted during different stages of this particular effort. Many of the contributing authors are recognized leaders in MST and applied environmental microbiology. While the information herein presented is contemporary, it should be noted that MST is a very intense and dynamic field, and therefore, the reader is encouraged to consult the scientific literature frequently.

Chapter 2. Decision Criteria

2.1 Introduction

A number of methods, both genomic and phenotypic, have been developed for use in microbial source tracking (MST). Some of these methods are library-dependent (i.e., rely on fingerprint databases of culturing microorganisms) and some are library-independent (i.e., normally performed by nucleic acid amplification techniques that do not require cultivation of microorganisms). Comparison studies have shown that no single method is clearly superior to the others (Griffith et al., 2003; Stewart et al., 2003; Stoeckel et al., 2004). Therefore, no single method has emerged as the method of choice for determining sources of fecal contamination in all fecally impaired wate bodies. The decision on which method to use depends on the unique set of circumstances associated with the specific study area in question, the results of sanitary surveys, as well as budgetary and time constraints. In some situations, a rather coarse method will suffice, particularly if it is only necessary to distinguish between human and animal fecal sources or between domestic animal and wildlife sources. In other situations, it may be necessary to identify the species of domesticated animal or even the specific herd or flock that is the major contributor of fecal pollution, both of which require more precise methods.

2.2 Choice of method

The microbial source tracking decision tree that appears in this Chapter (Figure 1) was created to assist state and local authorities in deciding whether or not MST methods are necessary to determine the sources of fecal pollution in their particular watershed or bathing beach and, if so, which group of methods might be most appropriate for their needs. Identification of an appropriate group of MST methods is the outcome of a series of decision points. A menu of methods appears at each decision point, allowing the potential users to make informed decisions. The reader is directed to Chapter 3 to evaluate the advantages and disadvantages of specific MST methods.

The following steps outline the process as shown in the decision tree. These steps are meant to serve as a guide to the reader as to the decision points in the tree.

2.2.1 Step 1: Is the problem adequately defined?

MST can be used in a number of circumstances. First, the problem to be addressed must be adequately defined and the desired outcomes considered. For example, if the problem is bacterial exceedences that result in beach advisories/closures, there are many variables to be determined. These include: the conditions under which exceedences are likely to occur, the bacterial indicator species of concern, and the desired outcome (removal of future advisories, determination if human pollution is a source, etc.)

Problem definition can vary for the same situation. In the case of TMDLs, the problem and desired outcome may initially be defined to determine if human feces are contributing to the exceedences so that a prioritization scheme can be fulfilled, i.e., if human feces are present, the area becomes a high-priority target for management action due to the known risks associated with this type of

contamination. Once a TMDL is scheduled, the definition may change to a desire to know every source that may be contributing and, if possible, to what degree.

Failure to adequately define the problem and desired outcomes prior to initiating the decision tree make it unlikely that MST will serve a useful function and achieve results that can be acted upon. Given this, readers are strongly cautioned about proceeding without this information.

2.2.2 Step 2: Has an adequate sanitary survey been conducted?

A sanitary survey can be used to evaluate and document sources of contaminants that might adversely affect public health. Although sanitary surveys are frequently associated with drinking water supply systems, they can be used to identify sources of pollution and to provide information on source controls and identification, persistent problems such as exceedance of water quality standards, magnitude of pollution from sources, and management actions and links to controls. A Registered Sanitarian or professional with experience in these areas should perform the survey. A sanitary survey can be an effective tool for protecting human health and can provide information that helps in designing monitoring programs and selecting sampling locations, times, and frequencies.

In this instance, the sanitary survey should be of sufficient rigor to identify all of the potential sources within the study area, as well as the conditions under which unacceptable contamination should occur. The spatial and temporal extent of the contamination is typically based on local conditions including tidal cycle, nearshore currents, dam releases, and rainfall. Lack of a sufficient survey will hinder the overall approach to identifying the source of pollution in the study area.

For information on how to conduct sanitary surveys, the reader is referred to EPA's *National Beach Guidance and Required Performance Criteria for Grants – June 2002* (EPA 823-B-02-004), Appendix G. This dcument is publically available at the following electronic site; http://www.epa.gov/waterscience/beaches/grants/guidance/factsheet.pdf

2.2.3 Step 3: How many sources were identified in the sanitary survey?

Single source: It is quite possible that the sanitary survey will identify a single, dominant source of contamination within the watershed. In this case, MST is likely unnecessary and remediation of the source is warranted. However, some resource managers may desire a confirmatory test to back up the result of the sanitary survey. In this case, one option would be to use a library independent method, assuming there is an available technique that targets the source identified by the sanitary survey. Use of a library independent method in this scenario is advantageous because these methods can confirm the findings of the sanitary survey without investing the time and money necessary to build a library. However, it may also be cost effective to employ a library dependent method if an appropriate local database already exists. If MST results confirm the findings of the survey, then remediation is again warranted. If the confirmatory test fails to substantiate the findings of the survey or remediation fails to fix the problem, this would indicate a failure in the sanitary survey. In such a case the sanitary survey should be repeated. In some cases a new survey strategy should be considered.

Multiple sources: Proceed to the next step.

2.2.4 Step 4: Is the watershed/study area of manageable size?

This is a rather subjective step, but experience in the field has shown that the smaller the watershed/study area under examination, the greater the chance of success in determining the cause of the exceedence and the likelihood of success at correcting the problem. In general terms, watersheds or study areas with drainage areas greater than 14 digit USGS hydrologic unit code in size are not amenable to using MST. An exception to this general statement is that non-library based methods may prove useful in larger area evaluations if the desired outcome is to know whether human fecal contamination is present. If previous steps have been performed on areas greater than the 14-digit zone, it is strongly recommended that the size of the affected watershed or drainage area be whittled down by use of extensive targeted sampling as previously documented by Kuntz et al. (2003). In addition, a new sanitary survey may be necessary as the original one applies specifically to the larger area.

2.2.5 Step 5: What is the desired level of discrimination?

As previously noted in Step 1, positive identification of a human source may be sufficient for some purposes. However, more detailed information about all fecal sources may be necessary to address a different set of objectives. Step 5 is meant to lead the reader to the set of methods which will provide the level of resolution necessary to fulfill the objectives of the study. Possible discriminations are: 1) humans vs. all other sources, 2) species specific results (humans vs. cows vs. horses vs. deer etc.), 3) group comparisons (humans vs. livestock vs. wildlife), and 4) specific individual hosts (cows from a certain farm vs. other farms vs. other livestock on farms vs. human etc).

2.3 Explanation of Resolution/Outcome Endpoints

#1 Humans vs. All Other Sources and #2 Species Specific Results

Both library independent and library dependent methods are amenable to the resolution of single species. Library independent methods may be appropriate if techniques have been developed that target the desired species. For example, methods that have been proposed to identify human fecal contamination include PCR for host-specific *Bacteroides* species, *E. coli* toxin genes, or human-associated viruses (see Chapter 3). Other species can likewise be targeted, although a limited number of methods currently exist for all species that may be desired (Dick et al., 2005a; Dick et al., 2005b). As a result of sequencing efforts of fecal bacteria as well as fecal microbial communities (Xu et al., 2003; Bäckhead et al., 2005; Eckburg et al., 2005), the number of host specific assays is likely to increase significantly in the near future. If there is not a library-independent method available to target the desired species, a new method may need to be developed or library-dependent methods should be considered. Likewise, if presence/absence results will not suffice to meet the study objectives, and the available library independent methods are not capable of providing quantitative results, then library-dependent methods should again be considered.

An adequate library must be available or developed in order to effectively utilize library dependent

methods. At present, it is not possible to provide generic guidance for what would constitute an adequate library for any MST study. Readers should examine Chapter 5 to determine the requirements for library based methods. Assuming that a library is available or developed, the level of discrimination should be determined to lead to the appropriate suite of methods.

As a caution, the use of 'weighed estimates' or 'quantitation' in these flow charts does not imply that an exact, quantitative assessment is provided by these methods. With changing conditions in a watershed, robustness of the base library, and other methodological considerations, the best that current technology can do is to give a general idea as to the level of contribution from sources at the time the assessment is done. Results from these types of analyses should be regarded as an estimate of contribution, rather than a well-defined fraction associated with each source. With continuing evolution of the technology and methods for source tracking, it is possible that precise quantitative results will be possible in the future.

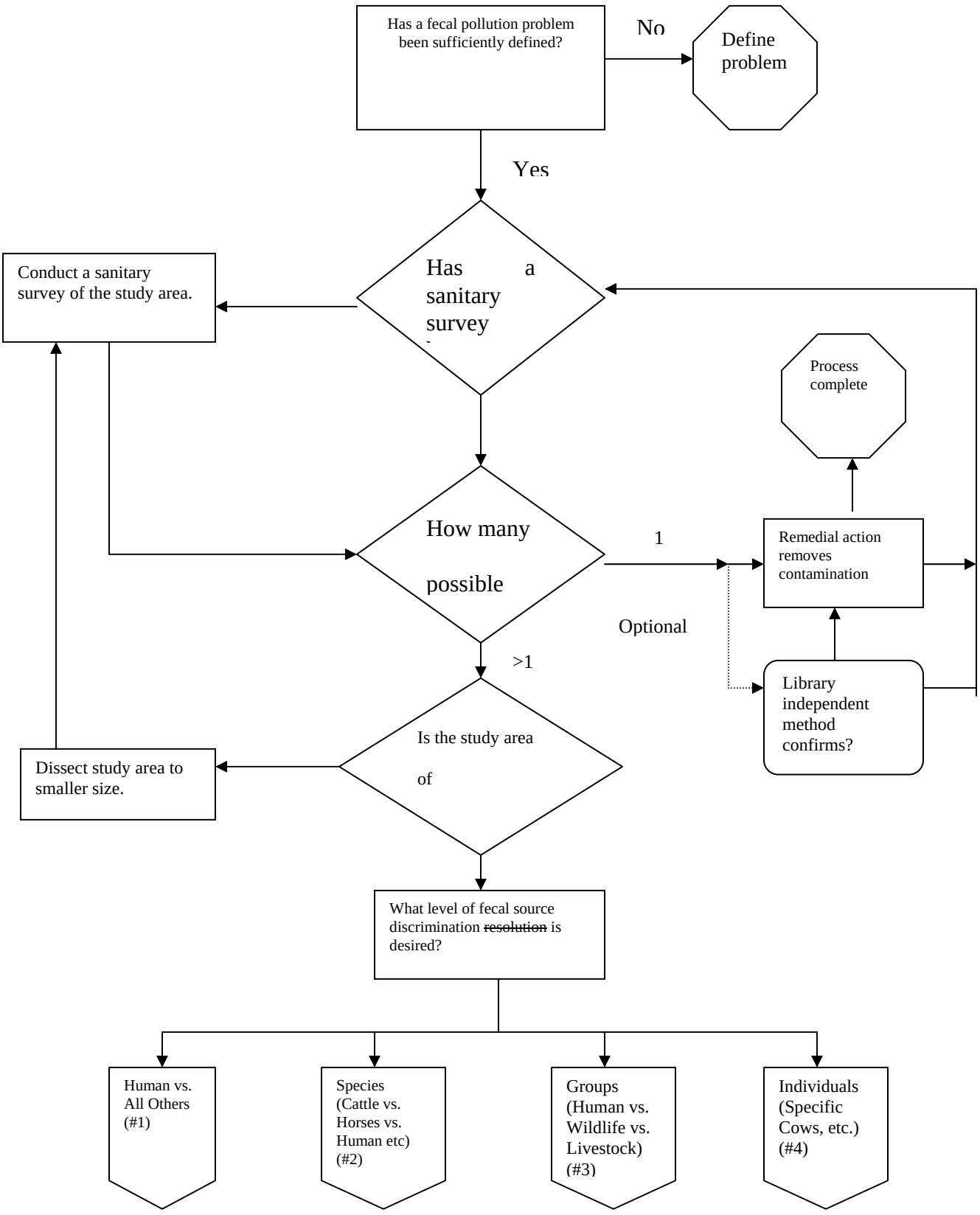
#3 Host Group Comparison (Humans vs. Livestock vs. Wildlife)

This track is very similar to #1 and #2; however, non-library based methods are not considered here because the resolution of these methods is insufficient to discriminate to the level required. Library-based methods only are applicable and come with the same caveat concerning a sufficient library as expressed in #1 and #2.

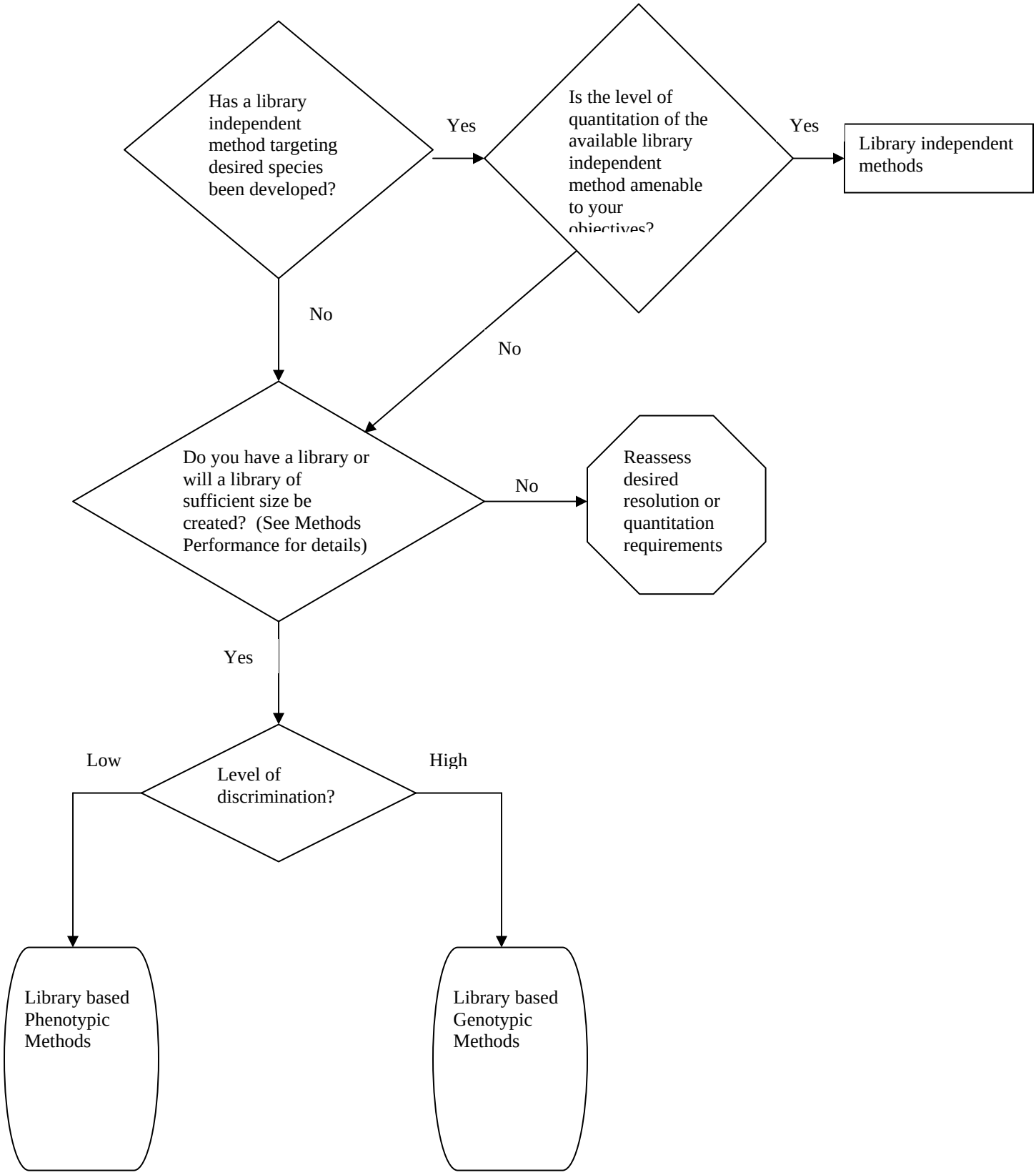
While there are non-library based methods that show promise for presence/absence analyses (e.g., the method that employs ruminant specific primers for *Bacteroides*), they do not currently offer the resolution necessary to make a group comparison. In the case of the ruminant primers, the method will detect cows equally as well as deer. More sensitive non-library-based methods continue to be developed and may become an option for this type of group comparison in the future.

#4 Individual Hosts

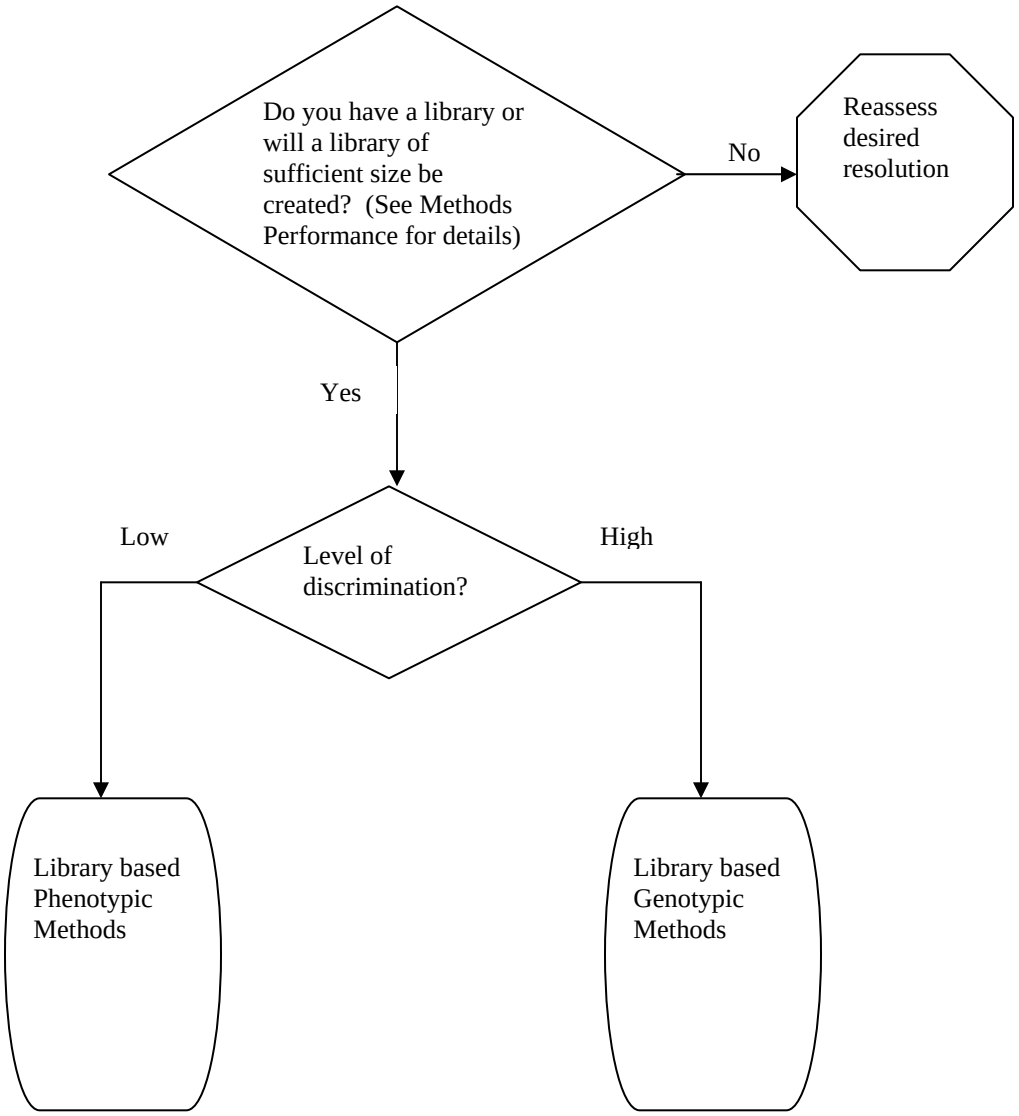
The only methods now available that produce this type of result are library based genotypic methods. Again, these methods come with the caveat that a sufficient library must be available in order to get substantive results. Ideally, the library should be developed at the time of the study to counteract temporal variations that have been observed in genomic libraries (Jenkins et al., 2003).


Has a fecal pollution problem been sufficiently defined?
No
Define problem
Yes
Has a sanitary survey
Conduct a sanitary survey of the study area.
Process complete
How many possible
1
Remedial action removes contamination
Optional
>1
Library independent method confirms?
Is the study area of
Dissect study area to smaller size.
What level of fecal source discrimination ~~resolution~~ is desired?
Human vs. All Others (#1)
Species (Cattle vs. Horses vs. Human etc) (#2)
Groups (Human vs. Wildlife vs. Livestock) (#3)
Individuals (Specific Cows, etc.) (#4)

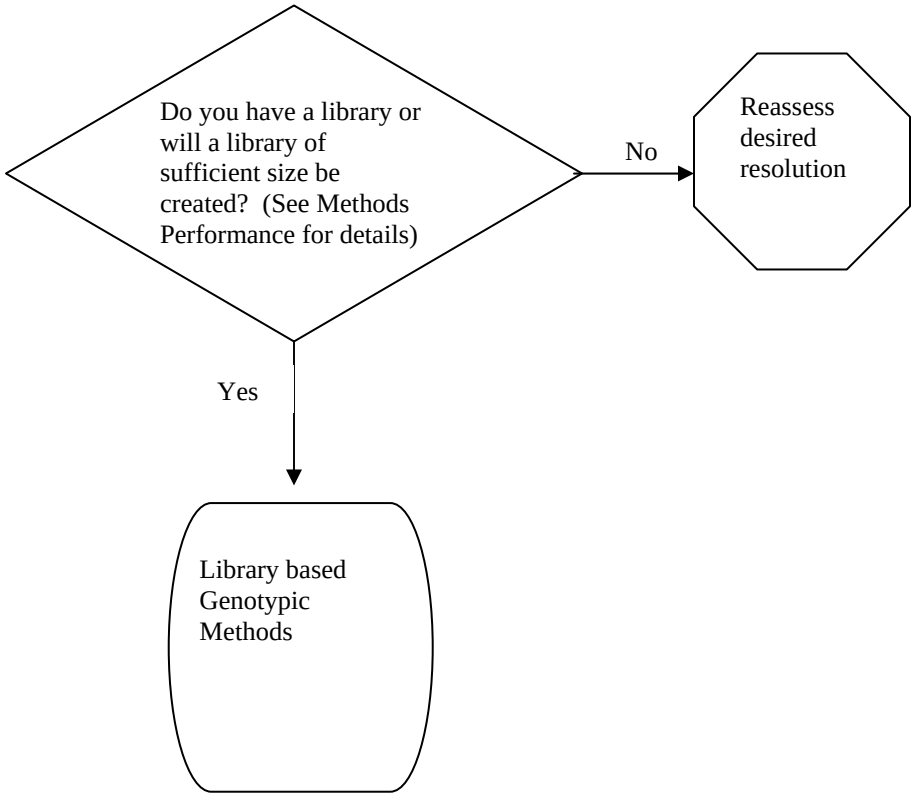
Human only #1 and Species Specificity #2



Define by Groups #3



Define to specific sources by type and location #4



Chapter 3. Microbial Source Tracking Approaches

3.1 Introduction

Numerous approaches have been used to determine potential sources of fecal contamination in the environment. These methods are at various stages of development and validation. Accordingly, this Chapter serves only as a resource for users to make an informed decision on the approach that best suits their needs and financial resources. Currently one method cannot answer all questions and it is likely that this will not change in the near future. This Chapter focuses on methods based on phenotypic and genotypic analysis of microorganisms that have been used for source tracking. A number of the methods described in this Chapter can be, or have already been, adapted for different target organisms. Chapter 6 reviews the target organisms and factors that must be considered when appropriate methods are being chosen. The Chapter on case studies provides more detail on the successful application of some methods. Methods for MST are dynamic with a number of new approaches are being developed, such as gene chips with toxin genes and/or fecal indicator sequences, and biosensors for the detection of target organisms.

Methods currently used for microbial source tracking fall into a few broad categories, genotypic versus phenotypic analysis of either cultivated target organisms, or cultivation-independent approaches by direct analysis of samples from the environment (Figure 1). Genotypic analyses are based on some aspect of an organism DNA sequence, whereas phenotypic analysis measures a trait that is expressed. Genotypic methods differ by targeting specific genes or by measuring genetic polymorphism (differences) in the genome. Genotypic methods that have been used for microbial source tracking are: strain specific PCR (e.g., 16S rRNA gene, host-specific toxin genes, or phage specific sequences), ribotyping, whole genome restriction fragment length polymorphism (RFLP) analysis using pulse field gel electrophoresis (PFGE), repetitive element sequence PCR (rep-PCR) fingerprint profiles, random amplification of polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP). All these methods require selective cultivation of indicator bacteria from water samples as well as from fecal sources that are used to construct a host reference library, with the exception of methods that detect bacterial host-specific genes (e.g., Bacteroides sp. 16S rDNA sequences) using PCR. The two most often used phenotypic methods for MST, antibiotic resistance and carbon source utilization, also require cultivation of the indicator bacteria. Each of these methods will be described in detail in the following sections.

3.2 Cultivation versus cultivation-independent microbial targets

Many of the methods first tested for microbial source tracking used a cultivation approach for *E. coli*, fecal streptococci/enterococci, and coliphage, as these organisms are used as indicator of fecal pollution in waters. Standard methods for the cultivation of *E. coli* and fecal enterococci (USEPA, 2000) and coliphage (USEPA, 2001a, b) have been previously described. Although EPA has standard cultivation methods, caution must be taken when comparing studies in the literature. Often different methods have been used to cultivate and confirm the target organism (Harwood et al., 2003; Myoda et al., 2003). For example, *E. coli* may be isolated on mTEC, MI, mFc, a combination of mENDO and NA-MUG, or by using commercial systems such as Colilert™ and Colitag™. Some *E. coli* confirmatory tests used either singly or in combination are: IMViC – indole production,

methyl red reaction, Voges-Proskauer test, failure to grow on citrate-minimal media, MUG (4-methylumbelliferyl-ß-D-glucuronide) hydrolysis (test for ß-glucuronidase), indole production, gas formation on lactose, failure to express urease, failure to express oxidase, and Analytical Profile Index (API) biotyping system. There is still a need for researchers to standardize detection and confirmation methods for all indicators to ensure the same organism is isolated and study results are comparable. The discriminatory power of each method may vary when different target organisms are used and therefore each target organism must be tested independently to assess the value of a method.

An alternative approach for studying microbial ecology has been prompted by research that estimated that only a small fraction (0.1 to 10%) of bacterial species have been cultivated from most environments (Ranjard et al., 2000; Staley and Konopka, 1985; Torsvik et al., 2002). Most relevant to microbial source tracking is the analysis of gastrointestinal microbes, which indicates that some 400 different species of bacteria may be found in animal intestines and populations are in the order of 10^{11} g^{-1} of contents (reviewed by Zoetendal 2004). Intestinal microflora have been well characterized in a number of animal hosts including humans (Suau et al., 1999), swine (Leser et al., 2002; Pryde et al., 1999) and cattle (Ramšak et al., 2000). Collectively, and when compared to cultivation-dependent methods, cultivation-independent methods suggest that the numerically dominant bacteria in animal colons are anaerobic and belong to the low G+C Gram-positive and Cytophaga-Flavobacter-Bacteroides bacterial phyla. Common genera in animal intestines are *Bacteroides*, *Eubacterium*, *Clostridium*, *Ruminococcus*, *Peptococcus*, *Peptostreptococcus*, *Bifidobacterium* and *Fusobacterium* (Matsuki et al., 2002). However, these bacteria are not readily cultivated in the laboratory, which has limited their use as fecal indicators in the past. In contrast, the more easily cultivated fecal indicator bacteria, *E. coli* and *Enterococcus*, are present in lower concentrations. A number of molecular genetic methods and kits have been developed to isolate nucleic acids from organisms or environmental samples without need for cultivation, making it possible to use alternative targets. After extraction, a number of methods can be used to examine DNA directly or indirectly after amplification by polymerase chain reaction (PCR). The polymerase chain reaction (PCR) technique is an extremely useful, sensitive and rapid method that can be applied to both laboratory-cultivated organisms and nucleic acids directly obtained from environmental samples. Nucleic acid replication via PCR is automated in the laboratory resulting in an approximately 10^6-fold amplification of a target nucleotide sequence. This approach provides a means to examine targets that are not readily cultivated and may not be in high numbers in the environment, but nevertheless serve as better indicators of fecal sources.

Want more details on PCR?

Polymerase Chain Reaction (PCR) is a method in which a target DNA sequence is preferentially replicated from a mixture of non-target sequences. All methods that involve PCR have some common requirements: (i) target primer(s); (ii) each of the four nucleotides (adenine, cytosine, guanine, thymine); (iii) thermal tolerant polymerase (e.g., Taq polymerase); (iv) nucleic acid template (e.g., DNA from cultures or environmental sample); and (v) appropriate buffers and co-factors to maintain the proper pH and optimize the enzymatic reaction. Primers are short lengths of nucleotides (oligonucleotides) that usually range from 8 to 24 basepairs in length depending on the desired specificity, which increases with primer length. DNA synthesis is initiated from

the primers, and therefore PCR amplifies the sequence between the primers. The specificity of the primers is dependent on the sequence information available at the time of their development. The polymerase enzyme catalyzes the synthesis of new DNA. Once all ingredients are combined, the solution temperature is cycled 25-35 times between three temperatures to control the amplification process. First it is raised (usually 94-95°C) to denature the two strands of DNA, then lowered to allow primers to anneal to the template DNA (typical range is 50°C to 65°C), and then raised to optimize activity of the thermal tolerant polymerase (e.g., 72-75°C) to synthesize new strands of DNA. After PCR is performed, the DNA products can be visualized after gel electrophoresis and staining. In addition to the basic molecular biology supplies this method requires access to a thermal cycler. The reagents (mainly the polymerase and nucleotides) and PCR disposables (e.g., micropipette tips, PCR tubes, etc.) are the major costs incurred by this method.

3.3 Cultivation-dependent/library-dependent methods

Many methods that rely on the cultivation and isolation of the target microorganisms also require the creation of a reference library. Reference libraries are built using isolates taken from known hosts or environmental sources. In most cases isolates are taken from fecal samples, if possible, collected directly from the animal or directly after excretion to ensure there is limited contamination from other sources. However, some investigators believe sewage lagoons and animal waste holding ponds provide isolates more representative of survivors that would most likely be found in the environment. Most libraries have been built using isolates taken from potential sources in the region being studied. Currently, there are conflicting opinions on the geographic and temporal stability of source libraries, likely arising from a number of factors including: differences in library sizes, sampling method, and data analysis method. Isolates with identical patterns from the same fecal sample, are presumed to be clones and should be discarded from the library otherwise an inaccurate statistical bias will occur. A number of approaches that have been used to determine the accuracy of libraries are discussed in the method performance Chapter. However, the most important unsolved factor is the size of library necessary to successfully identify host sources.

3.3.1 Phenotypic methods

Antibiotic resistance

Antibiotic resistance was developed as a method for source tracking based on the demonstrated phenomenon that bacteria from hosts exposed to antibiotics will develop resistance to those antibiotics, and on the hypothesis that this selective pressure would be a mechanism for discriminating among fecal bacteria from various hosts. Antibiotics are used to prevent and treat infections in humans and domestic animals and to increase growth rates in animal production. Bacteria resistant to antibiotics used in animal feed (Bryan et al., 2004) have been found in poultry litter (Kelley et al., 1998), cattle feces (Dargatz et al., 2003), and in swine manure (Smalla et al., 2000). Throughout the literature, different permutations of antibiotics and concentrations (range in μg/ml) have been used for antibiotic resistance tests including: amoxicillin (4-128), ampicillin (10), bacitracin (10-100), cephalothin (sodium salt) (10-50), chloramphenicol hydrochloride (4),

chlortetracycline hydrochloride (20-80), chlortetracycline (20-80), doxycycline hydrochloride (4), erythromycin (5-50), gentamicin (1-20), kanamycin monosulfate (3-50), monensin, (5-250), moxalactam-sodium salt (0.2-1), nalidixic acid-sodium salt (3-25), neomycin sulfate (3-50), norfloxacin (0.1), oxytetracycline hydrochloride (20-100), penicillin G-potassium salt (20-200), polymixin B (1-10), rifampicin (2-16), streptomycin sulfate (20-800), sulfathiazole (500), tetracycline hydrochloride (4-64), trimethoprim:sulfamethoxazole (1:19 ratio) (0.2- 5), and vancomycin (2.5-30). There is currently no standard suite of antibiotics and concentrations used for antibiotic resistance testing. Antibiotics are best chosen after determining potential animal fecal sources and antibiotics used in their treatment. Furthermore, the antibiotics chosen must be appropriate to the source identifier utilized, i.e., *E. coli* and other fecal coliforms are intrinsically resistant to vancomycin; therefore, its use with this class of source identifier is not informative.

This method has been used extensively because it is rapid, relatively simple, and relatively inexpensive. Furthermore, it requires less technical expertise than molecular methods and no specialized equipment. There are three approaches that have been used in MST studies antibiotic resistance analysis (ARA), multiple antibiotic resistance (MAR) and Kirby-Bauer antibiotic susceptibility. In MAR studies, bacteria are tested for resistance to different antibiotics (Parveen et al., 1997). ARA differs slightly by including different concentrations of each antibiotic being tested (Wiggins, 1996; Wiggins et al., 1999). The Kirby-Bauer antibiotic susceptibility test has been a standard method for use in clinical studies and uses small filter disks that have been impregnated with antibiotics. The zone of growth inhibition around the disks is used to quantify resistance. Some MST researchers believe that ARA provides the most information of the three antibiotic-based approaches. A potential problem when using antibiotic resistance as a phenotypic source tracking method is the transfer of resistance genes between bacteria. Genes conferring antibiotic resistance have been found on a variety of mobile genetic elements including plasmids, transposons, and conjugative transposons that provide a means for lateral transfer of the genes (Bass et al., 1999; Kruse et al., 1994; Ohlsen et al., 2003; Salyers et al., 1995; Smalla et al., 2000). Although indigenous bacteria have the potential to transfer antibiotic resistance genes to fecal bacteria after bacteria from fecal sources enter the environment, this would have to occur at very high frequency to affect the overall proportion of resistant cells in the fecal host population. Even if gene transfer frequencies were as high as 1%, which is much higher than has been reported (Smalla et al., 2000), their detection will be unlikely with current antibiotic resistance protocols unless there is extensive regrowth of the recipients in the environment.

Application of antibiotic resistance to MST

Among the different antibiotic resistance approaches available, ARA is the most common method in MST studies (Booth et al., 2003; Choi et al., 2003; Graves et al., 2002; Hagedorn et al., 1999; Harwood et al., 2000; Harwood et al., 2003; Whitlock et al., 2002; Wiggins, 1996; Wiggins et al., 1999, Wiggins et al., 2003), and has been utilized in many TMDL studies. Regardless of the specific method, they all first require cultivation of the target organism, and *E. coli*, fecal enterococci, and fecal streptococci have been tested with this method (Harwood et al., 2003; Parveen et al., 1997; Wiggins, 1996).

Basic antibiotic resistance methodology

For ARA and MAR antibiotic resistance analysis is carried out by first developing a database of antibiotic resistance patterns (ARPs) of indicator bacteria isolated from the feces or sewage of known animal sources. Colonies are isolated by membrane filtration or by streaking onto the appropriate selective-differential media. These isolates are transferred to a 96-well microplate filled with growth medium, incubated, and then replica-plated on a battery of antibiotic-containing media. Multiple concentrations of each antibiotic are used for ARA, while a single antibiotic concentration is used for MAR. The isolates are then scored positive or negative for growth on each plate. Plates with no antibiotic addition are used as positive controls. Typically, the ARP of each isolate consists of approximately 30 data points. The procedure for determining the ARPs of isolates requires four to five days.

ARPs of bacteria from known sources are then analyzed using discriminant analysis, a form of multiple analysis of variance. Discriminant analysis uses the ARPs from known sources to generate the predictive equations (the "classification rule") that will be used to classify unknown isolates by source. The accuracy of the database is assessed by using ARPs of the isolates from known sources as test data. This procedure generates a source-by-source matrix that provides the rate of correct classification for each source. Overall performance is measured by averaging the rates of correct classification (ARCC) for each source. Fecal bacteria isolated from polluted water are then processed in the same manner as the known isolates, and identified using discriminant analysis. More rigorous tests can be utilized to validate the predictive accuracy of the database, i.e., ARPs of isolates from samples that are not included in the library can be used to challenge the database's predictive capability.

The Kirby-Bauer disk diffusion antibiotic susceptibility test is performed following the NCCLS protocol (National Committee for Clinical Laboratory Standards 1999). In this method filter paper disks with known concentrations of antibiotics are placed into a Petri plate that has been heavily inoculated with the bacterium of interest. The antibiotics diffuse from the disks into the agar making a gradient of antibiotic concentrations. The plate is incubated usually for about 24 h, then the zone of inhibition surrounding the disk is measured, which indicates the antibiotic sensitivity of that isolate. This diameter is called the minimum inhibitory concentration (MIC) for that antibiotic. The size of the growth inhibition zones can vary due to: (1) the culture medium used; (2) incubation conditions; (3) the rate of antibiotic diffusion; and (4) the concentrations of the antibiotics used. All these factors must be kept constant to make between experiment comparisons. An antibiotic sensitive control must be used for comparisons (e.g., *E. coli* ATCC 25923). Tests of each isolate must be replicated to ensure reproducibility. Isolates are scored as sensitive, intermediate, or resistant compared to the control for each antibiotic used. For all three approaches isolates are classified based on a combination of the antibiotics (and concentrations if known) to which they are sensitive and resistant.

Carbon utilization

This method compares differences in the utilization of several carbon and nitrogen substrates by different bacterial isolates. Substrate utilization can be rapidly scored by the formation of a purple

color due to the reduction of a tetrazolium dye included with the substrates and automatically detected using a microplate reader. Isolates are typically classified using only a subset of indicative substrate, for example, Hagedorn et al. (2003) used only 30 of the 95 wells for their analysis.

This method was first investigated for potential use in MST because it is rapid, simple and requires little technical expertise. It has been most successfully used for identification of isolated clinical Gram-negative bacteria (Holmes et al., 1994). Its use in analysis of environmental samples has been questioned due to variability and poor reproducibility (Konopka et al., 1998; Tenover et al., 1995). It is possible to test substrate utilization of each isolate using an array of substrates in the laboratory but the method has been simplified by the availability of commercial microwell plates containing substrates. Most commonly used are Biolog microplates (Hayward, CA), and more recently PhenePlate (PhP plates; Stockholm, Sweden).

Application of carbon utilization to MST

This method has been tested for use in MST only at a small scale (Hagedorn et al., 2003; Harwood et al., 2003; Wallis and Taylor, 2003). In one study, 30 *Enterococcus* strains were isolated from stream sites where an obvious source of pollution was apparent and analyzed using the Biolog system. Using a 365 isolate source library, classification of sample isolates correctly matching the presumptive sources ranged from 86.6-93.3%. However, in another study using the PhenePlate system, a larger number of *Enterococcus* isolates (1,766) from six sources were compared, diversity was very high in wastewater samples (Simpson's Diversity Index= 0.95) and seabird feces (DI = 0.72) but much lower in animal feces such as cows (DI=0.32) (Wallis and Taylor, 2003). High diversity increases the size of library needed to differentiate hosts. In a controlled study, results of carbon utilization were compared to antibiotic resistance and found to be comparable (Harwood et al., 2003). Although positive identification was high (93%), there were also a number of false positives (51.5%). The authors speculated that the library was too small, resulting in the false positives.

Carbon utilization methodology

Very little preparation by the user is necessary since microwell plates with 95 different substrates may be purchased from Biolog, (Hayward, CA) or those with 24 substrates from PhenePlate (Stockholm, Sweden). Since the PhenePlate has only 24 substrates, one plate can be used for replication or different isolates. Different microwell plates are used for the analysis of Gram-positive bacteria (e.g., GP2 MicroPlate™, Biolog) and Gram-negative bacteria (e.g., GN2 MicroPlates™, Biolog). Isolates are first grown and a liquid suspension of cells at a standardized turbidity is used to inoculate the microplates. After incubation at 37ºC for 24 h, presence or absence of growth is indicated by purple dye formation and is assessed manually or automatically using a plate reader (MicroLog™ System, Biolog). Discriminant analysis of the binary data from known sources is then typically used to determine the substrate combination that best distinguishes the host.

3.3.2 Genotypic methods

Molecular (DNA) typing or fingerprinting tools are used to differentiate specific microorganisms. Bacteria and in particular *E. coli* strains have been analyzed by a variety of genotyping methods that vary in their sensitivity and technical complexity. Genotypic methods requiring a reference library fall into two categories, direct analysis of the genome or indirect analysis after PCR. The sensitive and rapid nature of the PCR method and its ability to amplify target sequences approximately 10^6-fold has made it an attractive method, and is commonly used in many of the newer source tracking approaches. This section discusses the cultivation-dependent library methods, but PCR is used both in cultivation dependent and independent approaches. The latter are discussed toward the end of this Chapter. In general, methods that employ PCR are usually more rapid than those that directly examine the genome. The advantages of using PCR based method are: only a small amount of starting DNA material is needed, often bacterial cells can be used without performing DNA extraction, some analyses can be automated reducing labor costs, and most produce highly reproducible and accurate fingerprint profiles. All the methods listed in this section require a laboratory and personnel with at least basic equipment and expertise in molecular genetics.

rep-PCR DNA fingerprinting

The repetitive element sequence-based PCR (rep-PCR) DNA fingerprinting technique (de Bruijn, 1992; Versalovic et al., 1991; Versalovic et al., 1994) uses the polymerase chain reaction and primers to amplify specific portions of the microbial genome, which are subsequently visualized following electrophoresis. The primers used for rep-PCR DNA fingerprinting are complementary to naturally occurring, multi-copied, conserved, repetitive, DNA sequences present in the genomes of most Gram-negative and Gram-positive bacteria (Lupski and Weinstock 1992). The repetitive elements are usually comprised of duplicated genes, interspersed repetitive extragenic palindromes (REP) and other palindromic unit sequences, intergenic repeat units (IRU), enterobacterial repetitive intergenic consensus (ERIC) sequences, bacterial interspersed mosaic elements (BIME), short tandemly repeated repetitive (STRR) sequences, and Box elements (Sadowsky and Hur, 1998). Three major families of repetitive sequences have been generally used for rep-PCR DNA fingerprinting: repetitive extragenic palindromic (REP) sequences (35-40 bp), enterobacterial repetitive intergenic consensus (ERIC) sequence (124-127 bp) and the 154 bp Box element (Versalovic et al., 1994). The use of these primer(s) coupled with PCR leads to amplification of the specific genomic regions located between adjacent REP, ERIC or Box elements. While the methods done using these sequences should be referred to as REP-PCR, ERIC-PCR and Box-PCR genomic fingerprinting, respectively, collectively the technique is referred to as rep-PCR genomic fingerprinting (Versalovic et al., 1991; Versalovic et al., 1994). The resulting mixture of amplified DNA fragments is resolved in agarose gels, producing a banding profile referred to as a rep-PCR genomic DNA fingerprint (Versalovic et al., 1994). Thus, the banding pattern serves as a "fingerprint" for strain identification or analysis of microbial populations. Bacteria having identical fingerprints are regarded as being the same strain, and those having nearly identical or similar banding patterns are regarded as being genetically related.

This method has been used extensively because it is rapid and relatively simple. Among the molecular genotyping methods, it is the least expensive and requires less technical expertise. Only the basic equipment present in most laboratories performing molecular genetic analyses is necessary

unless higher throughput and greater accuracy is desired and one chooses to use an automatic sequencer or fluorescence scanner (see HEFERP below).

Application of rep-PCR to MST

The rep-PCR DNA fingerprinting technique is relatively quick, easy, and inexpensive to perform, and lends itself to high throughput applications, making it an ideal method for microbial source-tracking studies (Carson et al., 2003; Dombek et al., 2000; Johnson et al., 2004; Lipman et al., 1995; McLellan et al., 2003). In studies where rep-PCR have been compared to other methods, it has been shown to give better predictions than ribotyping (Carson et al., 2003), and rRNA intergenic spacer region (ISR)-PCR (Seurinck et al., 2003).

Overview of rep-PCR methodology.

The rep-PCR DNA fingerprinting technique is amenable for use with DNA templates produced using a variety of methods. These include liquid cultures, colonies, and purified DNA. Using colonies directly instead of performing DNA extraction reduces the time and cost of using this method, particularly in comparison to other genetic fingerprinting methods. Among the primers used for rep-PCR, the Box primer A1R has proven to be the most useful in distinguishing environmental isolates of *E. coli* (Dombek et al., 2000). There are two general methods to perform rep-PCR DNA fingerprinting, differing in the way in which the DNA fragments are visualized. In the first more conventional method, the resulting DNA fragments in agarose gels are visualized following staining in ethidium bromide (de Bruijn, 1992). Despite careful attention to detail, it is often difficult to get rep-PCR gels to run consistently straight and avoid lane distortions, which makes alignment and comparisons within and between gels difficult. To overcome these major limitations, a second method has been developed, a horizontal, fluorophore-enhanced, rep-PCR (HFERP) technique (Johnson et al., 2004). The technique is similar to that previously described for use with a DNA sequencer (Rademaker and deBruijn, 1997; Versalovic et al., 1995). In HEFERP, however, a standard horizontal agarose gel electrophoresis system and a dual-wavelength scanner are used. HFERP is ideal for high throughput analyses of bacteria and the protocol can be geared for 96 well microplates using colonies (details at: http://www.ecolirep.umn.edu/a_hferpoverview.shtml).

Randomly Amplified Polymorphic DNA (RAPD) analysis

The Random Amplified Polymorphic DNA Analysis (RAPD), and Arbitrary Primed Polymerase Chain Reaction (AP-PCR) techniques (Welsch and McClelland, 1990; Williams et al. 1990) represent two independently developed, but conceptually-related methods that have found extensive use in studies of microbial epidemiology, diversity, population genetics, taxonomy, evolution, and ecology (Mathieu-Daudé et al., 1998). Both methods rely on the fact that PCR conditions done using arbitrary primers at low stringency (AP-PCR) or with non-selective primers at high stringency (RAPD) produce a series of strain specific PCR products that depend on both the primer and template used. When separated on agarose gels and stained with ethidium bromide, these PCR products produce a series of species- or strain-specific bands that act as a fingerprint of the bacterial

genome. A subsequently developed method, DNA Amplification Fingerprinting (DAF) (Caetano-Anollés et al., 1992) differs from AP-PCR and RAPD in that a polyacrylamide gel and silver staining is frequently used to visualize the PCR products.

RAPD analyses are relatively inexpensive when compared to other molecular methods like ribotyping and pulse field gel electrophoresis (PFGE), require no previous knowledge of the genome examined, are amenable to using colonies, boiled preps, or purified DNA, and can be scaled-up for high throughput analyses. However, it has been shown that RAPD analyses are susceptible to the buffers used, cycle number, primer choice, and method of DNA preparation (Hopkins and Hilton, 2000; Mathieu-Daudé et al., 1998, Wang et al., 1993). Consequently, it has been reported that RAPD analyses may not be reproducible and suffer from lab-to-lab variation (Hilton et al., 1997; Hopkins and Hilton 2000; Penner et al., 1993). Nevertheless, Wang and co-workers (1993) reported that RAPD analyses were more sensitive than multilocus enzyme electrophoresis in differentiating among *E. coli* strains. It has been suggested that some variation may be eliminated by the use of standardized reagents and kits (Hopkins and Hilton, 2000).

Application of RAPD to MST

RAPD analyses have been used to examine genetic diversity of *E. coli* obtained from animals (Aslam et al., 2003), feedlots (Galland et al., 2001), humans (Pacheco et al., 1997; Vogel et al. 2000), and in culture collections (Wang et al., 1993). There has been considerable interest in using RAPD analyses to detect and analyze *E. coli* O157:H7 (Galland et al., 2001; Hopkins and Hilton 2000; Radu et al., 2001) and enterotoxigenic *E. coli* (Pacheco et al., 1996; Pacheco et al., 1997). However, RAPD analyses have only been preliminarily tested for use in microbial source tracking (Ting et al., 2003). These authors reported that RAPD fingerprints might be useful for differentiating among human and non-human sources of *E. coli* contamination.

Overview of RAPD methodology.

While RAPD and AP-PCR DNA fingerprinting have sometimes been used synonymously, in AP-PCR a single or sometimes two arbitrary primers are used in PCR under low stringency conditions and priming is done with sequences having the best match, with some mismatches. In contrast, RAPD DNA fingerprinting is often done at high stringency conditions using primers with low selectivity that anneal at the T_m of the primer. This is thought to result in priming of genomic DNA with less mismatches than is seen with AP-PCR (Mathieu-Daudé et al., 1998). RAPD DNA fingerprinting is typically carried-out using 10-mer random primers. These primer sets are commercially available (e.g. Genosys Biotechnologies or Amersham Biosciences Ready-to-Go RAPD Analysis) and can be initially screened for discrimination ability using the organism of interest. For example, three and six primers have been found to be useful to differentiate among *E. coli* strains (Madico et al., 1995; Pacheco et al., 1997; Wang et al., 1993).

Amplified Fragment Length Polymorphism (AFLP) analysis

Amplified fragment length polymorphism (AFLP) is a powerful and sensitive DNA fingerprinting

technique, which was originally developed to map plant genomes (Blears et al., 1998; Lin and Kuo, 1995). It uses a combination of genomic DNA digestion with restriction enzymes and PCR. In this method short adaptors are ligated (attached) to the digested fragment ends to provide sufficient length of known sequence for primers to be used for PCR. To amplify all of the digested fragments by PCR would result in a multitude of products that would be too difficult to resolve. To overcome this problem, additional PCR primers are used for a second round of PCR. These primers differ from the initial primers by the addition of 1-3 nucleotide bases resulting in the amplification of just a subset of the initial fragments. The addition of more nucleotides to the end of the primers increases the specificity and decreases the number of resultant PCR products. Separate reactions using different primers sets are often used and the data combined providing a substantial number of data points to be used to discriminate isolates. If sufficient number of primers are used the entire genome can be accurately sampled using this approach (Arnold et al., 1999). However, in most cases only about three primer sets are needed to obtain sufficient resolution between isolates. Currently, there is no standard set of primers designated for MST or for any bacterial species.

The need to conduct genomic DNA digestion and PCR makes this method more time consuming and more expensive than other methods that use only PCR. Of all the PCR based methods this one can produce the most bands, which provides a better chance for distinguishing isolates but also increases the need to precisely discriminate bands. Using an automatic sequencer improves band discrimination, decreases time and labor but adds to the costs both in purchasing the equipment and supplies.

Application of AFLP to MST

The AFLP method has been used to fingerprint different bacterial species and is reported to be more sensitive in the detection of DNA polymorphism in them (Clerc et al., 1998; Lin and Kuo, 1995; Restrepo et al., 1999; Valsangiacomo et al., 1995). The majority of studies have been focused towards epidemiology and not MST. The number of MST studies to date is limited but suggests that its resolution is as good or better than most other genetic fingerprinting MST methods (Guan et al., 2002; Hahm et al., 2003a, b; Leung et al., 2004). AFLP was compared to MAR and 16S rRNA gene sequences in *E. coli* collected from livestock, wildlife, or human feces (Guan et al., 2002). Discriminant analysis indicated AFLP was better than MAR and rRNA gene sequence analysis at assigning isolates correctly to each source. Another study comparing *E. coli* isolates obtained from cattle, humans and pigs using AFLP and ERIC-PCR revealed similar results (Leung et al., 2004). There was greater than 90.6-97.7% correct classification using AFLP and 0-75% for ERIC-PCR. A third study compared a number of different methods but mainly examined *E. coli* serotype O157:H7 isolates and only a small number of environmental isolates (Hahm et al., 2003a). However, that study and some follow-up work (Hahm et al., 2003b) suggested that AFLP resolved strain differences in *E. coli* at the same level as PFGE. More fundamental research is still needed to determine the best primer sets to use for different levels of discrimination between isolates. At the same time, when considering this approach the expertise, time and cost factor should be compared to other genotyping methods for the accuracy achieved.

Overview of AFLP methodology

Isolates are analyzed using an AFLP fingerprinting kit following the instruction of the manufacturer (Gibco BRL). Briefly, DNA is extracted from cultures using any standard total genomic DNA isolation method. Purified DNA is then digested with a frequently cutting and a less frequently cutting restriction enzyme *Mse*I and *Eco*RI, respectively, and the fragments are ligated to *Eco*RI and *Mse*I adapters to generate template DNA for PCR amplification. This restriction-ligation mixture is diluted and amplified with *Eco*RI and *Mse*I core sequence primers for pre-selective amplification. Selective amplification is then performed using primer sets with additions of 1-3 arbitrary nucleotide sequences on the 3' end of each. Eight primers of each *Eco*RI and *Mse*I adapters are provided with the AFLP kit. A total of 64 combinations of primer pairs can be used for PCR amplification. Three commonly used selective primer sets are: *Eco*RI-A (FAM^{TM}) plus *Mse*I-C, *Eco*RI-0 (FAM^{TM}) plus *Mse*I-CG and *Eco*RI-C (NED^{TM}) plus *Mse*I-C. The primers used for PCR amplification are fluorescently labeled (e.g., FAM^{TM} and NED^{TM}) for automatic detection of the different size products using an automatic sequencer. This also allows high throughput analysis of AFLP patterns. Labeled size markers (DNA size markers) are included in each lane to ensure accuracy of band detection and differentiation. The typical size range of amplification products is between 50 and 4000 bp. The number of isolates that can be analyzed in a single run depends on the automated sequencer that is being used. All information collected from the sequencer is then transferred to a fingerprint analysis program (e.g., Bionumerics, Applied Maths). The data is binary, based on the presence and absence of bands in each profile (see data analysis Chapter).

Pulse Field Gel Electrophoresis (PFGE)

The most common genotyping method used in epidemiological investigations is pulse field gel electrophoresis (PFGE) of total genomic DNA after restriction enzyme digestion using an infrequently cutting enzyme (Tenover et al., 1995). It involves direct analysis of the microbial genome and PCR is not performed. Digestion of total genomic DNA by an infrequently cutting restriction enzyme, results in the production of 10 to 30 large fragments. These fragments are too large to be separated in a standard agarose get electrophoresis unit because the gel pore size limits their migration. To overcome this limitation PFGE was developed in which the orientation of the electric field is changed at different intervals allowing the large DNA molecules to re-orient themselves at regular intervals and "snake" through the pores. The most commonly used instruments apply a contour-clamped homogeneous electric field (CHEF) (Chu et al., 1986). To optimize separation it is often necessary to vary the angle, pulse time and voltage. The top of the line CHEF electrophoresis unit is computerized; the desired fragment size range is entered and optimal separation conditions are automatically obtained. Fragment sizes are determined by comparison to molecules of known size. This also provides a means to perform between gel fingerprint comparisons.

The PFGE technique is time consuming and very tedious, thus may not be suitable for rapid identification of large number of strains (Willshaw et al., 1997) often necessary for MST. While PFGE requires a specialized gel rig with multiple electrodes configured in a hexagonal design, a chiller and pump, and programmable power supply, the operator does not require special molecular skills. However, the PFGE aparatus is more expensive that conventional gel electrophoresis. Since

only a limited number samples can be processed per gel, the number of available apparatuses is the limiting factor for high throughput analysis using this method. Sample preparation does require some training but with experience many samples can be prepared daily.

Application of PFGE to MST

This method is described as "superior to most other methods for biochemical and molecular typing" (Olive and Bean 1999). The Centers for Disease Control and Prevention has adopted this method for their "National Molecular Subtyping Network for Foodborne Disease Surveillance" mainly to discriminate *E. coli* O157:H7 and other foodborne pathogens. They have developed a network for health agencies to quickly compare molecular PFGE genotype data at a centralized website called PulseNet (http://www.cdc.gov/pulsenet/). It has been used successfully to rapidly compare PFGE profiles of suspect culture with those in the national database at CDC. In the future, this could serve as a model if EPA adopts any of the genotypic fingerprinting approaches for MST. However, publication of MST studies are much more limited. In a beach study, PFGE of *E. coli* was better for discriminating host sources compared to the other fecal coliforms, *Klebsiella*, *Citrobacter*, and *Enterobacter* spp (McLellan et al., 2001). As mentioned in the previous section, Hahm et al. (2003a) found levels of discrimination using PFGE similar to AFLP. The same study and another by McLellan et al. (2001) found that methods such as rep-PCR were less discriminatory than PFGE. However, high resolution between fingerprint patterns is not always ideal when genetic diversity is high between isolates taken from the same host animal. Greater genetic diversity translates into an increase in reference library size needed to differentiate isolates from different hosts. Also, care must be taken in the restriction enzymes chosen for this analysis because no relationship between fragment pattern and source was seen when the restriction enzyme *Sfi*I was used (Parveen et al., 2001).

Overview of PFGE methodology.

There are no standardized methods for PFGE for MST, but protocols set for CDC studies can be used for *E. coli* isolates. This method can be used on any bacteria, but conditions for optimal DNA extraction must first be determined. Isolates are first grown using standard conditions then DNA is extracted using an agarose plug total genomic DNA isolation method, which minimizes undesired breakage of the DNA. The cells are pelleted by centrifugation then suspended in unmolten low melt agarose or equivalent agarose specialized for PFGE. Sufficient microbial biomass must be used to have at least 1 µg of DNA in the plugs used for digestion. While still liquid, the agarose/cell solution is transferred to plug mold where it is left to solidify. Once solid, the plugs are removed from the mold and put through a series of steps to lyse the cells, remove proteins and degrade RNA. Depending on the protocol used, this process can take from a few hours to two days. Purified DNA still embedded in the agarose plugs is then digested with a rare cutting restriction enzyme. The most commonly used restriction enzyme is *Xba*I but others have also been tested and show variable results. Electrophoresis performed at 14°C with 6 V/cm, angle 120°, linear ramping factor and 30 hr running time will separate digested DNA fragments ranging between 100 kb and 500 kb in size. Gels are stained with ethidium bromide after fragments have been separated. It is often necessary to destain for several hours to optimize band contrast. Gel images can be digitized and then entered into a fingerprint analysis program (e.g., Bionumerics, Applied Maths). The data is binary, based on the presence and absence of bands in each profile (see data analysis Chapter).

Ribotyping

Ribotyping is a version of restriction fragment polymorphism (RFLP)-Southern hybridization analysis (Demezas, 1998; Sadowsky, 1994) that has found wide application in the subtyping of a variety of Gram-negative and Gram-positive bacteria (Olive and Bean 1999). It is another method that does not include PCR, except in the making of the labeled rDNA probe. The technique has been broadly used in molecular epidemiology (Bingen et al., 1992, Bingen et al., 1996, Picard et al., 1991), and taxonomic identification (Brisse et al., 2000) studies, including those with *E. coli* (LiPuma et al., 1989, Stull, et al. 1988, Tarkka et al., 1994). RFLP patterns of bacterial genomic DNA made with moderate cutting enzymes contain too many fragments for easy analysis, but ribotyping takes advantage of selective hybridization of a limited number of fragments for strain differentiation. Ribotyping is based on the detection of genetic differences in the genomic sequences within or flanking the 16S and 23S ribosomal RNA genes. Since rRNAgenes exists in several copies (2-11) in the bacterial genome and is highly conserved among bacteria, (Grimont and Grimont, 1986), hybridization of restriction enzyme-digested genomic DNA with labeled rDNA probes produces a ladder of labeled fragments that resemble a bar code. In addition, it has been recognized that since ribotyping produces relatively few bands for each strain (~5-15 for *E. coli*, depending upon the enzyme used and the strain), the technique is amenable to computerized analyses (Lefresne et al., 2004, Machado et al., 1998). If greater discrimination between strains is desired, more than one restriction enzyme can be used to digest DNA, and the banding patterns produced by each enzyme are combined to form a composite pattern (Harwood et al. 2003; Jenkins et al., 2003).

Ribotyping is a relatively demanding procedure requiring multiple steps and some specialized equipment. The need for specialized training, high supply costs and the time required to complete the procedure are disadvantages of using this method. However, the recent development of an automated ribotyping instrument, the Riboprinter (DuPont-Qualicon, Wilmington, Delaware) has promoted renewed interest in using ribotyping as a molecular tool for epidemiological, microbial source tracking, and clinical studies (Ito et al., 2003). However, the instrument has limited throughput, analyses are relatively expensive, and there have been reports that automated riboprinting may not be as reliable as manual methods (Grif et al., 1998). Despite these shortcomings, several microbial source tracking studies have used automated riboprinters to examine genetic diversity and groupings of fecal bacteria from known animal sources and the environment.

Application of ribotyping to MST

Ribotyping has been widely used in microbial source tracking studies (Farag et al., 2001; Carson et al. 2001; Carson et al. 2003; Hartel et al. 1999; Hartel et al. 2002; Harwood et al., 2003; Jenkins et al., 2003; Parveen et al., 1999; Scott et al., 2003). While the authors of these studies used the same basic technique, different laboratories have used different restriction enzymes in their analyses, and some have used a two-enzyme scheme, making comparisons difficult. As with any genotypic method, lab-to-lab variation, issues of repeatability, within and between gel variability and methods of analysis often make comparison of results done in different laboratories difficult (Lefrense et al., 2004). Moreover, several different studies done using slightly different procedures have reported

variable results with respect to the ability of ribotyping to differentiate among bacteria isolated from different animal hosts (Carson et al., 2003; Hartel et al., 2002; Parveen et al., 1999; Scott et al., 2003). Furthermore, database size, geographic distribution of the isolated bacteria, and the presence of replicate isolates in the bacterial source library impact the ability of ribotyping to differentiate among bacteria at the host species level (Scott et al., 2002; Scott et al., 2003).

Overview of ribotyping methodology

The ribotyping method is carried out in multiple steps. The technique involves restriction enzyme digestion of genomic DNA, separation of fragments by gel electrophoresis, immobilization of DNA fragments to a solid matrix (e.g., nylon membrane) by Southern transfer and subsequent hybridization using a labeled probe of the *E. coli* rRNA genes or the entire operon (Grimont and Grimont, 1986). Several different procedures can be used to isolate bacterial DNA (see Sadowsky 1994) for ribotyping and several different restriction enzymes may need to be tried to show differences at the strain level (Lefresne et al., 2004; Martin et al., 1996; Parveen et al., 1999, Scott et al., 2002). However, while *Eco*RI, *Pvu*II and *Hin*dIII have frequently been used for source tracking studies (Carson et al., 2001; Hartel et al., 2003; Scott et al., 2003; Vogel et al., 2000) it has been suggested that two enzyme systems should be routinely used to increase the technique's discrimination ability (Scott et al., 2003). The probes used for subsequent hybridization analysis can vary in the different regions of the *E. coli* rRNA operon used, but most investigators use the entire *E. coli rrnB* rRNA operon (Atwegg et al., 1989), only the 16S and 23S rRNA genes from *E. coli*, or mixtures of oligonucleotides complementary to specific regions in the operon (Gustaferro and Persing, 1992, Lafresne et al., 2004). The probe is usually generated by PCR, but can also be generated by nick translation or random primer labeling (Ausubel et al., 2004) and labeled with ^{32}P-, DIG-, or chemiluminescent-labels (Gustaferro and Persing, 1992; Regnault et al., 1997). Next hybridized fragments that constitute the ribotype banding patterns are detected using autoradiography or color formation. When the Riboprinter (DuPont-Qualicon, Wilmington, Delaware) is used, the sample (typically one bacterial colony) is added into the first tube and the instrument automatically carries out subsequent steps.

3.4 Cultivation-dependent/library-independent methods

When the target for MST is typically found in low numbers, it is first necessary to enrich the sample or obtain isolates. Enrichments are typically performed under conditions that favor the target organism. These methods are based on presence or absence of the target organism or gene therefore a source library is unnecessary.

F$^+$RNA coliphage typing

F$^+$RNA coliphages can help distinguish human and animal waste contamination by typing isolates into one of four subgroups (Alderisio et al., 1996; Brion et al., 2002; Cole et al., 2003; Griffin et al., 2000). Ecology studies have demonstrated that groups I and IV are generally associated with animal feces, whereas groups II and III are more sewage-specific (Furuse, 1987). Schaper et al. (2002a) found these associations to be statistically significant but also noted that exceptions occur.

Serotyping or genotyping can be used for typing of F^+RNA coliphages. In serotyping, group-specific antisera are used whereas in genotyping, hybridization with group specific oligonucleotides is used (Beekwilder et al., 1996; Hsu et al., 1995).

Coliphage cultivation techniques are simple with low supply costs (only plates and media), but require an overnight incubation step. Molecular methods have also been developed that allow for more rapid characterization of coliphages. For example, Vinjé et al. (2004) have developed an RT-PCR and reverse line blot hybridization technique capable of rapid detection and genotyping of coliphages. Additionally, phage characterization studies are underway which may allow for identification of more refined and host-specific subgroups. These advances could lead to an improved and more specific phage genotyping system.

Application of F^+RNA coliphage typing to MST

The use of coliphage typing for microbial source tracking is library independent, but can only currently be used to broadly distinguish human and animal fecal contamination. Coliphages have been detected in domestic, hospital, and slaughterhouse wastewaters (Funderburg and Sorber, 1985) and from treated wastewaters (Gantzer et al., 1998) but there appears to be some limitation when individual samples are used (Noble et al., 2003). Quantitative source tracking using F^+RNA coliphage typing may be problematic owing to differential survival characteristics of the subgroups (Brion et al., 2002; Schaper et al., 2002b).

Overview of coliphage typing methodology

Methods for isolation of coliphages include two standard USEPA procedures. One is Method 1601, a two-step enrichment procedure (USEPA, 2001a). The second is Method 1602, the single agar layer procedure (USEPA, 2001b). Method 1601 is more sensitive than 1602, but may not be the best choice for isolation of F^+RNA coliphages meant to be subsequently typed for microbial source tracking. The enrichment step likely excludes or masks other strains that may have been present in the original sample, typically resulting in only one strain of phage isolated from any given sample. The single agar layer procedure, on the other hand, is a pour plate technique from which viruses can be easily isolated for subsequent typing.

Isolated viruses are grown in the presence of RNase A to distinguish F^+RNA coliphages from F^+DNA coliphages. F^+RNA coliphages cannot form plaques when RNase A is present. Then either a serotyping or genotyping method is used to identify the F^+RNA coliphages. For serotyping, virus infectivity is tested in the presence of group-specific antisera. Inhibition of infectivity in the presence of a particular antiserum identifies the group to which an isolate belongs. Coliphages are genotyped by using hybridization of group specific labeled probes. Nucleic acid isolation is not necessary and plaques can be used directly for hybridizations. Group I, II, III, or IV specific probe sequences are used for hybridization (Hsu et al., 1995; Beekwilder et al., 1996). Identification of human source contamination is indicated by the hybridization of group II or III and animal sources by group I or IV.

Gene specific PCR

Gene specific PCR methods have been developed for *E. coli* carried by humans (Oshiro et al., 1997), cattle and swine (Khatib et al., 2002; Khatib et al., 2003,); and it is anticipated that methods will soon be available for *E. coli* carried by several other species of mammals and by birds. These methods are based on the discovery that certain enterotoxin genes are carried almost exclusively by *E. coli* that infect individual species of warm-blooded mammals; the STIb gene, the LTIIa gene and the STII gene are carried only by *E. coli* of human, bovine and swine origin, respectively. Similarly, enterococci virulence genes have been used as targets for host specific markers (Scott et al., 2005).

This two-step approach is relatively simple and can be performed within two working days. The biggest advantages of these gene specific methods are that they are highly specific and they are library independent. The biggest disadvantage is that the toxin genes are carried only by a small number of isolates, which makes it necessary to perform a cultural enrichment step prior to testing by PCR.

Application of gene specific PCR to MST

This method is still in the developmental stages and there are no publications with its application for MST. However, there is some indication that the prevalence of these genes in animal waste systems is greater than previously expected (Chern et al., 2004) suggesting that it has potential in the future.

Overview of gene specific PCR methodology

Samples (1 L) are collected in sterile containers and shipped on ice to the laboratory and are processed within 24 hours. Samples are processed using membrane filtration, with filters being placed on mTEC agar and mTEC agar plus Congo Red. The mTEC plates are incubated for 1.5 hours at 35°C then at 44°C overnight. The 10^0 and 10^{-1} dilutions are harvested after 24 hours and the DNA is recovered. When samples contain sufficient particulate matter to clog the filters, six filters of the 10^{-1} dilution are used. DNA extracts are pooled and stored at – 80°C until nested PCR amplification. Two sets of primers are used for each toxin trait, an outer primer set and a second set. All PCR amplicons are visualized through gel electrophoresis. Confirmation may be done by restriction fragment analysis or Southern blot hybridization using probes previously designed specifically for each toxin.

Recently, magnetic beads were used to increase the sensitivity of the LTIIa biomarker for cattle. In this method, total DNA was extracted either from the mTEC medium colonies or directly from the environmental samples. Next, the LTIIa gene was removed from the DNA mixture by hybridization with magnetic beads containing the LTIIa probe. Finally PCR was used to amplify the LTIIa gene for detection by gel electrophoresis and staining. The combination of magnetic beads followed by PCR resulted in an increase in sensitivity over the nested PCR technique by as much as 10,000 fold, even in the presence of PCR inhibitors such as humic acids (Tsai et al., 2003).

3.5 Cultivation-independent/library-independent methods

Cultivation-independent methods for MST are primarily based on nucleic acid techniques arising

from the field of molecular microbial ecology. Molecular microbial ecology began in the 1980's with the development of a phylogenetic framework for the placement of any organism into one of three domains (Bacteria, Archaea, or Eukarya) based solely its ribosomal RNA (rRNA) gene sequences (Head, et al. 1998; Olsen et al., 1986). As rRNA gene sequences accumulated into publicly assessable databases (Ribosomal Database Project (RDP) rdp.cme.msu.edu, GenBank at the National Center for Biotechnological Information (NCBI) www.ncbi.nlm.nih.gov), the level of classification based on rRNA gene sequences increased. Today, most organisms can be classified from Kingdom to the genus-species level based on their rRNA gene sequences. Phylogenetic analysis of microbial communities based on rRNA gene sequences has been applied to many environments including soil, water, extreme environments and animal gastrointestinal tracts (Zoetendal, 2004). In molecular microbial ecology, methods can be broadly grouped into three categories: 1) those designed to characterize or identify the members of a bacterial community; 2) those designed to measure large changes in community structure; and 3) those designed to identify or quantify specific members of a community (for reviews see Head et al., 1998; Zoetendal, 2004).

Total community analysis

Identification using 16S rRNA gene clone libraries

Microbial communities from environmental samples are frequently analyzed by the construction of 16S rRNA gene clone libraries. Clone libraries can also be made from other genes but currently the gene with the most available information is the 16S rRNA gene. Clone library construction and analysis is one of the more expensive and time-consuming cultivation-independent methods. The generation of clone libraries requires the combination of several molecular biological techniques including, nucleic acid extractions, PCR, DNA ligation, bacterial transformation, and plasmid isolation, which may take up to a week to perform. In recent years, these methods have been simplified by the use of commercial kits. Therefore, laboratory technicians with minimal training can successfully generate clone libraries. DNA sequencing involves the use of costly equipment and many laboratories send their DNA to specialized facilities for sequencing at a cost ranging from around $4.00 to $20.00 a sequence. Thus a large portion of the total cost is based on the number of clones sequenced. DNA sequence analysis of clone libraries generates a large amount of electronically archival data, which may be time consuming to process. The analysis of this type of data requires, at a minimum, an understanding of the publicly available sequence matching databases and programs. Realistically, the time and cost to perform this method, one-month and $5-10K for 100 clones, does not make it an appropriate choice for MST. Its value lies in research and development of new approaches for MST.

Application of 16S rRNA gene clone libraries to MST

Construction of clone libraries from water samples for MST is not widely used because hundreds of sequences are needed to accurately profile an entire community. However, with regards to MST, the cloning and sequencing of microbial communities from contaminated sites is useful for research purposes. At least two studies (Cho and Kim, 2000; Simpson et al., 2004) demonstrated that the native microbial communities in water are changed by the addition of fecal contamination. In both these studies fecal bacteria indicative of the host source either bovine (Cho and Kim, 2000) or

equine (Simpson et al., 2004) were detected. Also, the construction and analysis of smaller clone libraries (< 50 sequences) from environmental samples can be used to verify the specificity of specific primers (such as *Bacteroides* specific primers) used in PCR assays or verify the presence of host-specific bacteria in the environmental sample.

Overview of16S rRNA gene clone library methodology

In this method, nucleic acids are extracted and then amplified using primers designed to match the 16S rRNA genes from as many bacterial species as possible (for a review of available general primers see Baker et al., 2003). The 16S rRNA genes from the microbial community are cloned into plasmids and transformed into *E. coli* to construct a library containing many individual *E. coli* colonies, each containing a different 16S rRNA gene. Individual *E. coli* colonies are propagated, and the 16S rRNA genes carried in the plasmid are isolated and sequenced. The 16S rRNA sequences representing the microbes from the environmental sample are analyzed by comparison with other sequences in available databases using the BLAST program at NCBI (www.ncbi.nlm.nih.gov/BLAST) or the Similarity program at RDP (rdp.cme.msu.edu/html). Additionally, taxonomic or similarity relationships can be determined using cluster analysis and tree construction programs based on the number of matching base pairs between the sequences (Olsen, et al., 1986). When phylogenetic trees are constructed, the relationships between microbial sequences are generally presented as OTUs (operational taxonomic units), clusters, or clades, because phenotypic information is needed to describe or confirm bacterial species.

Community structure by fingerprinting

Fingerprinting methods are often used to monitor changes in a community or to compare communities because the expense and labor involved in the construction and analysis of clone libraries limits the number of samples that can be analyzed (Table 1). Essentially, all of the cultivation-independent fingerprinting methods examine DNA size or conformation profiles generated from a microbial community after PCR amplification of rRNA genes, or randomly amplified DNA fragments. The amplicons may be separated based on sequence-specific melting behavior of amplicons by denaturing gradient gel electrophoresis (DGGE) or temperature gradient gel electrophoresis (TGGE) (Muyzer and Smalla, 1998). In addition, one of the primers used for PCR amplification may be labeled fluorescently, and amplicons can be separated by size before restriction enzyme digestion (Length Heterogeneity Restriction Fragment Length Polymorphisms, LH-RFLP) or after restriction enzyme digestion (Terminal Restriction Fragment Length Polymorphisms, T-RFLP) (Liu et al., 1997). The underlying principle for all of these fingerprinting methods is that differences in banding patterns result from differences in microbial species comprising the community. Amplification with generalized PCR primers from environmental samples usually results in a large number of bands, which are analyzed by band matching computer programs and statistically using cluster analysis. DNA bands can be extracted from the gels and sequenced to identify the key members of the microbial community.

PCR methods using standard thermocyclers are relatively inexpensive and easy to perform with minimal training. Most fingerprinting methods can be performed in about one day with the electrophoresis separation run overnight, thus allowing data analysis the next morning. Differences in cost between fingerprinting methods will occur depending on the type of post-analysis performed.

In general, electrophoresis methods with better resolution require more costly equipment. For instance, gel electrophoresis equipment designed to separate PCR products by temperature gradient gel electrophoresis are more expensive than standard gel electrophoresis equipment. LH-RFLP and T-RFLP are one of the most expensive fingerprinting methods because separation of DNA fragments differing by only a single base pair requires acrylamide gels and DNA sequencing equipment. As with DNA sequencing, the separation of the DNA fragments on an automated DNA sequencer may be subcontracted to a specialized facility.

Application of community structure to MST

Although, fingerprinting analyses of fecal samples have been used to demonstrate host-specificity of the microbial community with the animal host (Zoetendal et al., 2004), cultivation-independent community analysis by fingerprinting has not been widely applied to MST studies. This is probably in part because in water samples the portion of the community that can be linked to host specificity may be very small compared to indigenous microbial community. Fingerprinting methods can be linked with more specific primers to produce fewer DNA bands. In one study relevant to MST, LH-PCR methodology was used with *Bacteroides* primers to identify a band size distinctive of bovine specific *Bacteroides* (276 bp) (Field et al., 2003). Additional digestion of the PCR amplified sequences with restriction enzymes (T-RFLP) resulted in the detection of two additional markers for bovine-specific *Bacteroides* and one marker for human-specific *Bacteroides*. These researchers previously demonstrated that LH-PCR could be used with *Bifidobacterium* specific primers to detect a bovine-specific amplicon of 453 bp. Digestion of the *Bifidobacterium* amplicons with restriction enzymes resulted in human and bovine specific fragments (Bernhard and Field, 2000a).

Overview of community structure methodology

Detailed explanations of community structure analysis are available from other sources (Liu et al., 1997; Muyzer et al., 1996; Nakatsu and Marsh, 2005). Briefly, both methods use PCR to amplify the rRNA gene. Typically, universal primers targeting the small subunit, 16S rRNA gene in bacteria are used to amplify sequences directly from DNA or RNA extracted from environmental samples. However, primers that amplify specific groups such as *Bacteroides* is often more useful for MST. In general, primers selected for T-RFLP amplify almost the entire 16S rRNA gene whereas in DGGE primers generating PCR products less than 500 bp are selected to reduce the occurrence of artifacts. In DGGE, the PCR products are directly analyzed by gel electrophoresis whereas in T-RFLP the PCR products are first digested with frequently cutting restriction enzymes before electrophoresis. In T-RFLP either one or both primers are labeled with different fluorescent tags to allow visualization and distinction of the end fragments using an automatic sequencing system. In DGGE, the PCR products are separated in gels composed of a gradient of chemical denaturants that causes differences in DNA migration based on their sequence. In both methods differences in migration of PCR amplicons either because of fragment sizes or sequence composition, generate a fingerprint of the community and a view of its complexity.

Alternate targets

While most of the culture independent/library independent methods have targeted fecal bacteria and

viruses, eukaryotic cells have also been suggested as useful markers for fecal source identification. For example, species of the genus *Cryptosporidium* has been shown to exhibit some degree of host specificity based on sequence differences in the small ribosomal subunit (Xiao et al., 2004). These differences have been used to characterize the primary fecal sources of surface water and wastewater (Xiao et al., 2001). However, because *C. parvuum* is found in relatively low numbers in environmental waters with moderate level of fecal contamination, their use in MST will have the same problems as with enteric viruses, this is, the need of concentration steps from large volumes of water.

Recently, PCR-based assays targeting host mitochondrial genes were used to discriminate between human, bovine, porcine, and ovine fecal samples (Martellini et al., 2005). The assays were developed to produce PCR products of different length facilitating their use in a multiplex PCR approach. The use of host mitochondrial PCR approaches is based on the fact that as gut epithelial cells become senescent they are shed into the gut lumen, after which they become part of the animal feces. The presence of relatively large numbers mitochondrial genes per eukaryotic cell increases significantly the detection sensitivity of this method. This is a significant advantage over other gene specific PCR methods which normally target markers with less than five copies per cell. The expected limited survival of gut epithelial cells might limit the use of this approach to recent fecal contamination events in areas nearby fecal inputs.

3.6 Identification and quantification of specific bacteria

Identification and quantification of microbes in environmental samples by cultivation independent methods is dependent on sequence information derived from clone libraries (see above section) or sequencing of genes from cultivated organisms. Identification and quantification methods can be divided into direct probing methods not requiring PCR or PCR-based methods.

Direct probing of specific genes

Originally, direct probing methods were used to quantify microbes in cultivation-independent studies (Giovannoni et al., 1998; Stahl et al., 1998). Hybridization methods usually use small oligonucleotide sequences (less than 25 base pairs), called probes, designed to hybridize with target DNA sequences. Direct probing methods are moderately time-consuming and may require specialized training depending on the method used to label and detect the probe. In recent years, the use of radioactive probes, which require licensees and training to use, have been replaced by non-radioactive labels. Filter membrane hybridization methods, such as dot blot hybridization or Southern blot hybridization, require multiple handling steps including DNA extraction, blocking, hybridization and washing. The total process may take one to three days depending on the method used to measure the amount of probe bound to the filter. The cost for reagents is relatively inexpensive. Fluorescent in situ hybridization (FISH) using fluorescently labeled probes can also be performed directly on bacterial cells on a microscope slide. The total process of fixing the cells to a slide followed by hybridization and washing takes one to two days and the cost of the reagents is also relatively inexpensive. However, visualization of the fluorescent signal in bacterial cells requires the use of a high quality epifluorescence or confocal laser scanning microscope and specialized imaging software. This equipment is expensive and requires specialized training.

Application of direct probing to MST

This method has not been used directly in any MST studies. Although numerous probes for quantifying fecal bacteria have been designed for dot blot hybridization (Matsuki et al., 2002, Wang et al., 2002), the method is used infrequently because quantitative PCR (QPCR) methods have a detection limit 0.01% compared to 10% for dot blot hybridization (Malinen et al., 2003). FISH is an effective method for monitoring population changes in fecal samples (Franks et al., 1998) but has not been widely applied to MST because the concentrations of bacteria in water samples are generally too low to measure by FISH and fluorescent microscopy. However, the coupling of flow cytometry with FISH may improve the sensitivity of detection and the number of samples that can be processed (Rigottier-Gois et al., 2003) allowing future MST applications.

Overview of direct probing methodology

In dot blot hybridizations, DNA extracts are bound to nylon membranes and probes are labeled with radioactive ^{32}P or non-radioactive labels. After hybridization and washing, the amount of radioactivity remaining on the filter corresponds to the amount of target signal present in the sample. In Fluorescent In Situ Hybridization (FISH) the probe is labeled with a fluorescent compound. The probe is hybridized with whole cells that are treated to make them more permeable. Cells that hybridize to the probe fluoresce when viewed under a fluorescent microscope (DeLong et al., 1989). Results for FISH are generally reported as the percent of the population that is positive to each of the group-specific probes (Santo Domingo et al., 1998).

Target specific PCR-based methods

In the 1980's, several bacteria including *Bacteroides* (Fiksdal et al., 1985), *Bifidobacteria* (Resnick and Levin, 1981) and *Rhodococcus coprophilus* (Mara and Oragui, 1981) were suggested as alternative host-specific fecal indicators to *E. coli* and coliforms. Although several of these bacteria showed promise, most of them were difficult to cultivate and required lengthy incubations periods (up to 3 weeks for *Rhodococcus coprophilus*) before colonies could be enumerated, thus making them impractical for MST. With the advent of cultivation-independent methods, several of these bacteria have been and are being reevaluated for use with MST. *Enterococcus* has also been suggested as an alternative host-specific indicator and has been well studied by cultivation-dependent methods. Therefore, it is logical that cultivation-independent assays have also been developed for *Enterococcus*.

In addition to the basic PCR method described earlier, variations have been developed that include the detection of several target DNA's simultaneously (multiplex PCR), increasing the sensitivity of detection by using two amplification steps (nested PCR) (Yang and Rothman, 2004) and quantifying the initial template by quantitative PCR (QPCR) also known as real time PCR (RT-PCR). PCR assays and real-time PCR assays have also been designed to detect and quantify common fecal bacteria in both humans (Bartosch et al., 2004; Liu et al., 2003; Malinen et al., 2003; Wang et al., 1996; Wang et al., 1997) and cattle (Tajima, 2001). Ultimately, some of these assays may prove

useful for MST, but they need to be tested for host-specificity before they can be applied for MST because not all fecal bacteria reflect host-specificity. Because each assay is specific to one species or subset of microbes, multiple assays will be needed in environmental samples with the potential to identify several sources of contamination (e.g., both human and cattle). In addition, the combination of assays may strengthen the argument for the source of contamination. For instance, samples with positive results for ruminant-specific *Bacteroides*, *Rhodococcus coprophilus* and a *Streptococcus bovis* would indicate cattle as a source of fecal contamination. Similarly samples with positive results for human-specific *Bacteroides* (or *B. fragilis*), *Bifidobacterium adolescence* or *B. dentium* and *Enterococcus* would indicate human as a source of fecal contamination.

Target specific PCR-based methods are probably the least expensive of the cultivation-independent methods. PCR-based methods require minimal personnel training and can be performed within one day. Although minimal training is needed to perform PCR, laboratories routinely performing target specific PCR must incorporate quality control measures to prevent cross-contamination of samples and false positives. Presence-absence PCR assays are less expensive than QPCR assays because they can be performed using standard thermocyclers and inexpensive gel electrophoresis equipment. QPCR requires a thermocycler with a fluorescent detector that costs at least $20,000 more than the standard thermocycler. However, presence-absence PCR assays are more time consuming than QPCR assays requiring 2-3 hours for the PCR step and 1-3 hours for the gel electrophoresis step. In QPCR, the complete PCR assay and analysis can be performed in less than three hours. Some QPCR thermocyclers are designed to be used in the field and can provide data within 30 minutes. Individual QPCR assays are also slightly more expensive than presence-absence assays because an additional fluorescently labeled probe must be added to the reaction.

What is QPCR?

Theoretically, the amount of DNA synthesized during each cycle doubles so that millions of copies of the target DNA are generated after 40 cycles. However, in reality PCR amplification slows and plateaus as the nucleotides used in DNA synthesis are exhausted. Thus, a maximum amount of DNA is accumulated independent of the starting template concentration and the amount of target DNA in the sample cannot be inferred from the band intensity determined after gel electrophoresis. To overcome this limitation, quantitative PCR (QPCR) methods were developed to allow estimation of the amount of the starting template DNA in an unknown sample. The most widely accepted QPCR methods use a fluorescent signal generated from a fluorescent DNA-intercalating dye (SYBR green), or fluorescently labeled oligonucleotide probes to monitor the amount of DNA generated after each PCR cycle (Ginzinger, 2002). In the SYBR green method, the SYBR green binds to the double-stranded DNA as it is synthesized, resulting in an increase in fluorescence. Because the SYBR green will bind to all double-stranded DNA including primer-dimers or other non-specific products, a melting curve analysis is run at the end of the PCR reaction to verify the specificity of the reaction (Klein 2002). Alternatively, fluorescence can be measured by using a third oligonucleotide sequence or probe containing a fluorescent label on one end and a quencher on the other end such as hydrolysis probes (TaqMan) (Livak et al., 1995) or hybridization probes (Molecular Beacons) (for a review see Ginzinger 2002). TaqMan probes are designed to hybridize to one strand of the DNA target during the annealing step. When the TaqMan probe does not bind to the DNA target, it does not fluoresce because a quencher blocks the fluorescent signal. However, as DNA is synthesized, the TaqMan probe begins to bind to the single stranded DNA immediately after the denaturation step (heating to 95^{o}C) of the PCR reaction. As the Taq

polymerase synthesizes a new DNA strand, it digests the TaqMan probe bound to the template strand and thus releases the fluorescent label resulting in a fluorescent signal. Therefore, as more target DNA is synthesized, the fluorescent output increases, resulting in sigmoid shaped fluorescence curves with respect to the number of cycles.

The calculation of target DNA copies per reaction for any QPCR assay (SYBR Green or hydrolysis probes) begins with the determination of a cycle threshold (C_T) value for each PCR reaction. The threshold is the point at which the signal generated from the sample is significantly greater than the background fluorescence, and the C_T is the cycle at which this occurs. The C_T is linearly correlated to the log of the copies per reaction for a set of standards, so the C_T of the unknown sample can be used to calculate the number of target copies in that sample. For additional reviews on real-time PCR and application of real-time PCR to environmental samples see Ginzinger (2002), and Klein (2002).

Application of target specific PCR to MST

Bacteroides

Currently, *Bacteroides* assays are the most widely used cultivation independent host-specific microbial assays for MST. The use of *Bacteroides* as a potential indicator was proposed in the mid-1980's because the amount of *Bacteroides* that could be cultivated from human fecal samples was around 1,000 fold greater than the amount of *E. coli* that could be cultivated from human fecal samples (Fiksdal, 1985). Additional research using cultivation independent methods indicated that *Bacteorides-Porphyromonas-Prevotella* group comprised 10-60% of the intestinal population from many animals including humans (Franks et al., 1998, Harmsen et al., 2002), cattle (Wood et al., 1998) and horse (Daly and Shirazi-Beechey, 2003). Kreader (1995) developed PCR primers and specific hybridization probes to distinguish three *Bacteroides* species and demonstrated that the *B. fragilis* group (*B. distasonis* and *B. thetaiotaomicron*) and *B. vulgatus* were at higher concentrations in human feces than in farm animal species (cattle, swine, horses, goats and sheep, and poultry). Bernhard and Field (2000b) demonstrated that *Bacteroides* isolated from ruminant and humans were host-specific and designed PCR primers to distinguish human-specific and ruminant-specific *Bacteroides.* The human-specific *Bacteroides* presence/absence PCR assay was used as part of a tiered approach to identify fecal contamination as human or non-human (Boehm et al., 2002). Recently, QPCR assays have been developed for the detection of all *Bacteroides* species (Dick and Field, 2004), human-specific *Bacteroides* (Seurinck et al., 2005) and bovine-specific *Bacteroides* (Layton, unpublished). A QPCR assay for the detection of *B. fragilis* from human fecal samples has been developed, but this assay has not been tested for host-specificity against fecal samples from non-human sources (Malinen et al., 2003).

Bifidobacterium

Bifidobacterium are a well-studied group of beneficial intestinal bacteria that have also been proposed as fecal indicator species. Several *Bifidobacterium* species have been proposed as being human host-specific including *B. adolescentis* (Matsuki et al., 2004, Bonjoch et al., 2004), *B. dentium* (Nebra et al., 2003; Bonjoch et al., 2004) and *B. longum* (Matsuki et al., 2004). General

PCR primers have been developed to detect all *Bifidobacterium* (Kaufmann et al., 1997), and several PCR platforms have been designed to detect individual species. These include PCR amplification with genus-specific *Bifidobacterium* primers followed by hybridization with a species-specific probe for *B. dentium* (Nebra et al., 2003) and multiplex PCR for the detection of *B. adolescentis* and *B. dentium* (Bonjoch et al., 2004). QPCR assays have been designed to quantify *B. longum* (Malinen et al 2003, Matsuki et al 2004), *B. adolescentis* and *B. dentium* (Matsuki et al., 2004). Two concerns with the use of *Bifidobacterium* as an indicator may be their short survivability in water (50% reduction in 10 hours, Resnik and Levin 1981), and its lower concentration in human feces than *Bacteroides* (Sghir et al., 2000). The combination of these two factors may make it more difficult to detect in the environment than *Bacteroides*. However, both *B. dentium* and *B. adolescentis* have been found in human sewage but not animal wastewaters (Bonjoch et al., 2004). In addition, the detection of human associated *Bifidobacterium* in water samples may indicate recent contamination events.

Streptococcus Lancefield Group D

The taxonomic group, *Streptococcus* Lancefield Group D contains both *Streptococcus* and *Enterococcus*. These bacteria are routinely isolated from fecal samples and were named according to the host from which they were isolated implying host specificity. It was generally believed that *E. faecalis* and *E. faecium* (formerly *S. faecalis* and *S. faecium*) were associated with humans (Vancanneyt et al., 2002), whereas *S. bovis* were specific to ruminants (Whitehead and Cotta 2000). However, more recent literature indicates that *S. bovis* isolates may not be completely host-specific, as *S. bovis* isolated from clinical samples may cause approximately 24% of the streptococcal infections resulting in endocarditis, meningitis and septicemia (Whitehead and Cotta, 2000). Although not applied to MST, primers have been designed to differentiate *Streptococcus bovis* strains isolated from rumen and humans sources (Whitehead and Cotta, 2000). Several QPCR assays also have been developed to detect *Enterococcus* species for application to drinking water and recreational water regulations. Frahm and Obst (2003) published primers and a probe sequence that matches a range of *Enterococcus* species, whereas Santo Domingo et al. (2003) published primers and a probe sequence specific for *E. faecalis*. For MST applications, additional research is needed to confirm host-specificity of the *S. bovis* and *Enterococcus* groups (Vancanney et al., 2002).

Rhodococcus coprophilus

This target has not been used in any MST studies and is still being tested for its distribution among hosts. *Rhodococcus coprophilus* was proposed as an indicator of fecal contamination from farm animals (Mara and Oragui, 1981). This bacterium inhabits the digestive system of almost all grazing animals and is passed to other animals grazing on the contaminated grass via the fecal oral route. The design of a TaqMan-based QPCR assay by Savill et al. (2001) allows continued testing of this bacterium as an indicator. Additional information is needed on the prevalence of this bacterium in the U. S. and the amount of bacteria contained in feces. It is likely that this bacterium persists longer in the environment than either *Bacteroides* or *Bifidobacterium* as it is aerobic and is passed between grazing animals.

Overview of target specific PCR methodology

Application of target specific PCR assays to water samples generally requires concentration of water samples for two reasons. First, in a PCR reaction the amount of target-containing sample added is only a few microliters (µL). Given the dispersed and dilute nature of bacteria in water, larger samples are needed for representative sample. Second, assuming a worst-case scenario where the detection of one copy of DNA in a PCR reaction is equal to one culturable bacterium, very high concentrations of bacteria (e.g., approx. 10^6) would be needed in the environmental sample. This is a worst-case scenario because even for easily cultivated bacteria such as *E. coli* only about 1% of the population can be re-grown from an environmental sample, thus the actual number of target bacteria in the sample is higher. For most situations a 100 ml water sample is suitable for analysis. Water samples are often concentrated by filtering a 100 ml aliquot through a 0.45-µm membrane filter. After filtration the DNA can be extracted from the filter (Boehm et al., 2003; Frahm and Obst, 2003), the bacteria enriched in nonselective broth (Frahm and Obst, 2003) or selective agar (Santo Domingo et al., 2003), or the bacteria can be eluted or washed off the filter and PCR performed without DNA extraction (Fode-Vaughan et al., 2001).

Identification and quantification of specific viruses

Identification of enteric viruses with limited host ranges can help distinguish sources of fecal pollution in water (Noble et al., 2003). Human-specific adenoviruses (Jiang et al., 2001; Pina et al., 1998) and enteroviruses (Griffin et al., 1999; Noble and Fuhrman, 2001) are candidate indicators for human fecal contamination. Bovine enteroviruses (Ley et al., 2002) and bovine and porcine adenoviruses (de Motes et al., 2004) have been proposed for detection of animal-source fecal contamination. Similarly, Teschoviruses have been used as an indicator of porcine fecal contamination (Jiminex-Clavero et al., 2003). Additional viral targets could also be appropriate for MST depending on host specificity and pending development of molecular assays.

Application of host-specific viruses to MST

This method is in developmental stages and the number of studies applying this approach is still limited, although recently assays targeting enteric viruses were used to detect human and bovine fecal contamination in coastal waters (Fong et al., 2005). In addition, a microbial source tracking methods comparison study found that detection of human viruses has among the lowest false positive rates for tested methods (Griffith et al., 2003; Noble et al., 2003). That is, human viruses were not identified in samples that lacked human-source contamination. However, the study also demonstrated that this approach fails to always detect contamination from individual humans; human viruses were detected in samples seeded with sewage but not in samples seeded with fecal material from individual humans. These results are consistent with the low carriage rate of viruses in the human population (Payment and Hunter, 2001).

Overview of host-specific viruses methodology

Molecular methods such as PCR allow rapid detection of viruses. These assays also tend to be more

sensitive than traditional cell culture, which can be technically difficult, time consuming, and inefficient (Schwab et al., 1995). Concentration and purification of viral nucleic acids from environmental samples can be challenging, but advances are being made within research laboratories to address these issues. Quantitative PCR assays have been developed for some viruses, which allows levels of viral contamination from various sources to be quantified.

Table 3.1 Summary of Logistics of Methods Tested for MST

METHOD	Targets tested	Cultivation	Library	Major Equipment Needs‡	Major Costs	Time Required*
Antibiotic Resistance	• *Escherichia coli* • Fecal streptococci • *Enterococcus spp.*	Individual Isolates	Yes	None	Antibiotics 96-well microplates	4-5 days
Carbon Utilization Profiles	• *Escherichia coli* • Fecal streptococci • *Enterococcus spp.*	Individual Isolates	Yes	None Plate reader (optional)	Microplates with substrates (e.g., Biolog, Phene Plate)	2-5 days
rep-PCR	• *Escherichia coli*	Individual Isolates	Yes	Thermal cycler Agarose gel electrophoresis units Gel documentation system Fluorescence scanner for HEFERP	PCR reagents PCR disposable Gel electrophoresis	1 day
RAPD	• *Escherichia coli*	Individual Isolates	Yes	Thermal cycler Agarose gel electrophoresis units Gel documentation system	PCR reagents PCR disposable Gel electrophoresis reagents	1 day
AFLP	• *Escherichia coli*	Individual Isolates	Yes	Thermal Cycler Automated sequencer	DNA extraction kit AFLP kit ($5 per reaction)	5 days
PFGE	• *Escherichia coli* • *Enterococcus spp.*	Individual Isolates	Yes	Thermal cycler Pulse Field Gel Electrophoresis Gel documentation system	Plug prep. reagents Restriction enzymes Gel electrophoresis reagents	2-4 days
Ribotyping	• *Escherichia coli* • Fecal streptococci • *Enterococcus* spp.	Individual Isolates	Yes	Agarose gel electrophoresis units Gel blotting/Hybridization oven Gel documentation system	DNA purification reagents Gel electrophoresis reagents Restriction enzymes Hybridization/ detection solutions Labeled gene probe	1-3 days
Phage Sero- or Geno-typing	• F+ coliphage	Individual Isolates	No	Hybridization oven None if serotyping	Hybridization/ detection solutions Labeled gene probe or Phage specific antigen	1-3 days
Gene Specific PCR	• *E. coli toxin genes*	Sample Enrichment	No	Thermal cycler Agarose gel electrophoresis units	PCR reagents PCR disposables	2 days

Table 3.1 Summary of Logistics of Methods Tested for MST (Cont.)

Host-specific PCR	• *Bacteroides* • *Bifidobacteria* • *Enterococcus* • *Rhodococcus* • *F+ coliphage* • *Enterovirus* • *Adenovirus*	None	No	Thermal cycler Agarose gel electrophoresis units	Filtration units PCR reagents PCR disposable	6-8 hours
Host-specific QPCR	• *Bacteroides* • *Rhodococcus* • *Bifidobacteria*	None	No	Fluorescent Thermal Cycler	Filtration units PCR reagents/label PCR disposable	1-3 hours

‡ All methods require standard microbiological equipment, such as, micropipettors ($200-300 each), microcentrifuge ($1-2K), in methods requiring cultivation growth chambers (incubators) are needed

Major equipment costs are in the range of: Microcentrifuge ($1-2K), thermal cycler ($5K), thermal cycler with fluorescence detector for quantitative PCR ($25,000-$90,000), automated sequencer ($55K), submarine agarose gel unit with power supply ($1-2 K), PFGE unit ($11-25K), riboprinter ($175K), gel documentation system ($2-15K), statistical analysis software ($8-15K) needed for all library-dependent methods

Reagent costs: PCR ($2-$10/reaction including primers), filters to concentrate water samples ($4/ sample), all molecular method using gel electrophoresis require agarose and buffer solutions

*Time after enrichments or isolation performed, time for isolation dependent on target and method used for isolation and confirmation can vary considerably. Also, time required for data analysis for library dependent methods are not included because it is highly variable and dependent on available gel and data analysis software.

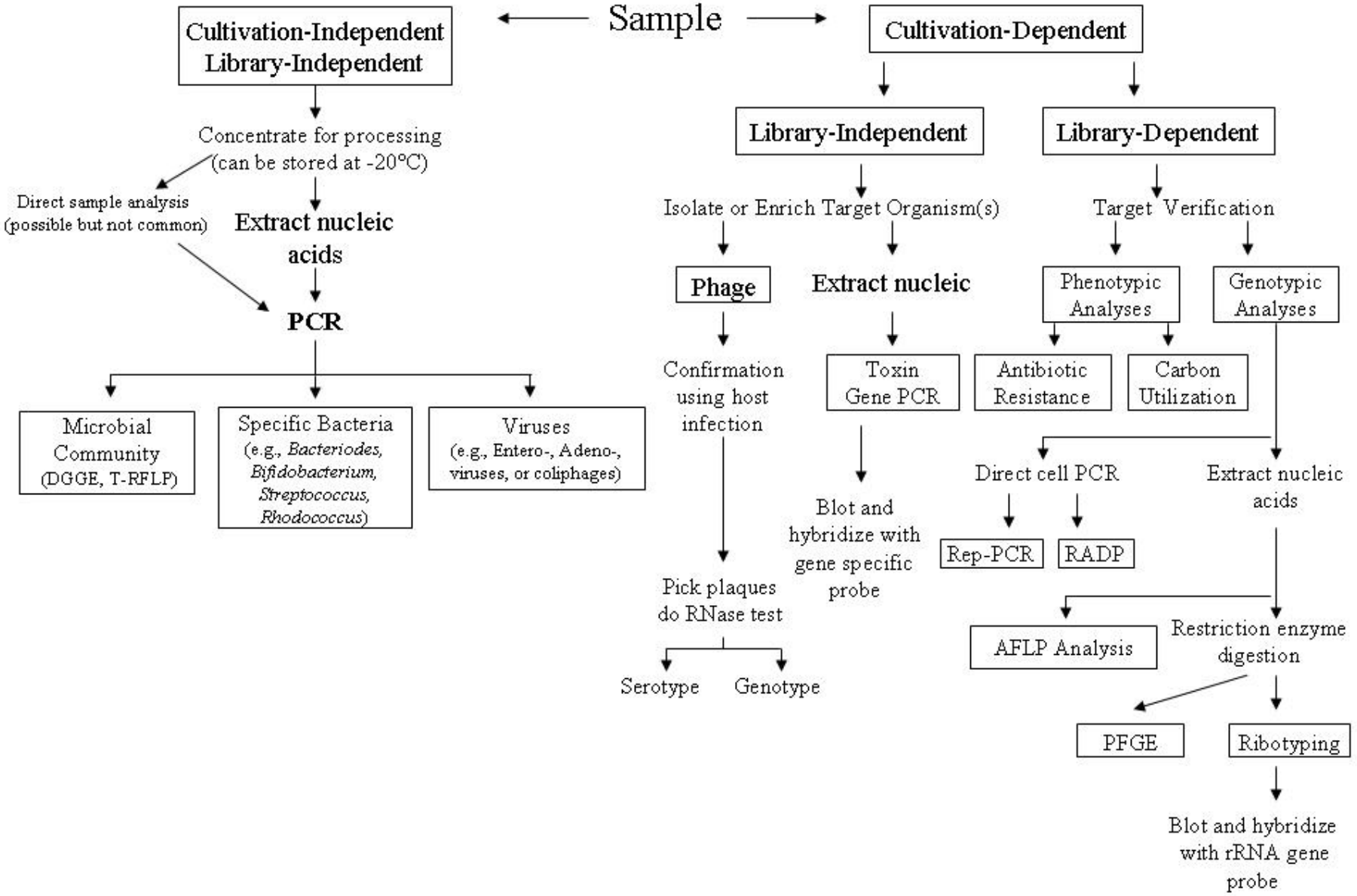
Table 3.2 Comparison of advantages and disadvantages of source tracking methods*

METHOD	ADVANTAGES	DISADVANTAGES
Antibiotic Resistance	• Rapid; easy to perform • Requires limited training • May be useful to differentiate host source	• Require reference library • Requires cultivation of target organism • Libraries geographically specific • Libraries temporally specific • Variations in methods in different studies
Carbon Utilization Profiles	• Rapid; easy to perform • Requires limited training	• Require reference library • Requires cultivation of target organism • Libraries geographically specific • Libraries temporally specific • Variations in methods in different studies • Results often inconsistent
rep-PCR	• Highly reproducible • Rapid; easy to perform • Requires limited training • May be useful to differentiate host source	• Requires reference library • Requires cultivation of target organism • Libraries may be geographically specific • Libraries may be temporally specific
RAPD	• Rapid; easy to perform • May be useful to differentiate host source	• Requires reference library • Requires cultivation of target organism • Libraries may be geographically specific • Libraries may be temporally specific • Has not been used extensively for source tracking
AFLP	• Highly reproducible • May be useful to differentiate host source	• Labor-intensive • Requires cultivation of target organism • Requires reference library • Requires specialized training of personnel • Libraries may be geographically specific • Libraries may be temporally specific • Variations in methods used in different studies

Table 3.2 Comparison of advantages and disadvantages of source tracking methods (Cont.)

PFGE	• Highly reproducible • May be useful to differentiate host source	• Labor-intensive • Requires cultivation of target organism • Requires specialized training of personnel • Requires reference library • Libraries may be geographically specific • Libraries may be temporally specific
Ribotyping	• Highly reproducible • Can be automated • May be useful to differentiate host source	• Labor-intensive (unless automated system used) • Requires cultivation of target organism • Requires reference library • Requires specialized training of personnel • Libraries may be geographically specific • Libraries may be temporally specific
F+ RNA coliphage	• Distinguishes human from animals • Subtypes are stable characteristics • Easy to perform • Does not require a reference library	• Requires cultivation of coliphages • Sub-types do not exhibit absolute host specificity • Low in numbers in some environments
Gene specific PCR	• Can be adapted to quantify gene copy number • Virulence genes may be targeted; providing direct evidence that potentially harmful organisms present • Does not require reference library	• Require enrichment of target organism • Sufficient quantity of target genes may not be available requiring enrichment or large quantity of sample • Requires training of personnel • Primers currently not available for all relevant hosts
Host-specific PCR	• Does not require cultivation of target organism • Rapid; easy to perform • Does not require a reference library	• Little is known about survival and distribution in water systems • Primers currently not available for all relevant hosts
Virus specific PCR	• Host specific • Easy to perform • Does not require reference library	• Low in numbers, requires large sample size • Not always present even when humans present

All methods require validation. All methods require personnel trained in basic microbiology and potentially basic molecular biology skills (e.g., PCR and agarose gel electrophoresis), and only those requiring specialized training are labeled.


Sample
Cultivation-Independent
Library-Independent
Concentrate for processing
(can be stored at -20°C)
Direct sample analysis
(possible but not common)
Extract nucleic acids
PCR
Microbial Community
(DGGE, T-RFLP)
Specific Bacteria
(e.g., Bacteriodes, Bifidobacterium, Streptococcus, Rhodococcus)
Viruses
(e.g., Entero-, Adeno-, viruses, or coliphages)
Cultivation-Dependent
Library-Independent
Library-Dependent
Isolate or Enrich Target Organism(s)
Target Verification
Phage
Extract nucleic
Phenotypic Analyses
Genotypic Analyses
Confirmation using host infection
Toxin Gene PCR
Antibiotic Resistance
Carbon Utilization
Direct cell PCR
Extract nucleic acids
Blot and hybridize with gene specific probe
Rep-PCR
RADP
Pick plaques do RNase test
AFLP Analysis
Restriction enzyme digestion
Serotype
Genotype
PFGE
Ribotyping
Blot and hybridize with rRNA gene probe

Chapter 4. Data Collection and Analysis in Library-dependent Approaches

4.1 Introduction

Data collection and analysis are two critical components of microbial source tracking (MST) that require careful attention in the planning stages of any study. Different approaches to MST produce different types of numerical data and consequently require different considerations and strategies in sampling and analysis. This Chapter highlights key issues in sampling design and data representation and discusses several statistical methods commonly used in various stages of MST. The discussion will be limited to library-dependent methods as they pose the most technical challenge from a statistical point of view. Library-independent approaches to MST use host-specific markers to identify contaminant sources (See Sections 4.3-4.4.). The existence of host-specific markers reduces the dependence on libraries, which are subject to geographic and temporal variability. However, while the presence of a host-specific marker enables source identification with near certainty, the absence of the same marker does not necessarily exclude any host from consideration. In addition, there are currently only a limited number of hosts for which such markers have been found. Therefore, at present, library-independent approaches to MST may need to be used in conjunction with a library-dependent approach and the associated statistical analysis would require similar adaptation. For example, a simple two-stage procedure could be used which first screens for host-specific markers and then resorts library-dependent methods and statistical analyses if none are found.

4.2 Data collection

Effective MST requires that appropriate data are collected to meet the objectives of the study. For example, an analysis that indicates cattle as the major source of fecal contamination to a stream on 70% of dates sampled may not be particularly meaningful if the stream did not exceed regulatory criteria on those days. Despite dominance by cattle contamination on most dates, humans could very well be the major source on exceedance dates and, therefore, the logical target of remediation efforts. The sampling plan must be designed around the objectives of the study.

Applications of source tracking could use various sample schedules to accomplish their objectives. For example, in applications to total maximum daily load (TMDL) water quality assessments, it might be essential to evaluate contributions at all concentration levels across all seasons, while for application to beach closures it might be more important to evaluate contributions when concentrations exceed regulatory limits during the recreational season.

Most water bodies, whether streams, lakes, or aquifers, are not well mixed so a single sample does not represent the entire water body. In moving water, in particular, short-term variability must be considered because a single enriched particle can greatly skew the results from that sample. Furthermore, transient animal populations mean that potential contributors change with season and hydrology creates different flowpaths from those contributors with weather and season. A single sample should rarely, if ever, be interpreted as a comprehensive indicator of pollution status across the entire water body, the entire year, or all flows.

Some general principles to follow in sampling watersheds of various kinds include:

1. Composite samples are preferred to single dip samples in order to include more of the entire cross-sectional area or volume of the sampled water body.
2. Taking several replicate samples or compositing samples over time helps to even out short-term variability. (Hyer and Moyer, 2003)
3. Existence of transient animal populations implies that the known-source library may not be useful in all seasons (Haack et al., 2003). This stresses the need that the known-source library should be collected concurrently with water samples.
4. Different sources of fecal contamination could be expected in storm flow from base flow and this should be taken into account in the sampling plan (Hartel, 2004). For instance, fecal pollution in base flow is generally considered to be from ground water seep (including leaky sewer lines and leach fields), direct deposition by wildlife, and various NPDES-permitted effluents. Fecal pollution in storm flow, on the other hand, is transported with overland flow (including field-spread manure), stormwater discharges (including combined-sewer overflows), and other flooded areas (Tian, 2004).

4.3 Numerical representation of isolate profiles

As has been stated in previous Chapters, the majority of currently applied approaches for microbial source tracking are library-dependent. That is to say, they rely on a collection of isolate profiles (fingerprints, banding patterns or discrete data) from each source category and the information contained in this library of isolate profiles forms the basis for classifying indicator organisms of unknown origin by source category. Both genotypic and phenotypic library-dependent approaches are currently employed for MST. Genotypic approaches characterize isolates based on DNA-based characteristics, often visualized as banding patterns of DNA fragments on agarose or polyacrylamide gels, whereas phenotypic approaches characterize isolates based on their observable physiology or growth characteristics on specific laboratory media, or via quantitative measurements of traits like cell surface antigens or resistance to antibiotics. Compilations of genotypic and/or phenotypic characteristics can be measured and used to define a reproducible profile or fingerprint for each isolate. However, there are usually several ways to represent an isolate's characteristic profile as numerical data, and decisions about data representation can have a significant impact on both sampling and analysis strategies and outcomes.

Genotypic data, such as an isolates DNA fingerprint, can generally be represented numerically as (1) a "continuous" intensity curve where peaks represent the location (fragment size) of bands and the heights (and/or areas) of the peaks are a quantitative measure of a bands intensity (See Figure 1), (2) a discrete listing of band locations and intensities, defining presence and magnitude of a finite set of bands from a list of possible band fragment sizes or (3) a discrete profile listing of band locations, as binary (presence/absence) data, defining only the presence of a finite set of bands from a list of possible band fragment sizes (See Figure 1). While the latter method has frequently been used for genotypic profiles having simple banding patterns (those having a limited number of bands), more complex band patterns are often analyzed using data derived from fragment location and band intensity. In cases where numerical differences in band intensity are theoretically meaningful, either

of the quantitative representations is preferable to the binary representation. For example, in the analysis of PFGE profiles, high band intensity may indicate the presence of multiple fragments

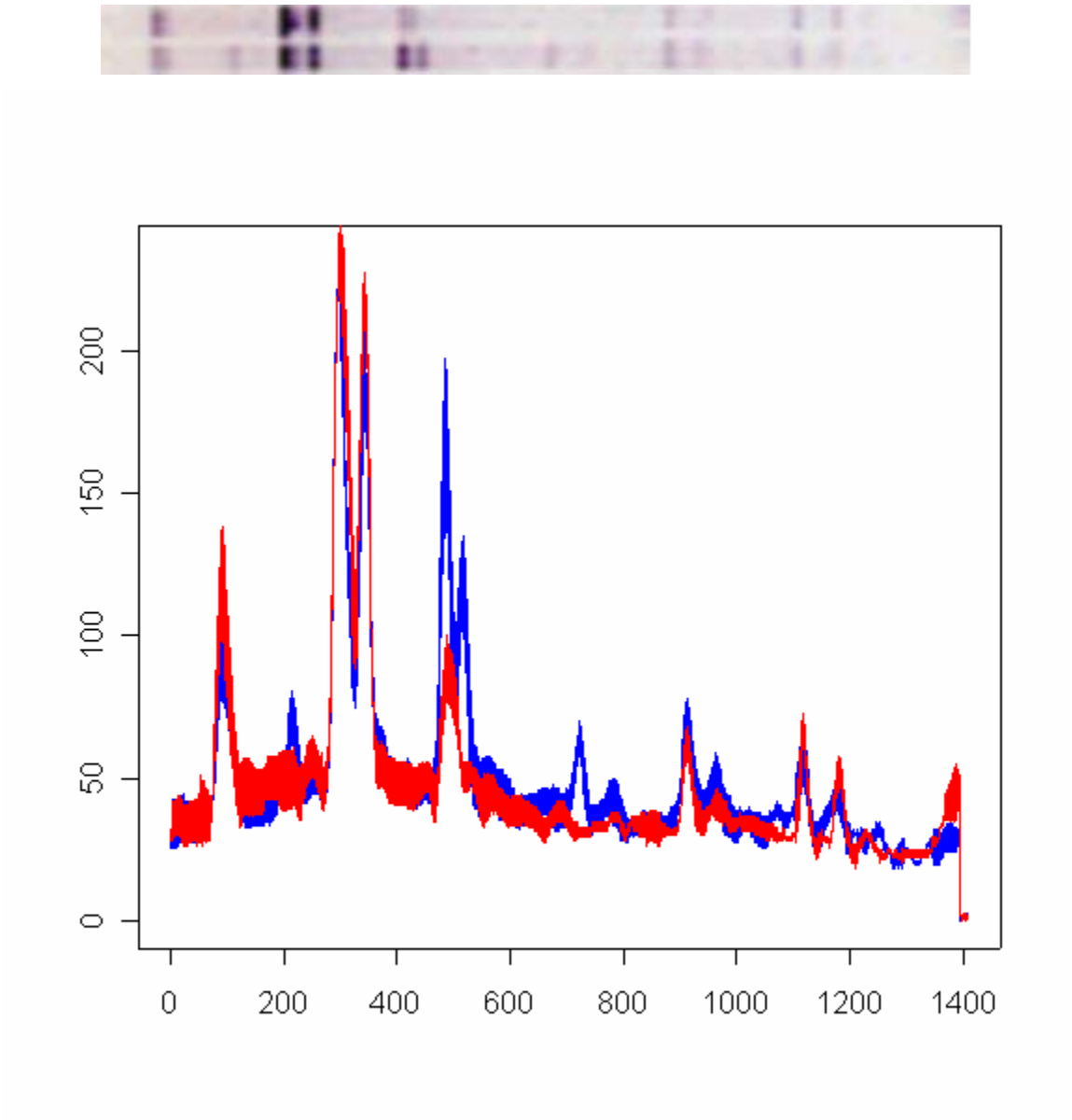


Figure 4.1 Two genotypic fingerprints from Figure 1C of Chapter 2 with corresponding curve representation

of similar length. Similarly, in PCR-generated DNA fingerprints, enhanced band intensity may be due to PCR bias, or target copy number. In extreme cases, alternate numerical representation that takes these factors into account might yield useful information.

Thus, a key factor to consider when deciding whether to use quantitative values or presence/absence character tables is the interpretability of the quantities measured by each variable in the numerical

profile. A confounding influence, however, is that laboratory and image processing protocols can also affect numerical representation and, consequently, the analysis of genotypic profiles. In particular, it is essential that data profiles for each isolate be carefully aligned such that all common bands are, in fact, positioned at the same location. This task is more difficult than it may at first seem. The default settings of software packages commonly used for genotypic fingerprint analysis are designed to aid, but cannot by themselves ensure, accurate alignment. Therefore, the incorporation of subjective judgments by an experienced analyst is required.

The numerical representation of phenotypic isolate profiles is more straightforward. Phenotypic profiles sometimes consist of a series of quantitative measurements of phenotypic traits, such as in antibiotic resistance analysis (ARA, see Chapter 3) in which growth in the presence of serial concentrations of antibiotics is tested. Phenotypic profiles can also consist of binary character tables, such as in multiple antibiotic resistance (MAR, see Chapter 3), where resistance to only one concentration of each of several antibiotics is measured and carbon utilization patterns (See Chapter 3), where a substrate may or may not support growth of a strain. In this case, the only numerical representation is a profile of binary variables which indicates growth or absence of growth under the test conditions. Although discrete data such as that collected in ARA could also be depicted as binary character tables, it is preferred to record quantitative values (maximum concentration at which growth was not inhibited for each antibiotic) because data for each concentration of antibiotic tested are not independent.

4.4 Library construction and validation

Sample size and library representativeness

Library-dependent MST studies require the creation of a known source database to which unknown field isolates are compared. Library size and the representativeness of strains in a known-source the library are two major considerations that need to be carefully assessed before embarking on any MST study. The same considerations must be given to MST studies done using phenotypic or genotypic data, although the final number of strains in a known source database may vary depending on the methods chosen. Moreover, based on usually empirical information, one must carefully weigh decisions on whether to take a large number of samples from a few animals or a lesser number of samples from a large number of animals. Generally speaking, a library needs to be large enough to (1) capture the total genetic diversity present within the population of indicator bacteria in a given host animal and (2) be of sufficient size so that environmental isolates can be reliably typed to host origin. The ultimate size of the known source database library is also linked to the size of the watershed under consideration and the number of potential sources in the watershed. For example, a smaller library will be needed if a watershed is primarily inhabited by a limited number of potential animal sources that occupy a limited geographic location.

The genetic diversity of indicator bacteria (most people use databases consisting of *E. coli* or eneterococci) in a given animal host is related to feeding habit, food sources, diet variation in a host animal group (Hartel et al., 2003), fecal contamination from other animals, temporal and geographic variation of bacterial genotypes within and between animal species (Gordon, 2001; Hartel et al., 2002; Scott et al., 2002; Jenkins et al., 2003) and the number of strains in a single animal (McLellan et al., 2003). Accordingly, estimates of library sizes are often difficult to make without empirical

data. Generally speaking, most genotypic-based MST studies that have been conducted to date have used relatively small host origin databases, containing between 35 and approximately 500 isolates (Johnson et al., 2004). A small library size makes comparisons to populations of *E. coli* and *Enterococcus* in the environment difficult, mostly due to the large number of unidentified strains that result from such analyses. Recently, Johnson and coworkers (Johnson et al., 2004) reported that library size and representativeness have a major influence on the accuracy of MST studies. In contrast, many phenotypic-based MST studies, mostly done using antibiotic resistance patterns, have used larger known-source libraries consisting of about 1,000 – 6,000 isolates (Johnson et al., 2004). In many cases, however, the strains examined have been isolated from the same source animal or sample, introducing biases due to the presence of multiple replications of the same bacterial genotype.

There are several methods available to measure the representativeness of known-source libraries. Many of these methods, however, are empirical in nature. Rarefaction analysis has been considered a useful tool for comparing species richness and diversity. This type of analysis has been used in MST studies and provides a statistical method for estimating the number of genotypes that are expected to be present in a random sample of individuals. The data requirements for the rarefaction analyses are not exacting and do not require abundance information (Koellner et al., 2004). Rarefaction analysis estimates the rarity of a given genotype in a population by calculating a series that approximates the number of genotypes present in randomly and successively drawn subsets of the original database. This method allows for the generation of a rarefaction curve that allows comparison of the observed richness (diversity) among randomized library entries by averaging randomizations of the observed accumulation curve (Heck et al., 1975). If a library is "saturated" with genotypes, the rarefaction curve will appear to have a horizontal asymptote, indicating that additional library entries do not appreciably increase the number of new genotypes uncovered. In contrast, rarefaction curves that appear linear indicate that the library is not saturated with respect to diversity of genotypes. As such, additional library entries are needed to be useful to type unknown environmental isolates. As a consequence, it has been suggested that a library size of tens of thousands of *E. coli* isolates may be needed to capture all the genetic diversity present in natural populations (Mansour Samadpour, personal communication).

The representativeness and fidelity of known-source libraries can also be ascertained by applying jackknife analysis and reporting the average rate of correct classification (ARCC). This method of analysis is frequently reported in MST studies (e.g., Harwood et al., 2000.). The ARCC simply calculates the number of library isolates assigned to the correct source group when the library is queried using "hold-out" or Jackknife analyses. To do this, each isolate is individually removed from the database. The degree of similarity of the removed isolate to those remaining in each source group is determined, and then the average rate of correct classification is determined. Library entries that are incorrect or small libraries containing insufficient entries to capture all the genetic diversity will have lower ARCC values.

4.5 Measuring spatial and temporal variability

There are several statistical techniques available to measure and compare patterns of spatial and temporal variability. Among these are exploratory graphical techniques such as multi-dimensional

scaling (MDS) or principal components analysis (PCA) and confirmatory analyses performed using statistical techniques such as multivariate analysis of variance (MANOVA). The goal of an exploratory analysis is to identify patterns of variation in the data relevant to assumptions and hypotheses (Chapter 3). The goal of a confirmatory analysis is to test the validity of specific assumptions and hypotheses which may have been formulated based on observations made during an exploratory analysis.

Multidimensional Scaling

MDS (Torgerson, 1958) is a technique for representing a dissimilarity (or distance) matrix in relatively few dimensions. To illustrate the usefulness of MDS, consider a data set like the distance tables at the back of a road atlas, giving driving distances between major cities. The MDS algorithm could accurately reconstruct a map of the United States from this matrix of distances. The distance between pairs of cities in a MDS map of the United States would be roughly proportional to the corresponding geographic distances. (However, the map itself might be rotated or inverted.) This is illustrated below in the Figure 2 which is an MDS map of the ten USEPA Regional Offices, which was constructed by applying MDS to the table of geographic distances in Table 4.1. If the user did not know the geographical relation between the cities listed in the road atlas, the MDS map would be helpful for identifying geographic relationships.

In the context of MST, MDS plots are based on a matrix of numerical inter-isolate dissimilarity measures (instead of driving distances). Patterns of inter-isolate variation can be represented in a two or three-dimensional plot in which distances between points are roughly proportional the dissimilarity between the isolates they represent. As with PCA, this technique allows the multi-dimensional isolate profile data to be plotted in two or three dimensions and aids in the identification of major sources of variation.

Table 4.1 Geographic distance between each pair of USEPA regional offices.

	Boston	New York	Philadelphia	Atlanta	Chicago	Dallas	Kansas City	Denver	San Francisco	Seattle
Boston	0									
New York	200	0								
Philadelphia	300	110	0							
Atlanta	1100	850	750	0						
Chicago	1000	810	790	710	0					
Dallas	1750	1560	1440	820	920	0				
Kansas City	1440	1230	1170	820	540	510	0			
Denver	2000	1790	1740	1430	1020	780	610	0		
San Francisco	3130	2930	2900	2480	2170	1750	1860	1260	0	
Seattle	3020	2840	2820	2630	2050	2130	1860	1340	810	0

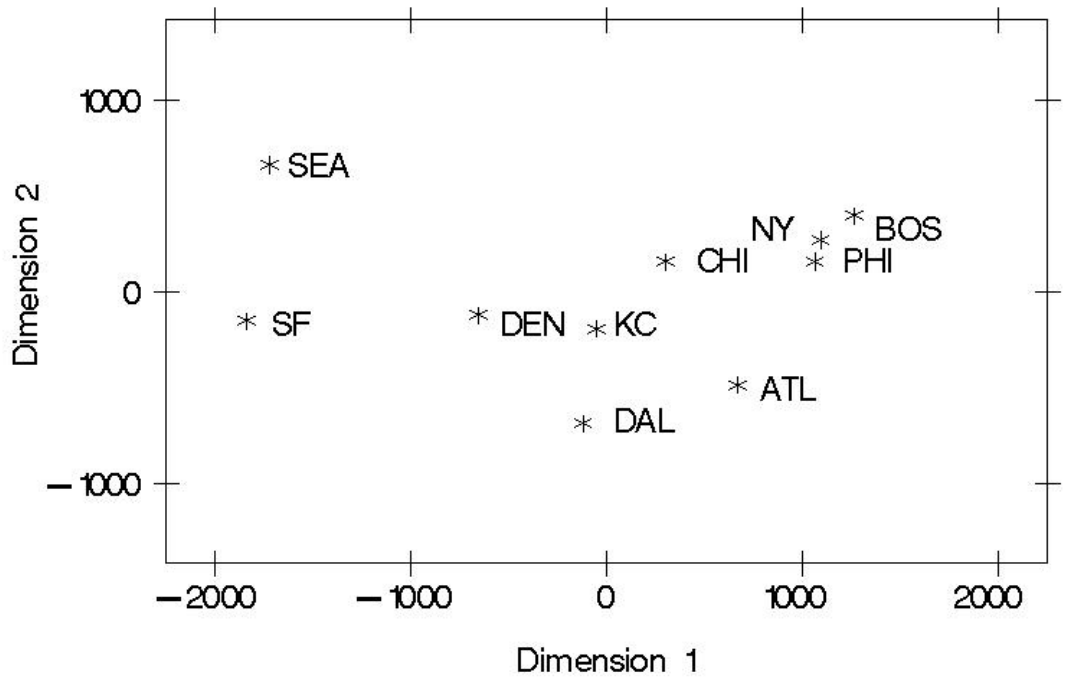


Figure 4.2 MDS map of the ten USEPA regional offices constructed from the table of geographic distances in Table 1.

More about similarity coefficients

An underlying assumption of many MST analyses, including MDS, is that distance (or dissimilarity) between isolate profiles can be measured numerically in some meaningful way. If the measure of dissimilarity is inappropriate or inaccurate then any inference drawn from MDS is invalid. Most software packages used for the analysis of MST data provide the user with several options for measuring inter-isolate similarity and distance.

Typical choices for similarity measurements among binary profiles include the similarity coefficients of Jaccard (1901), Dice (1935), Sokal and Michener (1958), Ochiai (1957) and Kulczynski (1928). Each of these coefficients can be expressed in terms of four quantities: (1) the number of bands common to both isolate profiles, denoted by *a*, (2) the number of bands present in the profile of the first isolate, but absent in the profile of the second isolate, denoted by *b*, (3) the number of bands present in the profile of the second isolate, but absent in the profile of the first isolate, denoted by *c*, and (4) the number of bands that are absent in both of the isolate profiles being compared, but present in at least one of the other isolates in the library, denoted by *d*. For simplicity of notation, the total number of distinct band locations for the entire library is often denote by $p=a+b+c+d$. The expression for six of the most commonly used measures of similarity between binary profiles are presented in Table 2 along with several properties of each that are relevant to the interpretability, and thus the selection, of a similarity measure.

		Isolate 2	
		pos	neg
Isolate 1	pos	a	b
	neg	c	d

Each of the six coefficients range in value from 0 to 1, where values near 0 indicate extreme dissimilarity and values near 1 indicate extreme similarity. All coefficients, with the exception of the simple matching coefficient of Sokal and Michener equal 0 when two isolate profiles contain no matching bands (i.e., $a=0$) and all coefficients except the Russell-Rao coefficient equal 1 when isolate profiles contain no mismatched bands (i.e., $b=c=0$). Both of these exceptions are due to the coefficients dependence on *d*, the number of bands not present in either of the isolate profiles being compared. Therefore, if this quantity is not meaningful in the context of an analysis, these two coefficients are inappropriate.

Another property of these coefficients displayed in Table 2 is whether or not its corresponding dissimilarity measure (usually one minus the similarity measure) satisfies the properties of a distance metric. In particular, it should be noted that the coefficients of Dice, Ochiai and Kulczynski (i.e., "Jeffrey's x") do not satisfy these conditions, in particular the triangle inequality. Therefore, while all these are valid measures of inter-isolate similarity, their use for MDS may result in some minor distortion. Shi (1993) makes a much more extensive comparison of these six coefficients and several alternative measures of similarity for binary profiles.

Table 4.2 Common similarity measures for binary profiles

Coefficient	Mathematical expression	Value when a=0 (no matches)	Value when b=c=0 (no mismatches)	1 – value is a distance metric	Suitability for MDS	Alternate names
Jaccard (1901)	a/(a+b+c)	0	1	Yes	High	Coefficient of community
Sokal and Michener (1958)	(a+d)/p	d/p	1	Yes	Moderate	Simple matching
Dice (1945)	2a/(2a+b+c)	0	1	No	Moderate	Sorensen (1948)
Ochiai (1957)	a/sqrt((a+b)(a+c))	0	1	No	Moderate	Coefficient of closeness
Kulczynski (1928)	a(2a+b+c)/[2(a+b)(a+c)]	0	1	No	Low	Jeffrey's x
Russell and Rao (1940)	a/p	0	a/(a+d)	Yes	Low	

Principal components analysis

PCA (Hotelling, 1933) is a statistical technique for dimension reduction and identification of dependence patterns among variables. PCA uses the interdependence between the original set of variables, as measured by correlation or covariance, to reduce the data set to a smaller set of variables called principal components. The principal components reproduce patterns present in the full set of variables and are easier to visualize. For example, two or three principal components can sometimes be used to summarize data for 50 or more of descriptive variables such as bands in a fingerprint. The major assumption of PCA is that the dependence between variables is fully described in terms of pairwise covariances and that this covariance structure is similar for the entire population.

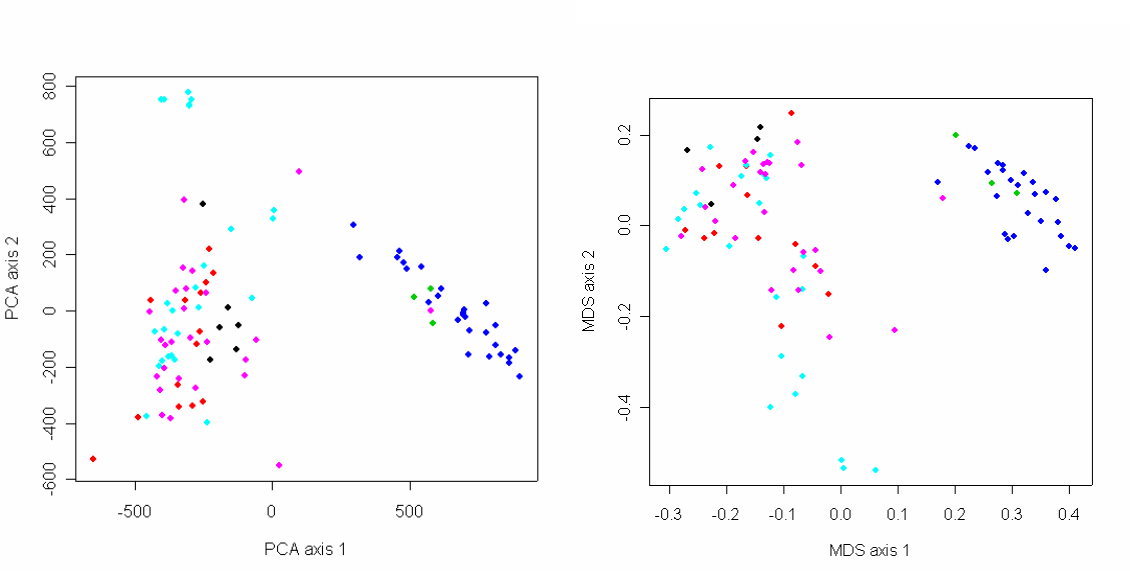


Figure 4.3 PCA and MDS plots of a small library of isolates where different colors indicate different source categories.

While PCA and MDS are both useful techniques for representing multivariate data in relatively few dimensions, it should be emphasized that both the underlying assumptions and objectives of these two methods are quite different. In particular, PCA is based on measurements and assumptions about variable interdependence and MDS is based on measurements and assumptions about inter-isolate similarity (See Table 3.). As a matter of practice, however, one often observes strong similarities between MDS and PCA plots (See Figure 4.3). If there is a natural similarity among profiles, whether based on host of origin, time frame, or geographic location, it may be detectable by either approach. Further discussion of both MDS and its connections to PCA can be found in Cox and Cox (2001).

More about Principal Components Analysis

Principal components are uncorrelated linear combinations of the original variables. That is to say, if the original variables are represented by $X_1, X_2, \ldots, X_p$, then each principal component is of the form $a_1X_1+a_2X_2+\ldots+a_pX_p$ where the variable weights $a_1,a_2,\ldots,a_p$ are based on the interdependence structure of the original variables. In particular, the variable weights are the normalized eigenvectors of the sample covariance matrix. This transformation allows the multi-dimensional isolate profile data to be plotted in two or three dimensions and major sources of variation to be identified. It should be noted that such plots will often display an arched pattern known as the *horseshoe effect* (Guttman, 1950). Plots of principal components that exhibit a severe horseshoe effect can be misleading and therefore caution should be exercised when drawing conclusions (See Figure 3). Further details on both the application of PCA and its underlying theory can be found in Jolliffe (2002).

Multivariate analysis of variance

If patterns of spatial or temporal dependence are suspected or observed in an exploratory analysis such as PCA or MDS, a multivariate analysis of variance (MANOVA) (Wilks, 1932) can be used to test for significant spatial or temporal effects. However, in order for valid conclusions to be drawn from a typical MANOVA, data must satisfy the assumption of multivariate normality. This assumption is rarely met by MST data and therefore resampling-based methods for MANOVA (Anderson, 2003), which do not assume, multivariate normality are preferable. Unfortunately, this methodology is not currently available in standard software packages. Therefore, it should be emphasized that results of standard MANOVA must be interpreted with caution.

Table 4.3 Techniques for identifying patterns of spatio-temporal variability in the isolate profiles

Method	Reference	Exploratory or Confirmatory	Objective	Relevant Assumptions
Principal components analysis (PCA)	Hotelling (1933)	Exploratory	Represent the variation in a large number of variables by a small number of principal components.	Covariance is an appropriate measure of variable interdependence. Dependence between variables is similar for all isolates.
Multidimensional scaling (MDS)	Turgorsen (1952)	Exploratory	Represent an interobject distance matrix in relatively few dimensions.	Selected inter-isolate distance metric is appropriate.
Multivariate analysis of variance (MANOVA)	Wilks (1932)	Confirmatory	Test for statistically significant differences between the means of specified groups of isolate profiles.	Multivariate normality Dependence between variables is similar for all observations
Non-parametric MANOVA	Anderson (2002)	Confirmatory	Test for statistically significant differences between the means of specified groups of isolate profiles.	Selected inter-isolate distance metric is appropriate.

Isolate identification

Once a library of isolate profiles from each potential source has been collected, a rule for identifying the most likely source of isolates of unknown origin must be constructed. Statistical methods to accomplish this are referred to as discriminants or classification rules. This section describes the general process of constructing and evaluating classification rules in the context of microbial source tracking and discusses the assumptions of several classification rules commonly used for microbial source tracking. Certain types of classification rules and their corresponding assumptions are more appropriate for different types of MST data, but the same general process should be followed in the construction and assessment of classification rules regardless of the type of data (Hastie et al., 2002).

1. Declone isolates from each feces sample by deleting identical patterns within each single feces sample in the database. These are essentially duplicate observations and it would be inappropriate to have replicate observations in the training, validation, and test sets (below).
2. Randomly divide data into training isolates (~50%), validation isolates (~25%) and test isolates (~25%).
3. Use the training isolates to construct various classification rules.
4. Estimate the accuracy of each rule by attempting to classify the validation isolates.
5. Select the most accurate rules and refine them. Refinement techniques include variable selection and the adjustment of tuning parameters.
6. Once a (single) best rule has been selected, use the test isolates to estimate generalizability of the rule. This step is important to predict how well the classification rule will work in real-world application.

4.6 Techniques for classification and discriminant analysis

There are several different techniques for classification of isolates to source categories based on the known-source library. This section attempts to give some details of a few commonly used procedures for discriminant analysis and identify situations where each is appropriate.

Linear and Quadratic Discriminant Analysis

As with most statistical techniques, at the foundation of the most commonly-used and well-known classification rules is the assumption that the training data (known-source library) are a random sample from a population for which the variation between samples is well-described by a normal distribution. In particular, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) assume that the data from each population follow a *multivariate normal distribution*. Implicit in this assumption is the notion that each individual variable follows a (univariate) normal distribution and that the dependence structure among the variables is fully characterized by a matrix of pairwise covariances.

In the case of binary profiles, the concept of a mean, or "consensus," fingerprint surrounded by a cloud of variants, distributed at similarity distances that follow a normal distribution, is not meaningful. Therefore, the assumption of multivariate normality makes it difficult to justify the use of LDA or QDA for classification of binary profiles resulting from genotypic fingerprints because these presence/absence data do not follow a normal distribution.

LDA makes the additional assumption that the matrix of pairwise covariances is the same for samples from each population. However, in the context of microbial source tracking, this may not always be true (e.g., two bands might be positively correlated for one source category and negatively correlated for another source category). Finally, classification rules based on the assumption of multivariate normality use estimates of covariance matrices that tend to be poor unless sample sizes are very large. Thus classification rules based on LDA and QDA can often perform poorly when sample sizes are not very large. In conclusion, LDA should be used with caution for MST, but QDA seems to be a somewhat reasonable approach for data resulting from phenotypic profiles when sample sizes are large.

Nearest-neighbor rules

In addition to LDA and QDA there are several other types of classification rules. When the assumptions required by LDA and QDA are inappropriate, nearest-neighbor rules are the most common alternative. There are several varieties of nearest neighbor rules, but they share the general characteristic of classifying objects based on the group membership of the most similar objects of known origin. These rules do not assume any explicit form for the data distribution, such as multivariate normality, but in order to provide reasonable classifications, similar objects, as measured by some distance or similarity coefficient, must come from the same population.

The three most common types of nearest neighbor rules in the MST literature are the (1) maximum similarity, (2) average similarity and (3) k-nearest-neighbor rules. Maximum similarity

classification simply assigns isolates of unknown origin to the source of the most similar isolate in the library. Average similarity measures the similarity between the isolate of unknown origin and all isolates of known origin and then assigns the unknown isolate to the source with which the unknown isolate profile has the highest average similarity.

The *k*-nearest neighbor rules (Fix and Hodges, 1951) are somewhat of a compromise between these two methods. For some specified value of *k*, the *k* most similar objects are identified and the isolate of unknown origin is assigned to the source with the largest representation among the *k* nearest neighbors. Surprisingly, little research has been conducted regarding the choice of the value of *k*. However, for the simple case of two multivariate normal populations of comparable group sizes, Enas and Choi (1986) recommend selecting *k* to be approximately between $n^{2/8}$ and $n^{3/8}$ depending on whether there are small or large differences between the group covariance matrices. So, even for sample sizes of n=1000 the recommended value of k is somewhere between 5 and 13. Thus a large number of neighbors are not advisable. Further information on the theory and implementation of nearest neighbor rules can be found in Dasarathy (1991).

Epidemiological matching

So called "epidemiological matching" is another approach that has been used for isolate identification. (Note: Statistically, this can be viewed as a generalization of a maximum-similarity classification rule.) This practice involves clustering isolate profiles into subtypes and assigning an isolate of unknown origin to a source category only if it is similar enough to all the isolates of a particular subtype, which themselves are all associated with the same source. Definition of subtypes is accomplished via complete linkage hierarchical cluster analysis, which establishes a minimum similarity for all isolate profiles within a subtype. Determining the value of this minimum similarity value depends on both the quality of the data and on the similarity being used. For example, Figure 4 illustrates the fact that the simple matching coefficient is always larger than the Russell-Rao coefficient, the Jaccard coefficient is always larger than the Russell-Rao coefficient and the Dice coefficient is always larger than the Jaccard coefficient. Therefore, it is difficult to establish guidelines for establishing subtypes beyond stating that the relative magnitudes of similarity measures should be considered.

$$\text{Russell} - \text{Rao} = \frac{a}{p} \le \frac{a+d}{p} = \text{Simple matching}$$

$$\text{Russell} - \text{Rao} = \frac{a}{p} = \frac{a}{(a+b+c)+d} \le \frac{a}{a+b+c} = \text{Jaccard}$$

$$\text{Jaccard} = \frac{a}{a+b+c} \le \left[\frac{a}{a+b+c}\right] + \left[\frac{a(b+c)}{(2a+b+c)(a+b+c)}\right] = \frac{2a}{2a+b+c} = \text{Dice}$$

Figure 4.4 Relationships between similarity measures for binary profiles.

Table 4.4 Summary of common classification rules

Rule	Assumptions	Suitability classification of genotypic profiles	Suitability classification of phenotypic profiles
Linear Discriminant Analysis (LDA)	Multivariate normality Common covariance structure for each group	Low	Low
Quadratic Discriminant analysis (QDA)	Multivariate normality	Low	Moderate
Average similarity	Appropriate similarity	Low	Low
Maximum similarity (1-nearest neighbor)	Appropriate similarity	High	Moderate
k-nearest neighbor	Appropriate similarity	High	High
Epidemiological matching	Appropriate similarity	High	Moderate

4.7 Practical issues and Chapter summary

Total cost of misclassification

In a formal decision-theoretic framework the cost of a classification error is factored into the evaluation of a classification rule. For example in the context of microbial source tracking, it might be more costly to identify a poultry farm as the source of contamination, when in fact wild geese are the true source of contamination, than it would be to make an error in the reverse direction. Additional examples include unnecessary human sewer upgrades using public money, BMP for livestock waste management at a portion of the farmer's personal income, or wildlife management plans at a small amount of public money. Incorrect identification of contamination sources is also likely to have political costs in addition to monetary costs. However, most software packages do not let users specify the costs of each type of misclassification error. An alternative protection against costly errors is requiring a threshold of evidence before any classification can be made, since it is often preferable to make no attempt at classification rather than classify incorrectly. An example of an analysis of MST data which includes the use of threshholding is Ritter et al. (2003).

Software

The techniques and tools for data management and analysis discussed in this Chapter require software for implementation. There are several software packages available for image processing, library management and data analysis. These packages vary in cost, capabilities and ease of user interface. Table 4.5 attempts to give some indication of the relative strengths and weaknesses of some commonly used software packages.

Table 4.5 Comparison of software packages commonly used for analyses associated with microbial source tracking

Software	Company	Capabilities	Ease of use	Flexiility	Cost
Bionumerics	Applied Maths, Belgium	Image analysis, data management and statistical analysis	High	Low	High
SAS	SAS Institute, Cary NC	Statistical analysis	Moderate	Moderate	High
R	CRAN, www.r-project.org	Statistical analysis	Low	High	Free for academic use
ImageQuant		Image analysis		Moderate	Low

Summary

In conclusion, we reemphasize that there are several critical decisions to be made regarding data collection and analysis in any MST study and reiterate the most important ideas below.

1. The sampling plan must be designed around the objectives of the study.

2. There are usually several ways to represent an isolate's characteristic profile as numerical data, and decisions about data representation can have a significant impact on both sampling and analysis strategies and outcomes.

3. The genetic diversity of indicator bacteria in a given animal host is influenced by several factors. Accordingly, estimates of library sizes are often difficult to make without empirical data.
    a. Generally speaking, most genotypic-based MST studies that have been done to date have used relatively small host origin databases, containing between 35 and about 500
    b. In contrast, many phenotypic-based MST studies, mostly done using antibiotic resistance patterns, have used known-source libraries consisting of about 1,000 – 6,000 isolates
    c. In some of the more extreme cases a significantly large library (i.e., fingerprints for 20,000 to 40,000 *E. coli* isolates) may be needed to capture all the genetic diversity present in natural populations

4. Certain types of classification rules and their corresponding assumptions are more appropriate for different types of MST data, but the same general process should be followed in the construction and assessment of classification rules regardless of the type of data
    a. LDA should be used with caution for MST, but QDA seems to be a somewhat reasonable approach for data resulting from phenotypic profiles when sample sizes are large.
    b. When QDA is inappropriate, nearest-neighbor rules are the most common alternative.

Chapter 5. Methods Performance

5.1 Introduction

The goal of Microbial Source Tracking (MST) is to associate a microorganism from a polluted site with an human or animal source to infer the origin of fecal pollution. This information is vital to managers, stakeholders, and other interest groups that play a role in contracting MST studies, water quality monitoring, risk assessment, and protection and restoration of U.S. surface waters. Decision makers require high quality data. Quality control strategies measure confidence in data and help ensure proper use of methods. As a result, researchers have developed quality measures to assess the performance of each MST method. A comparison of quality measures revealed a core group of performance criteria that all MST methods share in common. This Chapter will organize and define MST universal quality measures and provide an overview of method-specific performance criteria that can be used to evaluate the quality of data and overall performance of each MST approach.

5.2 Universal Quality Measures

Although MST researchers use a wide array of techniques to identify fecal pollution in surface waters, all methodologies should adhere to a strict set of quality measures. These measures are organized into five quality control issues including specificity, precision, control samples, quality assurance documentation, and minimum number of controls. Recommendations for each quality control issue are discussed below.

Specificity. Specificity refers to the ability of a particular MST method to discriminate between different animal fecal sources. The specificity of a method can be described as the proportion of samples that are negative [test negatives (TN) + false positives (FP)] that test negative [test negatives (TN)]. Specificity is mathematically expressed as:

$$\frac{TN}{TN + FP} \times 100\%$$

A specificity percentage should be reported for each animal fecal source included in a MST study. Although there is currently no consensus, specificity values below 80% percent reflect questionable discriminatory power. Managers should use data with caution and may need to consider data from an alternative MST approach. Specificity control standards should be prepared at concentrations easily detected by the respective MST method and should consist of a pool of fecal samples acquired from animal sources in the same geographic context as water samples. The minimum number of individual animal fecal samples will be dependent on the complexity of the watershed system (see Chapter 4). Currently, there is no agreement on how to calculate this number. A conservative estimate might be a minimum of ten individuals per animal source. Because specificity control standards are generated for each watershed, specificity must be established for each geographic location tested. It is also ideal to perform specificity controls before applying a particular MST method to test samples. Many researchers will collect test samples during specificity testing and archive samples until specificity is confirmed in the watershed of interest.

Data:

Source	Test Negatives	False Positives
Cow	900	100
Pig	850	150

Equation:

$$\text{Specificity} = \frac{\text{Test Negatives}}{\text{Test Negatives + False Positives}} \times 100$$

$$= \frac{900 + 850}{900 + 850 + 100 + 150} \times 100 = 87.5\%$$

Conclusion: **Human specific pattern detectable percentage is 87.5%**

Example 1: Library-dependent specificity calculation for human detection.

Precision. Precision or reproducibility is important for all MST applications and is measured through the use of replicates. Replicates are repetitions of an assay or part of an assay and fall into two categories: identical replicates and experimental replicates. Identical replicates are assays performed simultaneously using the same method preparations and same reagents (i.e. antibiotics, media, PCR reagents, etc.). Identical replicates serve two functions. They can preserve data. If one replicate fails, the other can potentially still provide data. They can also be used to monitor variability or low precision in a test sample batch. A sample batch is a set of test samples prepared and processed together through all steps of the MST method. Approximately 10% of all samples tested should be replicated. Replicate sample results should be in agreement. Experimental replicates are assays that share the same reagents, while the sample preparations come from similar, but not identical samples. They provide crucial information about the overall precision of the method. For example, if a researcher wishes to test the reliability of identifying human fecal pollution in a watershed, it is inappropriate to assay just one water sample. A number of samples must be analyzed to determine whether there is any variation in method response. If variability is prevalent, researchers can evaluate analyst performance, quality of reagents, proper equipment function, or sample matrix characteristics to increase precision.

Control Samples. Control samples are quality measures that monitor the proper performance of MST methods and screen for the presence or absence of extraneous microorganisms or nucleic acids introduced into a MST experiment. All MST methods should incorporate method positive controls and negative controls. Method positive controls verify whether a MST process is performing adequately. These controls should be obtained from a known source and should always yield a predefined result when the MST method is conducted correctly. For example, ARA laboratories commonly use enterococci and *E. coli* strains with known multiple antibiotic-resistant patterns as method positive controls. If the expected antibiotic-resistance pattern is not observed, the

researchers reject all data with the same sample materials and request immediate resampling. For culture-independent methods, control template should be tested at a concentration ten times above the limit of detection. Method positive controls should be performed for each batch of test samples.

Negative controls are used to monitor for the introduction of extraneous materials into an experiment. They are divided into two categories including field blanks and method blanks. Field blanks monitor for the introduction of extraneous material into MST experiments during field sample handling, transport, and storage. In the field, sterile water should be transferred to a sample collection tube and processed as a test sample. A positive result indicates the presence of contamination most likely due to poor aseptic technique in the field, contact with other samples, or damaged storage containers. The method blank is designed to screen for contamination throughout the entire MST process. This control determines whether glassware, filters, handling procedures, media, reagents, or lab environment introduce extraneous material into samples. In the laboratory, the control is processed in the same manner as a test sample except that sterile water is substituted for an environmental sample. At least one method blank should be performed for each sample batch.

Quality Assurance Documentation. The results of all quality measuring data and method validation should be thoroughly documented, published in future studies, and easily accessible to management personnel. In addition to laboratory standard record keeping procedures including equipment calibration and maintenance schedules, reagent catalogs, quality measure data, sample processing notes, and routine documentation back-up, MST researchers should pay careful attention to sample acquisition documentation. Information describing animal fecal sampling geographic location and date should be consistently documented for each MST experiment. Access to this information will be imperative for future research concerning library and genotypic target geographical and temporal stability. It may also be useful to record the diet of individual animals used for fecal sampling. All documentation should be reviewed by a laboratory supervisor for accuracy and completeness. Quality assurance documentation reviews ensure that all method quality requirements were met, and that any deficiencies are properly noted in the final report.

Minimum Number of Controls. The disadvantages of performing quality measure controls are that they result in additional cost, can occupy space in the laboratory, and can consume more sample. However, these controls are the only way to validate a MST method and ensure that data from test samples are genuine. Each researcher should weigh these disadvantages against the need for precise and accurate information when deciding how many controls to run in each experiment. Table 1 lists each quality measure control type, summarizes their importance, and lists recommended frequencies for a typical MST study.

Table 5.1 Summary of Quality Measure Controls

Description	Purpose	Frequency
Specificity	Verify ability to discriminate between animal sources	Establish for each MST geographic location tested
Identical Replicates	Monitor variability between test replicates in sample batch	10% of the number of field samples tested per batch
Experimental Replicates	Monitor method variability between sample batches	At least 10% of field samples tested per batch
Method Positive Control	Verify method process performing correctly	One control per sample batch
Field Blank	Verify that not contamination introduced during sample acquisition	5% of the number of field samples collected
Method Blank	Verify that no contamination introduced during entire method process	At on control per sample batch

5.3 Method-Specific Performance Criteria

5.3.1 Library-Dependent Methods.

Library-dependent methods compare traits from cultivated fecal isolates collected from water samples with a library of cultivated isolates from known fecal sources. The known source library acts as a predictive tool to determine the source of fecal pollution. Library-based methods include carbon utilization profiles, antibiotic resistance assays (ARA), ribotyping, pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), and repetitive PCR (rep-PCR). The utility of a particular method is directly related to the ability of a library to accurately represent and characterize fecal sources present in a watershed. Unsuitable libraries lead to inaccurate information and poor management decisions. Researchers evaluate the quality of libraries based on composition, size, continuity, sensitivity, and minimal detectable percentage.

Library Composition. The first step in library construction is to collect fecal samples from host species, then isolate bacteria from a number of different individuals. Most source tracking libraries are composed of either enterococci or *E. coli* isolates. Culture methods designed to isolate these microorganisms can sometimes allow the growth of other microorganism species. As a result many researchers perform additional tests to measure the percent of target organisms (enterococci or *E. coli*) in a library. Although there is no consensus, an adequate library should consist of at least 95% of the target indicator organism. The library should be comprised of isolates collected from source animals impacting the local watershed. Potential fecal pollution sources can be identified by performing a sanitary survey of the watershed.

Library Size. The ideal library should contain enough isolates from each host species to characterize the dominant traits of an indicator organism population. Some researchers suggest that small libraries misrepresent population diversity of indicator organisms in surface waters. However, it remains undefined what constitutes the optimal library size partly because few studies to date have rigorously evaluated this problem. Wiggins and colleagues (2003) conclude that a library should be

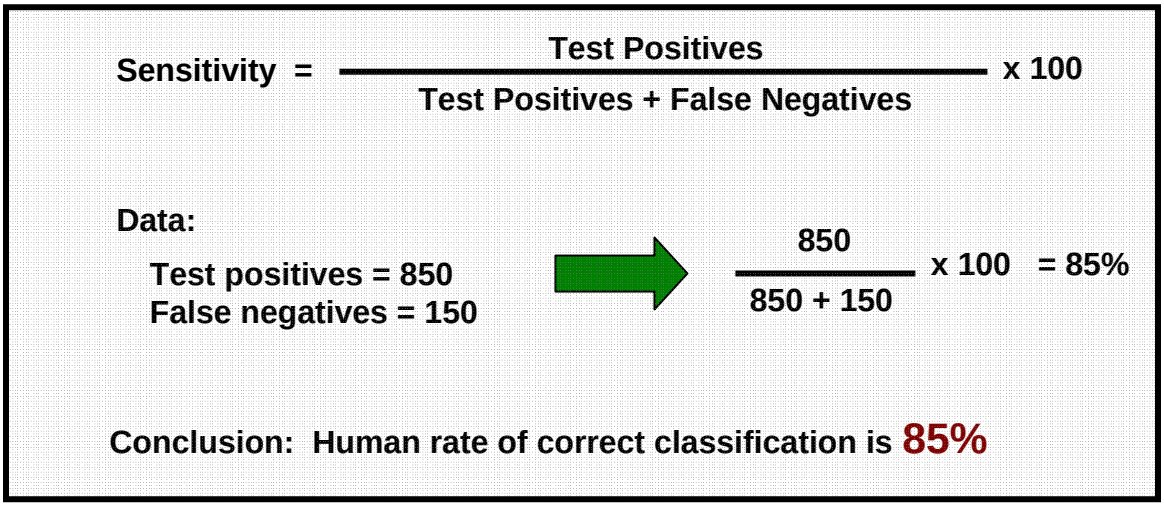
as large as it needs to be representative. Library representativeness is a measure of how well a library classifies the patterns found in a target microorganism from each of the host species found in a watershed. Representativeness is estimated by comparing the ARCC from a resubstitution analysis with the ARCC from a cross-validation analysis (see Chapter 4, Data Collection and Analysis for review). If the difference in ARCC values is less than 5%, then the library is representative. Researchers currently construct libraries based on sample accessibility, cost, and practical experience. As a general guideline, libraries should contain at least 1,000 isolates per host species of interest.

Library Continuity. The ideal library should be able to classify fecal isolates from numerous geographical areas and should be representative over time. However, factors such as season, diet, and horizontal gene transfer (movement of DNA from one bacterial cell to another) can create library discontinuity (Bryant, M.P., 1959; Hungate, R.E., 1966; Ogimoto, K. and Imai, S., 1981; Stewart and Bryant, 1988; Harmsen et al., 2000). Initial studies indicate that geographic variability can be high and that libraries should be constructed from local samples only (Hartel et al., 2002; Wiggins et al., 2003). The longest a library has been shown to be stable is 12 months (Wiggens et al., 2003). Thus, library continuity should be re-evaluated at least once a year until additional studies indicate otherwise.

Library Sensitivity. Library sensitivity measures the detectable percentage of isolated target microorganisms exhibiting a host-specific pattern. The sensitivity of a method is described as the proportion of samples that are positive [test positives (TP) + false negatives(FN)] that test positive [test positives (TP)]. Sensitivity is mathematically expressed as:

$$\frac{TP}{TP + FN} \times 100$$

A sensitivity value is also referred to as the rate of correct classification (RCC) and should be reported for each animal fecal source included in a MST study. In addition, researchers commonly report an average rate of correct classification value (ARCC) or mean of all RCC values. Sensitivity values should be determined from a set of characterized standards (from known fecal sources).



Example 2. Sensitivity or RCC calculation for human detection in a 100 ml control sample.

Minimal Detectable Percentage (MDP). The Minimal Detectable Percentage is a measure of the lower limit for considering that a source is present in a sample (Whitlock et al., 2002; Harwood et al., 2003; Wiggins et al., 2003). Its value is based on the average frequency of misclassification of the known sources in the library. The MDP can be used to estimate the likelihood that an isolate that is not from a given source will be classified into that source, and therefore provide the basis for a significance cut-off when predicting the sources of isolates in water samples (Harwood et al., 2003). Several methods of determining the MDP have been proposed, and although there is not yet consensus on the best method, all MST studies should present a value of the MDP and the method that was used to determine it.

5.3.2 Library-Independent Methods

Library-independent methods rely on genotypic traits to identify sources of fecal pollution. These methods do not require isolate cultivation. Library-independent methods include T-RFLP community analysis and the detection of host-specific DNA sequences. Host-specific strategies target 16S rDNA from Bacteroides, toxin and adhesion DNA sequences, and numerous phage loci. These methods rely on PCR technology and can detect small quantities of nucleic acids in a few hours. However, an increased limit of detection elevates the risk of amplifying extraneous nucleic acid templates. Inhibitory substances can co-extract with nucleic acids during sample purification and concentration (Wilson, 1997). In some cases, PCR inhibition may be the cause of false-negative reactions and can dramatically decrease the limit of detection.

Limit of Detection. The limit of detection is the minimum concentration or copy number of a control DNA target that routinely yields a PCR product. Detection limits are measured by adding a range of control DNA template concentrations (i.e. 1, 10, 10^2, 10^3, and 10^4 copies) to PCR test reactions. PCR control DNA templates can be any of the following: 1) purified total nucleic acid extract from a microorganism containing the sequence of interest, 2) the whole microorganism, which can be used when the DNA template is released by heating before or during amplification, 3) a specific DNA template containing the entire sequence to be amplified, including primer binding sites, or 4) a cloned DNA fragment containing a modified form of the DNA target (see Inhibition of Nucleic Acid Amplification section). After limit of detection is established for a MST method, researchers should include a control containing the minimum detectable quantity for each sample batch tested. This control will ensure that each PCR assay is performing at an optimal level.

Experiment: Test for detection of 1, 10, 10^2, 10^3, and 10^4 target copies. Each test reaction tested in triplicate.

Data:

Copy #	Trial 1	Trial 2	Trial 3
1	-	-	-
10	+	-	-
10^2	+	+	+
10^3	+	+	+
10^4	+	+	+

Conclusion: Human limit of detection is 10^2 copies. Assay will occasionally detect 10 copies.

Example 3. Measuring limit of detection for human host-specific PCR assay.

Confirmation of PCR data. Most host-specific PCR methods measure the presence or absence of a target DNA sequence in an environmental sample. For example, a water test sample is collected and concentrated on a filter. DNA from microorganisms adhering to the filter surface are extracted, purified, and amplified using primers that target a specific sequence or group of sequences. If the target DNA is present, the researcher will observe a PCR product on an agarose gel. Two strategies can be used to validate the authenticity of the resulting PCR product. First, the researcher should report the PCR product size (base pairs). For example, the human host-specific 16S rDNA PCR primer set HF134 and 708R (Bernard and Field, 2000) should yield a 574 base pair product. Second, the resultant PCR product can be cloned and sequenced. Sequencing is more time consuming and expensive, but it is the only way to definitively prove detection of target DNA. Sequencing will also help build a database of sequences that can be used to evaluate genetic variation of target DNA over time and in different geographic locales.

Extraneous Nucleic Acids. PCR methods that exhibit a low specificity may be contaminated with extraneous nucleic acids found in the laboratory environment or reagents. Nucleic acids from equipment, other samples, and previously synthesized amplicons can contaminate PCR reactions. Extraneous nucleic acids from these sources can be eliminated with physical barriers. Sample preparation, nucleic acid extractions, PCR cocktail assembly and amplifications, and post-PCR manipulations should occur in separate work areas. If laboratory space is limited, separation of pre-PCR (sample filtration, nucleic acid extraction, and PCR cocktail assembly) from post-PCR (i.e. gel visualization, molecular cloning, etc.) manipulations is most critical. Each area should contain dedicated equipment and be cleaned with 0.6% sodium hypochlorite (NaOCl) after each use. In addition to physical barriers, a unidirectional workflow between areas (i.e. sample preparation → extractions → PCR cocktail assembly and amplification → post-PCR analyses) should be used to reduce the potential for contamination.

PCR reactions may also amplify nucleic acids present in extraction and PCR reagents, which cannot

be eliminated with physical barriers. For example, several studies have documented the presence of eubacterial DNA in *Taq* DNA polymerase preparations (Hughes et al., 1994; Schmidt et al., 1991; Rand and Houck, 1990) and others suspect the presence of cow, pig, and chicken DNA in commercially prepared deoxynucleoside triphosphates (Shanks et al., in press). Reagents should be opened only in dedicated work areas and used exclusively for MST analyses. To screen for extraneous nucleic acids in PCR reagents, researchers should perform at least 20 no template PCR reactions with the reagents prior to the initiation of a study. Researchers should also do at least one method blank before environmental water samples are processed in the laboratory to monitor for extraneous nucleic acids in extraction reagents.

Inhibition of Nucleic Acid Amplification. PCR methods that exhibit a reduced limit of detection may be inhibited by substances that co-extracted with nucleic acids from water samples. Inhibition may be total or partial and can manifest as complete reaction failure or as a reduced limit of detection. Some inhibitory substances observed in environmental samples include detergents, humic acids, polysaccharides, fats, and other cellular debris (Wilson, 1997). To monitor the impact of inhibition, researchers can perform a matrix spike control for each suspected environmental sample. A matrix spike contains the minimum quantity of detectable control DNA template and is added directly into a PCR reaction containing sample extract. These controls are critical for quantitative PCR applications. The matrix control DNA template should be easily distinguished from wild-type sequences present in the sample extract. Matrix control DNA templates should be prepared from a cloned DNA fragment containing a modified form of the target sequence by size, by restriction mapping, and/or by an alternative probe recognition sequence. Modified control DNA can be prepared by *in vitro* generation of deletions, insertions, or other sequence changes. For example, a modified control DNA template engineered with a 20 bp insertion allows for gel visualization of wild-type and modified control DNA sequences simultaneously (Figure 1).

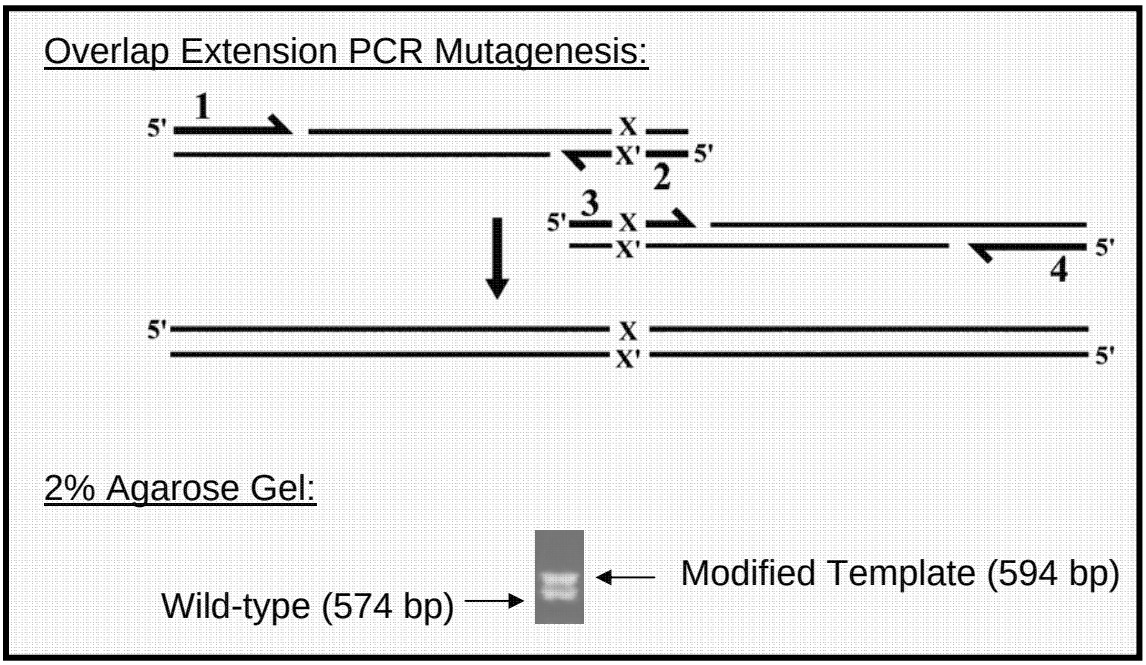


Figure 1. Panel A illustrates the construction of modified control template using

overlap extension PCR (Higuchi et al., 1988). Panel B shows discrimination of wild-type template (574 bp) from modified DNA template (594 bp) on a 2% agarose gel.

5.4 Conclusions

A comparison of quality measures for each available MST method uncovered a shared set of method performance criteria. These criteria are organized into five key quality measure issues. Specificity verifies the ability of an MST method to discriminate between different animal sources present in a watershed. Precision quality measures variability between test sample replicates and independent test sample batches. Control samples screen for the presence of extraneous microorganisms or nucleic acids introduced during the MST process and ensure that experimental technique, consumables, and equipment are functioning properly. Thorough quality assurance documentation of all parts of the MST process, especially method validation and sample acquisition encourage the accurate transfer of information from laboratory scientists to decision-making management. Finally, the incorporation of quality measures at recommended frequencies ensures the validation of high quality data and responsible data interpretation.

In addition to universal performance criteria, some MST methods require additional quality measures. Library-dependent methods must pay careful attention to library construction. Factors including library sensitivity, composition, size, and continuity directly impact the quality of MST data. Library-independent methods that utilize PCR strategies require rigorous adherence to quality standards that measure the limit of detection and that reduce contamination of MST experiments with extraneous nucleic acids originating from the laboratory environment, equipment, consumables, and reagents. Additional controls must also be included that monitor for the presence of inhibiting substances that often co-extract with nucleic acids recovered from environmental samples.

Accurate characterization of the source of fecal pollution in a watershed allows managers to identify the most appropriate management action to restore or protect an impaired waterway. Although it may not be feasible to include all of the recommended controls, the more controls used the more confidence a decision maker will have when evaluating MST data. In addition, quality measure recommendations will help bring more uniformity to MST research, will lead to more effective method evaluations, and the practice of sound science.

Chapter 6. Assumptions and Limitations of MST Methods

6.1 Introduction

Just as no "ideal" indicator organism for the assessment of water quality has been identified, an active body of research continues to seek the ideal source identifier (SI) for fecal contamination in environmental waters. This section will define the characteristics that MST practitioners seek in an ideal source identifier, which could theoretically be a chemical, a virus, a bacterium or other microorganism, or a gene(s). In many MST applications, the source identifier is subtyped ("fingerprinted") in order to discriminate between particular subtypes that are associated with various host sources. Many discriminatory characteristics of source identifiers are used in MST methods, including SI strain/species, fingerprint pattern, or genetic marker; therefore these will be grouped under the acronym SPM (species/pattern/marker). The ideal characteristics of SIs and SPMs will be compared with the more realistic expectations for good or useful SIs/SPMs.

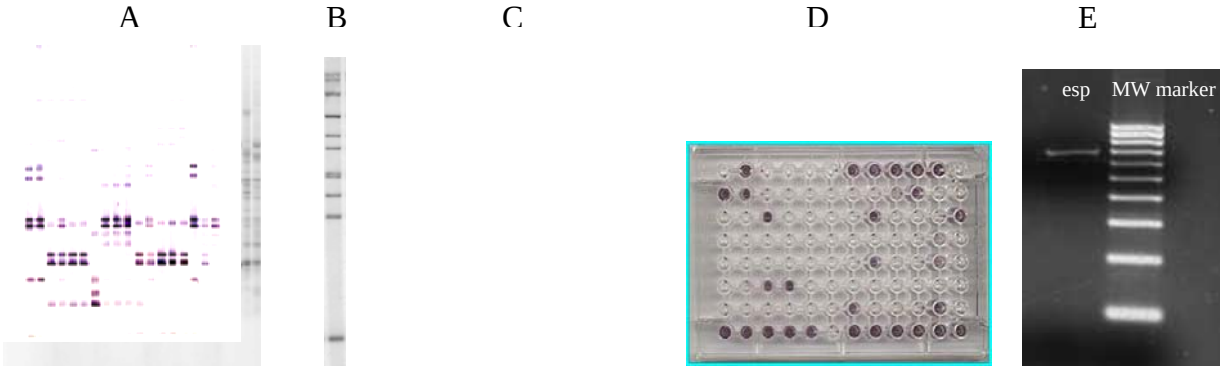
Every new field of scientific inquiry must make some practical assumptions about the effect of variables on the application of the method. Part of the process of maturation of that field is framing the assumptions as scientific hypotheses, followed by rigorous hypothesis testing. MST investigators are actively involved in this process, as highlighted in several recent reviews (Scott et al., 2002; Simpson et al., 2002; Stewart et al., 2003). In this Chapter, the assumptions made about various organisms and methods currently used for MST will be discussed in conjunction with the hypotheses that have been tested. Assumption/hypotheses that remain to be tested will be outlined, and the known limitations of and concerns about the methods will be presented.

Table 1 outlines the characteristics of a hypothetical, ideal source identifier, and contrasts them with the characteristics of a useful SI. MST investigators have identified many SI candidates, and MST approaches have focused on various SPMs, some of which are illustrated in Figure 1. None of the source identifiers currently in use have been demonstrated to have all the characteristics listed. Many methods are in an early stage of development, and further research may demonstrate that some possess all or most of these attributes.

Table 6.1 Characteristics of an ideal source identifier (SI) and those of a useful source identifier. Strain/pattern/marker is abbreviated SPM.

Characteristic	Ideal SI	Useful SI
Host specificity	Specific strain/pattern/marker (SPM) found only in one host species.	Specific SPM is differentially distributed among host species of interest.
Distribution in host	Found in all members of all populations of a host species.	Found in the waste streams from host species that could impact the study area
Stability of pattern/ marker	Not subject to mutation or methodological variability	Rarely subject to mutation; methodology has defined reproducibility[a]
Temporal stability in host	No temporal variability within host individuals or host populations	Temporal variability in individuals is balanced by temporal stability in host populations
Geographic range/stability	SPM associated with a particular host are constant across broad geographic ranges	SPM associated with a particular host can be consistently identified across the geographic area to be studied
Representative sampling	The diversity of the SI in host populations and in water is represented by a small sample size	The diversity of the SI in host populations and in water can be represented by a reasonable sample size
Survival in water		
A. Rate of decay	Consistent decay rate in various types of waters and habitats; no growth under any conditions	Predictable decay rate in various types of waters and habitats; no growth under the conditions of the study area; all SPMs decay at the same rate after leaving host
B. Abundance in 1^o vs. 2^o habitat	The distribution of SPMs in source material, i.e. feces, does not change after delivery to the water	The distribution of SPMs in water bears a significant resemblance to that found in contaminating fecal material
Quantitative assessment	The relative and absolute contribution of each host to SI concentration can be assessed	SI may not be quantitative, but accurately indicates presence/absence of source, e.g. conventional PCR markers
Relevance to regulatory tools	The SI itself is also used to regulate water quality, i.e. coliform, enterococci	The SI is correlated with a regulatory water quality parameter
Relevance to health risk	The SI itself constitutes a health risk	The SI is correlated with health risk

[a]Methodological reproducibility refers to the ability to generate the same pattern (i.e. a DNA or phenotypic profile) or result (i.e. PCR +/-) from independent assays.



(A) Courtesy Dr. Michael Sadowsky

(B,C) Courtesy Dr. Valerie J. Harwood

(D) Courtesy Dr. Charles Hagedorn

(E) Courtesy Dr. Troy M. Scott

Figure 6.1 Illustration of some strains/patterns/markers (SPMs) currently utilized in MST methods. (A) rep-PCR patterns of *E. coli* isolates – each vertical lane represents one pattern. (B) ribotype pattern of one *E. coli* isolate (C) ribotype patterns of *Enterococcus* isolates (D) carbon source utilization pattern of one *Enterococcus* isolate (E) specific genetic marker (*esp* of *E. faecium*) amplified by PCR of *Enterococcus* DNA

6.2 Host specificity of specific strain/pattern/marker (SPM)

The ideal source identifier (SI) would be unique to a host species, and have no alternative sources. Furthermore, the SI would be represented by variants, each of which would be unique to a host species that contributes contamination to water bodies. MST would be a much simpler field if all of the fecal microorganisms we use as indicator organisms were strongly and specifically associated with the gastrointestinal tract of their respective hosts; however, many fecal bacterial strains appear not to be host-specific. Strains that inhabit multiple host types have been termed "transient"(Harwood et al., 2003; Myoda et al., 2003), a term borrowed from earlier work on *E. coli* population dynamics. In the population dynamics literature the term "transient" had a different meaning, as it described subtypes that were not observed consistently in host individuals (reviewed in (Hartl and Dykhuizen, 1984)); i.e., they were sampled only once or infrequently from an individual. Other MST practicioners have utilized the term "cosmopolitan" to describe the multiple-host phenomenon (Field et al., 2003; Whitlock et al., 2002), which refers to the organism's ability to inhabit various host species, and implies nothing about the length of the habitation, which might be long-term or short-term, nor the geographic distribution. It should be noted that apparent lack of host specificity (observation of a SPM in more than one host) could be due to insufficient discrimination in the typing method; however, even very highly discriminatory methods such as PFGE identify cosmopolitan isolates.

The cosmopolitan distribution of some SPMs undoubtedly has a negative influence on MST applications, but efforts to understand the impact of this phenomenon are complicated by the fact that discrimination between microbial subtypes (strains with different fingerprints) depends upon the method utilized for subtyping (Guan et al., 2002; Johnson et al., 2004). The difference in discriminatory capability of the various MST methods has made comparison of studies that rely on

different analytical methods extremely difficult; however, cosmopolitan host distribution is well-documented for *E. coli*. Multilocus enzyme electrophoresis of *E. coli* revealed that 24 of 270 electrophoretic types were found in more than one (up to seven) distinct hosts (Ochman et al., 1983). Genotyping by REP PCR revealed some identical *E. coli* subtypes in gull feces and sewage (McLellan et al., 2003). A total of 22% of all distinct *E. coli* ribotypes (two-enzyme) isolated from cattle, chickens, horses and swine were shared by some combination of host species (Hartel et al., 2002), which represented 66% of all isolates tested. Absolute specificity was also lacking in F-specific coliphages; three serotypes (Type I, II and III) were found in municipal wastewater, and each of these was also found in animal feces (Cole et al., 2003). Only Type IV coliphages were specific to animal feces. No coliphage type was specific to human-derived wastewater, although Type II coliphages were the dominant serotype isolated from wastewater. Because *E. coli*, *Enterococcus* spp. and coliphages are commensal fecal indicators that are broadly distributed in feces and are widely used by the regulatory and MST community, we suggest that a better understanding of cosmopolitan distribution, and how profoundly it affects MST methods, is particularly important in these organisms. Furthermore, as new methods are developed their host specificity or host range should be fully explored.

It has been suggested (Simpson et al., 2002) that host specificity would be augmented if the MST target contributed to the specific interaction between host and fecal microbe. Candidates include the genes that code for microbial appendages such as pili and adhesins, which mediate attachment to cells of the host gastrointestinal tract. One method capitalizing on this approach is PCR amplification of the gene for the enterococcal surface protein (*esp*) of *E. faecium* (Scott et al., 2004), which, though promising, requires further validation. Enteric viruses, which rely on specific cell surface receptors to bind to host cells, are inherently species-specific and have been used to assess the presence of human fecal contamination in environmental waters (Griffin et al., 1999; Jiang et al., 2001).

6.3 Widespread distribution of SI and SPM in host populations

An MST tool that is adopted for water quality and total maximum daily load (TMDL) assessment and restoration throughout the U.S. will of necessity be widely distributed in host populations across the country. Thus, relatively rare markers such as some genes associated with pathogens are likely to be less useful than more common markers, even though they may be highly host-specific. For example, in Europe a bacteriophage (bacterial virus) that infects *Bacteroides fragilis* HSP40 was found only in human sewage and in sewage-contaminated waters (Tartera et al., 1989). This bacteriophage was considered a promising candidate for a human-specific fecal marker; however, its limited distribution in sewage (Scott et al., 2002) and the relative difficulty of the method (Leclerc et al., 2000) have probably contributed to its rare use status in the U.S. F-specific coliphages are common in sewage, but it has been estimated that only ~3% of humans carry this type of phage (reviewed in Leclerc et al., 2000).

The hypothesis that other proposed SPMs have widespread distribution in the gastrointestinal tracts of their respective hosts must be tested. Included among these are the species-specific genetic markers amplified from *Bacteroides* (Bernhard and Field 2000a; Bernhard and Field 2000b), the toxin genes of *E. coli* found in pigs and cattle (Chern et al., 2004; Khatib et al., 2002; Tsai et al.,

2003), and the *esp* gene of *Ent. faecium* (Scott et al., 2005). Some information is available for *E. coli* toxin genes LTIIA and STII, as the prevalence of species-specific forms of these markers was measured in animal waste from farms in several states (Khatib et al., 2002; Khatib et al., 2003). More than 93% of samples from cattle waste lagoons were positive for the cattle-specific LTIIA marker when $>10^3$ *E. coli* were screened, and the frequency of positive results rose to 100% when $>10^5$ *E. coli* were screened (Khatib et al., 2002). The swine-specific STII marker was found in 100% of samples when 35 *E. coli* were screened (Khatib et al., 2003).

Ideally, host-specific SPMs should be present at about the same density in separate populations of a given host species, which would provide greater confidence that sampling effort was adequate when using standardized protocols. Furthermore, it would be advantageous if host-specific SPMs were found at about the same density in various individual animals within a host population, which would facilitate accurate quantification. Very little is known about these concerns for any of the methods, except that the majority of animals in a herd do carry *E. coli,* but generally do not carry enterotoxigenic *E. coli* (Chern et al., 2004).

6.4 Stability of the signal

A required characteristic for a useful SPM is stability of the "signal", whether that signal is a phenotypic pattern, a genetic pattern, or a PCR amplicon. The assumption that genetic patterns/markers are a more stable type of signal than phenotypic patterns has appeared frequently in MST literature (Parveen et al., 1999; Scott et al., 2002; Simpson et al., 2002), due in part to the fact that bacterial phenotypes (traits such as antibiotic resistance or the ability to use a particular carbon source) are influenced by environmental conditions as well as the genetic makeup of the organisms. This assumption should, however, be tested in the context of a microbial source tracking study, in which the testing occurs in a controlled laboratory environment under near-optimal growth conditions. While it is known that some bacteria lose resistance to antibiotics when selective pressure (antibiotic presence) is removed, as occurs when bacteria are cultured from feces or water samples, it is unknown whether this phenomenon occurs often enough to significantly impact the accuracy of MST studies based on antibiotic resistance patterns. Similarly, the frequency and consequences of transfer of antibiotic resistance genes from one SI to another are not established for MST.

The gene(s) for ribosomal RNA (rRNA) are frequently targeted for MST studies (Carson et al., 2003; Parveen et al., 1999) because these genes mutate relatively rarely. Ribosmal RNAs are an integral component of the ribosome, the protein-synthesizing "machinery" of the cell, and certain regions of rRNA are very highly conserved (change very little, if at all, over thousands of generations). The low mutation rates of the rRNA genes do contribute to the stability of many types of fingerprints; in fact, sequencing of rRNA genes within a species such as *E. coli* generally results in very little strain discrimination (Guan et al., 2002). Ribotyping as it is used for MST should, perhaps, be clarified as "genomic ribotyping", since the chromosomal DNA is isolated, cut with restriction enzymes, and chromosomal fragments are separated by electrophoresis (see Chapter 3-Methods). Labeled fragments of the rRNA gene(s) are then used as probes to identify the gene loci on the chromosome. This method can be quite discriminatory, even within a species, because much of the variability in patterns is due to variation outside the conserved rRNA operons. Although it has

been assumed that ribotypes represent a very stable form of signal, no comparisons with phenotypic or other genotypic methods has been published.

A linked assumption of many MST methods is that mutations in host individuals that could change the specificity of the SPM are very rare. An individual could lose the ability to support the SPM if, e.g., a receptor in the gastrointestinal tract experienced decreased affinity for the SPM. Conversely, an individual from a different species might acquire the ability to support the SPM by mutation or horizontal gene transfer. While a recent mutation in a host population would not be a major concern, because few individuals would carry the mutation, over generations it could pose a problem, particularly in isolated host populations.

6.5 Transferable methodology

It is assumed that MST methods will be transferable across laboratories. The ability to successfully perform many of these methods will be dependent upon the relative expertise of laboratory technicians, the equipment and facilities available, and the extent to which protocols are standardized and made "user friendly." As protocols are being developed, every effort should be made to include rigorous controls and streamlined techniques into MST methods. The error associated with the method, whether described in terms of false-positives and false-negatives, or Type I and Type II error, should be thoroughly explored. An important aspect of the analytic parameters used for matching patterns is that as the similarity index required to call two patterns the same becomes more stringent, the number of distinct patterns (ribotypes, for example) identified increases (Hartel et al., 2002). The similarity values imposed for pattern matching must not be chosen arbitrarily, but should rely on measurements of the inherent variability of the method. For example, if *E. coli* isolate X is ribotyped ten times on ten separate occasions, what is the similarity of those patterns? The discriminatory power of the method cannot be greater than its inherent variability, i.e. if ten replicate measurements of the ribotype of *E. coli* X are 92% similar, only ribotypes that are less than 92% similar can legitimately be called different ribotypes. Ideally, a confidence interval should also be calculated to better define differences that should be considered significant, although this has not been practiced in the literature. It is important to keep in mind that development of any MST method that analyzes patterns based on band-matching algorithims requires confirmation of pattern matches and nonmatches by eye before one can rely on the matches called by the software.

6.6 Temporal stability within the host

The ideal SI should exhibit stability within individual host animals and within host populations over time. Although a good deal of information is available on the temporal stability of *E. coli* populations in host animals, very few studies have addressed the temporal stability of other SIs.

Previous studies on the temporal variability of *E. coli* established the concept of transient vs. resident populations of *E. coli* in the gastrointestinal tract. Caugant et al. (1981) defined a "transient" population as one observed at only one sampling point, while a "resident" population was one observed at more than one sampling point. Transient vs. resident populations are a

particularly relevant MST concern if the range of subtypes estimated in natural populations of *E. coli* (100-1000 per host species by multilocus enzyme electrophoresis) (Selander et al., 1987) are found to be comparable in other fecal indicator bacteria. Over an 11-month period, only 5.6% of the *E. coli* isolated from the feces of a single human host were considered "resident"(Caugant et al., 1981), and a total of 53 electrophoretic types were identified using multilocus enzyme electrophoresis. In another study, resident *E. coli* populations from multiple hosts accounted only for 8% of all the electrophoretic types identified (Ochman et al., 1983). A study on temporal stability of *E. coli* in humans, cattle and horses defined a "persistent" ribotype as one that was sampled from an individual in two consecutive sample events (Anderson 2003; Anderson et al., 2003). At least one persistent ribotype was observed per human, although only four of 36 (11%) of the ribotypes observed in the three humans were persistent. *E. coli* populations of horses and cattle tended to display higher diversity (more subyptes per host) than those of humans; however, they followed a similar trend in that most of the *E. coli* subtypes observed were not persistent (Anderson 2003; Anderson et al., 2003). These studies indicate a high probability that the *E. coli* subtype(s) obtained from a single host at a given time are not representative of the *E. coli* population in the animal's feces over time. Such a limitation has major repercussions in the establishment of host origin libraries, which may require continuous updating in order for a particular MST methodology to be able to track the host species (Jenkins et al., 2003) over an extended period of time.

While temporal stability of the SPM in individual host animals is an ideal characteristic for MST, temporal stability at the larger host population level is a characteristic of a useful SI. In a recent study on the temporal stability of *E. coli* ribotypes in cattle herds, individual cattle in the herds were sampled at random during four sample events (Jenkins et al., 2003). The *E. coli* ribotypes that were observed in more than one sample event ("residents") represented only 8.3% of 240 ribotypes. Among the 20 resident ribotypes, no ribotype was found at all four sampling times or in all of the steers sampled. Although many *E. coli* isolates were analyzed per cow (~11 to ~25), individual cattle were not resampled throughout the study. Thus, it could be argued that the observed variability was as likely due to undersampling of individuals in the herd as temporal variability. However, in support of the above results are data from an eight-month study of three beef cattle from one herd (Anderson 2003; Anderson et al., 2003) that were repeatedly sampled. *E. coli* ribotype variability in the feces of these animals was high, sharing between herd members was low, and temporal variability in the dominant ribotypes within each animal was consistently noted. Evidence of the temporal variability of *E. coli* populations in other species was observed in humans and horses (Anderson 2003; Anderson et al., 2003). Two humans that lived together tended to share *E. coli* ribotypes with each other, but not with a human working in the same room, while horses in the same herd shared very few subtypes (Anderson 2003; Anderson et al., 2003). However, investigations of temporal stability carried out on a larger scale (and with a different SI) were more encouraging, as the temporal stability of a large library of *Enterococcus* spp. subtyped by antibiotic resistance analysis was demonstrated for up to a year (Wiggins et al., 2003).

6.7 Geographic stability

Several assumptions based on the geographic distribution of an ideal SI can be identified: (a) SPMs sampled from one population of a host species will be similar to SPMs sampled from another population of the same host species, and a predictive relationship can be established between the

two; and (b) SPMs sampled from host populations separated by broad geographic ranges will exhibit a high similarity index and accurately track the host species. A hypothesis that could be contradictory to (a) and (b) has also been proposed: (c) SPMs exhibit geographic structure, that is, the similarity of SPMs in various populations of one host species is directly proportional to their geographic distance from one another (Gordon, 2001).

Studies indicate that hypothesis (c) regarding geographic structure for populations of the same host species is not met for *E. coli* populations; however, this assumption is probably the least important one for most MST applications. Very little of the variability in *E. coli* populations of humans seems to be attributable to geographic separation (Caugant et al., 1984; Whittam et al., 1983); which may be partly due to the mobility of human populations (Gordon 2001). (Caugant et al., 1984) reported that little geographic structure was observed in *E. coli* populations of families living within the same city, where only 6% of the variability was explained by geographic distance. Only 1% of the variability was explained by geographic distance for families living in different cities. Geographic structure accounted for only a small percentage of the variability in *E. coli* subtypes in mice (2%) (Gordon 1997). While studies that have compared *E. coli* population structure in various animals have found significant contributions to diversity from both geographic location and host source (Gordon and Lee 1999; Souza et al., 1999), only a small percentage of the variability (<20%) was accounted for by these factors. One study on livestock did find geographic structure in *E. coli* populations in cattle and horses, i.e. more ribotypes were shared in host populations in closer geographic proximity; however, no geographic structure was observed for *E. coli* from chickens and swine (Hartel et al., 2002).

Ideally, host populations in all parts of the U.S. would share similar SIs and SPMs so that nationwide (or more inclusive) databases could be constructed. Studies completed to date suggest that this ideal will not be met, at least for library-based methods. In a study performed across a relatively broad geographic area in Florida, *E. coli* from beef, dairy, poultry, swine and human hosts were ribotyped by a one-enzyme procedure (Scott et al., 2003). Although the method accurately differentiated *E. coli* originating from human vs. non-human hosts, it failed to distinguish among the different non-human host species across the broad geographic region. The diversity and distribution of *E. coli* ribotypes differed in captive vs. wild deer (Hartel et al., 2003), which was attributed to diet. The diets of host animals may differ significantly by geographic region, providing one of the drivers for geographic variability of commensal bacterial populations in one host species. *E. coli* and *Enterococcus* libraries from three geographic regions were assessed for broad geographic applicability (Dontchev et al., 2003). Subtyping methods used were antibiotic resistance analysis (ARA), ribotyping (one-enzyme) and pulsed field gel electrophoresis (PFGE). The regional sublibraries (Florida, Shenendoah Valley VA and southwest VA) identified isolates collected from within the region significantly more accurately than they identified isolates from outside the region. A three-region merged library identified the source of isolates much less accurately than each of the regional libraries, and this generalization held true for each of the methods and SIs.

The geographic applicability of an *Enterococcus* ARA library was broadened by increasing library size and representation of isolates from a number of watersheds in the Shenendoah Valley region of Virginia (Wiggins et al., 2003). Six watershed-specific libraries were merged to produce a library of 6,587 isolates, which identified the source of enterococci fairly accurately across the combined geographic area. The geographic range of the merged library was limited, as it identified isolates

from southwest Virginia and Florida significantly less accurately than isolates from the six-watershed region.

6.8 Representative sampling

One of the most important assumptions of any MST method is that the SI population can be adequately sampled so that all (or most) SPMs are represented. The assumption of representative sampling is extremely important with respect to sampling of both host fecal material and SPMs in water samples. Many factors impose limits on the amount of material or isolates that can be analyzed, including cost and time. Under-sampling of SI populations in fecal sources leads to nonrepresentative libraries, which may have high correct classification rates (internal accuracy) but low predictive accuracy for isolates that are not included in the library (Whitlock et al., 2002; Wiggins et al., 2003). Furthermore, nonreprepresentative libraries will display neither temporal nor geographic stability. Various estimates of *E. coli* subtype diversity within host populations have been advanced, e.g. between 100 and 1000 (Milkman, 1973; Selander et al., 1987). Rarefaction analysis of an *E. coli* rep-PCR library determined that a library size of 1535 isolates from humans and twelve animal species was not close to saturation (Johnson et al., 2004), which demonstrates the great diversity in *E. coli* genotypes. A 2:1 ratio of total isolates analyzed to estimated subtype richness has been suggested as a minimal requirement for capturing diversity (Jenkins et al., 2003; Parveen et al., 1999), further increasing the sampling effort needed. Complicating the issue is the fact that different host species and sample types (for example, human feces vs. sewage) contain *E. coli* populations of differing richness (Stoeckel et al., 2004), indicating that sampling effort should be adjusted based on host species and sample type. The apparently low frequency of "resident" *E. coli* subtypes compared to "transients" may be more a reflection of sampling limitations than it is a true characteristic of *E. coli* populations (Jenkins et al., 2003). Achieving representative sampling of *E. coli* populations in environmental waters will be affected by similar concerns; i.e. high-diversity *E. coli* populations were found in both pristine and anthropogenically impacted waters (Chivukula and Harwood 2004).

6.9 Persistence of SPMs in environmental waters

Microbial source tracking studies contain many implicit assumptions about the survival of a chosen source identifier. Many of these assumptions are based upon simplified or idealized views of microorganism survival characteristics. For example, it may be assumed that the decay rate of *E. coli* SPMs entering a stream directly from the feces of a herd of cattle will be exactly the same as the *E. coli* SPMs entering the stream from a failing septic system. However, a significant difference in decay rate might influence the relative numbers of *E. coli* SPMs recovered downstream from these fecal sources, which would in turn lead to inaccurate assessment of initial fecal loads. It is important that such assumptions be recognized and understood when choosing a SI and designing or interpreting MST studies.

The failure of the fecal coliform/fecal streptococcus (FC/FS) ratio for fecal pollution source tracking is a lesson to heed in the current pursuit of microbial source tracking methods. The lesson is particularly relevant when considering attempts to quantify source contributions for total maximum

daily load (TMDL) assessments. The FC/FS ratio has been criticized for aspects such as differential decay rates for fecal coliform and fecal streptococci in aquatic environments (American Public Health Association, 1995; Simpson et al., 2002). Initial assumptions about the comparable survival of coliforms and streptococci proved invalid after further study, and use of the FC/FS ratio as a microbial source tracking method has decreased in recent years. The lesson identifies the importance of testing survival assumptions for MST SIs before methods are widely applied to source tracking problems.

In order for a microorganism to be considered an ideal source indicator, it must meet a number of criteria pertaining to its survival in aquatic environments. An ideal SI would not exhibit any population growth upon entering aquatic environments. It would also have SPM decay rates that are constant over space and time. For example, SPM decay rates would not vary between water types (e.g. temperate freshwater lake or tropical saltwater beach) or across aquatic habitats within a watershed (e.g. lake water column or river sediment). In addition, an ideal SI would have SPM decay rates that would be constant between its primary fecal habitat and secondary aquatic habitats. Any variance from these ideal survival characteristics could have important implications for interpreting results from MST studies.

None of the currently used source identifiers are known to meet all of these ideal survival criteria. Therefore, it is important to understand their survival characteristics to determine where they can still be useful under the conditions of a specific MST study. The survival of some source identifiers has been better studied than others, and in these cases, their survival characteristics may be sufficiently predictable to make the microorganism a useful source identifier under the conditions of a specific MST study area and time. In other cases, important survival hypotheses remain untested and survival characteristics poorly known. This lack of information can compromise the value of the source identifier. The following section explores several assumptions about survival for three of the more commonly used source identifiers: *Escherichia coli*; *Enterococcus* spp.; and *Bacteroides* spp.

Escherichia coli : (i) The SPM decay rates are always negative after it enters water.

Escherichia coli has been regarded as a good practical indicator of fecal pollution that generally survives in aquatic environments between 4 and 12 weeks (Edberg et al., 2000). There are many studies indicating its decay rate is negative after entering water environments such as: river water (Grabow et al., 1975), groundwater (Filip et al., 1987), and seawater (Rozen and Belkin, 2001). However, there are a growing number of reports suggesting that some *E. coli* SPM decay rates may not be negative under certain conditions in aquatic environments.

A number of studies have provided evidence that suggests that *E. coli* can multiply in certain tropical and subtropical environments (Byappanahalli et al., 2003a; Carrillo et al., 1985; Desmarais et al., 2002; Hardina and Fujioka, 2001; Rivera et al., 1988; Solo-Gabriele et al., 2000; Byappanahalli and Fujioka, 2004). For example, one study found high levels of *E. coli* in Florida riverbank soils, and suggested that *E. coli* could be washed into the water during high tides (Solo-Gabriele et al., 2000). Associated laboratory experiments found that *E. coli* was capable of increasing by several orders of magnitude in these soils, and suggested the importance of soil properties and periodic wetting and drying as influential for *E. coli* multiplication. Microcosm experiments by Byappanahalli and

Fujioka (2004) indicated that *E. coli* has the capacity to multiply in tropical soils, but the bacteria require suitable nutrient and moisture conditions availability. A Tropical Water Quality Indicator Workshop in 2001 agreed upon a consensus statement that fecal indicator bacteria like *E. coli* can multiply and persist in soil, sediment, and water in some tropical/subtropical environments (e.g. Hawaii, Guam, Puerto Rico, south Florida) (Fujioka and Byappanahalli 2003).

The question of *E. coli* multiplication in certain temperate environments is also under investigation. For example, *E. coli* counts in sand and water gradually increased over the bathing season at a Lake Michigan beach (Whitman and Nevers 2003), which was attributed to higher survival rates (lower decay rates), and perhaps growth, in warmer temperatures. Growth of *E. coli* associated with the macro-alga *Cladophora* mats in the Great Lakes has also been investigated (Byappanahalli et al., 2003b; Whitman and Nevers, 2003; Whitman et al., 2003). *E. coli* survived over 6 months in Lake Michigan *Cladophora* algal mats (sun-dried and stored at 4°C) and then quickly multiplied when moisture was returned (Whitman et al., 2003). The authors suggested that *Cladophora* could be a secondary habitat and source for *E. coli* in certain beach areas, although the case for natural multiplication needed further validation.

(ii) The SPM decay rates are constant across aquatic habitats.
There have been numerous studies to investigate *E. coli* survival in aquatic environments and in laboratory microcosms simulating aquatic habitats. However, it can be difficult to compare survival studies across different microcosm designs and experimental conditions (e.g. many microcosm studies have been conducted under filtered water or sterile conditions). For this reason, the studies reviewed below are from field studies or laboratory experiments conducted under non-sterile conditions, and survival results are identified based upon whether they were obtained from field studies or from laboratory microcosms.

There are numerous studies to indicate that *E. coli* decay rates are not constant across aquatic habitats. Microorganisms entering aquatic habitats might generally be expected to survive longer under colder temperatures, or if they are attached to particles. For example, a number of studies have found significantly lower decay rates for *E. coli* in sediments than in the associated water column (Burton et al., 1987; Craig et al., 2004; Gerba and McLeod, 1976; LaLiberte and Grimes, 1982). Craig et al. (2004) found *E. coli* at $>5X10^3$ CFU/100 g after 28 days in the sediments of saltwater microcosms, while they were undetectable after 7 days in microcosms containing only water. *In situ* measurements of fecal coliforms in both water and sediment of their river and beach study showed that a rain event caused an intial peak of similar levels in both river water and sediment, which was followed by a more rapid decline of fecal coliform numbers in water than in sediments. Two days after the peak, levels of fecal coliforms were 100 times greater in river sediment compared to water. The authors concluded that there was extended persistence of fecal coliforms in the coastal sediments compared to water (Craig et al., 2004).

There is also evidence of variable *E. coli* decay rates across different types of sediment. *E. coli* decay rates varied according to sediment type, with the greatest rates of decay occurring in beach sediment microcosms consisting of large particle size and high organic carbon (Craig et al., 2004). Burton et al. (1987) found enhanced survival of *E. coli* in sediments with high proportions of clay and nutrients compared with sandy low-nutrient sediments.

These laboratory and field experiments are consistent with observations about *E. coli* persistence from field surveillance studies. Higher *E. coli* counts were consistently found in stream and bank sediments than in the stream water of a small Indiana watershed (Byappanahalli et al., 2003a). The authors suggested that the widespread and consistent occurrence of *E. coli* in the watershed could be attributable to long term persistence (and/or multiplication) of *E. coli* in soil and sediment, and the subsequent erosion and washing of sediment-borne *E. coli* into the water. Considerable progress is being made toward understanding the persistence of *E. coli* in beach habitats that may prove informative for MST studies, i.e. *E. coli* counts per unit weight were 3-17 times higher in sand than in the water column at 6 freshwater bathing beaches in the Great Lakes (Wheeler Alm et al., 2003). Similarly, *E. coli* counts in foreshore sand were typically several orders of magnitude higher than in the water at a Lake Michigan beach (Whitman and Nevers, 2003). These results strongly suggest that *E. coli* decay rates are lower in beach sand than in the water column, and that beach sand could be a significant reservoir for longer term persistence, and subsequent resuspension of *E. coli* into beach waters.

Elevated salinity has a detrimental effect on fecal coliform and *E. coli* survival, particularly in the water column. Numerous studies have shown that the decay rates of these organisms are much greater at marine/estuarine salinities compared to freshwater (Hood et al., 2002; Sinton et al., 2002). Solar radiation also increases *E. coli* decay rates (Sinton et al., 2002; Whitman et al., 2004).

Enterococcus: (i) The SPM decay rates are always negative after it enters water.

Many studies indicate that culturable enterococci decline after entering aquatic environments (Sinton et al., 1993; Sinton et al., 2002). However, there are also a growing number of reports suggesting that Enterococcus SPM decay rates may not be negative under certain conditions in aquatic environments.

Several studies have provided evidence indicating that *Enterococcus* spp. may be able to multiply in certain tropical and subtropical environments (Desmarais et al., 2002; Fujioka et al., 1999). A Tropical Water Quality Indicator Workshop in 2001 reached a consensus statement that fecal indicator bacteria like *Enterococcus* can multiply and persist in soil, sediment, and water in some tropical/subtropical environments (Hawaii, Guam, Puerto Rico, south Florida) (Fujioka and Byappanahalli 2003). However, microcosm experiments by Byappanahalli and Fujioka (2004) suggested that enterococci might require more complex nutrients than E. coli and, thus are less likely to multiply in tropical soils.

Enterococci may also be able to multiply under certain conditions in temperate aquatic environments. *Enterococcus* spp. on drift seaweed at recreational beaches in New Zealand exceeded seawater levels by 2-4 orders of magnitude (Anderson et al., 1997). The presence of genetically identical (clonal) enterococci dominating seaweed populations was suggested as evidence that active growth or selection was occurring, and that enterococci could be washed off into surrounding water. Similarly, a study in southern California suggested that a tidal saltwater marsh was serving as a source rather than a sink for *Enterococcus* contamination of nearby coastal beaches (Grant et al., 2001). The possible growth of *Enterococcus* spp. associated with *Cladophora* mats in the Great Lakes has also been investigated (Byappanahalli et al., 2003b; Whitman et al., 2003).

Enterococcus: (ii) The SPM decay rates are constant across aquatic habitats.

There have been fewer studies of *Enterococcus* spp. decay rates in aquatic ecosystems than *E. coli*; however, available information suggests that *Enterococcus* decay rates are not constant across aquatic habitats. Decay rates of enterococci from municipal waste stabilization pond effluents differed in river water depending upon salinity, season and sunlight exposure (Sinton et al., 2002), i.e. decay rates were higher in more saline waters, in the summer, and when exposed to increased sunlight. Some *Enterococcus* species have been associated with occurrence on plants (Mundt, 1961; Geldrich and Kenner, 1969) and in insects (Martin and Mundt 1972) which may suggest the possibility of more diverse SPM survival strategies which should be tested.

Bacteroides: (i) The SPM decay rates are always negative after it enters water.

Enteric anaerobes like *Bacteroides* spp. and *Bifidobacterium* spp. have been suggested as indicators of recent fecal pollution because they are believed to have predictably negative decay rates and survival times of hours in oxygenated waters (Carrillo et al., 1985; Fiksdal et al., 1985; Resnick and Levin, 1981). Bernhard and Field (2000a) suggested that ease of detection and longer survival in water made *Bacteroides-Prevotella* genetic markers superior to those of *Bifidobacterium*. *Bacteroides fragilis* did not maintain culturability as well as *E. coli* or *Enterococcus faecalis* in dialysis bags suspended in aerobic freshwaters (Fiksdal et al., 1985), but immunofluorescence assays demonstrated 18% persistence after 192 hours. Another study found that *Bacteroides* cells could survive for up to 6 days in drinking water under oxygen-stressed conditions (Avelar et al., 1998).

Like all MST methods, the usefulness of nonlibrary-based methods such as PCR detection of *Bacteroides* is based upon the assumption that these anaerobic bacteria do not multiply upon entering aquatic ecosystems. While the few studies conducted to date suggest this is the case, the fate and ecology of anaerobes like *Bacteroides* spp. in aquatic ecosystems remains poorly understood. It is possible that certain aquatic habitats, such as sediments, may provide suitable environments for anaerobic *Bacteroides* spp. to exhibit population growth. It is noteworthy that it has taken many years of studying the survival/growth of *E. coli* in diverse aquatic ecosystems to better understand some of its potential limitations as a source identifier. Hypotheses related to the possibility of *Bacteroides* sp. multiplication in certain unique aquatic habitats (e.g. anoxic sediments) need to be tested.

(iii) The SPM decay rates are constant across aquatic habitats.

There have been few studies of *Bacteroides* spp. survival in aquatic ecosystems, and so there is insufficient information to evaluate whether SPM decay rates are constant across aquatic habitats. One study (Kreader 1998) found that persistence of PCR-detectable DNA from the fecal anaerobe *Bacteroides distasonis* was dependent upon temperature and predation. Laboratory and *in situ* studies in river water found that *B. distasoni* was detectable by PCR for at least two weeks at 4° C, but for only 4-5 days at 14°C, 1-2 days at 24°C, and 1 day at 30°C. Although the PCR method detected both dead and living bacteria, predators were considered important factors in the decline of

both dead and living cells. The author stressed that seasonal variation in the *B. distasoni* decay rate would need to be considered for any water monitoring applications.

6.10 Persistence of SPM in primary vs. secondary habitats

Savageau (1983) advanced the concept that the gastrointestinal tract is a primary habitat for *E. coli*, while external environments such as soil and water are secondary habitats. In a recent review related to bacterial source tracking (BST), Gordon (Gordon 2001) asserted that for any bacterial species that is used to identify human and animal sources of fecal pollution in surface water, several assumptions must be validated. One of these assumptions is that "the clonal composition of the species isolated from soil and water [secondary habitat] represents the clonal composition of the species in the host populations responsible for the fecal inputs [primary habitat] to the environment". The rationale for this statement is clear: if the fecal SPM(s) that are used as the source identifier persists poorly in the water relative to other SPMs, the source-specific fecal signal will rapidly disappear.

Several studies on the distribution of *Escherichia coli* subtypes in the primary habitat vs. secondary habitat showed distinct differences in subtype distribution between the two. One hundred thirteen distinct *E. coli* electrophoretic types (determined by multi locus enzyme electrophoresis or MLEE) were isolated from bird feces and the litter on which they had defecated. Only 10% of the clones were found in both the primary and secondary habitat (Whittam 1989). Another study (Gordon et al., 2002) compared electrophoretic types (ETs) of *E. coli* from feces of two human couples, each representing a household, and *E. coli* from each household's septic tank. This study indicated that *E. coli* clones from a secondary habitat such as a septic tank can differ significantly from the primary habitat such as the couples' feces. Ribotyping of *E. coli* from dog feces, untreated wastewater and contaminated soil inoculated into water showed that the dominant subtypes in the primary habitat were distinct from those in the secondary habitat, and that certain "survivor" strains could be identified (Hood et al., 2003). *E. coli* clones isolated from swine manure slurry (a primary source, but secondary habitat) were compared to those isolated from soil inoculated with the same slurry by the genotypic method ERIC-PCR (Topp et al., 2003). Although a major shift in community structure was evident upon comparison of isolates from the secondary (manure slurry) vs. tertiary (soil) habitats, many subtypes were shared between the two habitats. However, one SPM that was prominent in manure was not recovered from soil, indicating differential survival of SPMs.

All types of F-specific RNA coliphages apparently do not have the same decay rate in water, as Type IV strains were less persistent than Types I, II and III in one study (Brion et al., 2002). F-specific RNA coliphages also may experience higher inactivation rates in warm waters compared to cooler waters (Cole et al., 2003).

Based on the above mentioned phenotypic and genotypic studies, *E. coli* appears to be a questionable candidate as a source tracking organism, although genetic fingerprinting of *E. coli* is the basis for some commercial source tracking enterprises. For library/culture independent methods such as the PCR for *Bacteroides* (Bernhard and Field, 2000b; Bernhard et al., 2003) and *E. coli* toxin genes (Khatib et al., 2002; Khatib et al., 2003) the primary habitat versus secondary habitat criterion for validity may be less stringent, but is still applicable. The library independent method is binary; a genetic signal specific for an animal host is either detected in an environmental sample or it is not.

However, if the signal (DNA in this case) is very short-lived in the water compared to indicator organisms and pathogens, it will not serve its purpose. Furthermore, efforts are underway to develop quantitative PCR protocols for some markers, and the efficacy of these methods will rely to a certain extent on the primary-vs.-secondary habitat hypothesis.

6.11 Relevance of SI to regulatory tools

Indicator bacteria such as coliforms have been used for over a century as indicators of fecal contamination in water. In the U.S., indicator bacteria (fecal coliforms, *E. coli* and enterococci) are the standard by which microbial water quality in environmental waters is measured. Currently, almost all MST studies, whether carried out on bathing beaches, in reservoirs, or for total maximum daily load (TMDL) assessments, are responses to exceedances of indicator bacteria standards. Understandably, water quality managers prefer a SI that is directly connected to the regulatory parameter (indicator bacteria) for assessment of fecal sources; however, as MST methods are tested and validated in the field, a method that utilizes one or more alternative SIs may show greater utility than methods that use conventional indicator organisms.

An assumption of MST methods that employ SIs other than conventional indicator bacteria is that the results generated by the SI will have some discernible relationship with indicator bacteria levels. The failure of indicator organism and SI to correlate is not *a priori* a reason to discard the SI, particularly if it is associated with human health risk (see below) (Field et al., 2003). The interpretation of the results may, however, prove more complex when the SI is not an indicator organism, particularly in the case of TMDL assessment. Very little is known about these relationships in environmental waters, making them an essential area for further study.

6.12 Relevance of SI to human health

The ultimate goal of MST is to determine the host species responsible for fecal pollution from among many possible candidates; however, simply discriminating human fecal material from nonhuman is of practical use for water quality managers (Harwood et al., 2003; Myoda et al., 2003; Stewart et al., 2003). The usefulness of human vs. nonhuman source discrimination is due in part to the assumption that human fecal material poses a greater human health risk than other types of fecal material (Scott et al., 2002). Although some of the rationale for this assumption is based on indirect evidence (i.e. the majority of gastroenteritis associated with recreational water use is caused by viruses, and human enteric viruses are highly host-specific), direct evidence also exists. Detection of enteric viruses, which are exclusively of human source, was correlated with gastroenteritis in swimmers in marine waters (Haile et al., 1999), and a meta-analysis of epidemiological studies showed that enteric viruses were strongly associated with gastroenteritis (Wade et al., 2003). Thus, SIs that can discriminate human vs. nonhuman fecal pollution should be useful, provided they have some association with human health outcomes and/or pathogens.

The indicator organism paradigm is based on the assumption that indicator organisms are predictive of human health risk. Much debate and many epidemiological studies have explored this assumption (reviewed in (Wade et al., 2003)). The 1986 standards for recreational water quality specify the use

of *E. coli* (and not fecal coliforms) for freshwater bodies, and enterococci for freshwater and marine water bodies (U.S. Environmental Protection Agency 1986). A meta-analysis of the epidemiological literature on gastroenteritis resulting from recreational water use found that *E. coli* was significantly associated with gastroenteritis in fresh water, and that enterococci were significantly associated with gastroenteritis in marine water (Wade et al., 2003), supporting the use of these organisms as indicators of human health risk. Coliphages were also predictive of gastroenteritis, although fewer studies were available for analysis. Much work remains to be conducted on the correlation of alternative SIs with human health risk in environmental waters.

6.13 Summary

- None of the source identifiers currently used meet the criteria for an ideal SI, including those that are indicator organisms recognized for regulatory uses.
- The ecology and population biology of some source identifiers, particularly fecal coliforms/*E. coli*, are much better understood than that of others, such as the enterococci and *Bacteroides* spp. While the high genetic diversity of *E. coli* allows great discrimination between subtypes, it also complicates development of known source libraries.
- The correlation of novel SIs such as *Bacteroides* with levels of conventional indicator organisms and/or with human health outcomes has not been determined, but should be if public health effects are under consideration.

Chapter 7. Application of MST Approaches

This Chapter presents a series of case studies involving application of several MST methods. The intent of this Chapter is to provide some real-world examples of how various MST methods have been applied. There have been far more studies than can be covered in this Chapter, so several have been chosen as examples. Many of these examples were compiled based on communications with the authors of the studies. However, others were been written based only on published reports and/or journal articles, and thus are not as complete. We have tried to include examples of studies using MST methods in current use, as well as some projects involving multiple techniques.

Each of the following case studies follows the same general outline. First is a general description of the watershed, with a statement of the problem and the goals and objectives of the project. Next is a brief description of the methodology used, including the classification method. (For more detailed information on the methods, please refer to Chapter 3 in this document.) For the studies that used a library-based MST method, a description of the library is included, with information on known source samples and evaluation methods. Following that is a section on sampling considerations, describing how and when the water samples were collected. Finally, a section on the outcomes of the study follows, with a summary of the major results and conclusions, and information on follow-up studies and implementation efforts.

This Chapter includes 8 case studies (presented in no particular order) which illustrate the use of many, but not all, currently applied MST methods:

Case 1. Saint Andrews Park (Georgia). Targeted sampling and *Enterococcus* speciation.
Case 2. Tampa Bay (Florida). ARA with fecal coliforms, ribotyping with *E. coli*, and human-pathogenic enterovirus detection.
Case 3. Vermillion River (Minnesota). rep-PCR with *E. coli*.
Case 4. Anacostia River (Maryland/District of Columbia). ARA and PFGE with enterococci.
Case 5. Accotink Creek, Blacks Run, and Christians Creek (Virginia). Two-enzyme ribotyping with *E. coli*.
Case 6. Avalon Bay (California). Host-specific *Bacteroides*/*Prevotella* markers and human-pathogenic enterovirus detection.
Case 7. Holmans Creek (Virginia). ARA with *E. coli*.
Case 8. Homosassa Springs (Florida). F+ RNA coliphage genotyping.

Several validation steps have been identified as being essential as part of the design of any new MST study (refer to Chapter 5 for details). These include precision measurements, positive and negative controls, external validation standards (including known source field samples to test library classification accuracy and primer specificity), spiked samples (including a matrix spike for PCR on community DNA extracts), and consideration of independent ancillary data (land use data, sanitary surveys, results by multiple methods, etc.). When reading these case studies, keep in mind that many of them were initiated several years ago, and the designers did not have the benefit of what has been learned in subsequent years. With 20/20 hindsight, it is easy to point out limitations of even the best contemporary studies. Every study could be improved on, given more time, more money, and better understanding of the approaches. The purpose of this Chapter is not to criticize, but to

learn from the past and to use these practical examples as guides when designing new source tracking projects.

Case 1. St. Andrews Park (Georgia)

Source of information: Hartel, P., K. Gates, and K. Payne. 2004. Targeted sampling of St. Andrews Park on Jekyll Island to determine sources of fecal contamination

A. **General description**

1. **Watershed description**. Saint Andrews Park is located on the southern tip of Jekyll Island facing St. Andrews Sound. The park beach is approximately 1.3 km long and is bounded by Beach Creek at the northern end and the tip of Jekyll Island at the southern end. Previous fecal coliform sampling of the park suggested that fecal contamination might have originated from a number of locations north of the park. A sampling of those creeks and pipes emptying into the Jekyll River, which flows north of the park into St. Andrews Sound, and of the sound itself, was conducted. Several creeks showed high counts of fecal enterococci. One broken sewer pipe, servicing a local restaurant, was observed and subsequently repaired.
2. **Problem definition**. Recently, high numbers of fecal coliforms were observed during beach monitoring of the park, and these numbers resulted in a beach advisory.
3. **Statement of objectives**. To use targeted sampling and enterococcal speciation to identify sources of fecal contamination to St. Andrews Park during calm weather conditions, and, if weather conditions in the one-month sampling period permit, during stormy weather conditions.
4. **Date of study**. Completed June 3, 2004

B. **Analytical approach**

1. **Method description**. The method chosen was targeted sampling followed by *Enterococcus* speciation. Targeted sampling has four steps. The first step is to divide the sampling into two conditions: base and storm. The second step is to conduct intensive sampling(s) of the contaminated waterway, collecting as many samples as possible in one day. Collecting the samples in this manner reduces temporal variability. The third step is to combine the fecal bacterial numbers with GPS data. The fourth step is to conduct MST at "hot" areas (i.e., those sites containing relatively high fecal bacterial numbers). The process is then repeated for storm conditions.

   Given the circumstances of St. Andrews Park, with its limited number of potential fecal-source categories (i.e., humans, pets, and wildlife) and the limited one-month sampling time, the simplest, quickest, and least expensive MST method was considered to be the one based on *Enterococcus faecalis*. In this phenotypic method, enterococci are speciated biochemically and the percentage of enterococci represented by *Ent. faecalis* determined. High percentages of *Ent. faecalis* are associated with humans and some wild birds (Wheeler et al., 2002).

All confirmed enterococci from Quanti-tray wells were speciated according to a modification of the Manero and Blanch (1999) protocol. The protocol was modified to identify only three fecal enterococcal species, *Ent. faecalis, Ent. faecium,* and *Ent. gallinarum*. In a further test for the presence of human-associated *Ent. faecalis*, approximately 100 isolates were spotted on each of two 0.45-micron membranes on 5-cm Petri plates containing mEI agar (Becton-Dickinson). The plates were incubated at 41±0.5 °C for 24 hours and were sent by overnight mail to Biological Consulting Service of North Florida (Gainesville, FL). Their proprietary method (Scott et al., in review) tests for the presence or absence of a human-specific factor in enterococci isolates.

2. **Target organisms**. Enterococci, recovered using Enterolert™ as primary cultivation (enrichment) medium with recovery and colony isolation using Enterococcosel agar. Confirmed enterococci were speciated (targeting *Ent. faecalis*) and tested for a human-specific marker as described above.
3. **Statistical approach/classification method**. A high proportion of enterococci as *E. faecalis* was taken to indicate presence of human or avian fecal sources. Presence of human-specific marker used to differentiate human from avian fecal contamination sources.

C. **Sampling considerations**

1. **Number and frequency of samples**. Targeted sampling, twice (21-22 April and 04 May, 2004). Number of samples not reported but indicated to be about 60 on April 21-22. Fifteen samples collected 04 May.
2. **Type of sample (depth-width integrated or a simple grab)**. Grab.
3. **When collected (season, flow conditions)**. Spring; one set under calm (low suspended sediments) and the other under windy (high suspended sediments) conditions
4. **Volume of sample and concentration factor.** 100 mL analyzed for each sample – *Enterococci* colonies evaluated for host-specific factor
5. **Evaluation and validation**
   a. **Spiked samples**. None reported
   b. **Blind samples**. None reported
   c. **Negative controls**. None reported
   d. **Comparisons to independent ancillary data**. Turbidity, land use pattern for one area (marsh)

D. **Outcomes**

1. **Summary of results and conclusions**. During calm weather, highest concentrations of enterococci were detected in the upper reaches of Beach Creek, the sediments of the creek, and the bathing area. Species composition in creek sediments and bathing area sediments were different, which was taken to indicate effects by different enterococci sources. The large proportion of *E. faecalis* in the upper reaches of Beach Creek was interpreted to implicate wild birds or humans as a source. The conclusion that wild birds, not humans, were a major source in the upper reaches of Beach Creek was supported by the marshy character of the area, which makes a human source unlikely at that location. Though there was no statistical correlation between turbidity and enterococci concentration, co-incidence of high enterococci concentrations and high turbidity in windy weather was taken as evidence that sediments were a source of elevated water-column numbers during windy

weather.

Human-specific adhesin factor was not detected in any of 200 isolates tested. This was interpreted as evidence that human sources were not major contributors of enterococci to the test area. However, the incidence rate of the human-specific marker in enterococci colonizing the human population is unknown, and there was no mention of a positive control in marker detection by the research method used, which might limit the interpretability of this result. Human population size, local approaches to control human waste, or proximity of human residences to the affected area, factors which were certainly considered in the study, were not mentioned in the report as further corroborating data.

2. **Implementation efforts based on the study**. None reported
3. **Follow-up monitoring**. None reported

Case 2. Tampa Bay (Florida)

Source of information: **J.B. Rose, J.H. Paul, M.R. McLaughlin, V. J. Harwood, S. Farrah, M. Tamplin, J. Lukasik, M. Flanery, P. Stanek and H. Greening**. 2000. Healthy Beaches Tampa Bay: Microbiological monitoring of water quality conditions and public health impacts. Final Project Report.

A. General description

1. **Watershed description**. Tampa Bay is located on the west central coast of Florida, opening to the Gulf of Mexico. This is a shallow subtropical estuary, one of the largest in the southeastern U.S. It is valued for its ecosystem, fisheries, recreational uses and as a port. The drainage basin is approximately 2300 square miles and includes 9 major and 76 minor sub-basins. The major tributaries in the Bay are the Hillsborough, Alafia, Little Manatee and Manatee Rivers, while minor systems include Alligator Creek, Joe's Creek (Pinellas County), Rocky Creek, Double Branch Creek, Sweetwater Creek (northwest Hillsborough County), Tampa Bypass Canal, Delaney Creek, Bullfrog Creek (central and south Hillsborough County), and Frog Creek (Manatee County). Freshwater inputs are very important to the Bay and are associated with rainfall, with about 60% of the annual precipitation occurring from June to September. Along with this freshwater input come contaminants originating from point and non-point sources.

2. **Problem definition**. Risk to swimmers using polluted beaches has been a major issue associated with the setting of ambient water quality standards and discharge limits to recreational sites. Prevention of disease depends on the use of appropriate fecal indicators. However, the finding that the most widely used fecal contamination indicator, fecal coliforms and more specifically *E. coli,* grow naturally on vegetation in warm climates clearly brings into question whether these or other indicators developed for temperate climates are applicable in Florida and other southeastern areas. In recent years, total and fecal coliform bacterial indicators have not been able to consistently indicate the persistence of pathogens, especially viruses in surface waters. F-specific RNA coliphage, enterococci and *Clostridium perfringens* have been suggested as better indicators of fecal contamination and public health risks in tropical and sub-tropical regions.

3. **Statement of objectives**. This study examined traditional and alternative pollution indicators, as well as the presence of pathogenic viruses, and their association with environmental variables (salinity, rainfall, stream flow) in fresh and marine water systems of the Tampa Bay area. The goals of this study were: 1) to determine appropriate indicators for microbiological water quality for recreational sites in area beaches and for Tampa Bay; and 2) to determine the occurrence of pathogens along with indicators in Tampa Bay watersheds and area beaches, their associated sources (animal vs human), public health risks and potential for management. The final goal of this project was to form the baseline for other studies and help to develop a long-term strategy for addressing or enhancing Florida water quality.

4. **Date of study**. Sampling began in June 1999 and ended in August 2000.

B. **Analytical approach**

1. **Method description**. ARA (using a combination of 32 antibiotics and antibiotic concentrations). Ribotyping was performed by the method of Parveen et al. (1999), using *Hin*dIII. Enterovirus counts were carried out on human cells lines.

2. **Target organisms**. Fecal coliforms for ARA, *E. coli* for ribotyping, enterovirus.

3. **Statistical approach/classification method**. Library-dependent methods used linear discriminant analysis. ARA: Classification was performed 6-way (chicken vs. cow vs. dog vs. human vs. pig vs wild). Ribotyping: Classification was performed 2-way (human vs. animal). Library-independent method used detection of human-pathogenic enterovirus or *Bacteroides fragilis* phage to indicate presence of human fecal contamination.

C. **Library considerations**

1. **When collected**. ARA: Not reported. Ribotyping: A previous isolate collection was used (Parveen 1997), plus 59 newly-collected isolates.

2. **Sources included**
   a. **numbers of samples (reference feces) of each source**. ARA: Not reported. Ribotyping: Not reported.
   b. **numbers of isolates from each sample (average)**. ARA: Not reported. Ribotyping: Not reported.
   c. **library size.** ARA: 3,309 fecal coliform isolates, of which 1,154 are from humans and the remainder are from chickens, cattle, dogs, pigs and wild animals (mostly wild birds and raccoons). Ribotyping: 238 isolates (114 human, 124 animal).

3. **Evaluation and validation**
   a. **testing for representativeness (cross-validation, holdouts, blind samples)**. ARA: Not reported. The ARCC of the ARA library was not reported. Ribotyping: Not reported. The ARCC of the ribotyping library was 82%
   b. **testing for random classification**. ARA: Not reported. Ribotyping: Not reported.
   c. **comparisons to independent ancillary data**. Compared to other fecal indicators including fecal coliforms, coliphage, *Bacteroides fragilis* phage, *Clostridium* and enterococci.

D. **Sampling considerations**

1. **Number and frequency of samples**. Twenty-two sites were chosen in Tampa Bay for this study. The final choices were based on watershed representation, areas of concern in regard to pollution, accessibility and previously sampled sites. Eleven sites of primarily rural or suburban land use were chosen in Hillsborough and Manatee counties. Six additional sites were located in highly urban areas, and 4 beach sites were chosen to represent various types, including urban, heavy boat use, recreational site in rural area, and pristine unpopulated

beach. A control site was located in the middle of the bay. Each site was sampled monthly for a period of approximately one year for traditional and alternative fecal indicators, which included fecal coliforms, *E. coli*, enterococci, *Clostridium perfringens* and coliphage. Ten of the sites were chosen for in-depth testing (including antibiotic resistance analysis of fecal coliform isolates, ribotyping of *E. coli* isolates, and enterovirus detection). These sites were monitored 6 times throughout the study.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. Sampling began in June 1999 and ended in August 2000.

4. **Number of isolates per sample**. ARA: 48. Ribotyping: 1-5.

E. **Outcomes**

1. **Summary of results and conclusions**. Perhaps one of the most striking findings of this study is the extent to which wild animals dominate as a source of fecal coliform and *E. coli* isolates. Over the course of the study, wild animal isolates dominated each site according to ARA. Ribotyping results were consistent; in 74% of all samples (n=53) the majority of isolates were identified as nonhuman. However, all sites displayed some level of human fecal pollution according to the three methods used (ribotyping, ARA and enterovirus counts). The three different methods did not always coincide on their detection of the presence or absence of human contamination, however the data collected over the course of the study unambiguously documents the presence of human fecal sources.

   Level of agreement among the two library-dependent methods (antibiotic resistance analysis and ribotyping) and enterovirus counts was assessed for each sampling event. Sites were scored positive for human impact when >20% of isolates were identified as human by ribotyping and by ARA, and when any enterovirus counts were detected. Sites were scored negative for human impact when <20% of isolates were identified as human by ribotyping and by ARA, and when no enterovirus counts were detected (<1/100 ml). Ribotyping and ARA results agreed for 31 of 53 samples (58%). Ribotyping and enterovirus results agreed for 29 of 52 (56%) samples. ARA and enterovirus results agreed most frequently, as positive results at the same sites were noted for 38 of 55 sampling events (69%). All three methods agreed for 21 of 51 samples (41%). There was no correlation between the percent of isolates identified as human by ribotyping and enterovirus counts. The Spearman rank correlation test (used for non-normally distributed data) showed a significant correlation between the percent of isolates identified as human by ARA and enterovirus counts ($p < 0.05$; $r = 0.324$).

   The percentage of isolates identified as human by ARA was significantly correlated with enterovirus counts, but the percentage of isolates identified as human by ribotyping was not significantly correlated with enterovirus counts. This discrepancy points to the need for including the fingerprints of more isolates from known, local sources in the respective databases. In the case of ARA, we have seen dramatic improvements in correct classification rates by adding fingerprints from local sources. The genetic and phenotypic variability of

indicator bacteria such as *E. coli* is quite great, therefore any information that can be obtained on the fingerprints of actual contamination sources to a watershed is extremely valuable. Encouragingly, ribotyping, ARA and enterovirus counts agreed on the presence/absence of human sources in 41% of samples. The probability of the three methods agreeing by chance alone is 0.125 (0.5 X 0.5 X 0.5), therefore the three methods agree on the presence of contamination far more frequently than would be predicted by a purely stochastic process.

2. **Implementation efforts based on the study**. None reported.

3. **Follow-up monitoring**. Improvements to the databases (ribotyping and ARA) are underway to increase accuracy.

Case 3. Vermillion River (Minnesota)

Source of information: **Sadowsky, M.** 2004. "Determination of Fecal Pollution Sources in Minnesota Watersheds". Technical Report prepared for the Legislative Commission on Minnesota Resources.

A. General description

1. **Watershed description**. The Vermillion River Watershed encompasses 372 square miles, mostly located through central Dakota County south of the Twin Cities metropolitan area. The main stem originates in Scott County to the west and flows generally northeast to the City of Hastings. Current land use in the watershed is still dominated by agriculture with suburban areas and smaller urban growth centers interspersed throughout the watershed.

2. **Problem definition**. In 1998, the Vermillion River main stem, from Empire Township to the dam in Hastings, was listed on the Federal Clean Water Act's 303(d) list of impaired waters for fecal coliform bacteria. The river was not meeting its designated use (primary contact – swimming) standard due to high bacteria levels. Also in 1998, the Vermillion River was placed on the Minnesota Pollution Control Agency's (MPCA) list of waters in need of a total maximum daily load (TMDL) assessments for fecal coliform. In 1999 the MPCA, with the help of local agencies and citizens, collected fecal coliform samples throughout the Vermillion River watershed to begin determining the extent of the bacterial problem. These data indicate that the river and its tributaries have bacteria levels in excess of the MPCA's state standard of 200 organisms/100 ml of sample.

3. **Statement of objectives**. The study was conducted to determine the major sources of fecal pollution in the watershed.

4. **Date of study**. April 2001 through December 2003.

B. Analytical approach

1. **Method description**. HFERP (Horizontal, Fluorophore-Enhanced Rep-PCR.)

2. **Target organisms**. *Escherichia coli*.

3. **Statistical approach-classification method**. A 4-way analysis was performed (domesticated vs human vs wildlife vs pets). Each test isolate was assigned to the group of the known-source isolate with which it had maximum similarity with 1% optimization using a curve-based (Pearson correlation coefficient) calculation as applied by BioNumerics software. Robustness of this classification was evaluated using the custom script ID Bootstrap within BioNumerics, and classifications were rejected for probabilities less than 90%.

C. Library considerations

1. **When collected**. July 1999 through November 2002, from known sources from central Minnesota, Duluth, and the far-western edge of Wisconsin.

2. **Sources included**

   a. **numbers of individuals of each source**. Cat (37), Chicken (86), Cow (115), Deer (64), Dog (71), Duck (42), Goat (36), Goose (73), Horse (44), Human (197), Pig (111), Sheep (37), Turkey (69).

   b. **numbers of unique isolates from each source**. Cat (48), Chicken (144), Cow (189), Deer (96), Dog (106), Duck (81), Goat (42), Goose (135), Horse (78), Human (210), Pig (215), Sheep (61), Turkey (126).

3. **Evaluation and validation**

   a. **testing for representativeness (cross-validation, holdouts, blind samples)**. Using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) calculation. The ARCC using this approach was 74%.

   b. **testing for random classification**. None

   c. **comparisons to independent ancillary data**. None.

D. **Sampling considerations**

1. **Number and frequency of samples**. Ten sites were sampled along the Vermillion River during each sampling event. Stream samples were collected on 07/11/01, 08/08/01, 09/05/01, 10/03/01, 03/27/02, 05/01/02, 06/05/02, and 07/02/02.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. Collected from 07/01-07/02. Samples were collected during periods of high and low flow.

4. **Number of isolates per sample**. The average number of isolates for each site on each sampling date was 25.

E. **Outcomes**

1. **Summary of results and conclusions**. **I**dentifications indicated that 14% of unknowns matched with geese, 12% with pigs, 12% with cats, 10% with cows, 9% with human, 9% with deer, 9% with sheep, and 9% with turkey. The remaining percentages (30%) then fall off to match with the other groups or remained unclassified. The conclusion was that geese, pigs, cats, cows, humans, deer, sheep, and turkeys were the dominant sources of fecal pollution in the watershed.

2. **Implementation efforts based on the study**. None.

3. **Follow-up monitoring**. None.

Case 4. Anacostia River (Maryland/District of Columbia)

Source of information: Hagedorn, C., K. Porter, and A. H. Chapman. 2003. Bacterial Source Tracking to Identify Sources of Fecal Pollution in the Potomac and Anacostia Rivers and Rock Creek, Washington, D.C. Final Project Report.

A. **General description**

1. **Watershed description**. The Anacostia River watershed is located in the Maryland counties of Montgomery (34%) and Prince George (49%), and in the District of Columbia (17%). It is a 456 km^2 drainage area and contains 15 km of river (plus an additional 25 km represented by two major tributaries), with 2% of the land in agricultural use, 28% in forest and park, and 70% in residential and industrial (urban). The possible/suspected sources of fecal contamination in the Anacostia River watershed are humans, waterfowl, seagulls and other shore birds, pigeons, starlings, dogs, and cats, deer, raccoons, muskrats, cattle and horses. The river is a tidal embayment with minimal recharge at its lower end where it empties into the Potomac River.

2. **Problem definition**. The Anacostia River does not meet the Clean Water Act national goal of "fishable or swimmable" standards. It is on the Priority List of impaired waters due to elevated fecal coliform levels and adversely affected benthic aquatic organisms.

3. **Statement of objectives**. The study was conducted to determine the major sources of fecal pollution in the stream, and especially to determine if human fecal pollution was present.

4. **Date of study**. July 2002 through May 2003.

B. **Analytical approach**

1. **Method description**. ARA, using 30 combinations of antibiotic x concentration, and PFGE using the restriction enzyme *Not*1.

2. **Target organisms**. *Enterococcus spp*.

3. **Statistical approach-classification method**. Linear discriminant analysis. Classification was performed 5-way (bird vs. human vs. livestock vs. pets vs. wildlife).

C. **Library considerations**

1. **When collected**. May 2002 through May 2003, from Four-Mile Run (Arlington County, Va), the Lower Potomac area (Coan and Little Wicomico Rivers), the area around Colonial Beach (Va), and from the Upper Potomac area (Harper's Ferry to Great Falls), from the Blue Plains Wastewater Facility, and the greater D.C. area.

2. **Sources included**

a. **numbers of samples of each source**. Bird: 40; Human: 31; Livestock: 23; Pets: 52; Wildlife: 22.

b. **numbers of isolates from each sample (average)**. Bird: 6; Human: 12; Livestock: 12; Pets: 6; Wildlife: 12.

c. **library size.** ARA = 1,806 isolates (248 bird, 430 human, 699 livestock, 168 pets, and 261 wildlife); PFGE = 750 isolates (150 per source for each of the five sources), all drawn from the samples in a (above), no more that 8 isolates per sample for the PFGE library (6 or less for most).

3. **Evaluation and validation**

a. **testing for representativeness (cross-validation, holdouts, blind samples)**. The ARCC of the ARA library was 89% and the ARCC of the PFGE library was 93%. The pulled-sample ARCCs were 74% for ARA and 81% for PFGE. Blind samples were all human isolates, as this was the most important source in the project. For ARA, the RCC for new sets of human isolates were 70% at the start of the study (when the library was roughly two-thirds completed), and 79% at the end of the study with the complete library. The RCC for blind samples with PFGE was 2% to 5% above the ARA values.

b. **testing for random classification**. Random rate of classification for ARA was 26%, or about 6% above the random expectation of 20% for a 5-way classification. Random rate of classification for PFGE was 24%, or about 4% above the random expectation of 20%.

c. **comparisons to independent ancillary data**. Seventeen combined sewer outflows (CSOs) are located on the Washington, D.C. portion of the river (10 km of the 15-km Anacostia River mainstem are located within the District). The city's NPDES permit allows 2.1 billion gallons of treated sewage per year to be discharged into the river. This limit is exceeded in most years, but information regarding the actual dates that the overflows occur or the amounts discharged are not readily available. What is known is that the discharges are almost always the result of storms and overflows.

D. **Sampling considerations**

1. **Number and frequency of samples**. Six sites were sampled along the Anacostia River. Samples were collected monthly between July 2002 and May 2003 (10 months). For quality control purposes, 10 duplicate samples were collected, one each month. An additional two sets of samples were taken immediately after heavy storms, one in the fall and one after a snowmelt in the winter for a total of 82 samples collected on 12 dates.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. Samples were collected during periods of high and low flow.

4. **Number of isolates per sample**. 24 for ARA, 8 for PFGE.

E. **Outcomes**

1. **Summary of results and conclusions**. The dominant sources over all 10 months of sampling were (using ARA) birds (31%), wildlife (25%), and humans (24%), followed by pets (20%). Livestock detections were essentially non-existent. There was a seasonality trend, as bird and wildlife sources dominated during the low–flow warm weather months (July, August, September, and October), whereas human and bird sources dominated during the high flow-cold weather months (January, February, March, and April). Storm events (both in October) elevated the human signature to levels found during high flow, even though the two storms occurred at the end of the low flow months. A March snowmelt event elevated the human signature even higher (42.4%), indicating that high flow events were related to an increased human signature (any major high-flow event in the city results in sewer overflows). The PFGE water sample results mirrored those from ARA, (except that wildlife and bird sources were each reduced by an average of 3% to 4% and human was higher by the same amount), and the two datasets had an r^2 value of 82.6%.

2. **Implementation efforts based on the study**. None.

3. **Follow-up monitoring**. None.

Case 5. Accotink Creek, Blacks Run, and Christians Creek (Virginia)

Source of information: Hyer, K. E. and D. L. Moyer. 2003. Patterns and Sources of Fecal Coliform Bacteria in Three Streams in Virginia, 1999-2000. USGS Water-Resources Investigations Report 03-4115

A. **General description**

1. **Watershed description**. Areas of three Virginia streams were chosen for evaluation in the reported study: Accotink Creek, drainage area 25 mi^2, human population greater than 110,000, was primarily urban; Blacks Run, drainage area 20 mi^2, human population about 34,700, was mixed urban and agricultural; Christians Creek, drainage area 107 mi^2, human population about 12,000, was primarily agricultural. Extensive base-flow, event-flow, and continuum sampling was done in each watershed over a period of 20 months. Microbial source tracking by use of ribotyping was performed on *E. coli* isolates collected at a single, state-determined water-quality compliance point for each watershed.

2. **Problem definition**. Surface-water impairment by fecal coliform bacteria is a water-quality issue of national scope and importance. In Virginia, more than 175 stream segments are on the Commonwealth's 1998 303(d) list of impaired waters because of elevated concentrations of fecal coliform bacteria. In Virginia, total maximum daily load (TMDL) assessments will need to be developed over the next 10 years for all impaired water bodies identified on the State's 1998 303(d) list. Establishment of TMDLs in waters contaminated by fecal coliform bacteria is difficult because the potential sources of the bacteria are numerous and the magnitude of their contributions is commonly unknown. Potential sources of fecal coliform bacteria include all warm-blooded animals (humans, pets, domesticated livestock, birds, and wildlife). The lack of information on bacteria sources makes it difficult to develop accurate load allocations, technically defensible TMDLs, and appropriate source-load reduction measures. Information about the major fecal coliform sources that impair surface-water quality would represent an improvement in the development of technically defensible TMDLs.

3. **Statement of objectives**. This study was performed to demonstrate the field application of a BST method and to identify the sources of fecal coliform bacteria in three streams on Virginia's 1998 303(d) list of impaired waters. The three streams sampled during this study were selected because they represent a range of land uses (urban, agricultural, and mixed urban/agricultural) and most of the potential fecal coliform sources that are likely to be encountered throughout the Commonwealth.

4. **Date of study**. 1999-2000.

B. **Analytical approach**

1. **Method description**. The known-source *E. coli* reference collection of Dr. Mansour Samadpour (Institute for Environmental Health, Seattle Washington; more than 50,000 isolates at the time) was used and supplemented by known-source samples in the studied

watersheds. Isolates were characterized by ribotyping using restriction enzymes *Eco*R1 and *Pbu*II.

2. **Target organisms**. *E. coli*

3. **Statistical approach/classification method**. 1:1 matching. The approach used in this study was that ribotypes (strains) of *E. coli* are specific to host species; therefore, any stream-isolated *E. coli* with a ribotype that matched a known-source isolate could be assigned to that host species as the source. Where there was a match to one source, isolates were classified to that source. Where there was a match to more than one source, isolates were classified as transient. Where there was no match in the library, isolates were classified as unknown.

C. **Library considerations**

1. **When collected**. Three sets of isolates were used as the known-source library: 1) 50,000 isolates from the IEH collection, collected over approximately 5-10 years prior to the current study, national coverage; 2) 450 isolates previously collected in Virginia, many by George Simmons, in the approximately 5-10 years prior to the current study; 3) 723 isolates collected in the three watersheds concurrently with water sample collection.

2. **Sources included**

   a. **numbers of samples of each source**. Though the distribution of samples among hosts was not noted in the manuscript, 723 source samples were collected during the study from a humans, pets, domestic animals, and wildlife.

   b. **numbers of isolates from each sample (average)**. One.

   c. **library size.** The overall known-source library comprised more than 50,000 isolates.

3. **Evaluation and validation**

   a. **testing for representativeness (cross-validation, holdouts, blind samples)**. 23 isolates from the known-source library were re-submitted as 66 blinds (some were submitted as duplicates or triplicates). The lab had prior knowledge of which 23 isolates were being used. Blind isolates were re-analyzed and matched in all cases to the correct identity among the 23 isolates used.

   b. **testing for random classification**. None.

   c. **comparisons to independent ancillary data**. Multiple lines of evidence were used to evaluate whether MST results were reasonable in these study streams. The authors began by evaluating populations and distributions of known fecal sources, and land-use patterns in each watershed. They conducted continuum sampling to evaluate longitudinal trends in fecal-indicator concentrations in the main stem, in tributaries, and in effluents discharged to the main stem. They also evaluated seasonal and flow-related

trends in fecal-indicator concentrations. These data were interpreted in terms of transport pathways and animal distributions in the watersheds to indicate expected sources of fecal-indicator bacteria.

Several quality control elements were considered to evaluate the interpretation of MST data in this study, and provided further information about some unexpected results. The unexpectedly high contribution by waterfowl in the urban Accotink Creek watershed was consistent with the results of a prior study in a neighboring urban watershed, Four Mile Run (Simmons et al., 1999). Contributions of bacteria from human sources were independently evaluated by sampling for wastewater organics compounds. In all three streams, detectable concentrations of caffeine and cotinine were present, consistent with MST-indicated contributions of human wastewater to the streams. The interpretation that poultry waste was in Christians Creek was supported by total arsenic data collected by Hancock et al. (2000). The poultry feed amendment Roxarsone contains arsenic, which is generally excreted by the birds. Arsenic-bearing poultry litter is ultimately land-applied on the surrounding agricultural fields. Total arsenic concentrations increased during a storm event, supporting the hypothesis that field-applied poultry litter was flushed into streams.

D. **Sampling considerations**

1. **Number and frequency of samples**. 400-450 water-isolated *E. coli* were evaluated for each of three watersheds. Samples were taken on two schedules – routine monitoring samples (2/3 of samples) were collected approximately every 6 weeks and event-oriented samples, targeted at storm flow, were collected as available (5 events, 1/3 of samples). For routine monitoring, 4-8 samples were collected 5 minute intervals to represent small-scale variability in concentration and sources. For event-oriented samples, 10 samples were collected across the hydrograph to represent small-scale variability in concentration and sources during rain events.

2. **Type of sample (depth-width integrated or a simple grab)**. Depth-width integrated samples using three depth-integrated transits (routine monitoring) and grab samples from the centroid of flow (storm flow samples)

3. **When collected (season, flow conditions)**. Samples were collected for 20 months over all seasons. Of the samples, 61% were taken during low-flow condition, 39% during storm-flow condition. Storm samples were collected across the hydrograph (10 samples).

4. **Number of isolates per sample**. 3-5 per water sample. Multiple samples from the same sites on the same dates were not composited.

E. **Outcomes**

1. **Summary of results and conclusions**. Overall, about 65% of isolates could be assigned to a source in this study. Of the remaining 35%, some had no match in the library (unknown) and others matched to multiple sources (transient). Classification was made to the species level with some exceptions (for example, some bird-origin feces could be classified to

species, but others could only be classified to "avian" or "poultry"). The MST results were a combination of the expected and the unanticipated. Fecal-indicator sources in Accotink Creek, the urban setting, were affected by human and pet feces, as expected, but were also strongly influenced by waterfowl. Blacks Run fecal-indicator bacteria were a mixture of human, pet, and livestock sources, as expected. Fecal-indicator concentrations in Christians Creek had a larger human and pet component than expected (about 25% of isolates), compared with livestock and poultry (about 50%). A further unexpected finding in all three watersheds was that relative contributions from each major source were about the same during both base-flow and storm-flow periods, despite the expectation that different transport pathways would dramatically change relative contributions from different sources. Lastly, the study detected seasonal patterns in the contributions of bacteria from cattle and poultry sources in Blacks Run and Christians Creek; this seasonal pattern was consistent with the land management strategies used in each watershed.

2. **Implementation efforts based on the study**. Volunteer implementation along with cost share implementation in support of the TMDL document. Exclusion fencing of cattle was one of the major implementation efforts.

3. **Follow-up monitoring**. Based on the results of this initial study, DEQ developed and submitted a TMDL to the USEPA in 2002 that included a goal to reduce the human sources of fecal coliform bacteria by 99%. The TMDL for Accotink Creek was approved by USEPA in July 2002. As a follow-up step to the TMDL, USGS initiated another study in cooperation with Fairfax County Stormwater Planning Division (SWPD), City of Fairfax, and DCR to help identify the distribution of fecal coliform and locate the precise sources of human fecal coliform inputs to Accotink Creek. This second study began in mid-to-late 2001 and will continue for 3 years. The field-work portion of the study is anticipated to be completed in late 2004. Staff from SWPD is currently assisting USGS field sampling efforts and laboratory analysis for some parameters.

Case 6. Avalon Bay (California)

Source of information: Boehm, A. B., Fuhrman, J. A., Mrše, R. D. and Grant, S. B. 2003. Tiered approach for identification of a human fecal pollution source at a recreational beach: Case study at Avalon Bay, Catalina Island, California. Environ. Sci. Tech. 37(4), 673-680.

A. General description

1. **Watershed description**. The impacted coastline is a 500-m stretch of sandy beach located in Avalon Bay, on the southeast side of Catalina Island, California (area 200 km^2). Avalon (area 6.9 km^2) is the largest town on the island with 3500 year-round residents. The city's primary source of revenue is tourism; on a typical summer day 17,500 tourists arrive via ferry, cruise ship, or personal vessel, and up to 400 vessels are anchored in the bay. Rainfall in this region occurs primarily from November through March, and consequently, during the summer-time study, there is no rainfall. As is the case for virtually any coastal community, there are many potential sources of fecal contamination in Avalon Bay. Sewer trunk lines run parallel to the beach, approximately 20 m from the shoreline. Nuisance runoff is directed into the sewer system by low-flow diverters; however, some of the runoff enters the ocean untreated through small drains that discharge to the sand, particularly during periods when streets are being washed down by City staff. Secondary treated sewage is released at a rate of approximately 2158 $m^3 d^{-1}$ southeast of the bay through an outfall that terminates 100 m from the coast, at a depth of 65 m. A pier with restrooms, restaurants, and recreational establishments extends from the shoreline near the southeast end of the beach. In addition, pigeons and sea gulls congregate to feed and nest near the shoreline.

2. **Problem definition**. During the summers of 2000 and 2001, water samples from Avalon Beach frequently exceeded the single sample standard for enterococci; thus, signs were posted at the beach warning swimmers not to enter the water. Based on historical data, this was not necessarily a new problem, but was magnified with the new, more stringent state water quality regulations that were instated in the summer of 1999.

3. **Statement of objectives**. City officials were not able to readily identify and remedy the pollution source, and thus the study was commissioned. At the outset of the study, it was not clear to what extent the following potential sources impacted water quality in Avalon Bay: effluent from the sewage treatment plant, nuisance runoff, feces of birds and other wild animals, contaminated subsurface water, and boat sewage collection tanks. The latter was not expected to contribute much to the pollution because the city has an aggressive dye program to reduce illicit discharges into the bay.

B. Analytical approach

1. **Method description**. A three-tiered approach for determining sources of human and nonhuman fecal indicator bacteria (FIB) at a recreation beach that utilizes both standard assays for FIB and novel detection techniques for human-specific *Bacteroides*/*Prevotella* and enterovirus. The first tier documents the spatiotemporal variability of the pollution signal and takes into account the possible influence of sunlight and tides on FIB

concentrations in coastal waters. The second tier consists of source studies. Studies in the first two tiers identify pollution sources and "hot spots" using only standard FIB tests. The third and final tier consists of selectively sampling FIB sources and hot spots for the enteric bacteria *Bacteroides*/*Prevotella* and enterovirus using nucleic acid detection techniques to determine if fecal contamination, indicated by FIB, is of human origin. This study illustrates how measurements made with traditional indicators, in conjunction with more novel indicators, can lead to source identification and mitigation.

2. **Target organisms**. *Bacteroides*/*Prevotella* and enterovirus.

3. **Statistical approach-classification method**. Presence/absence of PCR product. Sensitivity of the *Bacteroides*/*Prevotella* method was estimated at 1 µg/5-50 mL of seawater. Detection limit of the enterovirus method was approximately 1 PFU per 2-20 L of seawater.

C. **Sampling considerations**

1. **Number and frequency of samples**. 33 samples, collected between 9/19/2001 and 10/29/2001.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. Summer, no rainfall events included.

4. **Volume of sample and concentration factor**. For *Bacteroides*/*Prevotella*, bacteria from water samples were collected by filtration of 1-4 L. Most amplifications were from 1 and 10 ng of extracted DNA, equivalent to about 5-50 mL of seawater, chosen to provide a compromise between sensitivity and inhibition of the assay. For enterovirus, 2-20 L of water was filtered.

5. **Evaluation and validation**

   a. **Spiked samples**. All sets of assays included positive controls in which a small amount (1-100 pg) of human fecal DNA extract or cultured poliovirus was added to replicates of the field samples to see if reactions were inhibited by the matrix.

   b. **Blind samples**. Not done.

   c. **Negative controls**. All sets of assays included negative controls (no DNA or poliovirus added).

   d. **Comparisons to independent ancillary data**. Source tracking was performed on samples from locations which were identified using the first two tiers of the procedure.

D. **Outcomes**

1. **Summary of results and conclusions**. FIB in Avalon Bay appear to be from multiple,

primarily land-based, sources including bird droppings, contaminated subsurface water, leaking drains, and runoff from street wash-down activities. Multiple shoreline samples and two subsurface water samples tested positive for human-specific bacteria and enterovirus, suggesting that at least a portion of the FIB contamination is from human sewage.

2. **Implementation efforts based on the study**. Based on the results of the study, the city of Avalon slip-lined their sewer lines that run along the beach.

3. **Follow-up monitoring**. Not mentioned in report.

Case 7. Holmans Creek (Virginia)

Source of information: **Noto, M., K. Hoover, E. Johnson, J. McDonough, E. Stevens, and B. A. Wiggins.** 2000. "Use of Antibiotic Resistance Analysis (ARA) to Identify Nonpoint Sources of Fecal Contamination in the Holmans Creek Watershed". Technical Report prepared for the Lord Fairfax Soil and Water Conservation District.

A. **General description**

1. **Watershed description**. Holmans Creek is located in Shenandoah County, Virginia. It is a 11,988 acre drainage area and contains 12 miles of stream, with 72% of the land in agricultural use, 26% forested, and 2% mixed urban land use. All of the homes use septic systems and wells or cisterns. Holmans Creek feeds the North Fork of the Shenandoah River and flows eventually into the Chesapeake Bay. The possible/suspected sources of fecal contamination in the Holmans Creek watershed are beef and dairy cattle (cattle), chickens and turkeys (poultry), failing septic systems (human), and geese.

2. **Problem definition**. Holmans Creek does not meet the Clean Water Act national goal of "fishable or swimmable" standards. It is on the Priority List of impaired waters due to elevated fecal coliform levels and adversely affected benthic aquatic organisms.

3. **Statement of objectives**. The study was conducted to determine the major sources of fecal pollution in the stream.

4. **Date of study**. July 1999 through January 2001.

B. **Analytical approach**

1. **Method description**. ARA, using 16 antibiotics (51 concentrations total).

2. **Target organisms**. Enterococci.

3. **Statistical approach-classification method**. Linear discriminant analysis. Classification was performed 4-way (cattle vs poultry vs human vs geese).

C. **Library considerations**

1. **When collected**. July 1999 through January 2001, from known sources located within the watershed.

2. **Sources included**

   a. **numbers of samples of each source**. Cattle (3 animals/sample): 26; Poultry litter (multiple animals/sample): 11; Septic tanks (1 household/sample): 42; Geese (3 animals/sample): 7.

b. **numbers of isolates from each sample (average)**. Cattle: 18; Poultry: 23; Septic tanks 19; Geese 14.

3. **Evaluation and validation**

a. **testing for representativeness (cross-validation, holdouts, blind samples)**. The ARCC of the library was 73%. The Minimum Detectable Percentage (MDP) for each source type was determined to be 18% by averaging the percentages of other source types that were misclassified as that type. Further representativeness sampling was not done at the time, but subsequent cross-validation and holdout analysis showed that the library was reasonably representative for human and livestock sources, but was not representative for the wild (goose) samples.

b. **testing for random classification**. Not done.

c. **comparisons to independent ancillary data**. See section G.

D. **Sampling considerations**

1. **Number and frequency of samples**. Nine sites were sampled along Holmans Creek during each sampling event. Stream samples were collected on 7/23/99, 9/29/99, 11/18/99, 2/15/00, 2/19/00 (after a heavy storm), 7/20/00, 9/20/00, and 1/25/01.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. Collected over a year and a half. Samples were collected during periods of high and low flow. One set of samples were taken immediately after a heavy storm.

4. **Number of isolates per sample**. The goal for each sample was 46 isolates, but some samples had fewer. The average number of isolates per sample was 41.

E. **Outcomes**

1. **Summary of results and conclusions**. Human sources were dominant in five of eight sampling events, and at four of nine locations. In 53 of the 64 samples, the proportion of human was above the MDP, and human was the dominant source in 29 of the 64 samples. Cattle was the dominant source on three of eight sampling days, and at five of nine locations. The proportion of cattle was above the MDP in 52 of 64 samples, and cattle was the dominant source in 26 of them. Poultry and geese fecal contributions were low throughout the sampling period. The conclusions were that humans and cattle are the dominant sources of fecal pollution in the watershed.

2. **Implementation efforts based on the study**. Based on the results of this study, a septic system maintenance project was undertaken in the watershed. This project identified

numerous straight pipes discharging sewage directly into the stream, and found that approximately 25% of the septic systems in the watershed were failing. Through the use of cost-share funds, many of these systems have been repaired or replaced. The Implementation Plan for this watershed calls for removal of all straight pipes, all failing septic systems must be identified and corrected, and all livestock must be excluded from the stream.

3. **Follow-up monitoring**. Stream monitoring in this watershed has been continuing. Samples from the same sites have been collected quarterly from 2002 – 2004. The results from the newer sample indicate that the percentage of human pollution has decreased from the 2001 levels. Subsequent classification of the samples was performed using a larger, regional library that was determined to be representative for all sources (using cross-validation, holdouts, and random classification).

Case 8. Homosassa Springs (Florida)

Source of information: Griffin, D. W., R. Stokes, J. B. Rose, and J. H. Paul III. 2000. Bacterial indicator occurrence and the use of F^+ specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida. Microbial Ecology 39:56-64.

A. **General description**

1. **Watershed description**. The Homosassa Springs State Wildlife Park (HSSWP) is a 180-acre complex that surrounds Homosassa River's main spring (Homosassa Main). HSSWP is the home of numerous animals including birds, deer, bobcats, alligators, a hippopotamus, a permanent group of manatees, and fish. The Homosassa Main consists of three separate vents, each with its own distinct chemical signature, which have a combined average discharge of approximately 2,944 liters s^{-1}. To the southeast of the park is the Southeast Fork of the Homosassa River. The Southeast Fork is fed by a closely associated group of springs, which have a combined average discharge of approximately 1,953 liters s^{-1}. The waters of these two sources and the immediate region in the river receiving these waters appear clear.

2. **Problem definition**. Water quality issues in the Homosassa River system have received the attention of local citizen groups and the media. Of particular concern were the elevated levels of coliforms and fecal coliforms found in Homosassa River downstream of HSSWP, which have been attributed to Park animals. The Florida Department of Health (DOH), which has been monitoring water quality at a site just downstream of the park (an area which was to have been designated as a swimming site), found that fecal indicator concentrations consistently exceed recreational use standards (>200 fecal coliform colony forming units (CFU) 100 ml^{-1}).

3. **Statement of objectives**. This study was designed to assess microbial water quality and to differentiate fecal sources contributing to the contamination previously observed in HSSWP and its adjacent waters.

4. **Date of study**. November of 1997 and November of 1998.

B. **Analytical approach**

1. **Method description**. F^+ specific RNA coliphage genotyping. Types II and III coliphage are associated with human sources of fecal contamination and Types I and IV are associated with non-human sources.

2. **Target organisms**. F^+ specific RNA coliphage.

3. **Statistical approach/classification method**. Direct match of specific oligonucleotide probes.

C. **Sampling considerations**

1. **Number and frequency of samples**. Seven sites in November 1998 and nine sites in November 1998.

2. **Type of sample (depth-width integrated or a simple grab)**. Grab samples.

3. **When collected (season, flow conditions)**. November of two consecutive years.

4. **Volume of sample and concentration factor** 20-L samples concentrated by vortex flow filtration (>70% coliphage recovery) to 40-60 ml, of which 1 ml aliquots were used for coliphage analysis.

5. **Evaluation and validation**

   a. **Spiked samples**. None reported. No reference feces from local animals were positive for F+ RNA coliphage. A reference human-waste stream was positive for human-associated types II and III coliphage.

   b. **Blind samples**. None reported.

   c. **Negative controls**. None reported.

   d. **comparisons to independent ancillary data**. Several factions have attributed the fecal indicator prevalence to HSSWP animals. The watershed also contains many residences with older septic tanks.

D. **Outcomes**

1. **Summary of results and conclusions**. F^+ specific RNA coliphage analysis indicated that fecal contamination at all sites that had F+ RNA coliphage was from animal sources (mammals and birds). These results suggest that animal (either indigenous or residents of HSSWP) and not human sources influenced microbial water quality in the area of Homosassa River covered by this study.

2. **Implementation efforts based on the study**. None reported.

3. **Follow-up monitoring**. None reported.

Literature Cited

Alderisio, K. A., D. A. Wait, and M. D. Sobsey. 1996. Detection and characterization of male-specific RNA coliphages in a New York City reservoir. p. 133-42, In *Watershed Restoration Management*, J. J. McDonnell, D. L. Leopold, J. B. Stribling, and L. R. Neville (ed.), New York City Water Supply Studies. American Water Resources Association, Herndon, VA.

Altwegg, M., F. W. Hickman-Brenner, and J. J. Farmer III. 1989. Ribosomal RNA gene restriction patterns provide increased sensitivity for typing *Salmonella typhi* strains. *J. Infect. Dis*. 160:145-149.

American Public Health Association. 1995. In *Standard Methods for the Examination of Water and Wastewater*. Washington DC: American Public Health Association, Inc.

Anderson, M. A. 2003. Frequency distributions of *Escherichia coli* subtypes in various fecal sources over time and geographical space: Application to bacterial source tracking methods, pp. 117. Tampa, FL: University of South Florida.

Anderson, M. A., J. E. Whitlock, and V. J. Harwood. 2003. Frequency distributions of *Escherichia coli* subtypes in various fecal sources: Application to bacterial source tracking methods. In *American Society for Microbiology General Meeting*. Washington, DC

Anderson, S. A., S. J. Turner, and G. D. Lewis. 1997. Enterococci in the New Zealand environment: implications for water quality monitoring. *Water Science and Technology* 35:325-331.

Arnold, C., L. Metherell, J. P. Clewley, and J. Stanley. 1999. Predictive modelling of fluorescent AFLP: a new approach to the molecular epidemiology of *E. coli*. *Res. Microbiol*. 150:33-44.

Aslam, M., F. Nattress, G. Greer, C. Yost, C. Gill, and L. McMullen. 2003. Origin of contamination and genetic diversity of *Escherichia coli* in beef cattle. *Appl. Environ. Microbiol*. 69:2794-2799.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J.G. Seidman, J. A. Smith, and K. Struhl. 2004. Current Protocols in Molecular Biology. John Wiley and Sons, Hoboken, NJ.

Avelar, K. E., S. R. Moraes, L. J. Pinto, W. d. G. Silva e Souza, R. M. Domingues, and M. C. Ferreira. 1998. Influence of stress conditions on Bacteroides fragilis survival and protein profiles. *Zentralbl Bakteriol* 287:399-409.

Backhed F, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, J. I. Gordon. 2005. Host-bacterial mutualism in the human intestine. *Science* 307:1915-20.

Baker, G. C., J. J. Smith, and D. A. Cowan. 2003. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Meth*. 55:541-555.

Bartosch, S., A. Fite, G. T. Macfarlane, and M. E. T. McMurdo. 2004. Characterization of bacterial

communities in feces from healthy elderly volunteers and hospitalized elderly patients by using real-time PCR and effects of antibiotic treatment on the fecal microbiota. *Appl. Environ. Microbiol*. 70:3575-3581.

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

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

Bernhard, A. E., T. Goyard, M. T. Simonich, and K. G. Field. 2003. Application of a rapid method for identifying fecal pollution sources in a multi-use estuary. *Water Res* 37:909-13.

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.

Bingen, E., E. Denamur, N. Brahimi, and J. Elion. 1996. Genotyping may provide rapid identification of *Escherichia coli* K1 organisms that cause neonatal meningitis. *Clin. Infect. Dis*. 22:152-156.

Bingen, E. H., E. Denamur, B. Picard, P. Goullet, N.Y. Lambert-Zechovsky, N. Brahimi, J.-C. Mercier, F. Beaufils and J. Elion. 1992. Molecular epidemiology unravels the complexity of neonatal *Escherichia coli* acquisition in twins. *J. Clin. Microbiol.* 30:1896–1898.

Blears, M. J., S. A. De Grandis, H. Lee, and J. T. Trevors. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *J. Indust. Microbiol*. Biotech. 21:99-114.

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.

Bonjoch, X., E. Ballesté and A. R. Blanch. 2004. Multiplex PCR with 16S rRNA gene-targeted primers of *Bifodobacterium* 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. 129: 547-552.

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**:**2419-25.

Brisse, S., C. M. Verduin, D. Milatovic, A. Fluit, J. Verhoef, S. Laevens, P. Vandamme, B. Tummler, H. A. Verbrugh, and A. van Belkum. 2000. Distinguishing species of the *Burkholderia cepacia* complex and *Burkholderia gladioli* by automated ribotyping. J. *Clin. Microbiol*. 38:1876-1884.

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

Bryant, M. P. 1959. Bacterial species of the rumen. *Bacteriol. Rev*. 23:125-153.

Burton, G. A., Jr., Gunnison, D. and Lanza, G. R. 1987 Survival of pathogenic bacteria in various freshwater sediments. *Appl. Environ. Microbiol.* 53:633-8

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

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

Caetano-Anolles, G., B. J. Bassam, and P. M. Gresshoff. 1992. Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides. *Mol. Gen. Genet*. 235:157-165.

Carrillo, 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.

Carson, A. C., 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. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal Escherichia coli from humans and animals. *Appl. Environ. Microbiol.* 69:1836-1839.

Caugant, D. A., Levin, B. R. and Selander, R. K. 1984. Distribution of multilocus genotypes of Escherichia coli within and between host families. *J. Hyg. (Lond)* 92:377-84.

Chern, E. C., Y. L. Tsai, and B. H. Olson. 2004. Occurrence of genes associated with enterotoxigenic and enterohemorrhagic *Escherichia coli* in agricultural waste lagoons. *Appl. Environ. Microbiol.* 70: 356-362.

Chivukula, V. and V. J. Harwood. 2004. Impact of fecal pollution on the microbial diversity in natural waters. In *American Society for Microbiology General Meeting*. New Orleans, LA.

Cho, J.C., and S.J. Kim 2000. Increase in bacterial community diversity in subsurface aquifers receiving livestock wastewater input. *Appl. Environ. Microbiol* 66: 956-965.

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

Chu, G., D. Vollrath and R. W. Davis. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* 234:1582-1585.

Clerc, A., C. Manceau, and X. Nesme. 1998. Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae*. *Appl. Environ. Microbiol*. 64:1180-1187.

Clesceri, L. S., A. E. Greenberg, and A. D. Eaton. 1998. Standard methods for the examination of water and waste water. APHA AWWA WEF, Washington, DC.

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

Cox, T. F., and M. A. A. Cox. 2001. *Multidimensional Scaling*, 2nd ed., Chapman and Hall.

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.* 96:922-30.

Daley, K., and S.P. Shirazi-Beechey. 2003. Design and evaluation of group-specific oligonucleotide probes for quantitative analysis of intestinal ecosystems: their application to assessment of equine colonic microflora. *FEMS Microbiol. Ecol*. 44:243-252.

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

Dasarathy, B. V. 1991. Nearest Neighbor Pattern Classification Techniques. IEEE Computer Society Press. Los Alamitos, CA

de Bruijn, F. J. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl. Environ. Microbiol*. 58:2180-2187.

DeLong, E.F., G.S. Wickhan, and N.R. Pace. 1989. Phylogenetic strains: Ribosomal RNA-based probes for the identification of single cells. *Science*. 243:13600-1363.

Demezas, D. 1998. Fingerprinting bacterial genomes using restriction fragment length polymorphisms, pp. 383-398. In *Bacterial Genomes: Structure and Analysis*., F. J. de Bruijn, J.R. Lupski, and G. Weinstock (eds.). Chapman and Hall, New York.

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

Desmarais, T. R., Solo-Gabriele, H. M. and Palmer, C. J. 2002 Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl. Environ. Microbiol.* 68:1165-72.

Dice, L. R. 1945. Measures of the amount of ecological association between species Ecology 26:297-302

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

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

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.

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.

Dontchev, M., J. E. Whitlock, and V. J. Harwood. 2003 Ribotyping of *Escherichia coli* and *Enterococcus* spp. to determine the source of fecal pollution in natural waters. American Society for Microbiology General Meeting. Washington, DC.

Dufour, A. P. 1984. EPA health effects criteria for fresh recreational waters. Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina.

Eckburg, P. B. E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill , K. E. Nelson, D. A. Relman. 2005. Diversity of the Human Intestinal Microbial Flora. Science. 2005 Apr 14; [Epub ahead of print]

Edberg, S. C., E. W. Rice, R. J. Karlin, and M. J. Allen. 2000. Escherichia coli: the best biological

drinking water indicator for public health protection. *Symp. Ser. Soc. Appl. Microbiol.* 106S-116S.

Enas, G. C., and S. C. Choi. 1986. Choice of the smoothing parameter and efficiency of k-nearest neighbor classification. *Comput. Math. Applic.* 12A:235-244.

Farag, A. M., J. N. Goldstein, D. F. Woodward, and M. Samadpour. 2001. Water quality in three creeks in the backcountry of Grand Teton National Park, USA. *J. Fresh Water Ecol*. 16:135–143.

Field, K.G., A.E. Bernhard, and T. J. Brodeur. 2003. Molecular approaches to microbiological monitoring: Fecal source detection. *Environ. Moni. Assess*. 81: 313-326.

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

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.

Filip, Z., D. Kaddu Mulindwa, and G. Milde. 1987. Survival and adhesion of some pathogenic and facultative pathogenic microorganisms in groundwater. *Water Sci. Technol.* 19:1189-1190.

Fong TT, Griffin DW, Lipp EK. 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.

Fode-Vaughan, K. A., C. F. Wimpee, C. C. Remsen, and M. L. Perille Collins. 2001. Detection of bacteria in environmental samples by direct PCR without DNA extraction. *BioTechniques* 31:598-607.

Frahm, E., and U. Obst. 2003. Application of fluorogenic probe technique (TaqMan PCR) to the detection of *Enterococcus* spp. and *Escherichia coli* in water samples. *J. Microbiol. Meth*. 52: 123-131.

Franks, A. H., H. J. M. Harmsen, G. C. Raangs, G. J. Jansen, F. Schut, and G. W. Welling. 1998. Variations of bacterial populations in human feces measured by fluorescent *in situ* hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol*. 64:3336-3345.

Fujioka, R. S., and M. N. Byappanahalli. 2003. Proceedings and Report: Tropical Water Quality Indicator Workshop, pp. 90. Honolulu, HI, Water Resource Center, University of Hawaii at Manoa.

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

Funderburg, S. W., and C. A. Sorber. 1985 Coliphages as indicators of enteric viruses in activated

sludge. *Water Res*. 19:547-555.

Furuse, K. 1987. Distribution of coliphages in the environment: general considerations. In *Phage Ecology*, S. M. Goyal, C. P. Gerba, and G. Bitton (ed.), p. 87-124, John Wiley and Sons, New York

Galland, J. C., D. R. Hyatt, S. S. Crupper, and D. W. Acheson. 2001. Prevalence, antibiotic susceptibility, and diversity of *Escherichia coli* O157:H7 isolates from a longitudinal study of beef cattle feedlots. *Appl. Environ. Microbiol*. 67:1619-1627.

Gantzer, C., A. Maul, J. M. Audic, and L. Schwartzbrod. 1998. Detection of infectious enteroviruses, enterovirus genomes, somatic coliphages, and *Bacteroides fragilis* phages in treated wastewater. *Appl. Environ. Microbiol*. 64:4307-4312.

Geldreich, E. E., and B. A. Kenner. 1969. Concepts of fecal streptococci in stream pollution. *J. Water Pollut. Control Fed.* 41:R336-352.

Geldreich E. E., L. C. Best, B. A. Kenner, and D. J. Van Donsel. 1968 . The bacteriological aspects of stormwater pollution. *J. Water Pollut. Control Fed*. 40:1861-72.

Geldreich, E. E., and N. A. Clarke. 1966. Bacterial pollution indicators in the intestinal tract of freshwater fish. *Appl. Microbiol*. 14:429-37.

Geldreich, E. E., Kenner B. A., and P. W. Kabler. 1964. Occurrence of coliforms, fecal coliforms, and fecal streptococci on vegetation and insects. *Appl. Microbiol*. 12:63-9.

Gerba, C. P., and J. S. McLeod. 1976 Effect of sediments on the survival of Escherichia coli in marine waters. *Appl. Environ. Microbiol.* 32:114-20.

Ginzinger, D.G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematol.* 30:503-512.

Giovannoni, S. J., E. F. DeLong, G. J. Olsen, and N. R. Pace. 1998. Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J. Bacteriol.* 170:720-726.

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

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

Gordon, D. M., S. Bauer, and J. R. Johnson. 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology* 148:1513-1522.

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

Grabow, W. O. K., O. W. Prozesky, and J. S. Burger. 1975. Behavior in a river and dam of coliform bacteria with transferable or non-transferable drug resistance. *Water Res* 9:777-782.

Grant, S. B., B. F. Sanders, A. B. Boehm, J. A. Redman, J. H. Kim, R. D. Mrse, A. K. Chu, M. Gouldin, C. D. McGee, N. A. Gardiner, B. H. Jones, J. Svejkovsky, G. V. Leipzig, and A. Brown. 2001. Generation of enterococci bacteria in a coastal saltwater marsh and its impact on surf zone water quality. *Environ Sci Technol* 35:2407-2416.

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 a rural Virginia watershed. *J. Environ. Qual*. 31: 1300-1308.

Grif, K., H. Karch, C. Schneider, F. D. Daschner, L. Beutin, T. Cheasty, H. Smith, B. Rowe, M. P. Dierich, and F. Allerberger. 1998. Comparative study of five different techniques for epidemiological typing of *Escherichia coli* O157. *Diag. Microbiol. Infect. Dis*. 32:165-176.

Griffin, D. W., C. J. Gibson, E. K. Lipp, K. Riley, J. H. Paul, and J. B. Rose. 1999. Detection of viral pathogens by reverse transcriptase PCR and of microbial indicators by standard methods in the canals of the Florida Keys. *Appl. Environ. Microbiol*. 65:4118-4125.

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. Water Health*. 1:141-51.

Grimont, F., and P. A. D. Grimont. 1986. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Ann. Inst. Pasteur Microbiol*. 137B: 165-175.

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

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

Gustaferro, C. A., and D.H. Persing. 1992. Chemiluminescent universal probe for bacterial ribotyping. *J. Clin. Microbiol*. 30:1039-1041.

Guttman, L. 1950. The basis for scalogram analysis. In *Measurement and Prediction.* S.A.Stauffer et al. (eds.) Vol.**4**, Princeton University Press, Princeton, New Jersey.

Haack, S. K., L. R. Fogarty, and C. Wright. 2003. *Escherichia coli* enterococci at beaches in the Grand Traverse Bay, Lake Michigan: sources, characteristics, and environmental pathways. *Environ. Sci. Technol.* 37:3275-3282.

Hagedorn, C., S. L. Robinson, J. R. Filtz, S. M. Grubbs, T. A. Angier and R. B. Reneau. 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. 2003. Carbon source utilization profiles as a method to identify sources of faecal pollution in water. *J. Appl. Microbiol*. 94:792-799

Hahm, B. K., Y. Maldonado, E. Schreiber, A. K. Bhunia, and C. H. Nakatsu. 2003. Subtyping of foodborne and environmental isolates of *Escherichia coli* by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. *J. Microbiol. Meth*. 53:387-399

Hahm, B. K., A. K. Bhunia, and C. H. Nakatsu. 2003. Application of AFLP for discriminating *Escherichia coli* isolated from livestock, wildlife and humans. American Society for Microbiology Annual Meeting Abstracts. Washington D.C.

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. Wang. 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. *Epidemiology* 10: 355-63.

Hand, D. J., 1997. Construction and assessment of classification rules. John Wiley and Sons. Chichester, UK

Hardina, C. M., and R. S. Fujioka. 2001. Soil: the environmental source of *Escherichia coli* and Enterococci in Hawaii's streams. *Environ Tox Water Qual* 6, 185-195.

Harmsen, H. J. M., G. C. Raangs, T. He, J. E. Degener, and G. W. Welling. 2002. Extensive set of 16S rRNA-based probes for detection of bacteria in human feces. *Appl. Environ. Microbiol*. 68:2982-2990.

Harmsen, H. J., A. C. Wildeboer-Veloo, G. C. Raangs, A. A. Wagendorp, N. Klijn. 2000. *J. Pediatr. Gastroenterol. Nutr*. 30:61-70.

Hartel, P. G., E. A. Frick, A. L. Funk, J. L. Hill, J. D. Summer, and M. B. Gregory. 2004. Sharing of ribotype patterns of *Escherichia Coli* Isolates during baseflow and stormflow conditions. *USGS Scientific Investigations Report* 2004.

Hartel, P. G., W. I. Segars, N. J. Stern, J. Steiner, and A. Buchan. 1999. Ribotyping to determine the host origin of *Escherichia coli* isolates in different water samples. p. 377–382. In *Wildland hydrology*, D.S. Olsen and J.P. Potyondy (ed.). Am. Water Resour. Assoc., Herndon, VA.

Hartel, P. G., J. D. Summner, J. L. Hill, J. 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.

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

Hartigan, J. 1975. *Clustering Algorithms*. Wiley, New York, NY.

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

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*. 66:3698-704.

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. Water Health* 1:153-166.

Hastie, T., Tibshirani, R. and Friedman, J. 2002. *The Elements of Statistical Learning*. Springer Series in Satistiscs, Springer-Verlag, New York, NY, USA.

Head, I. M., J. R.Saunders, and R.W. Pickup. 1998. Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. *Microbial Ecol*. 35:1-21.

Hilton, A.C., J.G. Banks, and C.W. Penn. 1997 Optimization of RAPD for fingerprinting *Salmonella*. Lett. Appl. Microbiol. 24:243-248.

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

Hopkins, K. L., and A. C. Hilton. 2000. Methods available for the sub-typing of *Escherichia coli* O157. *World J. Microbiol. Biotechnol*. 16:741-748.

Hood, K. L., J. E. Whitlock, and V. J. Harwood. 2003. Factors that influence the persistence of bacterial indicator organisms in fresh and saline subtropical waters. American Society for Microbiology General Meeting, Washington, DC.

Hood, K. L., J. E. Whitlock, M. R. McLaughlin, J. B. Rose, and V. J. Harwood. 2002. Survival and fingerprint stability of indicator organisms in subtropical waters. American Society for Microbiology General Meeting, Salt Lake City, UT.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J. Educational Psychol*. 24:417-441, 498-520.

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.

Hughes, M. S., L. A. Beck, R. A. Skuce. 1994. Identification and Elimination of DNA Sequences in Taq DNA Polymerase. *J. Clin. Microbiol.* 32:2007-2008.

Hungate, R. E. 1966. The rumen and its microbes. Academic Press, New York, NY.

Hyer, K. E., and D. L. Moyer. 2003. Patterns and sources of fecal coliform bacteria in three streams in Virginia, 1999-2000., *USGS Water Resources Investigations Report* 03-4115.

Ito,Y., Y. Iinuma, H. Baba, Y. Sugino, Y. Hasegawa, K. Shimokata, S. Ichiyama, T. Hasegawa, and M. Ohta. 2003. Evaluation of automated ribotyping system for characterization and identification of verocytotoxin-producing *Escherichia coli* isolated in Japan. *Jpn. J. Infect. Dis*. 56:200-204.

Jaccard, P. 1901. Distribution de la flore alpine dans le Bassin des Dranes et dans quelques regions voisines. *Bull Soc. Vaud. Sci. Nat*., 37:241-272.

Jagals, P., W. O. K. Grabow, and J. C. De Villiers. 1995. Evaluation of indicators for assessment of human and animal faecal pollution of surface run-off. *Water Sci Technol*. 31:235-241.

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 coliphages in urban runoff-impacted coastal waters of Southern California. *Appl. Environ. Microbiol*. 67**:**179-84.

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**:**6311-5.

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 influence accuracy of determining sources of fecal pollution among natural populations of *Escherichia coli* from different animals. *Appl. Environ. Microbiol*. 70: 4478-4485.

Johnson, R. A. and Wichern, D. W. 2002. Applied Multivariate Statistical Analysis, Prentice-Hall.

Jolliffe, I. T. 2002. Principal Component Analysis, Springer-Verlag, New York.

Kaufmann, P., A. Pfefferkorn, M. Teuber, and L. Meile. 1997. Identification and quantification of *Bifidobacterium* species isolated from food with genus-specific 16S rRNA-target probes by colony hybridization and PCR. *Appl. Environ. Microbiol*. 63:1268-1273.

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

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.

Klein, D. 2002. Quantification using real-time PCR technology: applications and limitations. *Trends Mol. Med*. 8:257-260.

Koellner, T., Hersperger, A. M. and Wohlgemuth, T. 2004. Rarefaction method for assessing plant species diversity on a regional scale. *Ecography* 27:532-544.

Konopka, A., L Oliver, and R. F. Turco. 1998. The use of carbon substrate utilization patterns in environmental and ecological microbiology. *Micro. Ecol*. 35:103-115.

Kreader, C. A. 1998. Persistence of PCR-detectable Bacteroides distasonis from human feces in river water. *Appl. Environ. Microbiol.* 64, 4103-5.

Kreader, C. A. 1995. Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. *Appl. Environ*. *Microbiol*. 61:1171-1179.

Kruse, H., and H. Sørum. 1994. Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Appl. Environ. Microbiol*. 60: 4015-4021.

Kulczynski, S. 1928. Zespoly roslin w Pieninach *Bull Int Acad Po Sci Lettres Ser B Suppl* 2:57-203

Kuntz, R.L., P.G. Hartel, D.G. Godfrey, J.L. McDonald, K.W. Gates, and W.I. Segars (2003). Targeted sampling protocol as prelude to bacterial source tracking with Enterococcus faecalis. J Environ Qual. 32**:**2311-8.

LaLiberte, P., and D. J. Grimes, 1982. Survival of *Escherichia coli* in lake bottom sediment. *Appl. Environ. Microbiol.* 43, 623-628.

Leclerc, H., S. Edberg, V. Pierzo, and J. M. Delattre. 2000. Bacteriophages as indicators of enteric viruses and public health risk in groundwaters. *J Appl Microbiol* 88:5-21.

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., J.Z. Amenuvor, T.K. Jensen, R.H. Lindecrona, M. Boye, and K. Møller. 2002. Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisisted. *Appl. Environ. Microbiol*. 68:673-690.

Leung, K. T., R. Mackereth, Y.-C. Tien, E. Topp. 2004. A comparison of AFLP and ERIC-PCR

analyses for discriminating *Escherichia coli* from cattle, pig and human sources. *FEMS Microbio. Ecol.* 47:111-119.

Ley, V., J. Higgins, and R. Fayer 2002. Bovine enteroviruses as indicators of fecal contamination. *Appl Environ Microbiol*. 68**:**3455-61.

Lin, J.-J., and J. Kuo. 1995. AFLP: a novel PCR-based assay for plant and bacterial DNA fingerprinting. *Focus* 17: 66-70.

Lipman, J. A., A. de Nijs, T. J. G. M. Lam, and W. Gaastra. 1995. Identification of *Escherichia coli* strain from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. *Vet. Microbiol*. 43:13-19.

LiPuma, J. J., T. L. Stull, S. E. Dasen, K. A. Pidcock, D. Kaye and O. M. Korzeniowski. 1989. DNA polymorphisms among *Escherichia coli* isolated from bacteriuric women. *J. Infect. Dis.* 159:526–532.

Liu, C. X., Y. L. Song, M. McTeague, A.W. Vu, H. Wexler, and S. M. Finegold. 2003. Rapid identification of the species of the *Bacteroides fragilis* group by multiplex PCR assays using group- and species-specific primers. *FEMS Microbiol. Lett*. 222:9-16.

Liu, W. T., T. L. Marsh, H. Cheng, and L. J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol*. 63:4516-4522.

Livak, K. J., S. J. A. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Meth. Appl.* 4:357-362.

Lupski, J. R., and G. M. Weinstock. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol*. 174:4525-4529.

Machado, J., F. Grimont and P. A. D. Grimont. 1998. Computer identification of *Escherichia coli* rRNA gene restriction patterns. *Res. Microbiol*. 149:119–135.

Madico, G., N. S. Akopyants, and D. E. Berg. 1995. Arbitrarily primed PCR DNA fingerprinting of *Escherichia coli* O157:H7 strains by using templates from boiled cultures. *J. Clin. Microbiol*. 33:1534-1536.

Malinen, E., A. Kassinen, T. Rinttilä, and A. Palva. 2003. Comparison of real-time PCR with SYBR Green I or 5'-nuclease assays and dot-blot hybridization with rDNA-targeted oligonucleotide probes in quantification of selected faecal bacteria. *Microbiology* 149: 269-277.

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

Martellini, A., P. Payment, and R. Villemur. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water Res*. 39:541-548.

Martin, I. E., S.D. Tyler, K. D. Tyler, R. Khakhria, and W. M. Johnson. 1996. Evaluation of ribotyping as epidemiologic tool for typing *Escherichia coli* serogroup O157 isolates. *J. Clin. Microbiol*. 34:720-723.

Martin, J. D. and J. O. Mundt. 1972 Enterococci in insects. *Appl Microbiol* 24:575-580.

Mathieu-Daudé, F., K. Evans, F. Kullmann, R. Honeycutt, T. Vogt, J. Welsch, and M. McClelland, 1998. Applications of DNA and RNA Fingerprinting by the arbitrary primed polymerase chain reaction, pp. 414-436. In Bacterial Genomes, F. de Bruijn, J. R. Lupski and G. M. Weinstock (eds.), Chapman and Hall, New York, NY.

Matsuki, T., K. Watanabe, J. Fujimoto, Y. Miyamtoto, T. Takada, K. Matsumoto, H. Oyaizu, and R. Tanaka. 2002. Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl. Environ. Microbiol*. 68:5445-5451.

Matsuki, T., K. Watanabe, J. Fujimoto, Y. Kado, T. Takada, K. Matsumoto, and R. Tanaka. 2004. Quantitative PCR with 16S rRNA-gene-targeted species-specific primers for analysis of human intestinal *Bifidobacteria*. Appl. Environ. Microbiol. 70:167-173.

McLachlan, G. J. 1992. Discriminant Analysis and Statistical Pattern Recognition. John Wiley and Sons, New York, NY.

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

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

Milkman, R. 1973. Electrophoretic variation in Escherichia coli from natural sources. *Science* 182, 1024-6.

Muyzer, G., S. Hottentrager, A. Teske, and C. Wawer. 1996. Denaturing gradient gel electrophoresis of PCR-amplified 16S rDNA-A new molecular approach to analyse the genetic diversity of mixed microbial communities, p. 3.4.4: 1-23. In *Molecular microbial ecology manual*, A. Akkermans, van Elsas, J.D. and F. J. de Bruijn (ed.),. Kluwer Academic Publishers, Nowell, MA

Muyzer, G., and K. Smalla. 1998. Application of denaturing gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* 73:127-141.

Myoda, S. P., C. A. Carson, J. J. Fuhrmann, B.-K. Hahm, P. G. Hartel, R. L. Kuntz, C. H. Nakatsu, M. J. Sadowsky, M. Samadpour and H. Yampara-Iquise. 2003. Comparing genotypic bacterial source tracking methods that require a host origin database. *J. Water Health* 1:167-180.

Nakatsu, C. H., and T. L. Marsh. 2005. Analysis of microbial communities with denaturing gradient gel electrophoresis and terminal restriction fragment length polymorphism. In *Methods for general and molecular bacteriology*. C. A. Reddy, T. M. Schmidt (eds.) ASM Press, Washington D. C. (in press)

National Committee for Clinical Laboratory Standards. 1999. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically. Villanova, PA: NCCLS.

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.

Noble, R. T., and J. A. Fuhrman 2001. Enteroviruses detected by reverse transcriptase polymerase chain reaction from the coastal waters of Santa Monica Bay, California: low correlation with bacterial indicator levels. *Hydrobiology* 460**:**175-184.

Noble, R.T., S.A. 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**:**195-204.

Ochiai, A. 1957. Zoogeographical studies on the soleoid fishes found in Japan and its neighboring regions II, *Bull. Jap. Soc. Sci. Fish*. 22:526-530.

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

Ogimoto, K and S Imai. 1981. Atlas of rumen microbiology (Japan Scientific Society Press, Tokyo, Japan).

Olsen, G. J., D. J. Lane, S. J. Giovannoni, and N. R. Pace. 1986. Microbial ecology and evolution: A ribosomal RNA approach. *Ann. Rev. Microbiol*. 40:337-365.

Olive, D. M., and P. Bean. 1999. Principles and applications of methods for DNA-based typing of microbial organisms. *J. Clin. Microbiol*. 37:1661-1669.

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 (eds.), pp. 255-259. Royal Society of Chemistry, Cambridge, UK.

Pacheco, A. B., B. E. Guth, D. F. de-Almeida, and L. C. S. Ferreira. 1996. Characterization of

enterotoxigenic *Escherichia coli* by random amplification of polymorphic DNA. *Res. Microbiol*. 147:175–182.

Pacheco, A. B., B. E Guth, K. C. Soares, L. Nishimura, D. F. de Almeida, and L. C. S. Ferreira. 1997. Random amplification of polymorphic DNA reveals serotype-specific clonal clusters among enterotoxigenic *Escherichia coli* strains isolated from humans. *J. Clin. Microbiol*. 35:1521-1525.

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

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

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

Payment, P, and P. R. Hunter. 2001. Endemic and epidemic infectious intestinal disease and its relationship to drinking water. *In* L. Fewtrell and J. Bartram (ed.), Water Quality: Guidelines, Standards and Health. Assessment of risk and risk management for water-related infectious disease. IWA Publishing, on behalf of the World Health Organization, London.

Penner, G. A., A. Bush, R Wise, W. Kim, L. Domier, K. Kasha, A. Laroche, G. Scoles, S. J. Molnar, and G. Fedak. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Meth. Applicat*. 2:341-345.

Picard, B., N. Picard-Pasquier, R. Krishnamoorthy and P. Goullet. 1991. Characterization of highly virulent *Escherichia coli* strains by ribosomal DNA restriction fragment length polymorphism. *FEMS Microbiol. Lett*. 82:183–188.

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**:**3376-82.

Pryde, S. E., A. J. Richardson, C. S. Stewart, and H. J. Flint. 1999. Molecular analysis of the microbial diversity present in the colonic wall, colonic lumen and cecal lumen of a pig. *Appl. Environ. Microbiol.* 65:5372-5377.

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

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

John Wiley and Sons, New York, NY.

Radu, S., O. W. Ling, G. Rusul, M. I. Karim, and M. Nishibuchi. 2001. Detection of *Escherichia coli* O157:H7 by multiplex PCR and their characterization by plasmid profiling, antimicrobial resistance, RAPD and PFGE analyses. *J. Microbiol. Meth*. 46:131-139.

Ramšak, A., M. Peterka, K. Tajima, J.C. Martin, J. Wood, M.E.A. Johnston, R.I. Aminov, H.J. Flint, and G. Avguštin. 2000. Unraveling the genetic diversity of ruminal bacteria belonging to the CFB phylum. *FEMS Microbiol. Ecol*. 33:69-79.

Ranjard, L., F. Poly, and S. Nazaret. 2000. Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res.Microbiol*. 151: 167-177.

Rand, K. H., and H. Houck. 1990. Taq polymerase contains bacterial DNA of unknown origin. *Mol Cell Probes* 4:445-450.

Regnault, B., F. Grimont, and P. A. Grimont. 1997. Universal ribotyping method using a chemically labelled oligonucleotide probe mixture. *Res. Microbiol*. 148:649-659.

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

Restrepo, S., M. Duque, J. Tohme, and V. Verdier. 1999. AFLP fingerprinting: an efficient technique for detecting genetic variation of *Xanthomonas axonopodis pv. manihotis*. *Microbiology* 145:107-114.

Rigottier-Gois, L., A. -G. Le Bourhis, G. Gramet, V. Rochet, and J. Doré. 2003. Fluorescent hybridisation combined with flow cytometry and hybridisation of total RNA to analyse the composition of microbial communities in human faeces using 16S rRNA probes. *FEMS Microb. Ecol*. 43:237-245.

Ritter K. J., E . Carruthers, C. A. Carson, R. D. Ellender, V. J. Harwood, K. Kingsley, C. Nakatsu, M. Sadowsky, B. Shear, B. West, J. E. Whitlock, B. A. Wiggins, J. D. Wilbur. 2003. Assessment of statistical methods used in library-based approaches to microbial source tracking. *J. Water Health*. 1:209-223.

Rivera, S. C., T. C. Hazen,. and G. A. Toranzos. 1988. Isolation of fecal coliforms from pristine sites in a tropical rain forest. *Appl. Environ. Microbiol.* 54:513-517.

Rozen, Y., and S. Belkin. 2001. Survival of enteric bacteria in seawater. *FEMS Microbiol Rev* 25:513-529.

Savageau, M. A. 1983. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *American Naturalist* 122:732-744.

Russell, P. F., and T. R. Rao. 1940. On habitat an association of species of anopheline larvae in southeastern Madras. *J. Malaria Inst. India.* 3:153-178.

Sadowsky, M. J. 1994. Microbial DNA fingerprinting and restriction fragment length polymorphism analysis. pp. 647-664. In *Methods of Soil Analysis, Chemical and Microbiological Properties of Soils.* R. W. Weaver, J. S. Angle, and P. Bottomley (eds.), ASA-SSSA, Madison, WI.

Sadowsky, M. J., and H.-G. Hur. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA, 399-413. In *Bacterial genomes: structure and analysis,* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (ed.), Chapman and Hall, New York, NY.

Salyers, A. A., N. B. Shoemaker, A. M. Stevens and L.-Y. Li. 1995. Conjugative transposons: An unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* 59:579-590.

Santo Domingo, J. W., S. C. Siefring, and R. A. Haugland. 2003. Real-time PCR method to detect *Enterococcus faecalis* in water. *Biotech. Lett.* 25:261-263.

Santo Domingo, J. W. M. G. Kaufman, M. J. Klug, and J. M. Tiedje. 1998. Characterization of the Cricket Hindgut Microbiota with Fluorescently Labeled rRNA-Targeted Oligonucleotide Probes. *Appl. Environ. Microbiol*. 64:752-755.

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.

Schaper, M., J. Jofre, M. Uys, and W. O. Grabow. 2002a. Distribution of genotypes of F-specific RNA bacteriophages 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.

Schmidt, T. M., B. Pace, and N. R. Pace. 1991. Detection of DNA contamination in Taq polymerase. *Biotechniques* 11:176-177.

Schwab, K. J., R. De Leon, and M. D. Sobsey 1995. Concentration and purification of beef extract mock eluates from water samples for the detection of enteroviruses, hepatitis A virus, and Norwalk virus by reverse transcription-PCR. *Appl. Environ. Microbiol.* 61**:**531-537.

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* isolates 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.

Selander, R. K., J. M. Musser, D. A. Caugant, M. N. Gilmour, and T. S. Whittam. 1987. Population genetics of pathogenic bacteria. *Microb Pathog* 3:1-7.

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

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.

Sghir, A., G. Gramet, A. Suau, V. Rochet, P. Pochart, and J. Dóre. 2000. Quantification of bacterial groups within human fecal flora by oligonucleotide probe hybridization. *Appl. Environ. Microbiol.* 66: 2263-2266.

Shi, G. R. 1993. Multivariate data analysis in palaeoecolgy and palaeobiology – a review. *Palaeogeography, Palaeoclimatology, Palaeocology.* 105:1999-234.

Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner. 2004. Assessment of equine fecal contamination: the search for alternative bacterial source-tracking targets. *FEMS Microbiol. Ecol.* 47: 65-75.

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.

Sinton, L. W., Hall, C. H., Lynch, P. A. and Davies-Colley, R. J. 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. *Appl. Environ. Microbiol.* 68:1122-31.

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. Freshwat. Res.* 32:323-348.

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

Sneath, P. H. A., and R. R. Sokal. 1973. *Numerical Taxonomy*. Freeman, San Francisco, CA.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38:1409-1438.

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

Sorensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* 5:1-34.

Souza, V., M. Rocha, A. Valera, and L. E. Eguiarte. 1999. Genetic structure of natural populations of Escherichia coli in wild hosts on different continents. *Appl. Environ. Microbiol.* 65:3373-85.

Stackelberg, P. E., E. T. Furlong, M. T. Meyer, S. D. Zaugg, A. K. Henderson, D. B. Reissman. 2004. Persistence of pharmaceutical compounds and other organic wastewater contaminants in a conventional drinking-water-treatment plant. *Sci Total Environ*. 329:99-113.

Stahl, D. A., B. Flesher, H. R. Mansfield, and L. Montgomery. 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Appl. Environ. Microbiol.* 54:1079-1084.

Staley, J. T. and A. Konopka. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Ann. Rev. Microbiol.* 39: 321-346

Statsoft, Inc. 2004. *Electronic Statistics Textbook*. Tulsa, OK: Statsoft WEB: http://www.statsoft.com/textbook/stathome.html

Stewart, C. S. and M. P. Bryant. 1988. In *The rumen microbial ecosystem*, P. N, Hobson, (ed.) Elsevier Applied Science, London, United Kingdom.

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 from a methods comparison study. *J Water Health* 1:225-231.

Stoeckel, D. M., C. M. Kephart, V. J. Harwood, M. A. Anderson, and M. Dontchev. 2004 Diversity of fecal indicator bacteria subtypes: implications for construction of microbial source tracking libraries. American Society for Microbiology General Meeting. New Orleans, LA.

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

Stull, T. L., J. J. LiPuma, and T. D. Edlind. 1988. A broad-spectrum probe for molecular epidemiology of bacteria: ribosomal RNA. J. *Infect. Dis.*157:280–286.

Suau, A., R. Bonnet, M. Sutren, J. -J. Godon, G. R. Gibson, M. D. Collins, and J. Dóre. 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl. Environ. Microbiol*. 65: 4799-4807.

Tajima, K., R. I. Aminov, T. Nagamine, H. Matsui, M. Nakamura, and Y. Benno. 2001. Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. *Appl. Environ. Microbiol.* 67:2766-2774.

Tarkka, E., H. Ahman and A. Siitonen. 1994. Ribotyping as an epidemiologic tool for *Escherichia coli*. *Epidemiol. Infect.* 112: 263–274.

Tartera, C., F. Lucena, and J. Jofre. 1989 Human origin of *Bacteroides fragilis* bacteriophages present in the environment. *Appl. Environ. Microbiol.* 55:2696-701.

Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: Criteria for bacterial strain typing. *J. Clin. Microbiol*. 33:2233-2239

Tian, Y.Q., P. Gong, J. D. Radke, and J. Scarborough, 2002. Spatial and Temporal Modeling of Microbial Containments on Grazing Farmlands. *J. Environ. Quality* 31:860-869.

Ting, W. T., D. Johnson, A. Holler, K. Tran, and C. Tseng. 2003. A study of the sources of *E. coli* contamination at Marquette Park Beach by random amplified polymorphic DNA typing, Abstracts of the 103rd Annual Meeting of the American Society for Microbiology, Washington, DC.

Topp, E., M. Welsh, Y.-C. Tien, A. Dang, G. Lazarovits, K. Conn, and H. Zhu. 2003 Strain-dependent variability in growth and survival of *Escherichia coli* in agricultural soil. *FEMS Microbiol. Ecol.* 44:303-308.

Torgerson, W. S. 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17:401-419.

Torsvik, V., L. Ovreas and J. F. Thingstad. 2002. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* 296: 1064-1066.

Tsai, Y-L, J.Y. Le, and B.H. Olson. 2003. Magnetic bead hybridization to detect enterotoxigenic *Escherichia coli* strains associated with cattle in environmental water sources. *Can. J. Microbiol*. 49:391-398.

U.S. Environmental Protection Agency. 1986 Bacteriological ambient water quality criteria for marine and fresh recreational waters. Washington, D.C.: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 1984. Health Effects Criteria for Fresh Recreational Waters. Office of Research and Development, Washington, DC. EPA-600/1-84-004. 44 pp.

U.S. Environmental Protection Agency. 2000a. Improved enumeration methods for the recreational water quality indicators: Enterococci and *Escherichia coli*. Office of Science and Technology, Washington, DC. EPA/821/R-97/004. 55 pp.

U.S. Environmental Protection Agency. 2000b. Atlas of America's polluted waters. Office of Water (4503F), EPA 840-B00-002. United States Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2001a. Method 1601. Male-specific (F+) and somatic coliphage in water by two-step enrichment procedure. EPA Office of Water, Washington, DC. EPA 821-R-01-030. 40 pp.

U.S. Environmental Protection Agency. 2001b, Method 1602. Male-specific (F+) and somatic coliphage in water by single agar layer (SAL) procedure. EPA Office of Water, Washington, DC. EPA 821-R-01-029. 38 pp.

Valsangiacomo, C., F. Baggi, V. Gaia, T. Balmelli, R. Peduzzi, and J. -C. Piffaretti. 1995. Use of amplified fragment length polymorphism in molecular typing of *Legionella pneumophilia* and application to epidemiological studies. *J. Clin. Microbiol*. 33: 1716-1719.

Vancanneyt, T. M., A. Lombardi, C. Andrighetto, E. Knijff, S. Torriani, K. J. Björkroth, C. M. A. P. Franz, M. R. F. Moreno, H. Rvets, L. De Vuyst, J. Swings, K. Kersters, F. Dellaglio, W.H. Holzapfel. 2002. Intraspecies genomic groups in *Enterococcus faecium* and their correlation with origin and pathogenicity. *Appl. Environ. Microbiol*. 68:1381-1391.

Versalovic, J., V. Kapur, T. Koeuth, G. H. Mazurek, T. S. Whittam, J. M. Musser, and J. R. Lupski. 1995. DNA fingerprinting of pathogenic bacteria by fluorophore-enhanced repetitive sequence-based polymerase chain reaction. *Arch. Pathol. Lab. Med*. 119:23-29.

Versalovic, J., T. Koeuth, and J. R. Lupski. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucl. Acids Res*. 19:6823-6831.

Vesalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski. 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. *Methods Mol. Cell. Biol*. 5:25-40.

Vinjé, J., S. J. G. Oudejans, J. R. Stewart, M. D. Sobsey, and S. C. Long. 2004. Molecular detection and genotyping of male-specific coliphages by RT-PCR and reverse line blot hybridization. *Appl Environ Microbiol.* 70:5996-6004.

Vogel, L., E. van Oorschot, H.M. Maas, B. Minderhoud, and L. Dijkshoorn. 2000. Epidemiologic typing of *Escherichia coli* using RAPD analysis, ribotyping and serotyping. *Clin. Microbiol. Infect.* 6:82-87.

Wade, T. J., N. Pai, J. N. Eisenberg, and J. M. Colford, Jr. 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environ. Health Perspect.* 111:1102-9.

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. *Water Sci. Technol*. 47:27-32.

Wang, G., T. S. Whittam, C. M. Berg, and D. E. Berg. 1993. RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. *Nucl. Acids Res*. 21:5930-5933.

Wang, R. F., W. W. Cao, and C. E. Cerniglia. 1996. PCR detection and quantification of predominant anaerobic bacteria in human and animal fecal samples. *Appl. Environ. Microbiol*. 62: 1242-1247.

Wang, R. F., W. W. Cao, and C. E. Cerniglia. 1997. PCR detection of *Ruminiococcus* spp. in human and animal faecal samples. *Mol. Cell. Probes* 11:259-165.

Wang, R.F., S. J. Kim, L. H. Robertson, and C.E. Cerniglia. 2002. Development of a membrane-array method for the detection of human intestinal bacteria in fecal samples. *Mol. Cell. Probes* 16: 341-350.

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res*. 18:7213-7218.

Wheeler-Alm, E., J. Burke, and A. Spain. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. *Water Res.* 37, 3978-82.

Whitehead, T.R., and M.A. Cotta. 2000. Development of molecular methods for identification of *Streptococcus bovis* from human and ruminal origins. *FEMS Microbiol. Lett.* 182:237-240.

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.

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

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

Whitman, R. L., D. A. Shively, H. Pawlik, M. B. Nevers, and M. N. Byappanahalli. 2003. Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl. Environ. Microbiol.* 69:4714-4729.

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

Whittam, T. S., H. Ochman, and R. K. Selander. 1983 Geographic components of linkage disequilibrium in natural populations of Escherichia coli. *Mol Biol Evol* 1:67-83.

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. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *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.

Wilbur, J. D., J. K. Ghosh, C. H. Nakatsu, S. M. Brouder, and R.W. Doerge. 2002. Variable selection for high-dimensional multivariate binary data with application to microbial community DNA fingerprint analysis. *Biometric* 58:378-386

Wilks, S. S. 1932. Certain Generalization in the Analysis of Variance. *Biometrika* 24:471-494.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey. 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18:6531-6535.

Willshaw, G.A., H.R. Smith, T. Cheasty, P.G. Wall, and B. Rowe.1997. Vero cytotoxin-producing *Escherichia coli* O157 outbreaks in England and Wales, 1995: Phenotypic methods and genotypic subtyping. *Emerg. Infect. Dis.* 3:561-565.

Wilson, I. G. 1997. Inhibition and Facilitation of Nucleic Acid Amplification. *Appl. Environ. Microbiol.* 63:3741-3751.

Wood, J., K .P. Scott, G. Avgustin, C. J. Newbold, and H. J. Flint. 1998. Estimation of the relative abundance of different *Bacteroides* and *Prevotella* ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. *Appl. Environ. Microbiol*. 64: 3683-3689

Xiao L., R. Fayer, U. Ryan, S. J. Upton. 2004. *Cryptosporidium* taxonomy: recent advances and implications for public health. *Clin. Microbiol. Rev*. 17:72-97.

Xiao L., A. Singh, J. Limor J, T. K. Graczyk, S. Gradus, A Lal. 2001. Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater. *Appl. Environ. Microbiol.* 67:1097-101.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, J. I. Gordon.. 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299:2074-2076.

Yang, S., and R. E. Rothman. 2004. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis*. 4: 337348.

Zoetendal, E. G., C. T. Collier, S. Koke, R. I. Mackie, and H. R. Gaskins. 2004. Molecular ecological analysis of the gastrointestinal microbiota: a Review. *J. Nutr*. 134:465-472.

Glossary of Relevant Terms

16S and 23S rRNA – These are part of the ribosomal RNA genes that microbiologists use for the phylogenetic identification of bacteria. Due to the different levels of sequence conservancy they are also used in the development of methods to detect bacteria in complex samples. They are also known as 16S rDNA and 23S rDNA.

Antibiotic Resistance Analysis (ARA) – Method that uses resistance to antibiotics to generate phenotypic profiles of bacterial indentifier.

Clean Water Act (CWA) - An act passed by the U.S. Congress to control water pollution (formerly referred to as the Federal Water Pollution Control Act of 1972). Public Law 92-500, as amended. 33 U.S.C. 1251 et seq.

Clean Water Act Section 303(d) - annual report to Congress from EPA that identifies those waters for which existing controls are not sufficiently stringent to achieve applicable water quality standards.

Clean Water Act Section 305(b) - biennial reporting requires description of the quality of the Nation's surface waters, evaluation of progress made in maintaining and restoring water quality, and description of the extent of remaining problems by using biological data to make aquatic life use support decisions.

Clone - A population of identical microorganisms derived from the same genetic lineage. All of the bacteria in one culture, or one colony identical clones (unless a mutation occurs).

Coliphage – A bacterial virus (i.e., bacteriophage) that infects *E. coli.* Coliphages have been proposed as potential indicators for the presence of enteric viruses in fecally impacted waters.

Confined Animal Feeding Operation (CAFO) - A lot or facility where animals have been, are, or will be stabled or confined and fed or maintained for a total of 45 days or more in any 12 month period; and where crops, vegetation, forage growth, or post-harvest residues are not sustained over any portion of the lot facility in the normal growing season and more than 1,000 animal units are confined at the facility or from 301 to 1,000 animal units are confined at the facility and it also meets one of the specific criteria addressing the method of discharge.

Cosmopolitan – Describes strains that are found in more than one host species. “Transient” is sometimes used synonymously.

DNA - Deoxyribonucleic acid. Encodes for the genetic material of living organisms with the exception of some classes of viruses.

F^+RNA - RNA male-specific coliphages.

False-negative – A source is not identified when it is actually present.

False-positive – A source is identified when it is not actually present.

Genotype – The analysis is based directly on the DNA of the organism. Ribotyping and PCR are both genotypic analyses.

Microbial source tracking – Approach or approaches intended to identify the fecal sources impacting a water system. Other terms that relate to MST are bacterial source tracking (when bacteria is the target), microbial source identification, and fecal source identification.

Non-Point Source Pollution - pollution that occurs when rainfall, snowmelt, or irrigation runs over land or through the ground, picks up pollutants, and deposits them into rivers, lakes, and coastal waters or introduces them into ground water.

Point Source Pollution – Identifiable inputs of waste that are discharged via pipes or drains primarily (but not exclusively) from industrial facilities and municipal treatments plants into rivers, lakes, and ocean.

Phenotype – Characteristics of an organism that rely on translation of genetic information into proteins. Antibiotic resistance patterns and carbon source utilization patterns represent phenotypes, as they are mediated by enzymes and other proteins.

Quantitative PCR – Also known as real time PCR. The principles of QPCR are similar to those of conventional PCR techniques with the exception that in each round of amplification the accumulation of PCR products is quantified using a fluorescence detector.

Restriction fragment length polymorphism (RFLP) - A type of polymorphism detectable in a genome by the size differences in DNA fragments generated by restriction enzyme analysis.

Source identifier (SI) – A general category for the analytes used for MST. *E. coli*, enterococci, PCR bands and caffeine are all examples of SIs.

Species/pattern/marker (SPM) – A specific species, pattern or marker that is indicative of a particular host species. ARA patterns of enterococci, ribotypes of *E. coli* and the human-specific DNA band of *Bacteroides* are examples of SPMs.

Library – In MST is normally refered to the group of fingerprints generated from microbial isolates collected from the potential sources (i.e., animal feces) impacting a watershed. MST libraries should not be confused with gene cloning libraries. Fingerprints are based on phenotypic traits (e.g., antibiotic resistance analysis) or genotypic profiles (e.g., rep-PCR, ribotyping) of individual microbial strains

Library dependent methods (LDMs) - MST methods that require the development of a source library.

Library independent methods (LIMs) – MST methods that do not require the development of a

source library.

RNA – Ribonucleic acid. This polymer is primarily involved in protein synthesis.

Subtype – A microbial strain possessing a distinctive pattern or marker. Electrophoretic types, ribotypes, rep-PCR patterns and antibiotic resistance patterns all define bacterial subtypes. Coliphage types I-IV are also subtypes.

Total Maximum Daily Load (TMDL) – TMDL is a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and an allocation of that amount to the pollutant's sources. Water quality standards are set by States, Territories, and Tribes. They identify the uses for each waterbody, for example, drinking water supply, contact recreation (swimming), and aquatic life support (fishing), and the scientific criteria to support that use. A TMDL is the sum of the allowable loads of a single pollutant from all contributing point and nonpoint sources. The calculation must include a margin of safety to ensure that the waterbody can be used for the purposes the State has designated. The calculation must also account for seasonal variation in water quality. The Clean Water Act, section 303, establishes the water quality standards and TMDL programs.

Type I error – Occurs when a difference is identified that does not really exist (analogous to false-positive).

Type II error – Occurs when a difference that does exist is not identified (analogous to false-negative).

Glossary of Acronyms

AFLP	Amplified fragment length polymorphism
ARA	Antibiotic resistance analysis
ARCC	Average rate of correct classification
ARP	Antibiotic resistance profiling
BMP	Best management practices
BOX-PCR	Repetitive polymerase chain reaction using BOX primers
BST	Bacterial source tracking
CUP	Carbohydrate utilization profiling
DA	Discriminant analysis
DFA	Discriminant function analysis
DGGE	Denaturing gradient gel electrophoresis
rDNA	Ribosomal ribonucleic acid gene
EcoRI	Restriction endonuclease derived from *Escherichia coli*
ERIC-PCR	Enterobacterial repetitive intergenic consensus sequences polymerase chain reaction
FISH	Fluorescent in situ hybridization
HindIII	Restriction endonuclease derived from *Haemophilus influenzae*
ISR-PCR	Intergenic spacer region polymerase chain reaction
MLEE	Multilocus enzyme electrophoresis
MRA	Multiple resistance analysis
MST	Microbial source tracking
NOAA	National Oceanic and Atmospheric Administration
PCR	Polymerase chain reaction
PFGE	Pulse filed gel electrophoresis
PvuII	Restriction endonuclease derived from *Proteus vulgaris*
QPCR	Quantitative PCR
rep-PCR	Repetitive polymerase chain reaction
REP-PCR	Repetitive extragenic palindromic sequence polymerase chain reaction
RFLP	Restriction fragment length polymorphism
rRNA	Ribosomal ribonucleic acid

TMDL	Total maximum daily load
TRFLP	Terminal restriction fragment length polymorphism
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey