# Selection and Application of Microbial Source Tracking Tools For Water-Quality Investigations

Chapter 3 of
**Section A, Biological Science, of**
**Book 2, Collection of Environmental Data**

Techniques and Methods 2–A3

**U.S. Department of the Interior**
**U.S. Geological Survey**

# Selection and Application of Microbial Source Tracking Tools For Water-Quality Investigations

By Donald M. Stoeckel

Chapter 3 of
**Book 2, Collection of Environmental Data,**
**Section A, Biological Science**

Techniques and Methods Book 2, Chapter A3

**U.S. Department of the Interior**
**U.S. Geological Survey**

**U.S. Department of the Interior**
Gale A. Norton, Secretary

**U.S. Geological Survey**
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2005

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply

endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Citation:

Stoeckel, D.M., 2005, Selection and application of microbial source tracking tools for water-quality investigations: U.S. Geological Survey Techniques and Methods Book 2, Chapter A3, 43 p.

# Contents


Abstract ....1
1. Introduction—What is Microbial Source Tracking? ....1
2. Description of MST Protocols ....3
 2.1 Source Identifiers ....3
 2.2 Detection Methods ....4
 2.3 Analytical Approaches ....4
  2.3.1 Empirical Matching Approach ....4
  2.3.2 Population Biology Approach ....4
  2.3.3 Approach Based on Detection Rates ....5
 2.4 Methods-Comparison Studies ....5
3. Matching MST Protocols to Study Objectives ....7
 3.1 Single Household Scale—Ohio Septic Leach Field Study ....7
 3.2 Local Area, Small Population—”Possible Sources of Nitrate in Ground Water at Swine Licensed-Managed Feeding Operations in Oklahoma, 2001” ....8
 3.3 Medium Area, Medium Population—”Patterns and Sources of Fecal Coliform Bacteria in Three Virginia Streams” ....9
 3.4 Medium Area, Transient Population—”Tiered Approach for Identification of a Human Fecal Population Source at a Recreational Beach: Case Study at Avalon Bay, Catalina Island, California” ....8
 3.5 Large Area, Large Population—”Sharing of Ribotype Patterns of *Escherichia coli* Isolates During Baseflow and Stormflow Conditions” ....10
4. Assumptions Behind Choice of MST Protocol ....11
 4.1 Host Specificity ....12
 4.2 Even Distribution Within Host Population ....12
 4.3 Stability of Source Identifiers ....14
  4.3.1 Temporal Stability ....14
  4.3.2 Spatial Stability ....14
 4.4 Consistent Decay Rates in the Environment ....14
 4.5 Relevance to Study Objectives ....15
5. References Cited ....16
6. Glossary ....24
Appendix A: Source Identifiers ....26
 A.1 Regulatory Source Identifiers ....26
 A.2 Nonregulatory Source Identifiers ....28
Appendix B: Detection Methods ....29
 B.1 Conventional Typing Methods ....31
 B.2 Phenotypic Methods ....31
 B.3 Genotypic Methods ....32

## Figure

1. Decision points in the process of microbial source tracking application, and basic information and decisions that lead to effective use of MST protocols, sampling designs, and interpretation ....................2

## Tables

1. General sources of microbial source tracking information....................2
2. Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions of host specificity, distribution among the host population, stability in individual animal hosts, stability in the environment, and relevance to public health regulations or public health risk ....................37

## Conversion Factors and Abbreviated Water-Quality Units

| Multiply | By | To obtain |
|---|---|---|
| foot (ft) | 0.3048 | meter (m) |
| square mile ($mi^2$) | 259.0 | hectare (ha) |
| square mile ($mi^2$) | 2.590 | square kilometer ($km^2$) |
| gallons per day (gal/d) | 0.003785 | cubic meter per day ($m^3$/d) |

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F=(1.8\times°C)+32$$

Concentrations of chemical constituents in water are given either in milligrams per liter (mg/L) or micrograms per liter (μg/L).

Concentrations of bacteria in water are reported in colony-forming units per 100 mililiters (CFU/100 mL), colony-forming units per gram of sampled medium (CFU/g), or cells per gram of sampled medium (cells/g).

Concentrations of phage viruses and enterovirus are reported in plaque-forming units per 100 mililiters (PFU/100 mL).

# Selection and Application of Microbial Source Tracking Tools for Water-Quality Investigations

By Donald M. Stoeckel

## Abstract

Microbial source tracking (MST) is a complex process that includes many decision-making steps. Once a contamination problem has been defined, the potential user of MST tools must thoroughly consider study objectives before deciding upon a source identifier, a detection method, and an analytical approach to apply to the problem. Regardless of which MST protocol is chosen, underlying assumptions can affect the results and interpretation. It is crucial to incorporate tests of those assumptions in the study quality-control plan to help validate results and facilitate interpretation.

Detailed descriptions of MST objectives, protocols, and assumptions are provided in this report to assist in selection and application of MST tools for water-quality investigations. Several case studies illustrate real-world applications of MST protocols over a range of settings, spatial scales, and types of contamination. Technical details of many available source identifiers and detection methods are included as appendixes. By use of this information, researchers should be able to formulate realistic expectations for the information that MST tools can provide and, where possible, successfully execute investigations to characterize sources of fecal contamination to resource waters.

## 1. Introduction — What is Microbial Source Tracking?

Many of our Nation's resource waters fail to achieve water-quality standards for designated uses, as required by the Clean Water Act, even after implementation of the National Point-source Discharge Elimination System (NPDES; U.S. Environmental Protection Agency, 2000b). Therefore, contemporary (2005) state water-management plans emphasize control of nonpoint sources of contamination. Difficulty in allocating contributions of various nonpoint sources, however, poses a substantial obstacle to control of contaminant loads. The emerging field of microbial source tracking (MST) may help identify nonpoint sources of fecal contamination. This document describes the science and application of MST to water-quality investigations and identifies current technical and theoretical limitations associated with a variety of MST protocols.

The term "microbial source tracking" refers to a group of analytical protocols used to determine the source of fecal contamination. As evidence, MST uses host-associated characteristics of various microorganisms present in feces; that is, physiological differences in hosts are expected to select for specific characteristics (such as adhesion factors, antibiotic resistance, temperature optima, and other metabolic traits) in associated enteric microorganisms. The MST approach can be thought of as having several key common components and decision points, as illustrated in figure 1. In particular, the protocol used for MST (source identifier, detection method, and analytical approach) must be applicable to the scale and specific objectives of the study.

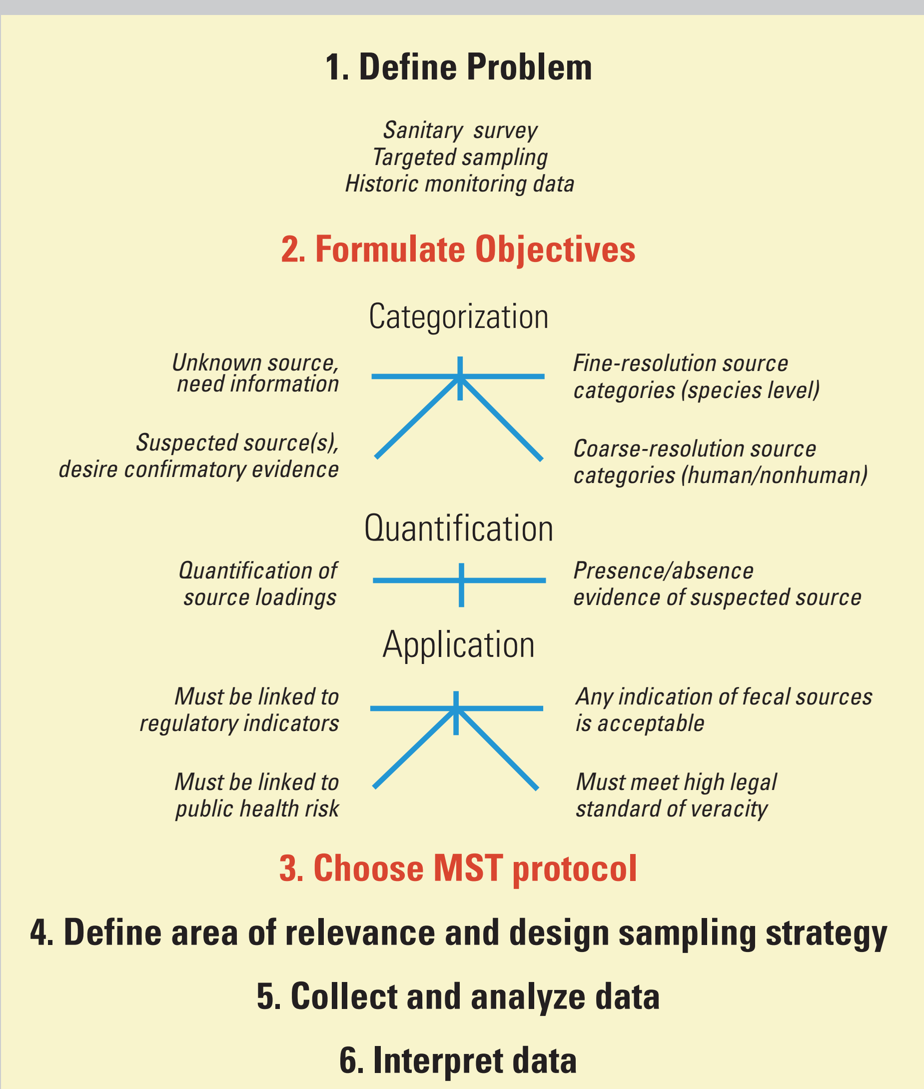


**Figure 1.** Decision points in the process of microbial source tracking application, and basic information and decisions that lead to effective use of MST protocols, sampling designs, and interpretation. General sources of information for each decision point can be found in table 1.

**Table 1.** General sources of microbial source tracking information (MST), by decision point in the MST process.

| Topic | References |
|---|---|
| 1. Problem definition | Ground Water Rule (U.S Environmental Protection Agency, 2000c); Beach guidance, Appendix G (U.S. Environmental Protection Agency, 2002) |
| 2. Formulation of objectives | This document |
| 3. Choice of protocol | This document; Scott and others, 2002; Simpson and others, 2002; Sinton and others, 1998 |
| 4. Design of sampling strategy | Protocol for developing pathogen TMDLs (U.S. Environmental Protection Agency, 2001) |
| 5. Data colection and analysis | Albert and others, 2003; Ritter and others, 2003 |
| 6. Interpretation | Hyer and Moyer, 2004; Long and Plummer, 2004 |

For the purposes of this report, a MST protocol is defined by a combination of source identifier, detection method, and analytical approach. Source identifiers are microbiological markers that can be used to discriminate feces from various sources. Examples are *Escherichia coli* (*E. coli*) genomic fingerprints, coliphage serotypes, and markers within a 16S rRNA gene carried by *Bacteroidetes*. Detection methods are laboratory methods used to detect or distinguish among source identifiers. Examples are antibiotic resistance analysis and ribotyping to distinguish *E. coli* types, traditional serotyping and line-blot hybridization to distinguish coliphage types, and length heterogeneity polymerase chain reaction (PCR) or terminal restriction fragment length polymorphism analysis to detect host-associated *Bacteroidetes* markers. Analytical approaches are strategies used to classify contamination in test samples by comparison to known-source fecal reference materials. Examples are direct matching of subtypes and clustering by fecal source.

Many MST protocols are available for research or practical application. None of the protocols, however, is completely accurate, nor is any one protocol universally applicable to all objectives. This report describes several prior applications of MST and presents information intended to assist in selection of appropriate MST protocols to meet design objectives of water-quality investigations. Information about formulation of objectives and choice of protocol is presented in the context of the framework portrayed in figure 1, but the other MST decision points from figure 1 are not discussed at length in this report. Other sources of information about each MST decision point are listed in table 1.

# 2. Description of MST Protocols

MST protocols are made up of three interrelated parts—a source identifier, a detection method, and an analytical approach:

- Source identifiers are the microbiological targets (that is, viruses, bacteria, or protozoa) on which the MST study is based. Examples are genotypes or phenotypes of *E. coli*, genotypes of F-specific RNA-based coliphage, and host-specific markers found in *Bacteroides/Prevotella* bacteria. (See Appendix A for detailed descriptions.)

- Detection methods are the techniques (generation of genotypic or phenotypic profiles, detection of molecular markers) used to recognize source identifiers from various sources. Examples are ribotyping, antibiotic resistance analysis, length-heterogeneity polymerase chain reaction (LH-PCR), or hybridization with host-specific sequences. (See appendix B for detailed descriptions.)

- Analytical approaches are the conceptual frameworks used to link results from water samples to fecal sources. Examples are library-dependent approaches (epidemiological matching, population biology clustering) and library-independent approaches (based on detection of host-specific markers). A library, in the context of MST, is an extensive collection of source identifiers (often bacterial isolates) that represents fecal material from a source.

## 2.1 Source Identifiers

The choice of source identifier depends, in part, on the objectives of the work. For example, MST in support of recreational water-quality objectives (such as total maximum daily load (TMDL) plans) frequently uses regulated indicator microorganisms as source identifiers. The U.S. Environmental Protection Agency (USEPA) has promulgated regulatory criteria for *E. coli* and enterococci concentrations in recreational waters (Dufour, 1984), and for total coliform and coliphage (proposed) presence as a measure of drinking-water quality (U.S. Environmental Protection Agency, 2000c). Similarly, the U.S. Food and Drug Administration uses fecal coliform concentrations as a regulatory measure of shellfish-harvesting water quality (U.S. Food and Drug Administration, 2003). Nonregulatory source identifiers, such

as *Bacteroidetes*, are used sometimes to confirm fecal sources when a sanitary survey is ambiguous. Alternately, nonregulatory source identifiers sometimes are used on the basis that any reduction in feces from detected sources (based on evidence from nonregulatory source identifiers) will reduce concentrations of regulatory fecal-indicator microorganisms. Descriptions of many available source identifiers are provided in Appendix A.

## 2.2 Detection Methods

Choice of detection method depends on the source identifier chosen, the objectives of the study, and the availability of analytical capabilities. Detection methods that have been developed and used for MST range from the conventional (based on host-associated species such as *Enterococcus faecalis*; Wheeler and others, 2002) through various typing methods of individual isolates (phenotypic or genotypic profiling). A new generation of MST methods based on detection of host-specific gene sequences from whole-water extracted DNA is under development (Field and others, 2003). Descriptions of many available detection methods are provided in Appendix B.

## 2.3 Analytical Approaches

A basic division in analytical approaches for library-dependent MST is between epidemiology and population biology. Epidemiology hypothesizes that specific microbial subtypes are present predominantly in one host category, whereas population biology hypothesizes that a more general combination of traits is common to source identifiers from a host category (such as antibiotic resistance patterns among the human population). A large number of cultivated reference isolates must be collected in the same time and space as the test samples to support classification by library-dependent approaches.

In contrast to library-dependent MST, inferred presence of a fecal source by library-independent MST is based on detection of (generally) a single host-associated marker. Presence of a single trait is anticipated among all or many individuals in a host category, and the trait is expected to be absent in individuals from other host categories. In the library-independent approach, reference feces samples must be collected to develop and validate the method, but the reference sample requirement from the same time and space as the test samples is dictated solely by quality-control needs.

### 2.3.1 Empirical Matching Approach

The empirical matching approach is often referred to as one-to-one matching within library-dependent MST. In this approach, each individual source identifier isolate is assigned to a type. Most types are expected to be exclusive to broad or narrow host categories, regardless of a functional basis describing why that type would be specific to a host category. A functional basis, which may indicate a host-microbe interaction that would support the hypothesis of host specificity, may be used to support the host specificity implied by empirical pattern matching, but it is not necessary: simple detection is sufficient. Epidemiologists use a wide variety of typing methods in an attempt to define useful microbial types in applications related to MST. (See Appendix B for details.) The most discriminating typing methods can separate even closely related lineages into defined types. Types are defined and tested for specificity to a host to allow association of unknown isolates with sources (Hyer and Moyer, 2003; Samadpour and Chechowitz, 1995).

### 2.3.2 Population Biology Approach

The population-biology-based approach can also be library dependent; users of this approach assume an underlying distribution of host-specific characteristics in enteric microorganisms. These characteristics could arise, for instance, during coevolution with specific

hosts (Gordon, 2001; Souza and others, 1999) and result in competitive advantage for specific strains in their coevolved hosts. For source tracking by use of a population biology model, it is assumed that functional characteristics such as phenotypes based on carbon utilization capability (Hagedorn and others, 2003) or sensitivity to antibiotics (Whitlock and others, 2002; Wiggins, 1996) can be used to generate rules to associate bacterial isolates with hosts. Genetic similarity measures, when used in library-dependent approaches, generally are not linked to a specific functional gene. In some experiments, however, functional-gene sequences have been used to define genotypes (Ram and others, 2004; Guan and others, 2002).

Because bacteria reproduce asexually yet often share genetic material, microbial population geneticists look at bacterial genes as a continuum of information in which "species" are nodes that represent stable combinations of genes. Within a species, finer-scale characteristics that may be advantageous within specific hosts are passed down through clonal succession. Statistical group separation methods, such as discriminant analysis, are used to identify combinations of characteristics that are found more commonly in one host over another (Harwood and others, 2000; Parveen and others, 1999; Wiggins, 1996).

### 2.3.3 Approach Based on Detection Rates

Interpretation of results from library-independent approaches is similar to the matching-type library-dependent approaches—a characteristic is either detected or not detected in a DNA extract. Methodological advances with quantitative PCR are showing some promise for quantitation of host-specific markers (Dick and Field, 2004; Seurinck and others, 2005); otherwise, frequency of detection or most-probable-numbers style analysis are the only ways to infer that feces from a particular host are dominant or not dominant. This approach is also different from the cultivation-based library-dependent approaches because (1) DNA contained in a large volume of water often has to be efficiently extracted because most of these methods are cultivation-independent, (2) extracted DNA has to be purified so that PCR-based detection can be done effectively, and (3) none of the source identifiers used in library-independent protocols have been directly correlated with concentrations of regulated fecal indicator microorganisms.

## 2.4 Methods-Comparison Studies

Many protocols are available for MST research. (See Appendixes A and B.) Two efforts to compare various protocols and approaches have been completed. A USGS-led study compared analytical protocols, all of which utilized a library-dependent, *E. coli*-based approach (Stoeckel and others, 2004). A USEPA-funded study led by the Southern California Coastal Water Research Project (SCCWRP) compared protocols utilizing a variety of approaches. The results of the USEPA study were published as a series of reports (see Griffith and others, 2003, for summary and other references).

In both studies, various researchers used MST protocols to determine whether they could correctly associate challenge known-source material with the correct host. In the USGS study, the same 630-isolate library of *E. coli* collected from animals in Berkeley County, W.Va., was distributed to each researcher for evaluation. The known-source library was used to classify isolates from a separate challenge collection of 200 isolates. The challenge isolates were collected from the same area but from different individual animals.

In the USEPA study, a collection of 36 reference feces samples (rather than isolates) and 8 reference composite waste samples was distributed to each researcher for characterization. Twelve challenge slurries were made from the reference materials, representing one to three sources. The USEPA study tested a broader range of MST protocols than the USGS study and, unlike the USGS study, full representation of the source population was achieved because the very same fecal reference materials added to test slurries were previously characterized by each researcher. Because feces—not isolates—were used as reference material, a variety of library-

dependent and library-independent methods could be compared with one another (Griffith and others, 2003).

The USGS study pointed out some limitations of protocols used in library-dependent, *E. coli*-based approaches to MST (Stoeckel and others, 2004). The evaluated protocols represented techniques often used for this type of MST at the time of the study (2001 to 2002):

- Antibiotic resistance analysis (ARA)
- Carbon utilization profiles
- Repetitive DNA element polymerase chain reaction (rep-PCR):
  - REP primer pair
  - Box A1R primer
- Ribotyping:
  - *Hin*dIII restriction enzyme
  - *Eco*R1 and *Pvu*II restriction enzymes
- Macrorestriction using *Not*I restriction enzyme, separated by pulsed-field gel electrophoresis

The seven protocols were evaluated for their ability to

- reproducibly generate data for a given *E. coli* isolate and
- accurately classify *E. coli* isolates to their hosts of origin:
  - eight-way classification to host species,
  - three-way classification to management categories of human, domestic animals, and wildlife, and
  - two-way classification to human or nonhuman.

Of the seven protocols evaluated, only pulsed-field gel electrophoresis was able to correctly classify all replicates, a group of challenge isolates subcultured from the known-source library. The other protocols classified fewer than 60 percent of replicates to the correct host species. Failure to correctly classify replicates indicated that either *E. coli* isolates in the known-source library lacked host specificity or tested protocols lacked precision.

When accuracy was evaluated, a clear tradeoff between correctness and completeness of classification was evident. One protocol correctly classified 90 percent (to species) of the 6 percent attempted; typical of other results was the protocol that correctly classified 27 percent (to species) of the 100 percent attempted. For the two-way classification, four protocols were 63–100 percent accurate when attempting classification of 6–100 percent of the isolates, but accuracy for three protocols was not significantly different from random classification to two sources (50 percent). These accuracy results indicated that tracking isolates to source categories may fail when, as in this study, feces from only a few individuals are used to characterize a much larger population (in this study, 20 humans from a population of 80,000). Failure was evident as both false classification and inability to classify many challenge isolates.

A fundamental conclusion of the USEPA study was that all current approaches to MST have both strengths and weaknesses. Library-dependent methods had low false-positive detection rates, indicating that they would best be applied in settings when large populations of individuals are expected to contribute fecal contamination (Field and others, 2003). One weakness of library-independent methods, however, is that the source identifier may not be present in the feces of most individuals (for example, viruses and *E. coli* toxin genes; see Appendix A for descriptions). This absence of the source identifier would lead to false-negative results. When applied in a setting with large populations, these methods would be expected to have a lower false-negative rate. Library-dependent methods, on the other hand, had low false-negative detection rates, in part because they tended to detect all sources in all samples. This tendency led to high false-positive rates that could be reduced by establishing “minimum detectable threshold” criteria to aid interpretation (Harwood and others, 2003; Myoda and others, 2003; Ritter and others, 2003). Setting a minimum detectable threshold involves estimation of classification error, then rejecting detection rates that are less than that classification error. This allows the

researcher to report only sources that are more certain to have contributed fecal contamination to the sample and to reject the uncertain ones.

Although every researcher had samples of all materials used to make the test samples, no method was completely accurate for all sources. Griffith and others (2003) concluded that length heterogeneity PCR, using markers for *Bacteroidetes* (see Appendix A, sec. A.2.4.1 and Appendix B, sec. B.3.3.2), had the lowest error rates of all protocols tested. This protocol is limited in the number of sources for which markers have been developed (human and ruminant) and the current lack of ability to quantify the detected sources. Library-dependent methods, though more prone to error, may be able to detect the presence of each source and to estimate relative contribution by each source. The results of the USEPA study, however, indicated that library-dependent methods as currently applied often assigned a large proportion of isolates to an "unknown" category, and there was no reliable relation between proportion of bacteria associated with each host and the known composition of the samples (Ritter and others, 2003).

# 3. Matching MST Protocols to Study Objectives

MST can be used to address a variety of objectives related to impacts of feces released to the environment. The following examples illustrate just a few of the ways in which MST protocols have been applied at various scales. This discussion is not intended to be a comprehensive review of past application of MST protocols, nor should these examples necessarily be taken as models of how MST should be applied. Following the description of each study is a brief discussion of alternative protocols that might have been applied to the problem.

## 3.1 Single-Household Scale—Ohio Septic Leach Field Study

Ohio's state guidelines for siting leach lines for household sewage-treatment system (septic systems) leach lines are based on the assumption that treatment of pathogens occurs within the first 2 ft of soil and treatment of nutrients occurs within the first 4 ft of soil (Duncan and others, 1994; Mancl and Slater, 2002). Curtain drains, where constructed to lower the local water table underlying a leach field, often are installed 10 ft from the leach line in Ohio. The hypothesis evaluated in this study was that septic systems installed in Ohio soils with a shallow depth to a limiting condition (the limiting condition often is a high water table or bedrock) provide sufficient treatment of waste. This hypothesis was tested by analyzing near-surface ground water collected by curtain drains and in piezometers at various distances from selected septic leach fields. *E. coli* were detected at horizontal distances of 4–20 ft from the septic leach line at a shallow depth range (4–8 ft), but not in deeper soil water. *E. coli* also were detected in one of two tested curtain-drain effluents.

The area of inference and source populations in this study were both strictly limited; thus, the approach used in this study was to detect direct matches between *E. coli* cultivated from potential sources and *E. coli* cultivated from contaminated water. There was no underlying assumption that *E. coli* genotypes were host-specific; rather, it was hypothesized (and tested) that human and alternate host populations did not carry the same genotypes. MST data were supplemented in this study by collection of nitrogen and oxygen stable-isotopes ratios, wastewater-organic-compound concentrations, and chloride concentrations (Dumouchelle and Stoeckel, 2005).

The objective of MST in this study was to determine whether *E. coli* were present in near-surface ground water and curtain-drain effluent. When present, *E. coli* were tested to determine whether types detected in ground water were the same as those detected in a particular septic system. The physical proximity of the sample water to the septic system makes support of this hypothesis likely, but *E. coli* could also come from droppings on the land surface (pets) or from widespread contamination of offsite, near-surface soil water. Populations of likely contributing hosts were small, consisting of a single family for the septic system and a few pets or wildlife for surface deposits. Contributing hosts were fully represented (household) or heavily represented (local outdoor pets and wildlife) in the known-source library. Widespread contamination of off-

site, near-surface soil water would be detected by observation of *E. coli* in water samples distant from the septic system. *E. coli* isolates were characterized by use of a rep-PCR protocol (see Appendix B, sec. B.3.1.3) similar to that used by Dombek and others (2000).

By use of MST tools (rep-PCR typing of cultivated *E. coli*), not only were *E. coli* detected in near-surface ground water, but the probable source of the contamination was indicated. Without the use of *E. coli* typing, it would have been possible to claim that the *E. coli* found in curtain drain effluent and other near-surface ground water samples came from surface scat, not from the septic system. Indeed, in one case where *E. coli* were detected at great distance (20 ft) from the leach field, *E. coli* typing indicated that the probable source of those *E. coli* was dog feces at the land surface rather than the septic leach field.

Most, if not all, available *E. coli* typing methods probably would have served the same purpose as rep-PCR in this case (van Belkum and others, 2001). Typing of enterococci probably also would have been sufficient to accomplish the objectives of this study. The low diversity of F-specific coliphage (only four types) would have limited its utility for this specific purpose; however, if found, F-specific coliphage and other library-independent markers may have been useful to indicate whether fecal contamination was of human or nonhuman origin. Dye tracers and wastewater organics would be sufficient in this setting to demonstrate a hydrologic link but could not demonstrate subsurface transport of particulates such as *E. coli* nor exclude other origins of *E. coli* detected at a distance from the septic leach field.

## 3.2 Local Area, Small Population—"Possible Sources of Nitrate in Ground Water at Swine Licensed-Managed Feeding Operations in Oklahoma, 2001"

The Oklahoma Department of Agriculture, Food, and Forestry collected ground-water samples underlying swine confined-animal feeding operations (CAFOs). Nitrate concentrations exceeding 10 mg/L as N were detected in 79 wells near 35 CAFOs. The nitrate Maximum Contaminant Level (MCL) for drinking source water is 10 mg/L (U.S. Environmental Protection Agency, 1996). Waste from the CAFOs was treated in waste-stabilization lagoons, and each lagoon contained high concentrations of nitrogen. Nitrate concentrations have been measured as high as 300 mg/L in swine waste-stabilization lagoons (cited in Becker and others, 2003). The hypothesis tested in this study was that nitrate-enriched water leaks from waste-stabilization lagoons and may contribute nitrate to nearby ground water. Support of this hypothesis required some test for subsurface flow between the lagoon and contaminated ground water through the lagoon liner.

MST was applied in this study to determine whether the types of *E. coli* in monitoring wells were the same as those in the adjacent waste-stabilization lagoon. Detection of the same types would provide evidence that *E. coli* traveled from the lagoon to the monitoring well and would support the hypothesis of a hydrologic link between the monitoring well and the lagoon. The objective was addressed by a design in which genotypes of *E. coli* in monitoring wells were compared with genotypes of *E. coli* in the associated swine lagoons. MST data were supplemented in this study by collection of data on surrounding land use, nitrogen stable-isotope ratios, and organic wastewater compound concentrations. *E. coli* collected in this study were analyzed by ribotyping at the Institute for Environmental Health by the protocol reported in Hyer and Moyer (2003; see Appendix B, sec. B.2.3.1). Selected isolates also were analyzed at the USGS Ohio Water Microbiology Laboratory (Columbus, Ohio) with a rep-PCR protocol similar to that reported by Dombek and others (2000; see also Appendix B, sec. B.3.1.3).

Fecal coliform concentrations in the lagoons were highly variable (16 to 210,000 CFU/100 mL), and *E. coli* were detected in 6 of 79 monitoring wells. Of the six wells where *E. coli* were detected, ribotyping characteristics indicated a swine source in five. Additional sources (human, avian, deer) were indicated in two of the five. The sixth had indications of deer as the primary source of fecal contamination (Becker and others, 2003). Additional evidence was collected by an alternate method (rep-PCR) for a well at which *E. coli* were detected[1] and its associated lagoon (data not reported in Becker and others, 2003). The well sample contained *E. coli* that

[1] Referred to as well 360618098024501 in Becker and others (2003).

were indistinguishable, by rep-PCR banding patterns, from *E. coli* in the lagoon. This evidence supported the interpretation based on ribotyping that swine were a source of *E. coli* to that well. The well also yielded water with a nitrate concentration (37 mg/L) above the MCL of 10 mg/L but with no detections of animal-waste-associated wastewater compounds (Becker and others, 2003).

As with the first study (section 3.1), this study required that enough different types of *E. coli* could be detected that a match with a type from the swine lagoon was strong evidence that the swine lagoon was the source of the fecal contamination. Many typing protocols could have been used to accomplish this purpose (see Appendix B); alternatively, enterococci could have been used as the indicator of fecal contamination (see Appendix A). An additional alternative approach would have been to search for swine-specific markers (such as the swine-specific toxin gene of *E. coli*; see appendix B, sec. B.3.3.2) in the monitoring wells, or to establish a hydrologic link between the lagoon and the monitoring well by use of dyes, detection of swine-associated fecal compounds, or other strategies.

## 3.3 Medium Area, Medium Population—"Patterns and Sources of Fecal Coliform Bacteria in Three Virginia Streams"

The objective of the study was to demonstrate field applicability of a MST approach and to identify the sources of fecal coliforms in the study streams (Hyer and Moyer, 2003). In general, the potential sources of fecal contamination in each watershed were considered to be humans, pets, domestic livestock and poultry, migratory birds, or other wildlife. Increased effort toward control of any one of the listed management categories by resource managers, without appropriate supporting data, likely would be contentious to the local government, the agricultural community, or the general public.

This study illustrates a different approach from the Ohio septic system study or the rep-PCR part of the Oklahoma swine lagoon study. Reseachers in the Ohio and Oklahoma studies attempted to compare a source population of *E. coli* (septic tank and swine lagoon, respectively) with the *E. coli* found in contaminated water. The specific identification of host was not attempted in those studies. Specific identification of host, as was done in the Virginia study, goes beyond "this was found both here and there, so there may be a link" to "wherever it is found, this *E. coli* type is associated with a specific host." The assumptions implicit in this approach are the main points of discussion in the chapter on assumptions underlying application of MST (chapter 4).

Many streams in Virginia do not meet applicable designated-use criteria because of microbiological impairment, despite control of known point sources of fecal contamination. Three streams were chosen for evaluation in the reported study: Accotink Creek, drainage area 25 $mi^2$, human population greater than 110,000, primarily urban; Blacks Run, drainage area 20 $mi^2$, human population about 34,700, mixed urban and agricultural; Christians Creek, drainage area 107 $mi^2$, human population about 12,000, primarily agricultural. Extensive base-flow, event-flow, and synoptic sampling was done in each watershed over 20 months. MST by ribotyping (Appendix B, sec. B.3.2.1) was done on *E. coli* isolates collected at a state-determined water-quality compliance point for each watershed. The known-source *E. coli* reference collection of Dr. Mansour Samadpour (Institute for Environmental Health, Seattle, Wash.; more than 50,000 isolates at the time) was used and supplemented by known-source acquisitions in the studied watersheds. The hypothesis 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.

Multiple lines of evidence were used to evaluate expectations for MST findings in these study streams. The authors began by evaluating populations and distributions of fecal sources in each watershed. Synoptic sampling was done to evaluate longitudinal trends in fecal-indicator concentrations in the main stem, in tributaries, and in effluents discharged to the main stem. Seasonal and flow-related trends in fecal-indicator concentrations also were evaluated. These data were interpreted in the context of transport pathways and animal distributions in the watersheds to indicate expected sources of fecal-indicator bacteria.

The MST results were a combination of the expected and the unexpected. Ribotype matching of fecal-indicator bacteria in Accotink Creek, the urban setting, indicated contributions by human and pet feces, as expected, but also a strong influence by waterfowl. Blacks Run fecal-indicator bacteria were of mixed sources (human, pet, and livestock) as expected. Fecal-indicator concentrations in Christians Creek had a larger human and pet component than expected for a primarily agricultural watershed (about 25 percent of isolates), compared with livestock and poultry (about 50 percent). A further unexpected finding in all three watersheds was that relative contributions from each major source were about the same during base-flow and stormflow periods, despite the expectation that different transport pathways would dramatically change relative contributions from different sources. Lastly, the study documented 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.

Several supporting elements were included to help interpret MST data in this study. The unexpectedly high contribution by waterfowl in the urban Accotink Creek watershed was supported by the results of a prior study in a neighboring urban watershed, Four Mile Run (Simmons and others, 2000). Contributions of bacteria from human sources were independently supported by detection of organic wastewater compounds. (Detectable concentrations of caffeine and cotinine were present in all three streams.) Detection of seasonal variability of poultry waste inputs to Christians Creek was supported by trends in total arsenic data collected by Hancock and others (2000). The poultry-feed amendment Roxarsone contains arsenic that is generally excreted by the birds, and the poultry litter eventually is applied to the surrounding agricultural fields. Total arsenic concentrations increased during storm runoff, supporting the hypothesis that field-applied poultry litter was flushed into streams.

Ribotyping of *E. coli* is one of several currently (2005) available approaches that could address the objectives of this study. Because of the size of the watershed (and the large contributing populations of sources), use of any library-dependent method would require a very large library of known-source bacteria (Wiggins and others, 2003; Yang and others, 2004). Because the data were used for development of TMDL plans, numerical approximations of contribution by each source was required; thus, library-independent methods would have been less appropriate to meet the goals of this study. One labor-intensive alternative would have been to conduct more intensive sampling of both flow and concentration in the streams and to calculate the load from each source on the basis of known population contributing to many small subdrainages.

## 3.4 Medium Area, Transient Population—"Tiered Approach for Identification of a Human Fecal Pollution Source at a Recreational Beach: Case Study at Avalon Bay, Catalina Island, California"

Avalon Bay is a tourist destination on Catalina Island, Calif. Boehm and others (2003) report that 17,500 tourists arrive on a typical summer-season day at the town of Avalon, population 3,500. At the time of the study, the beach was closed if any one of seven water-quality criteria was exceeded (single-sample or geometric-mean concentrations of total coliforms, fecal coliforms, or enterococci, plus the ratio of total coliform to fecal coliform concentrations). The authors report that in the summers of 2000 and 2001, the beach was commonly closed because of single-sample enterococci concentrations. The source of enterococci was not obvious. The community is sewered with trunk lines parallel to the beach at a distance of 20 m from the shoreline, and the wastewater-treatment plant outfall (2,200 $m^3$/d) is southeast of Avalon Bay, 100 m from the shore and 65 m deep. Other potential sources of fecal contamination include "nuisance water" from street cleaning and seabirds. The sewage collection systems of the 400 or so boats in the harbor are controlled by a dye program that limits the potential for undetected direct discharge by these vessels.

The authors of this study (Boehm and others, 2003) proposed an approach that combines intensive monitoring of fecal-indicator bacteria concentrations, observation of local hydrology and potential sources, and qualitative testing for presence of fecal material from specific sources to understand a particular contamination problem. This is one of the few published reports in which a library-independent, cultivation-independent protocol was applied to an environmental

setting. The researchers in this study used a combination of intensive sampling (similar to the targeted sampling approach proposed in Kuntz and others, 2003), a comprehensive sanitary survey that included sampling potential sources for fecal-indicator bacteria concentrations, and testing for human-specific genetic markers contained in *Bacteroidetes* (Bernhard and Field, 2000b) and human-specific enterovirus (Monpoeho and others, 2000).

The objective of MST in this study was to evaluate whether human feces contributed to water-quality-standard exceedances in Avalon Bay. The tiered approach described in the report was composed of three parts. First was a survey of contamination patterns, which, in conjunction with local hydrology, showed that contamination came from a localized nearshore source (effectively ruling out the wastewater-treatment-plant discharge as a source). Second was a survey of potential sources, in which a broken pipe that released fecal-indicator bacteria was found and subsurface beach water, nuisance water from street cleaning, and seabird droppings all were shown to carry high concentrations of fecal-indicator bacteria. A nonpermitted discharge pipe and boat cooling-water discharge were shown to carry negligible concentrations of fecal-indicator bacteria. The third and final step was to apply two library-independent tests (*Bacteroidetes* human-specific marker and human-specific enterovirus) to various water samples. Though human-specific markers were detected in both subsurface beach water and Avalon Bay water, the incidence of human-specific marker detection was not related to exceedance of the water-quality standards in the bay. Thus, although presence of human fecal material was detected in one source of high fecal-indicator bacteria, the relative contribution of that source and other sources of fecal-indicator bacteria to water-quality-standard exceedances in Avalon Bay remains unknown.

The objective of the reported study was to detect whether human-origin fecal material was present in Avalon Bay. By use of the approach described, the authors were able to detect fecal material of human origin in Avalon Bay—also indicated were that (1) one hydrologic pathway for that fecal material was by way of subsurface transport under the beach (recall that the sewage system trunk line runs 20 m from the shoreline), and that (2) the wastewater-treatment-plant outfall probably did not contribute fecal material of human origins to Avalon Bay. Had quantification of contributions been a requirement, the researchers may have chosen more expensive alternatives of calculating loads from each source to support a model for bay water quality or applying a library-dependent method in an attempt to classify the source of each colony-forming unit of fecal-indicator bacteria. Library-dependent methods would, however, be almost certain to fail in this setting because of the transience of the tourist population (17,500 different people every day).

## 3.5 Large Area, Large Population—"Sharing of Ribotype Patterns of *Escherichia coli* Isolates During Baseflow and Stormflow Conditions"

The upper Chattahoochee River flows through the city of Atlanta, Ga. Many reaches in the watershed in and around Atlanta do not meet water-quality standards for fecal-coliform bacteria concentrations. A study was designed to evaluate differences in *E. coli* detected during base flow and stormflow as a means of indicating management strategies to control fecal-coliform contamination in the study area. The hypothesis tested in the study was that distinct *E. coli* genotypes could be detected at sample sites during base flow in contrast to those detected during stormflow.

The objective of the study was to test the spatial and flow-related variability in *E. coli* genotypes. *E. coli* were differentiated by ribotyping by the protocol described in Hartel and others (2004), which is very similar to that used in the Oklahoma (Becker and others, 2003) and Virginia studies (Hyer and Moyer, 2003) described in sections 3.2 and 3.3. (See Appendix B, sec. B.3.2.1.)

This study demonstrated high diversity in *E. coli* ribotypes found in a watershed with many tributary streams and outfalls, a large catchment area, and large populations of people and other animals in an urban setting. From a collection of 262 isolates ribotyped, 163 distinct patterns were detected. Only 21 isolates (8 percent) were multiple detections of the same ribotype (6 ribotypes were detected multiple times). Because there were so many singleton detections

of ribotypes (only detected once during the study), the opportunity to characterize patterns in ribotype pattern occurrence with flow regime was very limited. It was apparent from this study that enormous sampling effort, particularly for a known-source library, would be necessary to do library-dependent MST in a system as complex as the Chattahoochee River watershed. An alternative approach in this setting might be to narrow the scope of the study area by use of targeted sampling (Kuntz and others, 2003) or to use some of the emerging library-independent PCR-based markers in a most-probable-numbers dilution series format to estimate the relative contributions by various sources in the watershed. (See appendix B, sec. B.3.3.)

# 4. Assumptions Behind Choice of MST Protocol

A comprehensive objective of MST studies is *quantitative* assessment of all fecal or fecal-indicator loads from *multiple source categories* affecting regulatory or other public health *standards for water quality* at a sample site. Reliable attainment of this objective is constrained by the following assumptions:

- *Host specificity*. The source identifier(s) is associated exclusively with a single host species or other grouping. There is no alternate source of the source identifier(s).
- *Even distribution within host population*. The source identifier(s) is found in every individual of the host population at the same concentration.
- *Stability of source identifiers*. The source identifier(s) is invariable on the spatial and temporal scale of the study's area of inference.
- *Consistent decay rate in the environment*. The source identifier(s) for each host category decays at the same rate and does not replicate outside of the host.
- *Relevance to study objectives*. The chosen source identifier(s) is amenable to specific objectives of the investigation, such as reduction of regulatory fecal-indicator bacteria concentrations or verifiable reduction in public health risk.

No MST protocol meets all of these assumptions (table 2 at back of report), nor do the objectives of every MST study require that each assumption be met. Choice of MST protocol is driven, in part, by the importance of each assumption to meeting the objective of the study and the degree to which the MST protocol meets each assumption.

## 4.1 Host Specificity

Host specificity is expected to be highest for microorganisms that coevolved with their hosts. The textbook examples of this type of host-specific association are symbioses, like the relationship in which the bacterium *Vibrio fischeri* produces light in a special organ under the eyes of the flashlight fish (*Photoblepharon palpebratus*) and the relationship in which bacteria of the species *Rhizobium* nodulate the roots of leguminous plants for nitrogen fixation (Atlas and Bartha, 1993). Examples of host-specific relationships more relevant to MST include pathogenic viruses, which must be able to recognize the correct host cell for infection. There is no expectation of strong host specificity for a microorganism that does not have a functional interaction with its host.

The regulatory fecal indicators *E. coli*, total and fecal coliforms, coliphage, and enterococci are examples of MST source identifiers that make up a small proportion of the enteric biota and have little or no functional interaction with their hosts. The widely studied species *E. coli* is considered to be an opportunist that has limited specificity for its enteric environment and, in warmblooded hosts, can be found in other body systems (such as urethra, causing urinary tract infections; udder, causing mastitis; and eyes, causing ocular conjunctivitis or pinkeye). *E. coli* is not strictly limited to warmblooded hosts: the species has been observed in the feces of reptiles

(Souza and others, 1999) and the water cups of bromeliads (Rivera and others, 1988), and it is suspected to sustain viable populations in warm, nutrient-rich environments such as beach sand (Whitman and Nevers, 2003) and beach wrack (Whitman and others, 2003). These observations about host specificity extend also to the total coliforms and fecal coliforms, of which *E. coli* is a member. Similarly, enterococci have been cultivated from a variety of warmblooded animals (Wheeler and others, 2002), cold-blooded animals (Ott, 2001), and plant surfaces (Kanoe and Abe, 1988). Though perhaps rare, these exceptions to the assumption that fecal-indicator microbes are exclusively associated with the feces of warmblooded animals illustrate that absolute specificity of fecal-indicator bacteria subtypes to species is unlikely.

Souza and others (2002) proposed that hosts act as selective sieves in which only the best adapted subtypes will succeed. The concept of selection does not, however, imply exclusivity. Regulatory fecal-indicator bacteria have been reported to survive and reproduce in a variety of environments, and none of the regulatory fecal-indicator microorganisms fully exhibits host specificity (table 2). Host specificity for regulated fecal-indicator microorganisms might be best described as "differential distribution." Accuracy of MST is limited, in part, by the extent to which host specificity is absolute. For example, detection of a subtype of *E. coli* that is most often found in cattle might be considered circumstantial evidence that cattle are a source of *E. coli* to that water. Human-specific adenoviruses and enteroviruses, in contrast, have been observed only in human hosts (Noble and others, 2003). Detection of these human-specific viruses provides strong evidence that humans were a source of fecal contamination to the water sample, but no evidence of how dominant the human source was to overall contamination. None of the alternative indicators of fecal contamination is nearly as well characterized as *E. coli* or enterococci (see Appendix A, sec. A.2), so there is the risk that limitations of alternative indicators will arise as they are studied more closely.

## 4.2 Even Distribution Within Host Population

Few comprehensive studies have been done to evaluate the distribution of source identifiers in their hosts. Perhaps not surprisingly, many of the studies that have been done have emphasized regulatory fecal-indicator microorganisms and pathogens of human health or economic importance. One such organism is *E. coli*. Gordon and Cowling (2003) measured the distribution of enteric coliform bacteria (of which *E. coli* is one species) among Australian mammals and found them to be only sporadically detected in component of gut populations (detection threshold >1 percent of bacteria capable of growth on MacConkey agar). In other studies, fecal coliform concentrations (of which *E. coli* is a major component) have been found to vary by orders of magnitude in feces coming from the same species (Geldreich, 1976). Use of *E. coli* concentrations to make interpretations about actual feces loads to the study area might lead to false conclusions. Where information is available, indications are that the other proposed source identifiers also tend to violate the assumption of presence at consistent concentrations in feces of their hosts (table 2).

## 4.3 Stability of Source Identifiers

### 4.3.1 Temporal Stability

Microorganisms are susceptible to genetic shuffling (genome plasticity). Genome plasticity occurs over timeframes relevant to this discussion because some microorganisms go through many generations in a matter of days (Savageau, 1983). Because bacterial reproduction is asexual, mechanisms for genetic rearrangement are generally within a single generation rather than during the reproductive process. Mechanisms include transfer of genetic material within and between species (conjugation), uptake and incorporation of naked DNA in the environment (transformation), incorporation of virus DNA (lysogeny), and transfer of mobile genetic elements (plasmids; Smith-Keary, 1989). Genome plasticity may affect subtypes defined by genetic fingerprints (pulsed-field gel electrophoresis (PFGE), ribotypes, rep-PCR and others) and phenotypic characteristics partially defined by plasmid-based traits (ARA) more than those defined by a single locus or other target sequence (length heterogeneity PCR, toxin genes, coliphage genotyping). Virus reproduction, through host-mediated replication, is subject to the same random genetic shuffling mechanisms as the host during replication. As with bacteria, some viruses are known to be more prone to genetic recombination and mutation than others. The mechanisms of genome plasticity have been defined and observed; however, actual rates of transformation and recombination in the environment are not well defined in most cases.

In addition to genome plasticity, there are constant shifts in dominance among the enteric microorganisms. Caugant and others (1981) reported finding few “resident” *E. coli* types and many ”transient” types in a human individual over about a year of repeated sampling. Overall, though, 62 percent of *E. coli* isolated from this individual were of the two resident types; thus, at any given time it would seem that a host is most likely to shed the resident type(s). Functional genetic markers may prove to be more temporally stable source indicators. Examples of functional genes include those encoding housekeeping genes (Maynard Smith and others, 2000) and adhesion proteins (Scott and others, 2005). Function-based markers have the dual benefit of conferring host specificity on the carrier organism and limiting fitness of alternates, thereby reducing the chance that individuals not carrying the marker will be shed persistently by the host.

### 4.3.2 Spatial Stability

There may be a geographic area to which each source identifier is limited. Studies of *E. coli* (Hartel and others, 2002; Scott and others, 2003) and enterococci (Wiggins and others, 2003) from known-sources indicate that libraries may be useful across watersheds within a region but not necessarily between regions. The dynamics of population distribution include not only suitability of habitat (host specificity, influenced by local diet and environment) but also transfer and colonization among geographically isolated host populations. An isolated population of host animals may contain a dominant marker, bacteria subtype, or other source identifier that, though suitable for colonization of other host populations, does not get the chance to do so because of geographic separation.

## 4.4 Consistent Decay Rates in the Environment

The failure of the fecal coliform/fecal streptococcus (FC/FS) ratio for use in source tracking was a lesson to heed in current pursuit of MST. This lesson is particularly relevant when attempting to quantify source contributions, which is the objective of many TMDL evaluations, and for less ambitious attempts to designate detected hosts as “major” or “minor” contributors. In the case of the FC/FS ratio, it was initially established that the distribution in fresh human feces is weighted toward fecal coliforms; thus, concentration ratios lower than 0.8 indicated predominantly nonhuman sources, whereas ratios greater than 4 indicated human sources of fecal contamination (Geldreich, 1976). Unfortunately, the relevance of the interpretations was found to be limited to recent contamination events because fecal coliform concentrations decay faster

than fecal streptococci concentrations; over the course of time, the FC/FS ratio shifts to high numbers (indicative of human sources) regardless of the source of fecal contamination (Clesceri and others, 1998).

Instability of source-indicating ratios (such as FC/FS), or shifts in the relative concentrations of source indicators, can occur through differential die-off rates or processes that change a source identifier's type (such as mutation, exchange of genetic material, or changes in gene expression because of environmental conditions). Selective pressures in the secondary (non-enteric) habitat may very well undermine interpretations based on observations of source identifier characteristics in the primary (enteric) habitat (Savageau, 1983; Gordon and others, 2002). This is a very important potential limitation to development of known-source isolate libraries by use of fresh feces as reference material and use of other microbial ratios to indicate fecal sources.

## 4.5 Relevance to Study Objectives

Study objectives may require direct correlation with regulated fecal-indicator bacteria, direct correlation with human or other health risk, or both. Research that indicates whether each source identifier meets these criteria is described in table 2.

# 5. References Cited

Aarestrup, F.M., Agerso, Y., Gerner-Smidt, P., Madsen, M., and Jensen, L.B., 2000, Comparison of antimicrobial resistance phenotypes and resistance genes in *Enterococcus faecalis* and *Enterococcus faecium* from humans in the community, broilers, and pigs in Denmark: Diagnostic Microbiology and Infectious Disease, v. 37, p. 127–137.

Albert, J.M., Munakata-Marr, J., Tenorio, L., and Siegrist, R.L., 2003, Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*: Environmental Science & Technology, v. 37, p. 4554–4560.

Atlas, R.M., 1988, Microbiology (2d ed.): New York, Macmillan Publishing Company, 807 p.

Atlas, R.M., and Bartha, R., 1993, Microbial ecology—Fundamentals and applications: New York, Benjamin/Cummings Publishing Company, 563 p.

Atwill, E.R., Johnson, E., Klingborg, D.J., Veserat, G.M., Markegard, G., Jensen, W.A., Pratt, D.W., Delmas, R.E., George, H.A., Forero, L.C., Philips, R.L., Barry, S.J., McDougald, N.K., Gildersleeve, R.R., and Frost, W.E., 1999, Age, geographic, and temporal distribution of fecal shedding of *Cryptosporidium parvum* oocysts in cow-calf herds: American Journal of Veterinary Research, v. 60, no. 4, p. 420–425.

Becker, M.F., Peter, K.D., and Masoner, J., 2003, Possible sources of nitrate in ground water at swine licensed-managed feeding operations in Oklahoma, 2001: U.S. Geological Survey Water-Resources Investigations Report 02–4257, 76 p.

Beekwilder, J., Nieuwenhuizen, A.H., Havelaar, A.H., and van Duin, J., 1996, An oligonucleotide hybridization assay for the identification and enumeration of F-specific RNA phages in surface water: Journal of Applied Bacteriology, v. 80, p. 179–186.

Bernhard, A.E., and Field, K.G., 2000a, Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes: Applied and Environmental Microbiology, v. 66, no. 4, p. 1587–1594.

Bernhard, A.E., and Field, K.G., 2000b, A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA: Applied and Environmental Microbiology, v. 66, no. 10, p. 4571–4574.

Bernhard, A.E., Goyard, T., Simonich, M.T., and Field, K.G., 2003, Application of a rapid method for identifying fecal pollution sources in a multi-use estuary: Water Research, v. 37, no. 4, p. 909–913.

Bettelheim, K., and Thomas, G., 2004, *E. coli* serotypes: University of Birmingham, England, accessed January 26, 2005, at *http://ecoli.bham.ac.uk/path/sero.html*.

Blackwood, A.D., Gregory, J.B., and Noble, R.T., 2003, Determination of the quantitative relationship between indicator bacteria and viral pathogens in recreational waters using quantitative PCR [abs.]: Washington, D.C., Abstracts of the General Meeting of the American Society for Microbiology, v. 103, Q–427.

Boehm and others, 2003, Tiered approach for identification of a human fecal pollution source at a recreational beach—Case study at Avalon Bay, Catalina Island, California: Environmental Science & Technology, v. 37, p. 673–680.

Bonjoch, X., Balleste, E., and Blanch, A.R., 2004, Multiplex PCR with 16S rRNA gene-targeted primers of *Bifidobacterium* spp. to identify sources of fecal pollution: Applied and Environmental Microbiology, v. 70, no. 5, p. 3171–3175.

Borchardt, M.A., Bertz, P.D., Spencer, S.K., and Battigelli, D.A., 2003, Incidence of enteric viruses in groundwater from household wells in Wisconsin: Applied and Environmental Microbiology, v. 69, p. 1172–1180.

Borrego, J.J., Morinigo, M.A., Vicente, A., Cornx, R., and Romero, P., 1987, Coliphages as an indicator of fecal pollution in water—Its relationship with indicator and pathogenic microorganisms: Water Research, v. 21, p. 1473–1480.

Brenner, K.P., Rankin, C.C., Roybal, Y.R., Stelma, G.N., Scarpino, P.V., and Dufour, A.P., 1993, New medium for the simultaneous detection of total coliforms and *Escherichia coli* in water: Applied and Environmental Microbiology, v. 59, no. 11, p. 3534.

Cabelli, V., Dufour, A.P., McCabe, L.J., and Levin, M.A., 1982, Swimming-associated gastroenteritis and water quality: American Journal of Epidemiology, v. 115, p. 606–616.

Calci, K.R., Burkhardt, W., III, Watkins, W.D., and Rippey, S.R., 1998, Occurrence of male-specific bacteriophage in feral and domestic animal wastes, human feces, and human-associated wastewaters: Applied and Environmental Microbiology, v. 64, no. 12, p. 5027–5029.

Carrillo, M., Estrada, E., and Hazen, T.C., 1985, Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed: Applied and Environmental Microbiology, v. 50, p. 468–476.

Carson, C.A., Shear, B.L., Ellersieck, M.R., and Asfaw, A., 2001, Identification of fecal *Escherichia coli* from humans and animals by ribotyping: Applied and Environmental Microbiology, v. 67, no. 4, p. 1503–1507.

Carson, C.A., Shear, B.L., Ellersieck, M.R., and Schnell, J.D., 2003, Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals: Applied and Environmental Microbiology, v. 69, no. 3, p. 1836–1839.

Caugant, D.A., Levin, B.R., and Selander, R.K., 1981, Genetic diversity and temporal variation in the *E. coli* population of a human host: Genetics, v. 98, p. 467–490.

Caugant, D.A., Levin, B.R., and Selander, R.K., 1984, Distribution of multilocus genotypes of *Escherichia coli* within and between host families: Journal of Hygiene, v. 92, p. 377–384.

Centers for Disease Control and Prevention, 1996, Standardized molecular subtyping of foodborne bacterial pathogens by pulsed-field gel electrophoresis (updated 2000 ed.): Atlanta, Ga., National Center for Infectious Diseases, Centers for Disease Control and Prevention.

Chalmers, R.M., Sturdee, A.P., Bull, S.A., Miller, A., and Wright, S.E., 1997, The prevalence of *Cryptosporidium parvum* and *C. muris* in *Mus domesticus*, *Apodemus sylvaticus* and *Clethrionomys glareolus* in an agricultural system: Parasitology Research, v. 83, no. 5, p. 478–482.

Chern, E.C., Tsai, Y., and Olson, B.H., 2004, Occurrence of genes associated with enterotoxigenic and enterohemorrhagic *Escherichia coli* in agricultural waste lagoons: Applied and Environmental Microbiology, v. 70, no. 1, p. 356–362.

Clesceri, L.S., Greenberg, A.E., and Eaton, A.D., eds., 1998, Standard methods for the examination of water and waste water: Washington, D.C., American Public Health Association [variously paginated].

Cole, D., Long, S.C., and Sobsey, M.D., 2003, Evaluation of F+RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters: Applied and Environmental Microbiology, v. 69, no. 11, p. 6507–6514.

Cooke, E.M., Ewins, S., and Shooter, R.A., 1969, Changing faecal population of *Escherichia coli* in hospital medical patients: British Medical Journal, v. 4, p. 593–595.

Delahoya, N.C., Kramer, M.F., Whitlock, J.E., Harwood, V.J., and Lim, D.V., 2002, Survey of *Enterococcus* species from various sources using API 20 strip and PCR methods for species identification [abs]: Abstracts of the General Meeting of the American Society for Microbiology, v. 102, p. 389.

Dick, L.K., and Field, K.G., 2004, Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR assay for 16S rRNA genes: Applied and Environmental Microbiology, v. 70, no. 9, p. 5695–5697.

Dick, L.K., Field, K.G., Brodeur, T., Simonich, M., and Pester, M., 2002, Phylogenetic analysis of fecal *Bacteroides-Prevotella* and development of host-specific PCR primers for source identification of fecal contamination [abs.]: Abstracts of the General Meeting of the American Society for Microbiology, v. 102.

Dombek, P.E., Johnson, L.K., Zimmerley, S.T., and Sadowsky, M.J., 2000, Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources: Applied and Environmental Microbiology, v. 66, no. 6, p. 2572–2577.

Dufour, A.P., 1984, Health effects criteria for fresh recreational waters: Research Triangle Park, N.C., U.S. Environmental Protection Agency, Office of Research and Development, EPA-600/1-84-004, 33 p.

Dumouchelle, D.D., and Stoeckel, D.M., 2005, Preliminary investigations of wastewater-related contaminants near home sewage treatment systems in Ohio: U.S. Geological Survey Open File Report 2005–1282, 31 p.

Duncan, C.S., Reneau, R.R.J., and Hagedorn, C., 1994, Impact of effluent quality and soil depth on renovation of domestic wastewater: Proceedings of the Seventh International Symposium on Individual and Small Community Sewage Systems, p. 219–228.

Farnleitner, A.H., Kreuzinger, N., Kavka, G.G., Grillenberger, S., Rath, J., and Mach, R.L., 2000, Simultaneous detection and differentiation of *Escherichia coli* populations from environmental freshwaters by means of sequence variations in a fragment of the beta-D glucuronidase gene: Applied and Environmental Microbiology, v. 66, no. 4, p. 1340–1346.

Fayer, R., Morgan, U., and Upton, S.J., 2000, Epidemiology of *Cryptosporidium*—Transmission, detection and identification: International Journal for Parasitology, v. 30, no. 12–13, p. 1305–1322.

Field, K.G., Chern, E.C., Dick, L.K., Fuhrman, J., Griffith, J., Holden, P.A., LaMontagne, M.G., Le, J., Olson, B., and Simonich, M.T., 2003, A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. Journal of Water and Health, v. 1, no. 4, p. 181–194.

Franke, S., Harmsen, D., Caprioli, A., Pierard, D., Wieler, L.H., and Karch, H., 1995, Clonal relatedness of shiga-like toxin producing *Escherichia coli* O101 strains of human and porcine origin: Journal of Clinical Microbiology, v. 33, p. 3174–3178.

Franks, A.H., Harmsen, H.J.M., Raangs, G.C., Jansen, G.J., Schut, F., and Welling, G.W., 1998, Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes: Applied and Environmental Microbiology, v. 64, no. 9, p. 3336–3345.

Furuse, K., 1987, Distribution of coliphages in the environment—general considerations, *in* Goyal, S.M., Gerba, C.P. and Bitton, G., eds., Phage ecology: New York, John Wiley & Sons, p. 87–124.

Geldreich, E.E., 1976, Faecal coliform and faecal streptococcus density relationships in waste discharges and receiving waters: Critical Reviews in Environmental Control, v. 6, p. 349–369.

Geldreich, E.E., and Kenner, B.A., 1969, Concepts of fecal streptococci in stream pollution: Journal of the Water Pollution Control Federation, v. 41, no. 8, Supplement p. R336–352.

Gerba, C.P., Gramos, D.M., and Nwachuku, N., 2002, Comparative inactivation of enteroviruses and adenovirus 2 by UV light: Applied and Environmental Microbiology, v. 68, no. 10, p. 5167–5169.

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

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

Gordon, D.M., and Cowling, A., 2003, The distribution and genetic structure of *Escherichia coli* in Australian vertebrates—Host and geographic effects: Microbiology, v. 149, p. 3575–3586.

Graves, A.K., Hagedorn, C., Teetor, A., Mahal, M., Booth, A.M., and Reneau, R.B., Jr., 2002, Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed: Journal of Environmental Quality, v. 31, p. 1300–1308.

Griffin, D.W., Gibson, C.J., III, Lip, E.K., Riley, K., Paul, J.H., III, and Rose, J.B., 1999, Detection of viral pathogens by reverse transcriptase PCR and of microbial indicators by standard methods in the canals of the Florida Keys: Applied and Environmental Microbiology, v. 65, no. 9, p. 4118–4125.

Griffin, D.W., Stokes, R., Rose, J.B., and Paul, J.H.,III, 2000, Bacterial indicator occurrence and the use of an F+ specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida: Microbial Ecology, v. 39, no. 1, p. 56–64.

Griffith, J.F., Weisberg, S.B., and McGee, C.D., 2003, Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples: Journal of Water and Health, v. 1, no. 4, p. 141–151.

Guan, S., Xu, R., Chen, S., Odumeru, J., and Gyles, C., 2002, Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources: Applied and Environmental Microbiology, v. 68, no. 6, p. 2690–2698.

Hagedorn, C., 1999, Developing methods for bacterial source tracking , *in* [Proceedings] 10th Annual meeting, Northwest On-Site Wastewater Treatment Short Course and Equipment Exhibition: Seattle, Wash., 64 p.

Hagedorn, C., Crozier, J.B., Mentz, K.A., Booth, A.M., Graves, A.K., Nelson, N.J., and Reneau, R.B., Jr., 2003, Carbon source utilization profiles as a method to identify sources of faecal pollution in water: Journal of Applied Microbiology, v. 94, no. 5, p. 792–799.

Hagedorn, C., Robinson, S.L., Filz, J.R., Grubbs, S.M., Angier, T.A., and Reneau, R.B., Jr., 1999, Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci: Applied and Environmental Microbiology, v. 65, no. 12, p. 5522–5531.

Hancock, T.C., Miller, C.V., Denver, J.M., and Riedel, G.F., 2000, Fate and transport of arsenical feed amendments in Chesapeake Bay watersheds [abst.]: Society of Environmental Toxicology and Chemistry Annual Meeting, 21st, Nashville, November 12–16, 2000.

Hartel, P.G., Frick, E.A., Funk, A.L., Hill, J.L., Summer, J.D., and Gregory, M.B., 2004, Sharing of ribotype patterns of *Escherichia coli* isolates during baseflow and stormflow conditions: U.S. Geological Survey Scientific Investigations Report 2004–5004, 10 p.

Hartel, P.G., Summer, J.D., Hill, J.L., Collins, V.J., Entry, J.A., and Segars, W.I., 2002, Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia: Journal of Environmental Quality, v. 31, p. 1273–1278.

Harwood, V.J., Whitlock, J., and Withington, V., 2000, Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis—Use in predicting the source of fecal contamination in subtropical waters: Applied and Environmental Microbiology, v. 66, no. 9, p. 3698–3704.

Harwood, V.J., Wiggins, B., Hagedorn, C., Ellender, R.D., Gooch, J., Kern, J., Samadpour, M., Chapman, A.C.H., Robinson, B.J., and Thompson, B.C., 2003, Phenotypic library-based microbial source tracking methods—Efficacy in the California collaborative study: Journal of Water and Health, v. 1, no. 4, p. 153–166.

Havelaar, A.H., 1993, Bacteriophages as models of human enteric viruses in the environment: American Society for Microbiology News, v. 59, no. 614, p. 619.

Heitman, T.L., Frederick, L.M., Viste, J.R., Guselle, N.J., Morgan, U.M., Thompson, R.C.A., and Olson, M.E., 2002, Prevalence of *Giardia* and *Cryptosporidium* and characterization of *Cryptosporidium* spp. isolated from wildlife, human, and agricultural sources in the North Saskatchewan River Basin in Alberta, Canada: Canadian Journal of Microbiology, v. 48, no. 6, p. 530–541.

Hsu, F., Shieh, Y.C., van Duin, J., Beekwilder, M.J., and Sobsey, M.D., 1995, Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes: Applied and Environmental Microbiology, v. 61, p. 3960–3966.

Hutchison, M.L., Walters, L.D., Avery, S.M., Synge, B.A., and Moore, A., 2004, Levels of zoonotic agents in British livestock manures: Letters in Applied Microbiology, v. 39, no. 2, p. 207–214.

Hyer, K.E., and Moyer, D.L., 2003, Patterns and sources of fecal coliform bacteria in three streams in Virginia, 1999–2000: U.S. Geological Survey Water Resources Investigations Report 03–4115, 76 p.

Hyer, K.E., and Moyer, D.L., 2004, Enhancing fecal coliform total maximum daily load models through bacterial source tracking: Journal of the American Water Resources Association, v. 40, no. 6, p. 1511–1526.

Jenkins, M.B., Hartel, P.G., Olexa, T.J., and Stuedemann, J.A., 2003, Putative temporal variability of *Escherichia coli* ribotypes from yearling steers: Journal of Environmental Quality, v. 32, p. 305–309.

Jiang, S., Noble, R., and Chu, W., 2001, Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California: Applied and Environmental Microbiology, v. 67, p. 179–184.

Jiménez-Clavero, M.A., Fernández, C., Ortiz, J.A., Pro, J., Carbonell, G., Tarazona, J.V., Roblas, N., and Ley, V., 2003, Teschoviruses as indicators of porcine fecal contamination of surface water: Applied and Environmental Microbiology, v. 69, no. 10, p. 6311–6315.

Jofre, J., Ollé, E., Ribs, F., Vidal, A., and Lucena, F., 1995, Potential usefulness of bacteriophages that infect *Bacteroides fragilis* as model organisms for monitoring virus removal in drinking water treatment plants: Applied and Environmental Microbiology, v. 61, no. 9, p. 3227–3231.

Kanoe, M., and Abe, T., 1988, Enterococcal isolates from environmental sources: Microbios Letters, v. 38, p. 15–20.

Khatib, L.A., Tsai, Y.L., and Olson, B.H., 2002, A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic *Escherichia coli*: Applied Microbiology and Biotechnology, v. 59, no. 1, p. 97–104.

Khatib, L.A., Tsai, Y.L., and Olson, B.H., 2003, A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*: Applied Microbiology and Biotechnology, v. 63, no. 2, p. 231–238.

Kopecka, H., Dubrou, S., Prevot, J., Marechal, J., and Lopez-Pila, J.M., 1993, Detection of naturally occurring enteroviruses in waters by reverse transcription, polymerase chain reaction, and hybridization: Applied and Environmental Microbiology, v. 59, no. 4, p. 1213–1219.

Kreader, C.A., 1998, Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water: Applied and Environmental Microbiology, v. 64, no. 10, p. 4103–4105.

Kuntz, R.L., Hartel, P.G., Godfrey, D.G., McDonald, J.L., Gates, K.W., and Segars, W.I., 2003, Targeted sampling protocol as prelude to bacterial source tracking with *Enterococcus faecalis*: Journal of Environmental Quality, v. 32, no. 6, p. 2311–2318.

Leclerc, H., Mossel, D.A., Edberg, S.C., and Struijk, C.B., 2001, Advances in the bacteriology of the coliform group—Their suitability as markers of microbial water safety: Annual Reviews in Microbiology, v. 55, p. 201–234.

Ley, V., Higgins, J., and Fayer, R., 2002, Bovine enteroviruses as indicators of fecal contamination: Applied and Environmental Microbiology, v. 68, no. 7, p. 3455–3461.

Long, S.C., and Plummer, J.D., 2004, Assessing land use impacts on water quality using microbial source tracking: Journal of the American Water Resources Association, v. 40, no. 6, p. 1433–1448.

Louws, F.J., Rademaker, J.L.W., and de Bruijn, F.J., 1999, The three Ds of PCR-based genomic analysis of phytobacteria—Diversity, detection, and disease diagnosis: Annual Review of Phytopathology, v. 37, p. 81–125.

Mallon, M., MacLeod, A., Wastling, J., Smith, H., Reilly, B., and Tait, A., 2003, Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*: Journal of Molecular Evolution, v. 56, no. 4, p. 407–417.

Maluquer de Motes, Carlos; Clemente-Cesares, P.; Hundesa, A.; Martín, M.; and Girones, R., 2004, Detection of bovine and porcine adenoviruses for tracing the source of fecal contamination: Applied and Environmental Microbiology, v. 70, no. 3, p. 1448–1454.

Mancl, Karen, and Slater, Brian, 2002, Suitability of Ohio soils for treating wastewater: Columbus, Ohio, The Ohio State University, Extension Bulletin 896, accessed 27 July 2005 at *http://ohioline.osu.edu/b896*.

Mara, D.D., and Oragui, J.I., 1983, Sorbitol-fermenting bifidobacteria as specific indicators of human faecal pollution: Journal of Applied Bacteriology, v. 55, p. 349–357.

Marteau, P., Pochart, P., Doré, J., Béra-Maillet, C., Bernalier, A., and Corthier, G., 2001, Comparative study of bacterial groups within the human cecal and fecal microbiota: Applied and Environmental Microbiology, v. 67, no. 10, p. 4939–4942.

Maynard Smith, J., Feil, E.J., and Smith, N.H., 2000, Population structure and evolutionary dynamics of pathogenic bacteria: BioEssays, v. 22, p. 1115–1122.

McLellan, S.L., Daniels, A.D., and Salmore, A.K., 2003, Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting: Applied and Environmental Microbiology, v. 69, no. 5, p. 2587–2594.

Miescier, J.J., and Cabelli, V.J., 1982, Enterococci and other microbial indicators in municipal wastewater effluents: Journal of the Water Pollution Control Federation, v. 54, p. 1599–1606.

Monpoeho, S., Dehee, A., Mignotte, B., Schwartzbrod, L., Marechal, V., Nicolas, J.-C., Billaudel, S., and Ferre, B., 2000, Quantification of enterovirus RNA in sludge samples using single tube real-time RT-PCR: BioTechniques, v. 29, no. 1, p. 88–93.

Myoda, S.P., Carson, C.A., Fuhrmann, J.J., Hahm, B.K., Hartel, P.G., Yampara Iquise, H., Johnson, L.A., Kuntz, R.L., Nakatsu, C.H., Sadowsky, M.J., and Samadpour, M., 2003, Comparison of genotypic-based microbial source tracking methods requiring a host origin database: Journal of Water and Health, v. 1, no. 4, p. 167–180.

Noble, R.T., Allen, S.M., Blackwood, A.D., Chu, W., Jiang, S.C., Lovelace, G.L., Sobsey, M.D., Stewart, J.R., and Wait, D.A., 2003, Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study: Journal of Water and Health, v. 1, no. 4, p. 195–207.

Noble, R.T., and Fuhrman, J.A., 2001, Enteroviruses detected by reverse transcriptase polymerase chain reaction from the coastal waters of Santa Monica Bay, California—Low correlation to bacterial indicator levels: Hydrobiologia, v. 460, p. 175–184.

Olive, D.M., and Bean, P., 1999, Principles and applications of methods for DNA-based typing of microbial organisms: Journal of Clinical Microbiology, v. 37, p. 1661–1669.

Olson, M.E., Goh, J., Phillips, M., Guselle, N., and McAllister, T.A., 1999, *Giardia* cyst and *Cryptosporidium* oocyst survival in water, soil, and cattle feces: Journal of Environmental Quality, v. 28, no. 6, p. 1991–1996.

Ong, C.S.L., Eisler, D.L., Alikhani, A., Fung, V.W.K., Tomblin, J., Bowie, W.R., and Isaac-Renton, J.L., 2002, Novel *Cryptosporidium* genotypes in sporadic cryptosporidiosis cases—First report of human infections with a cervine genotype: Emerging Infectious Diseases, v. 8, p. 263–268.

Oppegaard, H., Steinum, T.M., and Wasteson, Y., 2001, Horizontal transfer of a multi-drug resistance plasmid between coliform bacteria of human and bovine origin in a farm environment: Applied and Environmental Microbiology, v. 67, no. 8, p. 3732–3734.

Oshiro, R.K., and Olson, B.H., 1995, Occurrence of the STh toxin gene in animal feces, primary unchlorinated sewage and environmental water samples using PCR and hybridization techniques [abs.]: Abstracts of the General Meeting of the American Society for Microbiology, v. 95, p. 435.

Ott, E., Muller, T., Franz, C.M.A.P., Ulrich, A., Gabel, M., and Seyfarth, W., 2001, Population dynamics and antagonistic potential of enterococci colonizing the phyllosphere of grasses: Journal of Applied Microbiology, v. 91, p. 54–66.

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

Parveen, S., Murphree, R.L., Edmiston, L.E.E., Kaspar, C.W., Portier, K.M., and Tamplin, M.L., 1997, Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola Bay: Applied and Environmental Microbiology, v. 63, no. 7, p. 2607–2612.

Parveen, S., Portier, K.M., Robinson, K., Edmiston, L., and Tamplin, M., 1999, Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution: Applied and Environmental Microbiology, v. 65, no. 7, p. 3142–3147.

Persing, D.H., Smith, T.F., Tenover, F.C., and White, T.J., 1993, Diagnostic molecular microbiology—Principles and applications: Washington, D.C., American Society for Microbiology, 641 p.

Pickett, C.L., Weinstein, D.L., and Holmes, R.K., 1987, Genetics of type IIa heat-labile enterotoxin of *Escherichia coli*—Operon fusions, nucleotide sequence, and hybridization studies: Journal of Bacteriology, v. 168, p. 5180–5187.

Piña, S., Puig, M., Lucena, F., Jofre, J., and Girones, R., 1998, Viral pollution in the environment and in shellfish—Human adenovirus detection by PCR as an index of human viruses: Applied and Environmental Microbiology, v. 64, no. 9, p. 3376–3382.

Pourcher, A.M., Devriese, L.A., Hernandez, J.F., and Delattre, J.M., 1991, Enumeration by a miniaturized method of *Escherichia coli*, *Streptococcus bovis*, and enterococci as indicators of the origin of fecal pollution of waters: Journal of Applied Bacteriology, v. 70, no. 6, p. 525–530.

Puig, A., Queralt, N., Jofre, J., and Araujo, R., 1999, Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater: Applied and Environmental Microbiology, v. 65, no. 4, p. 1772–1776.

Rademaker, J., and de Bruijn, F.J., 1997, Characterization and classification of microbes by REP-PCR genomic fingerprinting and computer assisted pattern analysis, *in* Caetano-Anolles, G., and Gresshoff, P.M. eds. DNA markers—Protocols, applications, and overviews: New York, Wiley-Liss, p. 151–172.

Ram, J.L., Ritchie, R.P., Fang, J., Gonzales, F., and Selegean, J.P., 2004, Sequence-based source tracking of *Escherichia* coli based on genetic diversity of beta-glucuronidase: Journal of Environmental Quality, v. 33, p. 1024–1032.

Refsum, T., Heir, E., Kapperud, G., Vardund, T., and Holstad, G., 2002, Molecular epidemiology of *Salmonella enterica* serovar typhimurium isolates determined by pulsed-field gel electrophoresis—Comparison of isolates from avian wildlife, domestic animals, and the environment in Norway: Applied and Environmental Microbiology, v. 68, p. 5600–5606.

Reynolds, K.A., Roll, K., Fujioka, R.S., Gerba, C.P., and Pepper, I.L., 1998, Incidence of enteroviruses in Mamala Bay, Hawaii using cell culture and direct polymerase chain reaction methodologies: Canadian Journal of Microbiology, v. 44, no. 6, p. 598–604.

Rhodes, M.W., and Kator, H., 1997, Enumeration of *Enterococcus* sp. using a modified mE method: Journal of Applied Microbiology, v. 83, no. 1, p. 120–126.

Ritter, K.J., Carruthers, E., Carson, C.A., Ellender, R.D., Harwood, V.J., Kingsley, K., Nakatsu, C., Sadowsky, M., Shear, B., West, B., Whitlock, J.E., Wiggins, B.A., and Wilbur, J.D., 2003, Assessment of statistical methods used in library-based approaches to microbial source tracking: Journal of Water and Health, v. 1, no. 4, p. 209–223.

Rivera, S.C., Hazen, T.C., and Toranzos, G.A., 1988, Isolation of fecal coliforms from pristine sites in a tropical rain forest: Applied and Environmental Microbiology, v. 54, no. 2, p. 513–517.

Samadpour, M., and Chechowitz, N., 1995, Little Soos Creek microbial source tracking—A survey: Seattle, Wash., University of Washington, Department of Environmental Health, 49 p.

Satokari, R.M., Vaughan, E.E., Akkermans, A.L., Saarela, M., and de Vos, W.M., 2001, Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis: Applied and Environmental Microbiology, v. 67, no. 2, p. 504–513.

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

Savelkoul, P.H.M., Aarts, H.J.M., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J.L.W., Schouls, L., and Lenstra, J.A., 1999, Amplified-fragment length polymorphism analysis—The state of an art: Journal of Clinical Microbiology, v. 37, no. 10, p. 3083–3091.

Schaper, M., Duran, A.E., and Jofre, J., 2002, Comparative resistance of phage isolates of four genotypes of F-specific RNA bacteriophages to various inactivation processes: Applied and Environmental Microbiology, v. 68, no. 8, p. 3702–3707.

Schaper, M., Jofre, J., Uys, M., and Grabow, W.O.K., 2002, Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South Africa and Spain: Journal of Applied Microbiology, v. 92, p. 657–667.

Scott, T.M., Jenkins, T.M., Lukasik, J., and Rose, J.B., 2005, Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution: Environmental Science & Technology, v. 39, no. 1, p. 283–287.

Scott, T.M., Parveen, S., Portier, K.M., Rose, J.B., Tamplin, M.L., Farrah, S.R., Koo, A., and Lukasik, J., 2003, Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida: Applied and Environmental Microbiology, v. 69, no. 2, p. 1089–1092.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., and Lukasik, J., 2002, Microbial source tracking—Current methodology and future directions: Applied and Environmental Microbiology, v. 68, p. 5796–5803.

Sears, H.J., and Brownlee, I., 1952, Further observations on the persistence of individual strains of *Escherichia coli* in the intestinal tract of man: Journal of Bacteriology, v. 63, p. 47–57.

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S.D., 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: Environmental Microbiology, v. 7, no. 2, p. 249–259.

Simmons, G.M., Waye, D.F., Herbein, S., Myers, S., and Walker, E., 2000, Estimating nonpoint fecal coliform sources in Northern Virginia's Four Mile Run watershed, *in* Younos, T., and Poff, J., eds., Abstracts of the Virginia Water Research Symposium 2000: Blacksburg, Va., Virginia Water Resources Research Center Special Report SR–19, p. 248–267.

Simmons, O.D., III, and Sobsey, M.D., 2002, Survival of *Cryptosporidium* oocysts in environmental water samples [abs.]: Abstracts of the General Meeting of the American Society for Microbiology, v. 102, p. 427.

Simpson, J.M., Santo Domingo, J.W., and Reasoner, D.J., 2002, Microbial source tracking—State of the science: Environmental Science & Technology, v. 36, p. 5279–5288.

Sinton, L.W., Finlay, R.K., and Hannah, D.J., 1998, Distinguishing human from animal faecal contamination in water—A review: New Zealand Journal of Marine and Freshwater Research, v. 32, p. 323–348.

Sinton, L.W., Finlay, R.K., and Lynch, P.A., 1999, Sunlight inactivation of fecal bacteriophages and bacteria in sewage-polluted seawater: Applied and Environmental Microbiology, v. 65, no. 8, p. 3605–3613.

Smith-Keary, P., 1989, Molecular genetics of *Escherichia coli*: London, Guilford Press, 198 p.

Souza, V., Rocha, M., Valera, A., and Eguiarte, L.E., 1999, Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents: Applied and Environmental Microbiology, v. 65, no. 8, p. 3373–3385.

Souza, V., Travisano, M., Turner, P.E., and Eguiarte, L.E., 2002, Does experimental evolution reflect patterns in natural populations? *E. coli* strains from long-term studies compared with wild isolates: Antonie van Leeuwenhoek, v. 81, p. 143–153.

Stackebrandt, E., 2003, The richness of prokaryotic diversity—There must be a species somewhere: Food Technology and Biotechnology, v. 41, no. 1, p. 17–22.

Stern, M.J., Ames, G.F., Smith, N.H., Robinson, E.C., and Higgins, C.F., 1984, Repetitive extragenic palindromic sequences—A major component of the bacterial genome: Cell, v. 37, p. 1015–1026.

Stoeckel, D.M., Mathes, M.V., Hyer, K.E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T., Fenger, T.W., Samadpour, M., Strickler, K.M., and Wiggins, B.A., 2004, Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*: Environmental Science & Technology, v. 38, no. 22, p. 6109–6117.

Struelens, M.J., 1998, Molecular epidemiologic typing systems of bacterial pathogens—Current issues and perspectives: Memorias del Instituto Oswaldo Cruz, Rio de Janeiro, v. 93, p. 581–585.

Sturdee, A.P., Chalmers, R.M., and Bull, S.A., 1999, Detection of *Cryptosporidium* oocytes in wild mammals of mainland Britain: Veterinary parasitology, v. 80, no. 4, p. 273–280.

Tannock, G.W., 1995, Normal microflora—An introduction to microbes inhabiting the human body: London, Chapman and Hall, 115 p.

Tartera, C., Lucena, F., and Jofre, J., 1989, Human origin of *Bacteroides fragilis* bacteriophages present in the environment: Applied and Environmental Microbiology, v. 55, no. 10, p. 2696–2701.

Tenover, F.C., Arbeit, R.D., and Goering, R.V., 1997, How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections—A review for healthcare epidemiologists: Infection Control and Hospital Epidemiology, v. 18, p. 426–439.

Toivanen, P., Vaahtovuo, J., and Eerola, E., 2001, Influence of major histocompatibility complex on bacterial composition of fecal flora: Infection and Immunity, v. 69, p. 2372–2377.

Turbow, D.J., Osgood, N.D., and Jiang, S.C., 2003, Evaluation of recreational health risk in coastal waters based on enterococcus densities and bathing patterns: Environmental Health Perspectives, v. 111, no. 4, p. 598–603.

U.S. Congress, 2000, Beaches environmental assessment and coastal health (BEACH) act: U.S. Congress, 106th, Public Law 106-284, 114 STAT. 33 USC 1251 p. 870-877 accessed July 27, 2005 at *http://www.epa.gov/waterscience/beaches/bacteria-rule.htm#final*

U.S. Environmental Protection Agency, 1986, Ambient water quality criteria for bacteria—1986: Washington, D.C., Office of Water Regulations and Standards, EPA440/5–84–002, 18 p.

U.S. Environmental Protection Agency, 1996, Drinking water regulations and health advisories: Washington, D.C., EPA 822–B–96–002 [24 p].

U.S. Environmental Protection Agency, 2000a, Improved enumeration methods for the recreational water quality indicators—Enterococci and *Escherichia coli*: Washington, D.C., Office of Science and Technology, 27 p.

U.S. Environmental Protection Agency, 2000b, Atlas of America’s polluted waters: Washington, D.C., Office of Water (4503F), EPA 840–B00–002, 53 p.

U.S. Environmental Protection Agency, 2000c, National primary drinking water regulations—Ground Water Rule; proposed rules: Washington, D.C., U.S. Government Printing Office, 40 CFR parts 141 and 142, p. 30194.

U.S. Environmental Protection Agency, 2001, Protocol for developing pathogen TMDLs (1st ed.): Washington, D.C., Office of Water (4503F), EPA 841–R–00–002 [variously paginated].

U.S. Environmental Protection Agency, 2002, National beach guidance and required performance criteria for grants; Appendix G—Conducting a sanitary survey: Washington, D.C., Office of Water (4305 T), EPA–823–B–02–004, p. G1–G7.

U.S. Food and Drug Administration, 2003, National shellfish sanitation program guide for the control of molluscan shellfish: Washington, D.C., U.S. Government Printing Office, accessed July 27, 2005 at *http://www.cfsan.fda.gov/~ear/nss2-toc.html*

van Belkum, A., Struelens, M., de Visser, A., Verbrugh, H., and Tibayrenc, M., 2001, Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology: Clinical Microbiology Reviews, v. 14, p. 547–560.

Vaneechoutte, M., and Heyndrickx, M., 2001, Application and analysis of ARDRA patterns in bacterial identification, taxonomy, and phylogeny, *in* Dijkshoorn, L., Towner, K.J., and Struelens, M., eds. New approaches for the generation and analysis of microbial typing data: Amsterdam, Elsevier Science, p. 211–247.

Vinje, J., Oudejans, S.J.G., Stewart, J.R., Sobsey, M.D., and Long, S.C., 2004, Molecular detection and genotyping of male-specific coliphages by RT-PCR and reverse line blot hybridization: Applied and Environmental Microbiology, v. 70, no. 10, p. 5996–6004.

Wheeler, A.L., Hartel, P.G., Godfrey, D.G., Hill, J.L., and Segars, W.I., 2002, Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking: Journal of Environmental Quality, v. 31, p. 1286–1293.

Whitlock, J.E., Jones, D.T., and Harwood, V.J., 2002, Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Research, v. 36, no. 17, p. 4273–4282.

Whitman, R.L., and Nevers, M., 2003, Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach: Applied and Environmental Microbiology, v. 69, no. 9, p. 5555–5562.

Whitman, R.L., Shiveley, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N., 2003, Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan: Applied and Environmental Microbiology, v. 69, no. 8, p. 4714–4719.

Whittam, T.S., 1989, Clonal dynamics of *Escherichia coli* in its natural habitat: Antonie van Leeuwenhoek, v. 55, p. 23–32.

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: Applied and Environmental Microbiology, v. 62, no. 11, p. 3997–4002.

Wiggins, B.A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P., Eppard, J.R., Knupp, S.R., Limjoco, M.C., Mettenburg, J.M., Rinehardt, J.M., Sonsino, J., Torrijos, R.L., and Zimmerman, M.E., 1999, Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution: Applied and Environmental Microbiology, v. 65, no. 8, p. 3483–3486.

Wiggins, B.A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M., Han, J., Henry, B.L., Hoover, K.B., Johnson, E.L., Jones, K.C., McCarthy, J.G., McDonough, J.A., Mercer, S.A., Noto, M.J., Park, H., Phillips, M.S., Purner, S.M., Smith, B.M., Stevens, E.N., and Varner, A.K., 2003, Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries: Applied and Environmental Microbiology, v. 69, no. 6, p. 3399–3405.

Winfield, M.D., and Groisman, E.A., 2003, Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*: Applied and Environmental Microbiology, v. 69, no. 7, p. 3687–3694.

Wood, J., Scott, K.P., Avguštin, G., Newbold, C.J., and Flint, H.J., 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: Applied and Environmental Microbiology, v. 64, no. 10, p. 3683–3689.

Xiao, L., Singh, A., Limor, J., Graczyk, T.K., Gradus, S., and Lal, A., 2001, Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater: Applied and Environmental Microbiology, v. 67, no. 3, p. 1097–1101.

Yang, H.H., Vonopal, R.T., Grasso, D., and Smets, B.F., 2004, High diversity among environmental *Escherichia coli* isolates from a bovine feedlot: Applied and Environmental Microbiology, v. 70, no. 3, p. 1528–1536.

Yoder, J.S., Blackburn, B.G., Craun, G.F., Hill, V., Levy, D.A., Chen, N., Lee, S.H., Calderon, R.L., and Beach, M.J., 2004, Surveillance for waterborne-disease outbreaks associated with recreational water—United States, 2001–2002: MMWR Surveillance Summaries, v. 53, no. 8, p. 1–22.

Zelles, L., Bai, Q.Y., Beck, T., and Beese, F., 1992, Signature fatty acids in phospholipids and lipopolysaccharides as indicators of microbial biomass and community structure in agricultural soils: Soil Biology and Biochemistry, v. 24, no. 4, p. 317–323.

# 6. Glossary

The following are definitions of some terms used in this document. Definitions presented are not intended to be universally applicable—several terms are defined in a document-specific sense.

**Approach, MST approach** – The choice of library-dependent or library-independent strategies to characterize sources of fecal contamination. Within library-dependent MST, the choice of population-based classification or one-to-one matching to assign individual isolates to a host species or other source category. (*See also* library and associated definitions, empirical type matching, population biology-based matching.)

**Challenge isolates, challenge feces** – Quality-control samples composed of cultivated source identifiers or known-source feces. The source of the challenge isolates or feces is known to the compiler of the challenge but not to the analytical laboratory. Challenge materials are used to measure the internal accuracy of MST detection methods and can be used, for instance, to calculate the rate of correct classification for known-source libraries. Challenge isolates or feces differ from spikes in that they are in the same condition as the known-source library or reference material collection: the environmental matrix is excluded.

**Clone** – Individuals arising from a common lineage through asexual reproduction of mother cell to two daughter cells. Clones are genetically identical and, therefore, inseparable by various genetic typing methods (though they may demonstrate phenotypic differences depending on conditions).

**Cosmopolitan, cosmopolitan source identifier or type** – A source identifier that is found in individuals belonging to multiple host categories. Cosmopolitan source identifiers or types violate the MST assumption of host specificity.

**Detection method, MST detection method** – The laboratory methods used to detect or distinguish among source identifiers as part of a microbial source tracking protocol.

**Empirical type matching, one-to-one matching** – The strategy of assigning each isolate to a type, such that isolates of the same type appear the same by use of the typing method. Isolates of unknown source are typed and assigned to a source only if every isolate of the same type came from the same source. If the type is detected in multiple sources, the isolate of unknown origin is classified as cosmopolitan. (S*ee also* cosmopolitan.)

**Fecal indicator, regulatory fecal-indicator microorganism** Microorganisms that are typically found only in feces. Fecal indicators generally are not pathogenic themselves, but they indicate the potential presence of fecal pathogens. Regulatory fecal indicators are those that are written into legislation designed to establish acceptable levels of fecal contamination.

**Functional gene** – A gene found in a microorganism's genetic material that codes for some expressed characteristic mediating a host-microbe interaction in the preferred host but not in other hosts. An example functional gene used as a source identifier in microbial source tracking is the *esp* gene, encoding an adhesion factor that allows carrier strains of *Enterococcus faecium* to stick in the human intestine.

**Genotype, genotypic profile** – Characteristics that are specifically defined by the genetic sequence of the target organism, such as presence of a genetic marker or possession of identical DNA-fragment profiles.

**Host category, fecal source** – The category to which microbial source tracking protocols attempt to classify fecal contamination. Commonly used host classification schemes are by host species or by management category (human, wildlife, domestic livestock, or pets; human, wildlife, or domestic animals; human or nonhuman).

**Inhibitor, PCR inhibitor** – Chemical constituents, either from the environment (humic acids) or from the process of laboratory manipulation (ethanol) that stop the PCR from amplifying a source identifier genetic sequence and result in a false-negative result. PCR inhibitors are controlled for in laboratory analysis by adding a known PCR target (positive control) to an aliquot of the sample—if the known target is not amplified, then PCR inhibitors are present and a negative result for the source identifier does not mean that the source identifier is not present in that sample.

**Isolate** – A laboratory-generated culture of identical microorganisms that originated from a single cell or, in the case of conglomerates, from a colony- or plaque-forming unit.

**Library, known-source library** – An extensive collection of source identifiers (often bacterial isolates) used in library-dependent microbial source tracking approaches to represent fecal material from a source.

**Library dependent, library-dependent MST approach** MST that requires an extensive collection of reference source identifiers from the same time and geographic location as the test samples for effective classification of test isolates.

**Library independent, library-independent MST approach** MST that does not require further collection of reference fecal material after the host-specificity of the source identifier has been established. The library-independent approach does, however, necessitate quality-control reference fecal material to be evaluated to ensure that the source identifier is relevant to the time and geographic location of the study setting.

**Marker, host-associated or host-specific marker** – A single characteristic, generally genotypic but sometimes phenotypic, used to classify results. Examples are presence of human-specific *Bacteroidetes* sequence and the presence of human-associated serotypes of coliphage.

**Operational taxonomic unit** – A measure of genetic similarity that represents some pragmatic separation of isolates but does not necessarily have phylogenetic or taxonomic meaning. In one-to-one matching, an operational taxonomic unit includes all isolates that are genetically similar at or above the level of the typing method's precision.

**Phenotype, phenotypic profile** – Characteristics that are specifically defined by the physiological expression of traits, such as metabolism (carbon utilization) and resistance to antibiotics.

**Polymerase chain reaction, PCR** – The molecular-biology method that allows amplification (reproduction) of very specific sequences of genetic material. Use of the PCR allows detection of genes or other genetic characteristics that would otherwise be undetectable because of low concentrations. The PCR is used to amplify target DNA to the point where it can be detected by other laboratory techniques.

**Population-biology-based matching** – A MST approach based on the hypothesis that competition in hosts with various metabolisms, basal temperatures, and gut physiologies will select for a subset of microorganisms that share some characteristics. A common example of this hypothesis is antibiotic resistance, for which only those animals that have been exposed to antibiotics (humans and domestic animals) are expected to shed gut bacteria that are resistant to particular antibiotics.

**Primary habitat** – The host gut; the major environment to which the source identifier is adapted to live and reproduce.

**Protocol, MST protocol** – A combination of source identifier, detection method, and analytical approach chosen to accomplish microbial source tracking.

**Reference material, known-source reference feces** – Feces, either fresh (primary habitat) or treated (secondary habitat), for which the host of origin is known. Reference material is used to build known-source libraries, to validate library-independent methods, and to estimate classification error in MST.

**Resident, resident source identifier or type** – A source identifier that is consistently found in an individual over time.

**Sanitary survey** – The critical first step in a MST evaluation, in which the contributing area to a contaminated water body is inspected for sources of the contamination. A thorough sanitary survey is necessary to determine which sources need to be included or can be excluded from a MST evaluation. In some cases, the results of the sanitary survey may be convincing enough that the MST evaluation does not need to be done.

**Secondary habitat** – The environment outside of the host gut; the habitat which the source identifier must survive in order to be reintroduced to a primary habitat. The secondary habitat can be either a source of fecal contamination (such as a septic field or waste lagoon) or the resource water that is contaminated. (*See also* tertiary habitat.)

**Source identifier** – Microbiological marker used as part of a microbial source tracking protocol to discriminate feces from various sources.

**Speciation** – Classification of microorganisms, particularly bacteria, to species as a microbial source tracking detection method. Some species (such as *Enterococcus gallinarum* and *Rhodococcus coprophilus*) were named for the habitats in which they were commonly found.

**Spike sample, challenge spike** – A sample submitted to the analytical laboratory in which feces or isolates from known sources are mixed into an environmental matrix (such as autoclaved river water). The source of the feces or isolates is not made known to the analytical laboratory. Spike samples are used to estimate the accuracy of microbial source tracking protocols to classify isolates to host categories after passage from the primary habitat (source environment) to the secondary habitat (resource water).

**Tertiary habitat** – A term sometimes used in conjunction with primary and secondary habitat. When the primary habitat is the host gut and the secondary habitat is a waste-control facility, then the resource water which is contaminated is sometimes referred to as the tertiary habitat to distinguish it from the secondary-habitat source of fecal contamination.

**Total Maximum Daily Load** – A provision of the Clean Water Act under which non-point sources of contamination to a water body must be identified and controlled if the water body fails to meet regulatory standards even after all point sources of contamination have been controlled.

**Transient, transient source identifier or type** – A source identifier that is not consistently found in an individual over time.

# Appendix A: Source Identifiers

## A.1 Regulatory Source Identifiers

### A.1.1 *Escherichia coli*

*E. coli* is a member of the normal intestinal flora for most warmblooded animals (Gordon and Cowling, 2003). These bacteria constitute approximately 0.1 percent of total bacteria within an adult's intestines (on a western diet; Tannock, 1995). Detection of *E. coli* is a confirmatory step in the total coliform rule for ground water, and criteria for surface water designated use attainment in many states are based, in whole or in part, on *E. coli* concentrations. The USEPA required that all states initiate *E. coli* or enterococci-concentration-based monitoring for recreational waters by April 2004 (U.S. Congress, 2000). Though they have been recommended by the USEPA since 1984 for recreational water-quality assessment (for freshwaters), *E. coli* are bacteria; therefore, they may not adequately represent transport and survival characteristics of protozoan and viral pathogens.

*E. coli* are gram-negative, facultatively anaerobic bacteria that ferment lactose and grow at temperatures up to 44.5˚C. They are commonly detected by colony growth on the media mTEC, modified mTEC, MI, or the combination of mENDO and NA-MUG agars after membrane filtration. *E. coli* also can be detected in a most-probable-number format by use of, for instance, the commercial Colilert and Colitag systems. Immunological and PCR-based tests also have been proposed for *E. coli* detection and quantification. Confirmation of *E. coli* is typically achieved through biochemical tests or enzymatic assays. In clinical applications, the most commonly used confirmatory tests are collectively known as IMViC—indole production, methyl red reaction, Voges-Proskauer test, and failure to grow on citrate-minimal media. In research labs, it is common to use the Analytical Profile Index (API) biotyping system (22 biochemical tests) or Biolog GN plates (95 sole-source substrates) or simply to test for MUG (4-methylumbelliferyl-ß-D-glucuronide) hydrolysis (a test for ß-glucuronidase activity). For recreational waters, tests for indole production, gas formation on lactose, failure to express urease, failure to express oxidase, and failure to grow on citrate-minimal media are the recommended confirmation tests (U.S. Environmental Protection Agency, 2000a). There is a need for standardization of *E. coli* detection and confirmation methods among MST researchers because each of the listed media and confirmation tests excludes some portion of the entire *E. coli* population.

### A.1.2 Enterococci

Like *E. coli*, enterococci are members of the normal intestinal flora for most, if not all, warmblooded animals. The enterococci are non motile, gram-positive cocci that are found in mammals, birds (Wheeler and others, 2002), plants (Ott and others, 2001), fish, and insects (Kanoe and Abe, 1988). They predominate in the gastrointestinal tract and occur less commonly in the genitourinary tract and the oral cavity of animals. Enterococci survive longer in marine environments (Cabelli and others, 1982) and through water treatment processes (Miescier and Cabelli, 1982) than coliforms do. Numbers in recreational waters correlate with the risk of gastrointestinal illness better than *E. coli* (Cabelli and others, 1982; Turbow and others, 2003). Though they have been recommended by the USEPA since 1984 for recreational water-quality assessment (for both fresh and marine waters) enteroccocci are bacteria; therefore, they may not adequately represent transport and survival characteristics of protozoan and viral pathogens.

Enterococci are differentiated from other fecal streptococci by their ability to grow in 6.5 percent NaCl, at pH 9.6, and at 10˚C and 45˚C; they fail to express catalase, and they create a black precipitate on esculin iron agar (EIA agar; U.S. Environmental Protection Agency, 2000a) or exhibit ß-D glucosidase activity (mEI agar; Rhodes and Kator, 1997; U.S. Environmental Protection Agency, 2000a). The enterococci were previously classified as Group D streptococci,

because they contain a group-D-specific cell wall carbohydrate (glycerol teichoic acid) linked to the cytoplasmic membrane. The group includes as members the species *Enterococcus faecalis*, *E. faecium*, *E. avium*, *E. gallinarum*, and *E. durans.*

Enterococci are commonly detected by colony growth after membrane filtration using mEI agar or the combination of mE and EIA agars. Enterococci also can be detected in a most-probable-number format by use of, for example, the commercial Enterolert system. Standard maintenance growth conditions include incubation on brain heart infusion (BHI) or similar broth or agar at 35–37˚C without aeration (U.S. Environmental Protection Agency, 2000a).

### A.1.3 Coliphage

Coliphage are bacteriophages (viruses) that infect *E. coli* and, to a lesser extent, other coliform bacteria (Furuse, 1987). Because their presence is directly related to the presence of *E. coli* (Borrego and others, 1987), coliphage have been proposed as superior fecal-indicator microorganisms. Coliphage are generally divided into somatic and F-specific types, based on their site of infection on the *E. coli* cell. F-specific RNA coliphage serve as model indicators of enteric viruses in freshwater (Havelaar, 1993), an area where *E. coli* is deficient as a model organism (Borchardt and others, 2003). Coliphage may be promulgated formally as regulatory fecal indicators for ground water in the United States under the proposed Ground Water Rule (U.S. Environmental Protection Agency 2000c).

Coliphage is an operational taxonomic unit that includes distinct bacteriophage of multiple taxa. As noted above, the taxa correspond to somatic and F-specific coliphage. Somatic coliphage (at least four families, all DNA based) use a lipopolysaccharide on the *E. coli* cell outer surface as a recognition factor to initiate infection. F-specific coliphage use antigens on the F-pilus (used for sexual exchange of genetic material between *E. coli* cells) to initiate infection. F-specific coliphage include those that use DNA as their genetic material (Inoviridae) and those that use single-stranded RNA as their genetic material (Leviveridae; Vinje and others, 2004). F-specific RNA-based coliphage have been subdivided into four groups based on serotyping (Furuse, 1987) and genotyping (Beekwilder and others, 1996; Hsu and others, 1995; Vinje and others, 2004). Groups I and IV are most commonly found in nonhuman hosts, and groups II and III are most commonly found in human hosts (Cole and others, 2003).

Coliphage are detected by the presence of plaques (areas of no growth) on a lawn of *E. coli* host bacteria. Different host strains of *E. coli* are used for somatic and F-specific coliphage. Detection methods typically follow USEPA method 1601 (presence-absence) or method 1602 (quantification). In method 1601, host *E. coli* are added to a water sample to allow coliphage reproduction, and the resulting enrichment is spotted onto a plate previously inoculated with the same host *E. coli*. Formation of a plaque around the spot indicates the presence of coliphage in the sample. In method 1602, the water sample is combined with host *E. coli* and molten nutrient agar. During incubation, a plaque forms in the *E. coli* lawn where an infectious coliphage particle was embedded in the agar gel, leading to a count of plaque-forming units.

### A.1.4 Fecal Coliforms

Fecal coliform is an operational taxonomic unit that includes *E. coli*. Many states historically used fecal coliform concentrations as their primary indicator of bacterial water quality. As "thermotolerant coliforms," the fecal coliform group was established as an alternative to total coliforms to detect organisms exclusively of fecal origin. The group includes species other than *E. coli*, including some that have environmental origins and some with a potential for regrowth in the environment. In a landmark epidemiology study, USEPA (Dufour, 1984) did not find a good relation between health risk and fecal coliform concentrations; therefore, since 1986, USEPA has recommended the use of other indicator organisms to monitor for bacteriological water quality (U.S. Environmental Protection Agency, 1986). Fecal coliforms are all gram-negative, facultatively anaerobic bacteria that ferment lactose with gas production and grow at temperatures up to 44.5˚C. They are commonly detected by colony growth after membrane filtration using mFC agar.

### A.1.5 Total Coliforms

Total coliform is an operational taxonomic unit that includes *E. coli* and fecal coliforms. Total coliform concentrations have been used to monitor drinking-water sources for almost 100 years because they are seen as a more conservative indication of fecal contamination than *E. coli* or fecal coliforms. The total coliform group encompasses 19 genera (including *Escherichia, Klebsiella, Enterobacter*, and *Citrobacter*) and 80 species (Leclerc and others, 2001). Coliforms reside in the intestinal tract of humans and animals, but also occur naturally in soil, surface water, fish, mollusks, leafy plants, insects, and rodents. This is one reason that total coliforms are generally considered poor indicators of fecal contamination and health risk (Borrego and others, 1987). Total coliforms are used in the Ground Water Rule as an initial indication of fecal contamination (U.S. Environmental Protection Agency, 2000c). Total coliforms are gram-negative, aerobic or facultatively anaerobic, nonsporulating bacteria that ferment lactose at 35˚C within 48 hours. Total coliforms are commonly detected by colony growth after membrane filtration using mENDO or MI agar. Total coliforms also can be detected in a most-probable-number format by use of, for example, the commercial Colilert system.

## A.2 Nonregulatory Source Identifiers

### A.2.1 Ratios of Microbes

One of the earliest approaches to MST involved use of the ratio of fecal coliforms to fecal streptococci (Geldreich and Kenner, 1969). Use of these ratios eventually was deemed unreliable because of the variable survival rates of the bacterial species involved (Clesceri and others, 1998). Refer to section 4.4 for further details.

### A.2.2 Host-Specific Bacteria

Several source identifier bacterial species currently are used for MST. *Enterococcus faecalis* may have a host range limited to humans and a few other animals (dogs and chickens; Pourcher and others, 1991). Wheeler and others (2002) concluded that combining speciation of *Ent. faecalis* with ribotyping was a useful approach for identifying human-source contamination. *Rhodococcus coprophilus* is found primarily in wheat straw and, therefore, is most commonly detected in the feces of grazing animals such as cattle (Sinton and others, 1998). Two *Bifidobacterium* species, *B. adolescentis* and *B. dentium,* have a narrow host range that is dominated by humans (Bonjoch and others, 2004).

### A.2.3 Host-Specific Viruses

Adenoviruses have been proposed as an index of human viruses (Jiang and others, 2001; Piña and others, 1998). Enteroviruses have been found in activated sludge, sewage outfalls, and fresh and marine waters associated with human fecal contamination (Kopecka and others, 1993; Reynolds and others, 1998; Griffin and others, 1999; Noble and Fuhrman, 2001; Jiang and others, 2001). Both adenoviruses and enteroviruses have been evaluated and recommended for MST; however, it appears that they are limited to identifying human contamination originating from populations, rather than from individuals (Noble and others, 2003).

Bovine enteroviruses (Ley and others, 2002) and bovine and porcine adenoviruses (Maluquer de Motes and others, 2004) also have been proposed to detect animal-source fecal contamination. Similarly, teschoviruses have been used as an indicator of porcine fecal contamination (Jiménez-Clavero and others, 2003).

### A.2.4 Host-Specific Genetic Markers

Genetic markers are sometimes used to detect host-specific microorganisms or types. These markers are often genes or other fragments of DNA that can be multiplied from within a mix with other DNA by the polymerase chain reaction. Once the marker has been multiplied, it can be detected by gel electrophoresis or with labeled probes. Many references describe these and other molecular techniques. The reader is directed to Persing and others (1993) as one such reference.

#### A.2.4.1 DNA Coding16S Ribosome of *Bacteroides/Prevotella* Group Anaerobes

Source-specific *Bacteroidetes* sequences have been identified by means of terminal restriction fragment length polymorphism (t-RFLP), length heterogeneity PCR, or subtractive hybridization, and used to design PCR primers for MST (Bernhard and Field, 2000a,b). The length heterogeneity PCR primer pairs have been field tested (Bernhard and others, 2003) and they produced relatively accurate results for identification of human and ruminant contamination in a methods comparison study (Field and others, 2003). Primer pairs for additional host species are under development (Dick and others, 2002).

#### A.2.4.2 Toxin Genes of Pathogenic *Escherichia coli*

PCR assays have been developed that target the heat-stable enterotoxin STIb to identify human fecal contamination (Oshiro and Olson, 1995), the heat-labile enterotoxin LTIIa to identify cattle fecal contamination (Khatib and others, 2002), and the heat-stable enterotoxin STII to identify swine fecal contamination (Khatib and others, 2003).

# Appendix B: Detection Methods

## B.1 Conventional Typing Methods

### B.1.1 Speciation

#### Method

Microbiologists first divided bacteria into species on the basis of phenotypic information such as cell morphology and physiology. Suites of biochemical tests that facilitate classification of bacteria to species are described in “Bergey’s Manual of Systematic Bacteriology” (Atlas, 1988). Classical approaches to bacterial systematics have been challenged by the advent of phylogenetics, which reveals an evolutionary basis for species definition (Stackebrandt, 2003). Still, the guiding principle for species definition is 70 percent DNA-DNA homology, a threshold that leaves considerable room for variability within species (Maynard Smith and others, 2000).

Viruses, on the other hand, were referred to until recently simply by their common names (for example, poliovirus). Formal classification is now based on nucleic acid type, particle size, shape and structure, and mode of replication (Atlas, 1988). Classification of protozoa follows more closely the model of macrobiology: as eukaryotes, protozoa fit within the traditional definition of species.

#### Information Content

Speciation has a long history of use in classification. Host specificity has influenced speciation with effects ranging from incidental in the case of commensals to integral in the case of some pathogens.

Speciation is significant to the many forms of MST that start with a basic taxonomic unit (often the species) as a means of separating out extraneous information. Reports indicate that *Bacteroides fragilis* bacteriophage (Tartera and others, 1989), at least one species within *Enterococcus* (Wheeler and others, 2002), and *Cryptosporidium* species (Ong and others, 2002) are specific to host sources.

## B.1.2 Serotyping

### Method

Serotyping was among the first subtyping methods developed for identification of pathogenic strains. Serotyping has been applied extensively to organisms that are now used for source tracking, including *E. coli*. In serotype analysis, antibodies to cell-surface antigens are added to a suspension. If the antibody matches the target antigen, the suspension agglutinates (clumps), and the reaction is positive. For many gram-negative bacteria, series of antibodies have been defined against cell-surface lipopolysaccharides (O antigens) and flagellar proteins (H antigens). These results often are reported in conjunction, as in *E. coli* O157:H7. In addition to the O and H antigens, capsular (K antigens) sometimes are used in serotyping. The method for serotyping coliphage differs in mechanism but not principle (Hsu and others, 1995).

### Information Content

The information gained from serotyping is limited by the finite number of tested antigens and the reactions expressed by that isolate against the suite of antigen tests. Expression of cell-surface antigens is a stable characteristic. The O:H scheme of serotyping *E. coli* includes 170 O antigens and 55 H antigens, with more than 9,000 possible serotypes (Bettelheim and Thomas, 2004). Not all strains can be serotyped.

Surface antigens are an important component of host-microbe interactions. As such, there may be a functional link between surface antigen expression and adaptation to succeed in a host (Toivanen and others, 2001). Enterotoxigenic *E. coli* associated with specific hosts have been defined by serotype (O:H) or serogroup (O only) conventions (Bettelheim and Thomas, 2004). Early studies of the stability of *E. coli* lineages in the human gastrointestinal tract also made use of serotyping (Sears and Brownlee, 1952). See the review by Scott and others (2002) and the investigation by Parveen and others (2001) for descriptions of bacterial serotyping as used in source tracking investigations. Serotypes of the viral fecal-indicator $F^+$ RNA coliphage also have been evaluated for use in source tracking (Furuse, 1987).

## B.1.3 Multilocus Enzyme Electrophoresis

### Method

Multilocus enzyme electrophoresis (MLEE) has been used extensively in ecological studies to track populations of organisms. To perform MLEE, pure cultures of organisms are grown out, and enzymes are separated by electrophoretic mobility. Isolates demonstrating different enzyme mobilities are termed electrophoretic types (ETs).

### Information Content

The enzymes used in MLEE are generally housekeeping enzymes, those essential for cell function. Substitutions of amino acids in the primary enzyme structure may not affect enzyme function but can result in altered electrophoretic mobility. Because primary enzyme structure is directly related to genetic information, MLEE data often are reported as representing genotypic rather than phenotypic information.

MLEE may be useful for MST using either the 1:1 matching strategy or the population genetics strategy because there is a known basis for similarity among ETs. MLEE data have

been used to evaluate *E. coli* in studies related to source tracking (Caugant and others, 1984; Gordon, 2001; Gordon and others, 2002; and Souza and others, 1999)

## B.2 Phenotypic Methods

### B.2.1 Carbon Utilization Profiles

#### Method

To generate carbon utilization profiles, pure cultures of a strain are inoculated into test solutions containing a single carbon source. Growth of the test strain causes some observable change in the medium, which is scored as presence or absence of substrate utilization.

#### Information content

Presence or absence of carbon utilization ability has been criticized as an unstable or variable characteristic in bacterial strains (Tenover and others, 1997). Commercial substrate suites were developed to discriminate among species but may not have ecological significance, hindering data interpretation. Carbon utilization profiles using pure cultures of indicator organisms have been proposed as a viable approach for MST (Hagedorn and others, 2003).

### B.2.2 Membrane Fatty Acids

#### Method

The membranes of bacteria contain phospholipids that are crucial factors controlling cell permeability. To analyze the phospholipids, cell cultures are first saponified, then resultant fatty acids are extracted into an organic phase. Free fatty acids are then methylated, purified, and concentrated. Concentrated extracts are analyzed by gas chromatography. This variant of phospholipid fatty acid analysis is called gas chromatography fatty acid methyl ester (GC-FAME) analysis.

#### Information Content

Phospholipid composition information has been applied to the evaluation of microbial community composition in agricultural soils and other settings. Signature fatty acids have been proposed that indicate the presence of specific groups of microorganisms (Zelles and others, 1992) on the basis of fatty acid function in the cell membrane.

Parveen and others (2001) included GC-FAME analysis in their continuing characterization of a collection of *E. coli* used for source tracking and found no relation between FAME profile and isolate source. No other published reports have been found in the literature.

### B.2.3 Antibiotic Resistance

#### Method

Bacteria routinely are characterized by their sensitivity or resistance to antibiotics in clinical use. Many permutations of clinical antibiotic resistance methods have been developed in research laboratories. In clinical use, two standardized methods are the Kirby-Bauer test and multi-antibiotic resistance profiling using the National Committee for Clinical Laboratory Standards (NCCLS) to define clinical resistance.

The Kirby-Bauer test is done by placing a panel of paper disks impregnated with different antibiotics on the surface of a plate seeded with a test strain. Results are recorded as diameter of

growth inhibition after incubation for a standardized amount of time at a specified temperature. The diameter of growth inhibition is related to the degree of antibiotic sensitivity because of a diffusion-induced concentration gradient of antibiotic in the agar.

Clinical profiling for multiple antibiotic resistance is done by use of Mueller-Hinton agar fortified with known concentrations of chosen antibiotics. Each concentration of each antibiotic in the test panel is represented in a separate agar plate. Standardized inocula of test isolates are placed on each plate, usually with a replica plater.

#### Information Content

Information from antibiotic-resistance-based methods has logical value when source categories are exposed to different antibiotics. An example might be where differentiation was needed between feces from a population of cattle, which had been exposed to one antibiotic, and a population of swine, which had been exposed to a different antibiotic.

Many researchers have reported the ability of experimental antibiotic resistance methods to separate isolates obtained from animal sources (Graves and others, 2002; Hagedorn and others, 1999; Harwood and others, 2000; Parveen and others, 1997; Wiggins and others, 1999). Two experimental variants on antibiotic sensitivity testing have been used extensively in pursuit of source tracking. Multiple antibiotic resistance analysis (MAR; Parveen and others, 1997) tests each isolate against a single concentration of each antibiotic in the panel. The antibiotics and concentrations in the panel were developed to reflect antibiotic use in humans and domestic animals. Antibiotic resistance analysis (ARA; Wiggins, 1996) challenges each isolate against multiple concentrations of antibiotics, the concentrations of which are chosen on the basis of ability to separate isolates from different hosts rather than the NCCLS standards. Information generated by the Kirby-Bauer method has been demonstrated in one study to be ineffective for MST using *E. coli* (Hagedorn, 1999); however, its use continues to be investigated (J. Mott, Texas A&M University, Corpus Christi, Tex., oral commun., November 2004). A modification of ARA, in which multiple concentrations of antibiotic are used to calculate an antibiotic resistance index representative of a fecal-source population, also has been used for MST (D.S. Francy, U.S. Geological Survey, Columbus, Ohio, oral commun., December 2004).

## B.3 Genotypic Methods

Genotypic methods are represented by multiple approaches in three major categories: evaluation of similarity based on the whole genome, evaluation of similarity based on genes present in all target organisms, and direct evaluation of source based on presence or absence of a DNA sequence. Within these three major categories exist a wide array of sometimes overlapping protocols. Where possible, each is discussed separately.

Many of the genotypic methods were developed in the context of epidemiology or systematics. Several reviews present in-depth discussion of these methods, including those by Louws and others (1999), Olive and Bean (1999), Struelens (1998), Tenover and others (1997), and van Belkum and others (2001).

### B.3.1 Similarity Based on Whole Genome Analysis

#### B.3.1.1 Macrorestriction Digest Pulsed-Field Gel Electrophoresis (PFGE)

##### Method

In PFGE, extracted chromosomal DNA is purified and digested by an infrequently cutting restriction enzyme, resulting in some 10 to 30 large fragments. The sum of all fragment lengths can be used to estimate the size of the chromosome. Fragments are separated by PFGE in a specialized rig. Banding patterns are visualized by staining, and fragment sizes are estimated by comparison with a ladder of known molecular weight DNA standards.

### Information Content

PFGE is the standard against which other methods are compared in many epidemiological investigations. The method is described as "superior to most other methods for biochemical and molecular typing" (Olive and Bean, 1999). Interpretation of similarity is possible because each change in banding pattern can be related to a genetic event.

Parveen and others (2001) evaluated PFGE data by means of discriminant analysis and found no relation between fragment pattern and source when the restriction enzyme *Sfi*I was used to digest. Simmons and others (2000), however, reported success when applying PFGE typing to MST using 1:1 association of unknown patterns with source patterns. Current usage is to restrict with *Not*I following Simmons and others (2000) or *Xba*I following the CDC PulseNet protocol for PFGE (Centers for Disease Control and Prevention, 1996).

## B.3.1.2 Amplified Fragment Length Polymorphism Analysis (AFLP)

### Method

AFLP is a partial-genome analogue to PFGE. Whole chromosomal DNA is digested by both a frequently cutting and an infrequently cutting restriction enzyme. An adapter is ligated onto the sticky end left by the infrequently cutting restriction enzyme, and PCR is done using a primer complementary to the adapter sequence. If fewer bands are desired, the primer can be designed to overhang the cut site by a single base, effectively reducing amplification to one in four possible products. PCR products are separated by agarose gel electrophoresis and visualized by staining (Savelkoul and others, 1999).

### Information Content

The information generated by AFLP may represent the whole-genome similarity of two strains but may not be indicative of phylogenetic relatedness. Use of this method has not been frequently reported in the source-tracking literature. One report on AFLP utility for MST studies (Guan and others, 2002) is available in the literature, and further work is ongoing.

## B.3.1.3 Repetitive Element PCR (Rep-PCR)

### Method

Repetitive elements have been observed and described in many bacteria. Some of these repeated elements flank transposable elements (insertion sequences; Smith-Keary, 1989), and others are present at high frequency but have unknown function (Stern and others, 1984). The rep-PCR method uses primers for these repetitive sequences to generate PCR products representing DNA between proximal, facing repetitive sequences. Besides the insertion sequences, primers for three conserved motifs in bacterial repetitive elements are used for rep-PCR; namely, REP (repetitive extragenic palindromic sequence), ERIC (enterobacterial repeating intragenic consensus sequence), and BOX (named after the boxes drawn around candidate sequences in early research).

### Information Content

Bands generated by rep-PCR represent cryptic DNA fragments—the only thing that is known is that each band has the repetitive element at each end. Changes in banding patterns could be caused by a variety of genetic events, hindering interpretation of band pattern similarity. DNA amplification products represent a small fraction of the genome.

Variants of the rep-PCR method should be applicable to any microorganism having repeated sequences in its genome. In the past, the methods have been used to type plant pathogens (Rademaker and de Bruijn, 1997). Olive and Bean (1999) reported that rep-PCR was quickly becom-

ing a widely used typing protocol because of its discriminatory power, ease of application, and low cost.

Dombek and others (2000) reported that rep-PCR using the BOXA1R primer had utility for source tracking. Carson and others (2003) reported performance of rep-PCR using the BOXA1R primer that was equal or superior to the performance of ribotyping by use of the *Hin*dIII restriction enzyme.

## B.3.2 Similarity Based on Analysis of Specific Genes

Methods similar to the whole-genome-based methods are sometimes used with defined genes in an attempt to generate more readily interpreted similarity data. The genes chosen tend to have functional relevance, such as ribosomal RNA operons and toxin genes.

### B.3.2.1 Ribotyping and Other Restriction Fragment Length Polymorphism (RFLP)-Type Methods

#### Method

The ribotyping procedure begins by extracting and purifying whole chromosomal DNA from a pure culture. Purified DNA is digested by a frequently cutting restriction enzyme. Digested DNA fragments are separated by agarose gel electrophoresis. If the DNA is visualized at this stage, it appears as a smear of DNA fragments on the gel. Digested, separated DNA fragments are blotted from the gel onto a nylon membrane (the Southern blot procedure). The DNA is fixed in space on the nylon membrane by UV cross-linking. Fragments containing DNA encoding 5S, 16S, and/or 23S rRNA are detected by hybridization with radiolabeled universal rDNA probes or digoxigenin-labeled probes. Ribotype exposures can be inspected visually or scanned into image analysis software (Olive and Bean, 1999).

Other variants of the RFLP protocol are accomplished similarly with probes for other genes of interest. For more information about alternate RFLP protocols, refer to Olive and Bean (1999) and Tenover and others (1997).

#### Information Content

Some of the information generated by ribotyping is based on differences within the rRNA operon and is therefore conserved in related strains. The terminal fragments, however, extend outside of the operon into the nonconserved region of the genome. As such, similarity of ribotypes does not necessarily relate directly to whole-genome or phylogenetic similarity.

Ribotyping protocols using *E. coli* are among the most common means of accomplishing source tracking (Carson and others, 2001; Jenkins and others, 2003; Parveen and others, 1999; Samadpour and Chechowitz, 1995; Scott and others, 2003). In past MST studies, the enzyme used generally has been *Hin*dIII for the single digest technique (for example, Parveen and others, 1999) or *Eco*RI and *Pvu*II for the double digest technique (for example, Samadpour and Chechowitz, 1995).

### B.3.2.2 Amplified Ribosomal DNA Restriction Analysis (ARDRA) and Other PCR-RFLP Methods

#### Method

In PCR-RFLP methods, variation in a specific gene is detected by gene amplification using PCR followed by restriction of the amplicon with a frequently-cutting restriction enzyme. Minor variations in the sequences of functional genes can be detected by PCR-RFLP. The ARDRA analysis is the variant of PCR-RFLP based on amplification of the rRNA operon. (Olive and Bean, 1999; Vaneechoutte and Heyndrickx, 2001).

#### Information Content

Interpretation of differences in the amplified gene is facilitated by the presumed or known function of the gene sequence evaluated. Though theoretically suitable, use of ARDRA and other PCR-RFLP methods is not commonly reported in the MST literature.

### B.3.2.3 Terminal Restriction Fragment Length Polymorphism (t-RFLP)

#### Method

The t-RFLP method can be thought of as a further specialized application of PCR-RFLP. As in PCR-RFLP, a gene sequence of presumed or known function is amplified by PCR. In this case, however, one primer is labeled such that only a terminal restriction fragment is visualized, resulting in only one bit of information for each gene copy within each strain evaluated. The terminal fragment can be sequenced, or the pattern of terminal fragments can be observed by capillary electrophoresis in the sequencer. This method more often is used to characterize communities of microorganisms rather than individual strains (Bernhard and Field, 2000a).

#### Information Content

Data are generated by t-RFLP without the need for cultivation of isolates; therefore, this method has found utility for community comparisons. Fragment size information generated by t-RFLP does not have functional or phylogenetic significance, though sequencing information may have significance.

Bernhard and Field (2000b) sequenced signature t-RFLP peaks prevalent in host sources to enable their design of host-specific PCR primers against *Bacteroides-Prevotella*. Ongoing research by others in the field investigates the use of *Bacteroides-Prevotella* marker profiles to indicate the origin of fecal contamination (Seurinck and others 2005).

## B.3.3 Typing by Marker Sequence Presence or Absence

### B.3.3.1 Coliphage Genotyping and Other Dot Blot Analyses

#### Method

To genotype coliphage, RNA is fixed to a nylon membrane and challenged against each of four labeled probes. The probes could be labeled either radioactively or with digoxigenin for chemiluminescent detection. Successful hybridization with one of the four probes indicates the type of the coliphage (Hsu and others, 1995). Comparable probes could be developed for other applications. A reverse line blot hybridization method, in which the probes are fixed to the nylon membrane and tested against each isolate, uses essentially the same principle and promises to have higher throughput than traditional dot blot hybridization (Vinje and others, 2004)

#### Information Content

A great deal of prior knowledge is required to ensure that hybridization with a probe adequately represents the type. In the example of coliphage, types were defined on the basis of serological reactions, and the probes developed are surrogates for that information. Genotyping coliphages was reported to be a useful method of categorizing coliphage into the groups used for source tracking (Calci and others, 1998; Schaper, Duran, and Jofre, 2002; Schaper, Jofre, and others, 2002).

### B.3.3.2 Length-Heterogeneity PCR (LH-PCR) for *Escherichia coli* Toxin Genes and *Bacteroides-Prevotella* Group Markers

#### Method

LH-PCR is a method whereby a sequence of interest is amplified by PCR and resolved by agarose gel electrophoresis. Prior knowledge about specificity of the primer and expected amplicon size is required for interpretation of amplified product (Simpson and others, 2002).

#### Information content

Production of the desired product using a specific primer or primer set indicates the presence of the target sequence. Confirmation can be done by sequencing or probing the product by dot blot analysis.

This method should be applicable to a wide range of source identifiers provided that sufficient prior knowledge of products is available. LH-PCR provides an alternative to dot blotting to detect specific gene sequences from mixed DNA. Primers have been developed to detect *Bacteroides-Prevotella* types associated with different hosts (Bernhard and Field, 2000b) and to detect the presence of host-specific toxin genes in *E. coli* (Khatib and others, 2002).

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| ---------- Virus-based source identifiers ---------- | | | | | |
| Adenoviruses | Porcine and bovine adenovirus not detected in 12 human sewage samples or pooled samples (3–5 individuals) from other alternate hosts (Maluquer de Motes and others, 2004).<br><br>Human adenovirus detected in 16 of 18 treated and untreated human sewage samples and 4 of 17 slaughterhouse sewage samples (3 pig and 1 chicken; Piña and others, 1998).<br><br>Human adenovirus not detected in 3 of 3 samples seeded with nonhuman feces (Noble and others, 2003). | Porcine and bovine-specific viruses detected in about 70% of pooled samples (3–5 individuals; Maluquer de Motes and others, 2004).<br><br>Human adenovirus detected in 16 of 18 human sewage samples, but concentrations were not determined (Piña and others, 1998).<br><br>Human adenovirus detected in 3 of 4 samples seeded with sewage and 1 of 4 samples seeded with individual feces (Noble and others, 2003). | No report of the timespan over which adenoviruses occur or variation in concentration within individual hosts was found. | More stable in the environment than enteroviruses (in Piña and others, 1998).<br><br>Most resistant to UV irradiation of any pathogenic enteric virus tested (Gerba and others, 2002). | Frequently asymptomatic but may cause disease (Maluquer de Motes and others, 2004), particularly childhood gastroenteritis (in Noble and others, 2003).<br><br>Detected with greater frequency than pathogenic enterovirus and hepatitis A virus in human waste streams; detected in 25 of the 30 samples in which 1 of the other viruses was detected. Not correlated with fecal coliform concentrations (Piña and others, 1998). |
| *Bacteroides fragilis* phage | Specificity dependent upon the *B. fragilis* strain used to recover phage. One strain (HSP40) only recovered human-associated phage, whereas 5 others recovered phage from other animal hosts (Puig and others, 1999).<br><br>HSP40 phage strongly associated with human-origin fecal contamination (Tartera and others, 1989). | Detected in 12 of 12 urban sewage streams. Concentrations ranged from 3,200 to 19,000 PFU/100 mL (Puig and others, 1999).<br><br>Concentrations ranged from 210 to 46,000 PFU/100 mL in 20 sewage samples (Tartera and others, 1989).<br><br>Concentrations ranged from 2,300 to 20,900 PFU/100 mL in 18 samples of treated or untreated sewage (Piña and others, 1998). | No report of the timespan over which *B. fragilis* phage occur or variation in concentration within individual hosts was found. | *B. fragilis* phages generally survive as long as do coliphages and for longer than do fecal coliforms and enterococci (Sinton and others, 1999).<br><br>Generally more resistant to treatment and inactivation than coliphages (Jofre and others, 1995).<br><br>Cannot reproduce in natural waters and sediments (Tartera and others, 1989). | *B. fragilis* phages may be a more conservative indicator of fecal contamination than enterococci because of their lower dieoff rate in the environment (Tartera and others, 1989).<br><br>*B. fragilis* phages strongly correlated with human-pathogenic viruses concentrations in raw urban sewage (Piña and others, 1998). |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| Coliphage (genotypes) | Types reported to indicate human (II, III) and nonhuman (I, IV) sources of fecal contamination (Furuse, 1987).<br><br>Types differentially distributed among hosts (Cole and others, 2003).<br><br>Human types II and III found in 16 of 16 sewage samples and 17 of 29 pig samples. Nonhuman types I, IV found in 0 of 16 sewage samples and 74 of 91 chicken and pig samples (Hsu and others, 1995).<br><br>Human type III coliphages detected (23 of 31 isolates were type III) in composited gull feces (Noble and others, 2003) and poultry feces (Schaper, Jofre, and others, 2002).<br><br>Human type II coliphages were detected in human, cattle, poultry, and pig feces, though nonhuman type I and IV were not detected in human feces (Schaper, Jofre, and others, 2002). | Concentration and distribution both vary by host; where detected, concentration varies by 2 orders of magnitude (Calci and others, 1998).<br><br>Detected in 5 of 14 hosts tested (human, swine, cattle, gull, goose); where detected, concentration varies by 3 orders of magnitude (Cole and others, 2003).<br><br>Never detected in feces of nine hosts, though present in local water (Griffin and others, 2000).<br><br>Isolated from about 10% of individual humans and at generally higher rates (to 70%) from nonhumans (Schaper, Jofre, and others, 2002).<br><br>Human types II and III detected in 5 of 7 samples seeded with sewage but only 1 of 9 samples seeded with individual feces (Noble and others, 2003). | No report of the timespan over which coliphages occur or variation in concentration within individual hosts was found. | Type I more resistant to inactivation than Type II; Types III and IV more susceptible to inactivation than types I and II (Schaper, Duran, and Jofre, 2002).<br><br>F+ coliphages generally survive longer in the environment than do fecal coliforms and enterococci (Sinton and others, 1999). | The proposed Ground Water Rule uses detection of any coliphages to indicate fecal contamination, not specifically F+ RNA coliphages. The incidence or concentration of F+ RNA coliphage was not consistently related to the concentration of coliphage in general in the supporting information for the Ground Water Rule (U.S. Environmental Protection Agency, 2000c). |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| Enterovirus | Human enterovirus not detected in 4 of 4 samples seeded with nonhuman feces (Noble and others, 2003).<br><br>Bovine enterovirus found in significant proportions of cattle, Canada geese, and deer but other animals (including humans) not tested (Ley and others, 2002). | Human enterovirus detected in 3 of 4 samples seeded with sewage and 0 of 4 samples seeded with individual feces (Noble and others, 2003).<br><br>In 1 sample, enterovirus were present at 70 PFU/100 mL in human sewage (Piña and others, 1998).<br><br>Bovine enterovirus found in 76% of cattle, 1 of 3 Canada geese, and 38% of white-tailed deer individuals (Ley and others, 2002). | No report of the timespan over which enterovirus occur or variation in concentration within individual hosts was found. | Less resistant to UV than adenovirus (Gerba and others, 2002).<br><br>Less stable in the environment than adenovirus (in Piña and others, 1998). | No correlation between regulated fecal indicators and pathogenic enterovirus (Blackwood and others, 2003; Piña and others, 1998).<br><br>Bovine enteroviruses are generally asymptomatic (Ley and others, 2002). |
| Teschovirus | Putative specificity to swine, not found in feces of cattle, sheep, goats (Jiménez-Clavero and others, 2003). | No extensive evaluation of teschovirus distribution and/or concentration in various hosts was found | No report of the timespan over which teschovirus occur or variation in concentration within individual hosts was found. | Detectable 3 km (1.8 mi) downstream from a known source of swine feces (Jiménez-Clavero and others, 2003). | No study of the correlation between teschovirus concentration and regulated fecal indicators was found.<br><br>No report of human health risk from teschovirus was found. |
| -------------------------------------- Bacteria-based source identifiers -------------------------------------- | | | | | |
| *Bacteroides/ Prevotella* (markers) | Of 4 probes by 2 researchers, sequences detected in slurries made from other hosts in 3 of 18 cases (4% false detect; Field and others, 2003).<br><br>Ribotypes of *Bacteroides/ Prevotella* are differentially distributed among various hosts (Wood and others, 1998). | Present at concentrations of 1–10 x $10^{10}$ cells/g dry human feces (Franks and others, 1998), or 8 x $10^{10}$ cells/g wet human feces (Marteau and others, 2001).<br><br>Present in all individual cows and humans tested (Bernhard and Field, 2000b). | No report of the timespan over which these fecal anaerobes persist or variation in concentration within individual hosts was found. | Survival in the environment is limited to days (Kreader, 1998). | No description of the relation between *Bacteroides/Prevotella* incidence or concentration with regulatory fecal-indicator microbe incidence or concentration was found. |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| *Bifidobacterium* species | Twelve species are known to inhabit the human gastrointestinal tract (in Satokari and others, 2001).<br><br>*B. dentium*, *B. adolescentis*, and *B. longum* found exclusively in human-origin sewage; 5 other species differentially distributed. Limited (3 of 19) detections in slaughterhouse waste streams (Bonjoch and others, 2004).<br><br>Sorbitol-fermenting types found only in humans and pigs of 11 hosts tested (Mara and Oragui, 1983). | Present at concentrations of 0.3–1.5 x $10^{10}$ cells/g dry human feces (Franks and others, 1998) or 8 x $10^{10}$ cells/g wet human feces (Marteau and others, 2001).<br><br>Present in all 20 animals tested; *B. dentium* and *B. adolescentis* found in 12 of 12 humans and 22 of 22 human waste streams tested, and *B. longum* found in 7 of 12 humans tested (Bonjoch and others, 2004).<br><br>Sorbitol-fermenting *Bifidobacterium* concentrations range across 3 orders of magnitude, where present (Mara and Oragui, 1983). | Stable populations indicated over 4 weeks of monitoring (Satokari and others, 2001). | Short survival time in the environment and very sensitive to chlorination (in Mara and Oragui, 1983). | No description of the relation between *Bifidobacterium* spp. incidence or concentration with regulatory fecal-indicator microbe incidence or concentration was found. |
| *Enterococcus* (adhesin genes) | Only found in human-associated sewage sources (Scott and others, 2005). | Found in 94% of human-associated sewage sources (Scott and others, 2005). | See entry for *Enterococcus* genotypes and phenotypes. | See entry for *Enterococcus* genotypes and phenotypes. | May be relevant to enterococci-concentration-based regulations. |
| *Enterococcus* (genotypes and phenotypes) | Differentially distributed among hosts (Aarestrup and others, 2000). | Human feces contain $10^5$–$10^8$ CFU/g fecal streptococci (of which enterococci are a subset). Other mammals contain approximately $10^6$ CFU/g (Geldreich, 1976). | No report of the time span over which various *Enterococcus* types persist or variation in concentration within individual hosts was found. | No report of the relative survival of various *Enterococcus* genotypes and phenotypes was found. | Directly relevant to enterococci-concentration-based regulations. |
| *Enterococcus* (species) | Differentially distributed among hosts (Aarestrup and others, 2000; Delahoya and others, 2002).<br><br>Of 61 *E. faecalis*, 30 were from humans and the remainder from dogs and chickens (Wheeler and others, 2002). Note that the number of individuals sampled was 2–3 per host. | One of 4 human, 1 of 2 dog, and 2 of 3 chicken individuals sampled hosted *E. faecalis* (Wheeler and others, 2002). | No report of the timespan over which various *Enterococcus* species persist or variation in concentration within individual hosts was found. | No report of the relative survival of various *Enterococcus* species was found. | Indirectly relevant to enterococci-concentration-based regulations. |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| *Escherichia coli* (genotypes and phenotypes) | Genotype distribution minimally (6%) attributable to host species (Gordon, 2001); however, "geographic effects and host taxonomic group accounted for most of the genetic differentiation (of *E. coli*)" (Souza and others, 1999).<br><br>Same genotype found in multiple hosts (Gordon, 2001; McLellan and others, 2003; Scott and others, 2003); about 10–20% of types (40–65% of isolates) shared among hosts (Hartel and others, 2002).<br><br>Gene transfer has been demonstrated among *E. coli* between host species (Oppegaard and others, 2001). | Prevalent in mammals, less prevalent in birds, and uncommon in fish, amphibians, and reptiles (>1% of cultivable aerobic enteric bacteria; Gordon and Cowling, 2003). Found in areas with no fecal contamination expected (Carrillo and others, 1985).<br><br>Genotype distribution is minimally (1–10%) attributable to spatial separation (Gordon, 2001).<br><br>Concentrations can vary by as much as 6 orders of magnitude in specific hosts ($1.9x10^3$ – $6.3x10^9$ CFU/g wet wt poultry dropping); geometric mean concentrations range from $1.5x10^3$ in horses to $2.1x10^9$ CFU/g wet feces in dogs (Mara and Oragui, 1983). | Genotypes are temporally unstable in the host (Cooke and others, 1969; Gordon, 2001).<br><br>About half of *E. coli* genotypes are resident (stable) and the remainder are transient (unstable) in an individual (Caugant and others, 1981; Whittam, 1989). | Different genotypes detected in bird feces and underlying (aged) litter (Whittam, 1989).<br><br>Not considered to have survival strategies in the environment as a component of its life cycle (Winfield and Groisman, 2003).<br><br>Known to be stable in tropical environments, though thought to be unable to reproduce in temperate environments (Rivera and others, 1988; Winfield and Groisman, 2003); however, there exist habitats in temperate climates in which *E. coli* can survive and perhaps reproduce (Whitman and Nevers, 2003; Whitman and others, 2003). | One of two recommended fecal indicator bacteria for coastal recreational waters (U.S. Congress, 2000). |
| *E. coli* (glucuronidase gene *uid*A) | Several alleles are host specific; however, the majority are carried by more than one host (Ram and others, 2004). | Approximately 95% of *E. coli* have *uid*A–coded glucuronidase (Brenner and others, 1993).<br><br>Diversity of *uid*A alleles has been demonstrated (Farnleitner and others, 2000). | See entries for *E. coli* genotypes and phenotypes. | See entries for *E. coli* genotypes and phenotypes. | Should be directly relevant to *E. coli*-based regulations because the gene is nearly ubiquitous among *E. coli.* |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| *E. coli* (toxin genes) | The gene *lt*-IIa has been found exclusively in cattle (Pickett and others, 1987); the gene *st*-II (pig associated) is differentially distributed among several hosts (Chern and others, 2004); the gene *slt*-IIe has been found in pigs and humans (Franke and others, 1995); the gene *st*-Ib (human-associated) was detected in 8% of non-human individuals and 9% of human individuals (100% of composite human sources and 25% of composite non-human sources; Field and others, 2003)<br><br>Of three probes used to detect toxin genes, detection rate in alternate sources was 2 of 9 (7% false detect; Field and others, 2003) | Fewer than 1% of *E. coli* carry toxin genes in most waste streams (Chern and others, 2004).<br><br>Pathogenic *E. coli* are absent from most healthy individuals (Chern and others, 2004; Gordon, 2001). | No report of the timespan over which pathogenic *E. coli* persist or variation in concentration within individual hosts was found. | See entries for *E. coli* genotypes and phenotypes. | Indirectly relevant to *E. coli*-based regulations. Directly related to health risk for susceptible animals |

**Table 2:** Summary of how various microbial source tracking (MST) source identifiers meet generalized assumptions for application.—Continued

[More extensive details are presented in the text]

| Source identifier | Host specificity | Distribution | Stability in host | Stability (survival) in environment | Relevance to regulations, health risk |
|---|---|---|---|---|---|
| *Salmonella enterica* serovar Typhimurium | Avian types more similar to one another than to those from other hosts. Avian types can colonize other hosts and, specifically, cause disease in humans (Refsum and others, 2002). | No report of the distribution among various hosts was found. | No report of the timespan over which *Salmonella enterica* serovar Typhimurium persists or variation in concentration within individual hosts was found. | Considered to have survival strategies in the environment as a component of its life cycle, leading to long survival under adverse conditions ($t_{50}$ up to 1 year in soil; Winfield and Groisman, 2003). | Human outbreaks of salmonellosis (in Refsum and others, 2002). |
| ---------------------------------- Protozoan-based source identifier ---------------------------------- | | | | | |
| *Cryptosporidium* (species) | *C. parvum* genotype I is exclusive to humans, type II readily transfers between humans and other mammals, and various other host-associated types occasionally are detected in humans (in Ong and others, 2002).<br><br>Nine species other than *C. parvum* are host associated but not exclusive (Fayer and others, 2000).<br><br>Only five genotypes of *Cryptosporidium* spp. are known to infect humans (Xiao and others, 2001).<br><br>Some populations appear to be host-specific, but others move between host species (Mallon and others, 2003). | Incidence in fresh livestock waste ranged from 5% (cattle) to 29% (sheep; Hutchison and others, 2004).<br><br>*C. parvum* detected in 1% of wildlife (6 of 19 species), 5% of sewage influent samples, and 4% of agricultural waste (Heitman and others, 2002).<br><br>Detected in 4% of cattle (Atwill and others, 1999).<br><br>*C. parvum* detected in 13–22% of rodents; *C. muris* detected in 2–10% of rodents (Chalmers and others, 1997).<br><br>*C. parvum* detected in 12% of wild mammals (Sturdee and others, 1999).<br><br>Over 100 species have been reported infected with *C. parvum* and *C. parvum*-like strains (Fayer and others, 2000). | Known to pass through and be transmitted by various hosts (Fayer and others, 2000). | Can remain viable for greater than 6 months at 20 degrees Celsius in water (Fayer and others, 2000.<br><br>In three environmental fresh waters, 0.1% survival over 10 weeks at 25 degrees and higher rates of survival at 4 degrees (Simmons and Sobsey, 2002).<br><br>Survival and infectivity in freshwater detected after more than 12 weeks (Olson and others 1999). | Pathogenic agent of cryptosporidiosis, caused 50% of reported gastroenteritis cases due to consumption of treated water in 2001–02 (Yoder and others, 2004). |