

Using DNA Microarrays To Identify Library-Independent Markers for Bacterial Source Tracking

Marilyn Soule,¹ Edward Kuhn,¹ Frank Loge,² John Gay,³ and Douglas R. Call^{1*}

Department of Microbiology and Pathology, Washington State University, Pullman, Washington¹; Department of Civil and Environmental Engineering, University of California at Davis, Davis, California²; and Department of Veterinary Clinical Sciences, Washington State University, Pullman, Washington³

Received 28 June 2005/Accepted 27 December 2005

Bacterial source tracking is used to apportion fecal pollution among putative sources. Within this context, library-independent markers are genetic or phenotypic traits that can be used to identify the host origin without a need for library-dependent classification functions. The objective of this project was to use mixed-genome *Enterococcus* microarrays to identify library-independent markers. Separate shotgun libraries were prepared for five host groups (cow, dog, elk/deer, human, and waterfowl), using genomic DNAs (gDNAs) from ca. 50 *Enterococcus* isolates for each library. Microarrays were constructed (864 probes per library), and 385 comparative genomic hybridizations were used to identify putative markers. PCR assays were used to screen 95 markers against gDNAs from isolates from known sources collected throughout the United States. This validation process narrowed the selection to 15 markers, with 7 having no recognized homologues and the remaining markers being related to genes involved in metabolic pathways and DNA replication. In most cases, each marker was exclusive to one of four *Enterococcus* species (*Enterococcus casseliflavus*, *E. faecalis*, *E. hirae*, or *E. mundtii*). Eight markers were highly specific to either cattle, humans, or elk/deer, while the remaining seven markers were positive for various combinations of hosts other than humans. Based on microarray hybridization data, the prevalence of host-specific markers ranged from 2% to 45% of isolates collected from their respective hosts. A 20-fold difference in prevalence could present challenges for the interpretation of library-independent markers.

“Microbial source tracking” refers to a suite of procedures that can be used to test water samples and apportion fecal pollution to putative sources. The procedures can target any number of microbial markers (viral, bacterial, or protozoan), but they are all intended to assist watershed managers in complying with total maximum daily load requirements and with mitigation plans for microbiologically impaired waters. The identification of host-specific bacteria, also known as bacterial source tracking (BST), has been the subject of recent reviews (16, 19, 23, 24, 27).

BST methods can be characterized as library dependent and library independent. Library-dependent methods begin with characterization of a large collection of bacteria that are isolated from known sources (e.g., humans or cows). Only one species or genus of bacteria is usually considered in a given study (e.g., *Escherichia coli*, *Enterococcus*, or *Streptococcus*), and the traits of interest might be antibiotic resistance profiles (10, 12, 14, 30, 31), carbon utilization profiles (11), DNA fingerprints (5, 8, 13, 15, 29), or other DNA polymorphisms (17, 20). Once a large number of strains have been characterized, these data are used as “training” data to develop a classification equation using multivariate statistics. The reliability of the resultant classification function is evaluated both by its ability to classify the original training data and by its ability to correctly classify independent isolates from known sources, with the latter being the most legitimate means to assess the quality

of the classification equation. In both cases, the analysis reports the percent correct classification rates for the various host animals.

A number of challenges can arise from the library-dependent strategy. The first challenge is that a large number of isolates must be characterized before a suitable classification equation can be generated. No guidelines exist to help determine how many isolates should be characterized per host, how many hosts should be sampled, or how many hosts should be included in the sampling frame (15). More representation is clearly a preferred strategy, but actual guidelines may be difficult to generate because this will be a function of the variance in the system under study (27). Also, there is a spatial component to the analysis. Because most of the traits considered by these methods are unlikely to represent selective loci that are specific to the host organisms (e.g., pulsed-field gel electrophoresis patterns are unlikely to be related to fitness), the libraries may have geographic limitations. While a library may perform within acceptable limits for a given watershed, it may not be applicable to an adjacent watershed (13, 22, 31). There is probably a temporal component leading to potential misclassification errors because we can expect selectively neutral traits to drift with time. The coefficients for classification functions are not reported in most papers, and the classification databases are effectively proprietary information, which is understandable because the host institution often carries the burden of maintenance of classification libraries. Nevertheless, when raw data and classification functions are not available, the reported classification system is effectively unavailable to the broader community. Finally, suitable methods are lacking

* Corresponding author. Mailing address: Department of Microbiology and Pathology, Washington State University, 402 Bustad Hall, Pullman, WA 99164-7040. Phone: (509) 335-6313. Fax: (509) 335-8529. E-mail: drcall@wsu.edu.

for assigning valid confidence intervals for estimates originating from classification functions.

Library-independent methods require the development of an initial library of host-specific strains that are then characterized to identify host-specific genetic or phenotypic markers. The key difference from library-dependent methods is that these markers can be strongly associated only with bacterial strains from a single host animal. Thus, the markers can be used in a binary context (presence or absence) without reference to a classification function or library. Ideally, library-independent markers are "universal" in both space and time. To meet these criteria, the markers must target either an organism that appears specific to a given host or a gene product or gene sequence that is specific to bacteria from a given host. In the latter case, it is likely that the trait of interest confers a selective advantage in a specific host, which would explain a strong association between the marker and host. While several papers report polymorphisms in the 16S rRNA gene for bacterial species closely associated with specific hosts (1, 2, 6, 7, 9), there is only one published example of a host-specific genetic marker for a common, facultative aerobe such as *Enterococcus* (21), and this marker is only useful for detecting human fecal pollution.

The lack of progress in the identification of DNA markers suitable for identifying bacteria from specific hosts is probably related to our lack of understanding about which unique genes would be required to persist in different hosts and our inability to screen large numbers of potential markers. For this project, we attempted to circumvent the latter limitation by using a custom DNA microarray to screen a large number of genetic sequences from bacteria originating from different host animals. The microarray was composed of cloned DNA fragments from a large collection of isolates of *Enterococcus* from known sources (hence the term "mixed-genome" microarray) (4). This process identified a number of genetic markers that could be used in a library-dependent context, and these have been partially validated using *Enterococcus* isolates from known sources from across the United States.

MATERIALS AND METHODS

Bacterial isolates. Two sets of *Enterococcus* sp. isolates were collected for this study. The first set (set A) was used both to construct and to screen the mixed-genome microarray. Set A comprised 413 isolates from five different host sources, namely, cows ($n = 200$), dogs ($n = 61$), waterfowl ($n = 43$), humans ($n = 54$), and elk and deer ($n = 55$). Isolates in set A were collected locally in Pullman, Wash., Moscow, Idaho, and Lewiston, Idaho, where a single bacterial isolate was collected per fecal sample. The second set (set B) of *Enterococcus* sp. samples was used for PCR screening of putative host-specific markers and comprised samples from a variety of host sources, namely, cows ($n = 150$), dogs ($n = 36$), waterfowl ($n = 24$), humans ($n = 6$), chickens ($n = 12$), horses ($n = 18$), swine ($n = 40$), and deer ($n = 4$). Two individual isolates and two composites (collected by scraping colonies off agar plates) composed of approximately 200 isolates each were collected per fecal sample for set B. Set B isolates originated from throughout the United States.

Isolation of *Enterococcus* spp. Fecal samples were collected with a sterile tongue depressor or swab and stored in a sterile plastic bag at 4°C until processed. All samples were processed within 7 days of collection. A sterile swab containing fecal material was used to inoculate 3 ml salt tolerance solution (2.5% [wt/vol] brain heart infusion broth [Becton-Dickinson, Sparks, MD], 0.5% dextrose, 6% NaCl, 0.0016% bromocresol purple) and incubated overnight at 45°C. Positive samples (yellow) were plated onto M-Enterococcus agar (Remel, Lenexa, KS) and incubated at 37°C for 72 to 96 h. This isolation procedure selects for *Enterococcus*, and to verify that selection was effective, we confirmed a repre-

sentative sample of 52 isolates (from set A) as *Enterococcus* spp. by using the API Strep test (bioMérieux, Inc., Hazelwood, MO).

Fresh isolates were picked from the M-Enterococcus agar into brain heart infusion broth (Becton-Dickinson) and grown overnight at 37°C. Three milliliters of broth culture was used for genomic DNA (gDNA) extraction, and 1 ml was banked at -80°C after the addition of 330 μ l phosphate-buffered glycerol (45 mM Na₂HPO₄, 34 mM NaH₂PO₄, 58.8% [vol/vol] glycerol).

Construction of mixed-genome shotgun libraries. A separate shotgun library was constructed for each of the following hosts: cow, dog, human, waterfowl, and elk/deer. gDNAs were extracted from all isolates in set A by using a DNeasy tissue kit (QIAGEN, Valencia, CA) and were quantified by electrophoresis and spectroscopy. An equal amount of gDNAs from 50 isolates per host (only 43 for waterfowl) was mixed to make five mixed-genome pools containing 3 to 5 μ g DNA in 160 μ l elution buffer (supplied with the DNeasy tissue kit). The gDNAs were divided into 40- μ l aliquots and sonicated to obtain 500- to 700-bp fragments. This included up to four sonication treatments (45 s to 1 min each at level 7) with a Misonix cup-horn sonicator (Misonix Inc., Farmingdale, NY). Samples were cooled on ice between treatments. Sonicated gDNA fragments were separated in a 1% agarose gel, the 600-bp region was excised, and the DNA was extracted using a Montage gel extraction kit (Millipore, Billerica, MA) followed by ethanol precipitation. A TOPO Shotgun subcloning kit (Invitrogen, Carlsbad, CA) was used according to the manufacturer's instructions to repair fragment ends and clone the fragments into the pCR4-TOPO vector. Recombinant plasmids were electroporated into *E. coli* Top 10 cells (Invitrogen) and selected on LB agar plates supplemented with 100 μ g/ml ampicillin (Fisher Scientific, Fairlawn, NJ) and 40 μ g/ml X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside; Sigma-Aldrich, Milwaukee, WI) (LB_{Amp100}X-Gal40). For each library, 864 (9 \times 96) white colonies were selected, placed into a 96-spot grid on fresh LB_{Amp100}X-Gal40 agar plates, and incubated overnight at 37°C. A 96-pin replicating tool was used to inoculate a 96-well U-bottomed plate (plate A) containing 150 μ l LB_{Amp100} broth per well, which was incubated overnight at 37°C. This plate was then used to inoculate two additional 96-well plates (plates B and C) containing LB_{Amp100} broth, which were incubated overnight at 37°C. Glycerol (50 μ l of 50% glycerol) was added to plates A and B, which were stored at -80°C for routine use and long-term storage, respectively. Plate C was stored at -80°C without glycerol for use as a template for PCR.

Mixed-genome microarray fabrication. Clone inserts were PCR amplified from whole cells containing recombinant plasmids (5 μ l broth from plate C [described above]) in 50- μ l PCR mixtures as described for the TOPO Shotgun subcloning kit (Invitrogen), using T7 and T3 primers. PCR products were purified by isopropanol precipitation and resuspended in 22 μ l H₂O, and 2 μ l was evaluated in a 1% agarose gel. To the remaining 20 μ l PCR product, 12.5 μ l 4 \times print buffer (0.4 M Na₂HPO₄, 0.8 M NaCl, 0.04% sodium dodecyl sulfate; pH 11.6) and 17.5 μ l H₂O were added. Single spots of each PCR product were deposited as eight subarrays on Superfrost/Plus slides (Fisher Scientific) by using a MicroGrid II arrayer (Genomic Solutions, Ann Arbor, MI). Each subarray contained four replicate spots of a region of the 16S rRNA gene from an *Enterococcus* sp. isolate amplified with primers 16S_008Fwd and 16S_517Rvs (28) and four replicate spots of an arbitrary, biotinylated oligonucleotide. The former served as a control for target labeling and hybridization efficiency, and the latter served as a control for detection chemistry (4).

Sample hybridization. gDNAs were extracted from target strains using a DNeasy tissue kit (QIAGEN), and 0.5 μ g was nick translated for 2 h in the presence of biotin-dATP (BioNick labeling system; Invitrogen). Labeled gDNAs were purified by ethanol precipitation and resuspended in 225 μ l hybridization buffer (4 \times SSC [60 mM NaCl, 0.6 mM sodium citrate; pH 7.0] and 5 \times Denhardt's solution [0.1% {wt/vol} Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin]). Biotinylated gDNAs (75 μ l) were heat denatured, applied to the slide, and incubated overnight in a humidified chamber at 60°C. Slides were preblocked at 23°C for 30 min with TNB buffer (100 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.5% blocking reagent [TSA biotin system; Perkin-Elmer, Boston, MA]). The remaining detection steps were carried out as previously described (3), with 75 μ l of the appropriate reagent applied to the slide at each step. Images were captured with an arrayWoRx^e scanner (Applied Precision, Issaquah, WA).

Image and data analysis. Microarray images were quantified using softWoRx Tracker software (Applied Precision). The final output included median pixel values (range, 0 to 65,535) that were imported into a custom relational database (MS Access; Microsoft Corp., Redmond, WA) for further processing. Data from each slide were normalized by first calculating the mean intensity for the replicated 16S rRNA gene probes and then dividing every probe intensity value by this mean. When normalized values exceeded 0.5, the probe was considered positive (and thus present in the genome). We then generated frequency tables

TABLE 1. Primers used in this study

Marker no.	Forward primer sequence	T_m (°C)	Reverse primer sequence	T_m (°C)	Product length (bp)
15	AGATACCTTAACACATTACACCCAG	50.6	TGTGAAACTATCAATCAAACATGAT	50.1	357
19	AATCGTATTCATTGCCCCAAG	51.2	GGGGATTTTGTACGAAAGAAG	51.5	375
40	TCATGATCTGTGCAATATTCGAC	51.9	AGCAATTAATGAAGAAGAAGTTTG	50.6	326
48	AGTTCTTATAGAGTAAATCACGTTTCATAG	50.3	AAAACCTAGTATAATCAATTTTCCATC	50.5	319
66	TATTAATAAACTGAAAATGGCAAATG	51.4	AAGAAGAAATGAGATTCGTGGAC	51.1	308
67	TTTAAATGACAGGGATCAAACCTG	50.6	TTACCACAGAGAGAATCTTCGC	50.6	301
68	ATACCCGTACAGTTATTTCTTTACC	51.0	TTTAGCTATGTTCTACTAATTGCAG	51.0	322
77	GTAGATTTAGTGTACCGTTTGATTTC	50.0	ATATTGCTTGCTTTGGATTGAC	50.6	314
81	AAAATGTAACAACCGCAAAGC	51.5	TCCAGTTTCTAAAGTCCCG	51.0	314
89	CTTATGAAGATTCTGAAACTGG	50.5	CAAGTAATGCCCGTTGATCTC	51.4	388
90	TTGACAGTTTGTGGAAGTGTCTG	50.6	CGAAAACCTCTCCGAATACACTATC	51.2	419
91	AAAGGACAAGCTATTTAATGATGAG	51.9	TTTCCCTAAAGCTAAGCCATC	50.1	395
93	GTA AAACTCAATCTGATGAGGCAG	51.8	AGATGATCTGCCATCCGC	50.3	422
94	GAAGAATTCGGAACCCGTC	50.9	AGTTGTACTGACACATCAGAAACG	50.5	402
107	TCTTTTCTCACTACGCTAAGTG	50.4	CCTCTCCACTGTAAGGTCAAATC	51.4	401
7+	GAGAGAATCCTAAGGTGAGCG	50.0	AGACAGTGCCCCAAATCGTTAC	50.9	340

to identify probes that were exclusive or nearly exclusive to a specific host (chi-square test). We only considered probes to be specific to a host if the probes originated from the mixed-genome library for that host. In cases where no probes could be identified by contingency analysis, we selected the best possible match to the host. NCSS 2004 software (Number Cruncher Statistical Systems, Kaysville, UT) was used for statistical tests, with a P value of 0.05 as the threshold for significant findings. We did not employ a correction factor for multiple tests because our goal at this stage was to be liberal in the selection process and rely on subsequent PCR validation testing to narrow the list of potential BST markers.

Plasmid insert sequencing and primer design. Clones identified as potentially specific for a single host were retrieved from the glycerol stock bank and sequenced. Plate A (glycerol stock) was thawed on ice, and 20 μ l of each selected clone was used to inoculate 3 ml LB_{Amp100} broth. Cultures were grown overnight at 37°C, and a QIAGEN Plasmid mini spin kit was used to isolate plasmid DNA. Extracted plasmids were quantified by spectrometry, diluted to 25 ng/ μ l, and sequenced with T7 and T3 primers using an ABI BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). Each sequencing reaction contained 2 μ l Terminator mix, 3 μ l 5 \times dilution buffer, 3.4 μ l H₂O, 1.6 μ l 2 μ M primer, and 10 μ l of 25 ng/ μ l plasmid template. Thermal cycling parameters for sequencing were set as described by the manufacturer (Applied Biosystems). Sequencing reaction products were purified by ethanol precipitation, resuspended in 15 μ l HiDi formamide (Applied Biosystems), and analyzed with an ABI Prism 3100 genetic analyzer (Applied Biosystems). Vector NTI 9.0.0 (InforMax, Frederick, MD) software was used for base calling, to generate sequence alignments, and to design primers (Table 1). Database searches for similar protein sequences were performed by using the BLASTx network server (NCBI; www.ncbi.nlm.nih.gov) to assign a putative function to each probe sequence.

Preparation of templates for marker validation by PCR screening. The templates for validation screening, made from gDNAs isolated from set A and set B samples, are described in Table 2. The sewage and dairy templates (Table 2) were prepared from liquid samples taken from a municipal wastewater treatment plant (Pullman, WA) and from the WSU Dairy sewage lagoon, respectively. Liquid samples were placed on ice immediately after collection and centrifuged at 4°C. The supernatants were discarded, and the pellets were placed on ice. Sterile swabs were used to collect some of the cell pellet and inoculate four test tubes containing 3 ml salt tolerance solution each. Samples were incubated at 45°C for 24 h. A sample culture (1.5 ml) was taken from each test tube, and the cells were pelleted by centrifugation. Cells were washed twice in phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄ · 7H₂O, 1.4 mM KH₂PO₄, pH 7.3) before proceeding to gDNA isolation using a QIAGEN DNeasy tissue kit.

Validation PCR screening. All screening PCRs were performed in 50- μ l reaction mixtures with approximately 5 ng plasmid DNA or 10 to 50 ng gDNA template. The cycling program included a 2-min initial denaturation at 95°C, followed by 35 cycles of 95°C for 30 s, 48°C for 45 s, and 72°C for 40 s and a final extension at 72°C for 2 min. PCR with the *esp* primer pair was carried out as previously described (21).

Sequencing and analysis of 16S rRNA genes from selected isolates. We identified two or three isolates (32 total) that harbored one of the putative markers

from this study and sequenced a portion of the 16S rRNA gene to aid in their identification. A portion of the 16S rRNA gene was PCR amplified using primers 16S_008Fwd and 16S_517Rvs (28), with reaction and thermal cycling conditions as described previously. Two units of exonuclease I (New England Biolabs, Ipswich, MA) and 4 units of shrimp alkaline phosphatase (New England Biolabs) were added to each PCR mixture, and reactions were incubated at 37°C for 20 min followed by 80°C for 15 min. PCR products were then diluted to 4 ng/ μ l, and 10 μ l was added to a standard one-quarter sequencing reaction mix as described above. Sequencing was completed as described above. Vector NTI 9.0.0 (InforMax) software was used for base calling and to generate sequence alignments. Database searches for similar DNA sequences were performed by using the BLASTn network server (NCBI; www.ncbi.nlm.nih.gov) to assign a putative species to each isolate.

RESULTS

Identification of putative host-specific markers. We constructed a mixed-genome microarray composed of 4,320 probes that were derived from five distinct shotgun libraries representing *Enterococcus* isolates collected from cattle, dogs, humans, waterfowl, and cervids (elk and deer). Genomic DNAs from 50 *Enterococcus* isolates per host (only 43 for waterfowl; 243 isolates total) were used to construct these libraries. Genomic DNAs from all isolates in set A ($n = 413$; see Materials and Methods for a description) were hybridized to the microarray, and data from 385 successful hybridizations were used to identify probes that were deemed specific for a single host. Eighty-six probes hybridized exclusively to isolates from a single host, although probes were present in only 8% to 19% of isolates from the specific host. This set of 86 probes included 46 from cows, 5 from dogs, 25 from elk/deer, and 10 from humans. We did not identify any probes as specific to waterfowl using these selection criteria. We identified an additional 20 probes that were present in a significantly larger number of isolates from one specific host than isolates from the remaining four hosts (chi-square test, $P < 0.05$); these probes were present in 10% to 42% of the specific host isolates from set A and included six from cows, two from dogs, six from elk/deer, and six from humans. Finally, two additional probes from dog isolates and two probes from waterfowl isolates with the highest (but not statistically significant) specificities to their specific hosts were chosen for further screening. All 110 putative host-

TABLE 2. Templates used for validation PCR screening

Source of <i>Enterococcus</i> sp. isolates in validation template	Validation template name	Composition of validation template (per template) ^a	Source set ^{a,b}	Geographic source (U.S. state)
Cow	Cow pool	50 isolates from 50 hosts	B	NY, OH, AK, CA, KS
	CP1 through CP20	7–10 isolates	A	WA, ID
	CP-all (all CP pools)	179 isolates	A	WA, ID
	Composites ($n = 10$)	Approximately 400 isolates from a single host	B	KS, CA, AK, OH, NY, TN
	CCP (all composite pools listed above)	Approximately 4,000 isolates from 10 hosts	B	KS, CA, AK, OH, NY, TN
Dog	Dog pool	36 isolates from 18 hosts	B	IN, KS, CA, TN
	DP1 through DP6	7–10 isolates	A	WA, ID
	DP-all (all DP pools)	56 isolates	A	WA, ID
	Composites ($n = 9$)	Approximately 400 isolates from a single host	B	IN, KS, CA, TN
	DoCP (all composite pools listed above)	Approximately 3,600 isolates from 9 hosts	B	IN, KS, CA, TN
Human	Human pool	6 isolates from 3 hosts	B	OH
	HP1 through HP6	7–10 isolates	A	WA, ID
	HP-all (all HP pools)	47 isolates	A	WA, ID
	Composites ($n = 3$)	Approximately 400 isolates from a single host	B	OH
	HCP (all composite pools listed above)	Approximately 1,200 isolates from 3 hosts	B	OH
Elk/deer	Deer pool	4 isolates from 2 hosts	B	OH
	EDP1, EDP2, EDP3	7–10 isolates	A	WA, ID
	EDP-all (all EDP pools)	28 isolates	A	WA, ID
	Composites ($n = 2$)	Approximately 400 isolates from a single host	B	OH
	DeCP (all composite pools listed above)	Approximately 800 isolates from 2 hosts	B	OH
Waterfowl	Waterfowl pool	24 isolates from 12 hosts	B	IN, OH, TN
	WP1 through WP4	7–10 isolates	A	WA, ID
	WP-all (all WP pools)	39 isolates	A	WA, ID
Goose	Goose pool	8 isolates from 8 hosts	B	IN, OH, TN
Swine	Swine pool	20 isolates from 20 hosts	B	IN, KS, CA, OH, TN
Horse	Horse pool	9 isolates from 9 hosts	B	IN, NC, TN
Chicken	Chicken pool	6 isolates from 6 hosts	B	IN, OH
Sewage	Sewer2/5, Sewer2/22, Sewer3/2	Isolates from sewage sample	NA	WA
Dairy lagoon	Dairy2/7, Dairy2/22, Dairy3/2	Isolates from dairy wastewater sample	NA	WA

^a The collection of isolates and source sets are described in Materials and Methods.

^b NA, not available.

specific probes were retrieved from glycerol stocks, and we successfully sequenced 95 of the inserts. Primers were designed for these putative host-specific sequences, and those retained after the validation screening are shown in Table 1.

Screening putative host-specific markers using PCR. Putative host-specific primers were used to screen gDNA pools generated from multiple isolates of *Enterococcus* spp. (Table 2). We rejected all potential markers for waterfowl and dogs but identified 15 putative markers distributed between humans, cattle, and deer/elk (Table 3). Two cattle markers were identified as positive for only cow isolates and dairy lagoon samples. The elk/deer markers varied in their specificities, from specific for only elk/deer isolates to positive detection for several combinations of nonhuman templates. The human-specific primer pairs were mostly very specific, although only marker 77 amplified a band from the sewage samples (Table 3).

This screening procedure only provides information about the presence or absence of individual markers within collections of isolates. To better gauge the prevalence of these markers within a population of isolates, we examined the original microarray hybridization data (Table 4). By this analysis, the cattle markers were exclusive to cattle and were present in 6.8% to 8.7% of isolates tested. Two of the elk/deer markers

were exclusive to this host, while five remaining markers were reactive with isolates from all other hosts. Importantly, despite this apparent lack of specificity, the prevalence of these markers was much greater for elk/deer isolates (>25%) than for isolates from other hosts (2% to 7.1%). Four of the six human markers were relative rare (<4%), while the fifth marker was detected in 9.8% of isolates. The sixth marker was detected in 45% of human isolates, with occasional detection in isolates from other sources (1.1% to 5.3%).

Marker identity. Seven of the markers had either no significant matches with GenBank entries or the closest matches were hypothetical or uncharacterized proteins (Table 5). The remaining eight sequences had BLASTx matches that could be assigned putative functions. We also selected one probe that was positive for all microarray hybridizations as a positive control, and sequence analysis demonstrated that this probe was a close match to the 23S rRNA from *Enterococcus faecium* (marker 7+).

Identification of positive control strains for each primer pair. After selecting the 15 markers described in this report, we screened individual isolates ($n = 32$) of *Enterococcus* by PCR to identify positive control strains for each marker (Table 6). The 15 markers were distributed between four species of *Enterococcus*. The two cattle markers were exclusive to *E. hirae*, while the

TABLE 3. Validation PCR screening results

Source of <i>Enterococcus</i> sp. isolates in validation template	Validation template ^a	PCR result ^b															
		Cow markers		Elk/deer markers						Human markers							
		15	19	40	48	89	90	91	93	94	66	67	68	77	81	107	<i>esp</i>
Cow	Cow pool	+	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-
	CP-all	+	+	-*	-	+	+	+	+	+	-	-	-	-	-	-	-
	CCP	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	CP1-20	11	15	0*	0	ND	ND	ND	ND	ND	ND	0	0	0	ND	0	ND
	Composites (n = 10)	1	1	0	0	ND	ND	ND	ND	ND	0	0	0	0	0	ND	ND
Dog	Dog pool	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DP-all	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DoCP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Human	Human pool	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
	HP-all	-	-	-*	-	-*	-	-	-	-	+	+	+	+	+	+	-
	HCP	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
	HP1-6	0	0	0*	0	0*	0	0	0	0	5	1	4	5	1	2	ND
	Composites (n = 3)	0	0	0	0	ND	ND	ND	ND	ND	0	0	0	1	0	0	ND
Elk/deer	Elk/deer pool	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	EDP-all	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-
	DeCP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	EDP1-3	0	0	3	3	3	3	3	3	2	ND	0	0	0	0	ND	ND
	Composites (n = 2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ND	ND
Waterfowl	Waterfowl pool	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	WP-all	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-
Goose	Goose pool	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Swine	Swine pool	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-
Horse	Horse pool	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chicken	Chicken pool	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sewage	3 samples	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	3
Dairy	3 samples	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0

^a The validation templates are described in Table 2.

^b +, positive PCR result; -, negative PCR result; ND, not determined; integer, number of positive PCR results in the set of validation templates. Asterisks indicate that negative PCR results were obtained in validation screening but that one or more single positive control isolates gave a positive PCR result (see Table 6).

deer/elk markers were detected in three species (*E. hirae*, *E. mundtii*, and *E. casseliflavus*). Five of the six human markers were from *E. faecalis*, and the *esp* marker was negative for this panel of positive controls (the *esp* marker originated from *E. faecium*) (21).

TABLE 4. Distribution of markers among 385 isolates hybridized to the mixed-genome microarray

Host source of marker	Marker no.	% of isolates that hybridized to the marker probe				
		Cow isolates (n = 183)	Dog isolates (n = 61)	Human isolates (n = 51)	Elk/deer isolates (n = 52)	Waterfowl isolates (n = 38)
Cow	15	8.7	0	0	0	0
	19	6.8	0	0	0	0
Elk/deer	40	0	0	0	19.2	0
	48	0	0	0	15.4	0
	89	3.3	1.6	5.9	32.7	2.6
	90	3.8	3.3	2.0	32.7	2.6
	91	3.8	3.3	2.0	26.9	2.6
	93	7.1	1.6	2.0	38.5	2.6
Human	94	6.8	4.9	5.9	32.7	2.6
	66	0	0	2.0	0	0
	67	0	0	2.0	0	0
	68	0	0	9.8	0	0
	77	0	0	3.9	0	0
	81	0	0	2.0	0	0
	107	1.1	4.9	45.1	1.9	5.3

DISCUSSION

BST is a tool that can be used to apportion fecal pollution to putative sources, and BST methods can be library dependent or library independent. Library-dependent methods have a number of shortcomings that have been reviewed elsewhere (15, 17, 19, 24, 26, 27). Consequently, there is considerable interest in identifying library-independent markers, but the challenge is to find an efficient means to identify these markers. Efforts have ranged from rational selection of single genes, such as the virulence-associated *esp* gene from *E. faecium* (21), to the detection of polymorphisms within a single gene that can be used as host-specific markers (1, 2, 6, 7).

Our effort focused on a different strategy whereby we used microarrays to screen for the presence or absence of a large number of DNA fragments (n = 4,320) in an effort to detect host-specific markers. We focused our efforts on the *Enterococcus* genera in part because there is a considerable history of using these organisms for water quality testing and because these organisms are relatively simple to culture. We have no means to gauge the efficiency of our approach, but given that only 0.35% of the cloned inserts appeared to be applicable to BST, other approaches should be considered to enrich the libraries for host-specific markers (e.g., using suppression subtraction hybridization). While library construction and microarray hybridization consumed considerable effort, we found that a significant amount of time was also required for PCR

TABLE 5. Description of selected markers

Marker no. (GenBank accession no.)	Marker host origin	Accession no. of similar protein, organism, and BLASTx E value	Predicted function or property	Control isolate species containing marker
15 (DQ071632)	Cow	NP_951820.1, <i>Geobacter sulfurreducens</i> PCA, 8e-23	Putative helicase	<i>E. hirae</i>
19 (DQ071647)	Cow	NP_976695.1, <i>Bacillus cereus</i> ATCC 10987, 4e-13	Hypothetical protein	<i>E. hirae</i>
40 (DQ071633)	Elk/deer	NP_816591.1, <i>Enterococcus faecalis</i> V583, 6e-51	MutS2 family protein	<i>E. hirae</i> , <i>E. casseliflavus</i> , <i>E. hirae</i>
48 (DQ071634)	Elk/deer	None		<i>E. casseliflavus</i>
89 (DQ071637)	Elk/deer	None		<i>E. mundtii</i> , <i>E. hirae</i>
90 (DQ071635)	Elk/deer	ZP_00287232.1, <i>Enterococcus faecium</i> , 1e-19	Polyribonucleotide nucleotidyltransferase	<i>E. mundtii</i>
91 (DQ071638)	Elk/deer	ZP_00286134.1, <i>Enterococcus faecium</i> , 2e-53	Predicted SAM-dependent methyltransferase	<i>E. mundtii</i>
93 (DQ071636)	Elk/deer	ZP_00285574.1, <i>Enterococcus faecium</i> , 1e-55	Response regulator	<i>E. mundtii</i>
94 (DQ071639)	Elk/deer	ZP_00286604.1, <i>Enterococcus faecium</i> , 3e-29	Uncharacterized protein conserved in bacteria	<i>E. mundtii</i>
66 (DQ071640)	Human	NP_756371.1, <i>Escherichia coli</i> CFT073, 5e-19	Hypothetical protein	<i>E. faecalis</i>
67 (DQ071641)	Human	None		<i>E. hirae</i>
68 (DQ071642)	Human	ZP_00233409.1, <i>Listeria monocytogenes</i> strain 1/2a F6854, 3e-11	Carbohydrate kinase PfkB family	<i>E. faecalis</i>
77 (DQ071643)	Human	NP_346799.1, <i>Clostridium acetobutylicum</i> ATCC 824, 6e-17	Putative transcriptional regulator	<i>E. faecalis</i>
81 (DQ071644)	Human	NP_996712.1, bacteriophage phi LC3, 6e-60	Major tail protein	<i>E. faecalis</i>
107 (DQ071645)	Human	NP_816971.1, <i>Enterococcus faecalis</i> V583 plasmid pTEF1, 6e-23	Hypothetical protein	<i>E. faecalis</i>
7+ (DQ071646)	Waterfowl	X79341 (DNA), <i>Enterococcus faecium</i> , 0	23S rRNA gene	

validation testing, and this would be true regardless of how one goes about selecting putative markers.

The 15 markers identified in this study originated from human, cow, and cervid isolates. It is possible that our screening process invalidated additional markers that could be useful for source tracking. For example, markers identified by microarray hybridization that failed the PCR screening process could represent useful genetic polymorphisms. With enough polymorphisms (typically >10% difference between microarray probe and target sequences), we would see a reduced hybridization efficiency, but if our PCR primers were conserved for multiple alleles, then we would detect the presence of the gene in many hosts, regardless of host-specific internal sequence polymorphisms. Nevertheless, our simple screen for the presence or absence of markers was the most efficient means to identify relatively conserved markers for source tracking.

Of the 15 markers described here, the two cow markers (15 and 19) met all of our screening criteria to be defined as specific to cattle, and from both the hybridization and PCR screening results, we estimate that each of the cow markers is present in about 10% of cow isolates. One marker (48) was unambiguously associated with elk/deer, and we estimate that it is present in ca. 15% of *Enterococcus* isolates from cervid sources. We identified five human-specific markers. Four of these were present in an estimated 2% to 10% of human isolates, while the fifth was present in a larger proportion of isolates but may be present in isolates from other hosts. The remaining seven markers were specific for various combinations of hosts, and in most cases we found them to be clearly positive for nonhuman sources and negative for human sources.

We anticipated that our search would identify at least some host-specific markers that are involved in host recognition, such as cell surface proteins, or markers involved in the metabolism of host-specific nutrients. Three of the markers appear to be involved in metabolic pathways (markers 68, 77, and 93), but none are obvious outer membrane proteins that would be involved in host recognition and adhesion. Nevertheless, seven of the markers encode unknown and hypothetical proteins, and it is possible that these serve functional roles important to survival in specific hosts. It is very interesting that four of the markers appear to be involved in DNA replication and repair (markers 15, 40, 90, and 91), which are basic housekeeping functions. This implies that the host-specific strains are divergent enough to have evolved strain-specific housekeeping gene sequences. Others have found that host specificity can correlate with housekeeping gene divergence (e.g., *Flavobacterium psychrophilum*) (25). Finally, marker 81 encodes a major tail protein from a lactococcal bacteriophage. There is no documented evidence of the presence of this bacteriophage in *Enterococcus* spp.

The 15 markers from this study partition among discrete *Enterococcus* species. Cervid markers are present only in *E. mundtii* and *E. casseliflavus*, while *E. faecalis* harbors only human markers. *E. hirae* harbors a wider variety of markers, with human, cow, and cervid markers represented by this species. Nevertheless, the host-specific markers that we identified in *E. hirae* are still host specific, indicating the presence of host-adapted strains within a single species. This assumes that our 16S rRNA sequencing efforts correctly reflected species identification. In some cases, simply identifying the species present in compromised water may be useful as a preliminary and

TABLE 6. Distribution of markers among 32 positive control isolates

Host source	Positive control species	Isolate no.	Presence of marker ^a															
			Cow		Human						Elk/deer							
			15	19	66	67	68	77	81	107	40	48	89	90	91	93	94	
Cow	<i>E. hirae</i>	46	Y	Y									Y					
		144	Y	Y									Y					
		171	Y	Y									Y					
		184	Y	Y									Y					
		185	Y	Y									Y					
		186	Y	Y									Y					
		189	Y	Y									Y					
	190	Y	Y															
	<i>E. mundtii</i>	5											Y	Y	Y	Y	Y	
		192											Y	Y	Y	Y	Y	
Human	<i>E. faecalis</i>	401			Y		Y	Y			Y							
		403			Y													
		407			Y		Y	Y										
		409								Y								
		421			Y		Y	Y										
		424					Y											
		428								Y								
		432					Y											
		433					Y			Y								
		437			Y		Y	Y			Y							
		451								Y								
		453			Y													
		455			Y													
			<i>E. hirae</i>	423				Y					Y		Y			
				447				Y							Y			
Elk/deer	<i>E. casseliflavus</i>	502									Y							
		513									Y							
		514									Y	Y						
		516									Y	Y						
			<i>E. mundtii</i>	503										Y	Y	Y	Y	Y
				512										Y	Y	Y	Y	Y
				523										Y	Y	Y	Y	Y

^a The *esp* marker was not detected in any of the control isolates.

cost-effective means to test for sources of fecal contamination (29). We do not know what proportions of these species are represented in the normal flora of humans, cattle, and cervids.

Validating markers for bacterial source tracking is an open-ended process, and there are no universally accepted criteria to determine when a marker can be considered host specific. For example, it is possible that a host-adapted strain with a specific marker is present transiently within a nonspecific host, whereas it is found much more frequently in a target host (e.g., markers 89, 90, 91, 93, and 94; Table 4). It is also possible to have markers that are very host specific but that are so rare that they are not particularly useful for bacterial source tracking. The conclusion about the utility of the marker is also dependent on the assay that is employed. In our case, we tried to validate markers by testing them against a geographically diverse collection of specific and nonspecific isolates, and the tests involved a series of individual isolates or pools of gDNAs isolated from individual isolates or from broth enrichments. These isolates were obtained opportunistically from cooperators located around the United States. While we are not proposing validation criteria, we propose that any screening procedure should include isolates from diverse geographic origins.

As more library-independent markers are identified, there remain questions about how to implement these in a manner that provides an accurate, meaningful, and cost-effective assay

for BST. One strategy is to test material that has been cultivated from water samples. For example, Scott et al. (21) filtered water samples, incubated the filters on selective media for *Enterococcus* spp., and then briefly enriched the colonies in broth, extracted gDNAs, and tested this material for the presence or absence of the *esp* marker. This approach is highly desirable because of its simplicity, but it becomes problematic when markers from more than one host species are detected. Simple detection of a marker's presence or absence does not help to enumerate the proportion of each marker in the sample. Furthermore, if multiple species of *Enterococcus* are being detected (e.g., four species in the present work), then there could be problems with differential recovery with the enrichment media (although this can be tested). The latter complication could be circumvented by directly extracting gDNAs from membrane filtrates without enrichment (18). Depending on the filtration method that is employed, however, this strategy can encounter significant sensitivity limitations, and filtering large volumes of water can be expensive.

Another approach being considered in the BST field is to develop quantitative real-time PCR assays for library-independent markers. In theory, this will permit the practitioner to enumerate each marker either in a multiplex reaction or through a series of individual reactions for each marker of interest. This analysis usually assumes that quantification is feasible based on

a standard curve. For environmental samples, we know that each gDNA extraction can have different levels of PCR inhibition (18), so unless the standard curve is incorporated into the sample extract, there can be a fair amount of uncertainty for quantification. This limitation could be circumvented by using a relative ratio analysis where real-time PCR results could be expressed as ratios within each sample. Even this method, however, does not circumvent a larger question about marker representation, that is, if accurate enumeration or ratios can be estimated, how does this relate to the original contribution from each host animal? For example, suppose that we detect equal numbers of cattle and human targets within a sample. If the cattle marker is found in 1% of all cattle isolates and the human marker is found in 10% of all human isolates, then our original conclusion about a 50:50 contribution is incorrect by a 10-fold difference. More work will be needed to assess the proportion and variance for markers that are shed by host animals. More work is also needed to assess the sample-to-sample variance and differential survivorship of bacterial strains that harbor different BST markers.

Regardless of the method that is chosen, we advocate simplicity so that institutions that need to address total maximum daily load requirements can obtain accurate and timely information at the lowest possible cost. If we recognize that there is considerable variance in the temporal and spatial distributions of host-specific markers, then one strategy may be to use simple presence-absence detection (by PCR), but conclusions would be based on multiple site visits. That is, conclusions would be based on positive "events" rather than investing considerable effort to carefully enumerate individual single samples. For example, if a site is visited 10 times, and on three occasions markers are detected for human feces, but cattle fecal markers are detected for all 10 samples, then one can draw the reasonable conclusion that the contribution from cattle is the first concern.

ACKNOWLEDGMENTS

We gratefully acknowledge assistance with known-source fecal collections (M. Sanderson, W. Sisco, C. Calloway, D. Stoeckel, D. Nydam, J. Farmer, and G. Bright). D. Stoeckel and J. W. Santo Domingo provided useful discussions and suggestions. M. Oatley, M. Krug, A. Khachatryan, and R. McClanahan provided invaluable technical assistance.

This project was funded by USDA NRI contract 2002-35102-12374 and by the Agricultural Animal Health Program at the College of Veterinary Medicine, Washington State University, Pullman, Wash.

REFERENCES

- Bernhard, A. E., and K. G. Field. 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* **66**:1587-1594.
- Bernhard, A. E., and K. G. Field. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **66**:4571-4574.
- Call, D. R., M. K. Bakko, M. J. Krug, and M. C. Roberts. 2003. Identifying antimicrobial resistance genes with DNA microarrays. *Antimicrob. Agents Chemother.* **47**:3290-3295.
- Call, D. R., M. K. Borucki, and T. E. Besser. 2003. Mixed-genome microarrays reveal multiple serotype- and lineage-specific differences among strains of *Listeria monocytogenes*. *J. Clin. Microbiol.* **41**:632-639.
- Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw. 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Appl. Environ. Microbiol.* **67**:1503-1507.
- Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field. 2005. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:3184-3191.
- Dick, L. K., and K. G. Field. 2004. Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR assay for 16S rRNA genes. *Appl. Environ. Microbiol.* **70**:5695-5697.
- Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl. Environ. Microbiol.* **66**:2572-2577.
- Field, K. G., E. C. Chern, L. K. Dick, J. Fuhrman, J. Griffith, P. A. Holden, M. G. LaMontagne, J. Le, B. Olson, and M. T. Simonich. 2003. A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. *J. Water Health* **1**:181-194.
- Graves, A. K., C. Hagedorn, A. Teetor, M. Mahal, A. M. Booth, and R. B. Reneau, Jr. 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. *J. Environ. Qual.* **31**:1300-1308.
- Hagedorn, C., J. B. Crozier, K. A. Mentz, A. M. Both, A. K. Graves, N. J. Nelson, and R. B. Reneau, Jr. 2003. Carbon source utilization profiles as a method to identify sources of faecal pollution in water. *J. Appl. Microbiol.* **94**:792-799.
- Hagedorn, C., S. L. Robinson, J. R. Filtz, S. M. Grubbs, T. A. Angier, and R. B. Reneau, Jr. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Appl. Environ. Microbiol.* **65**:5522-5531.
- Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segars. 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. *J. Environ. Qual.* **31**:1273-1278.
- Harwood, V. J., J. Whitlock, and V. Withington. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. *Appl. Environ. Microbiol.* **66**:3698-3704.
- Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky. 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. *Appl. Environ. Microbiol.* **70**:4478-4485.
- Lemarchand, K., L. Masson, and R. Brousseau. 2004. Molecular biology and DNA microarray technology for microbial quality monitoring of water. *Crit. Rev. Microbiol.* **30**:145-172.
- Leung, K. T., R. Mackereth, Y. C. Tien, and E. Topp. 2004. A comparison of AFLP and ERIC-PCR analyses for discriminating *Escherichia coli* from cattle, pig and human sources. *FEMS Microbiol. Ecol.* **47**:111-119.
- Loge, F. J., D. E. Thompson, and D. R. Call. 2002. PCR detection of specific pathogens in water: a risk-based analysis. *Environ. Sci. Technol.* **36**:2754-2759.
- Meays, C. L., K. Broersma, R. Nordin, and A. Mazumder. 2004. Source tracking fecal bacteria in water: a critical review of current methods. *J. Environ. Manage.* **73**:71-79.
- Ram, J. L., R. P. Ritchie, J. Fang, F. S. Gonzales, and J. P. Selegan. 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. *J. Environ. Qual.* **33**:1024-1032.
- Scott, T. M., T. M. Jenkins, J. Lukasik, and J. B. Rose. 2005. Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. *Environ. Sci. Technol.* **39**:283-287.
- Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik. 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. *Appl. Environ. Microbiol.* **69**:1089-1092.
- Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik. 2002. Microbial source tracking: current methodology and future directions. *Appl. Environ. Microbiol.* **68**:5796-5803.
- Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner. 2002. Microbial source tracking: state of the science. *Environ. Sci. Technol.* **36**:5279-5288.
- Soule, M., K. Cain, S. LaFrentz, and D. R. Call. 2005. Combining suppression subtractive hybridization and microarrays to map the intraspecific phylogeny of *Flavobacterium psychrophilum*. *Infect. Immun.* **73**:3799-3802.
- Stewart, J. R., R. D. Ellender, J. A. Gooch, S. Jiang, S. P. Myoda, and S. B. Weisberg. 2003. Recommendations for microbial source tracking: lessons from a methods comparison study. *J. Water Health* **1**:225-231.
- Stoekel, D. M., M. V. Mathes, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. L. O'Brien, T. W. Fenger, M. Samadpour, K. M. Strickler, and B. A. Wiggins. 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. *Environ. Sci. Technol.* **15**:6109-6117.
- Warsen, A. E., M. J. Krug, S. LaFrentz, D. R. Stanek, F. J. Loge, and D. R. Call. 2004. Simultaneous discrimination between 15 fish pathogens by using

- 16S ribosomal DNA PCR and DNA microarrays. *Appl. Environ. Microbiol.* **70**:4216–4221.
29. Wheeler, A. L., P. G. Hartel, D. G. Godfrey, J. L. Hill, and W. I. Segars. 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. *J. Environ. Qual.* **31**:1286–1293.
30. Wiggins, B. A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman. 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Appl. Environ. Microbiol.* **65**:3483–3486.
31. Wiggins, B. A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner. 2003. Use of antibiotic resistance analysis for representativeness testing of multi-watershed libraries. *Appl. Environ. Microbiol.* **69**:3399–3405.