EPA/600/R-07/123 | December 2007 | www.epa.gov/athens

Evaluation of Selected DNA-based Technology in Impaired Watersheds Impacted by Fecal Contamination from Diverse Sources

Ecosystems Research Division, Athens, GA 30605
National Exposure Research Laboratory
Office of Research and Development

EPA/600/R-07/123
December 2007

Evaluation of Selected DNA-based Technology in Impaired Watersheds Impacted by Fecal Contamination from Diverse Sources

by

Marirosa Molina
Ecosystems Research Division
National Exposure Research Laboratory
Athens, GA 30605

U.S. Environmental Protection Agency
Office of Research and Development
Washinton, DC 20460

Notice

The information in this document has been funded by the United States Environmental Protection Agency. It has been subjected to the Agency's peer and administrative review, and it has been approved for publication as an EPA document. Mention of trade names of commercial products does not constitute endorsement or recommendation for use.

Abstract

Fecal pollution of surface waters is a top reason for impairment, as reported in the U.S. Environmental Protection Agency's report on the quality of the Nation's waters. To be able to develop and implement TMDLs for impaired aquatic resources, it is imperative to determine the sources of the contamination. One tool used to determine the sources of bacterial fecal contamination is to apply a microbial source tracking approach to the system of interest. Microbial source tracking (MST) approaches are based on the assumption that specific strains of bacteria, genetic fingerprints, or DNA-based markers are associated with specific host species. Because accurate source identification of fecal contamination is essential in MST, more sensitive, selective and reliable molecular markers are required. The two types of genotypic methods that have been applied widely in a variety of environments can be classified as library-independent (LI) and library-dependent (LD). For both types, the temporal and spatial stability of selected genotypes are aspects that need to be evaluated, and these aspects are often times missing when applying MST to environmental samples. LD-MST methods require the development of large databases comprised of source-specific isolates. Once a source-specific fingerprint has been identified, the temporal and spatial variability of that particular genotype still needs to be validated. LI-MST is based on the application of culture-independent methods such as amplification of DNA from environmental samples using 16S rDNA markers in combination with polymerase chain reaction (PCR). However, cross-reactivity of some of the 16S rDNA markers used in field studies has prompted the development of alternative PCR assays using metagenomic markers specific for bovine feces. In this study, we report on the comparison of selected LD and LI methodologies, their usability as rapid and reliable methods for developing and applying markers to various environmental scenarios, and the stability of these markers under various spatial and temporal conditions. From our results, we concluded that library production is highly time and resource consuming. Its application is probably appropriate in very specific scenarios where discrimination among a few, selective sources is necessary. In contrast, application of DNA, PCR-based markers yielded fairly rapid results and has the capability to screen multiple scenarios in a short period of time. Once stability and cross-amplification aspects have been addressed, this latter method can be a highly efficacious approach to determine sources of contamination in a variety of scenarios.

Table of Contents

ENVIRONMENTAL ISSUE **6**

RESEARCH GOALS **8**

DESCRIPTION OF METHODS USED IN THIS RESEARCH **8**

- **AMPLIFIED FRAGMENT LENGTH POLYMORPHISM** 8
- **BOX-PCR ANALYSIS** 10
- **HOST SPECIFIC 16S-rDNA MARKERS** 11
- **METAGENOMIC MARKERS** 12

GENERAL RESEARCH APPROACH **13**

METHODOLOGY **15**

- **SAMPLING LOCATIONS FOR 16S-RDNA AND METAGENOMIC MARKERS** 15
- **SAMPLE COLLECTION** 18
- **PHYSICO-CHEMICAL AND MICROBIOLOGICAL METHODS** 18
- **DNA EXTRACTION AND PCR AMPLIFICATION** 19
- **SOURCES FOR AFLP AND BOX-PCR ANALYSIS** 21
- **AFLP AND BOX-PCR PROCEDURES** 21

RESULTS AND DISCUSSION **22**

- **EVALUATION OF LIBRARY-INDEPENDENT METHODS** 22
  - ***Comparison of 16S rDNA-based vs. metagenomic marker performance in farms impacted by cattle contamination*** *22*
  - ***Relationship between enterococci enumeration and the occurrence of molecular markers*** *24*
  - ***Evaluation of human specific 16S-rDNA markers in freshwater streams impacted by rural non-point sources in Puerto Rico*** *28*
- **EVALUATION OF LIBRARY-DEPENDENT METHODS** 34
  - ***Seasonal Distribution of Enterococci Isolates*** *34*
  - ***Some methodological considerations developing the phylogeny of Enterococcus strains using AFLP*** *35*
  - ***Phylogeny of E . faecalis, E. hirae ,and E. casseliflavus strains using AFLP*** *36*
  - ***Comparison of AFLP and BOX-PCR analysis*** *39*

CONCLUSIONS **41**

SIGNIFICANCE OF RESEARCH **43**

FUTURE DIRECTIONS **44**

REFERENCES **45**

List of Figures

Figure 1 Main steps of the AFLP Procedure 10

Figure 2 Experimental scheme to perform 16S rDNA and metagenomic marker analyses..14

Figure 3 General experimental design to isolate and fingerprint enterococcal species...15

Figure 4 Four sites were sampled in Farm 1, three sites were located along the stream while site 4 was located in a pond used by the cattle for bathing and drinking. Aerial photo courtesy of GlobeXplorer.com. 16

Figure 5 Sampling sites related to WS2. Sites 1-7 are located within the farm boundary, 8 and 9 are located in a buffer zone between the farm and a subdivision, and 10-12 are located within a subdivision. Aerial photo courtesy of GlobeXplorer.com. 17

Figure 6 Relationship between the monthly enterococcal counts and the average frequency of the DNA markers per month in WS1 27

Figure 7 Relationship between monthly enterococcal counts and the average frequency of the DNA markers per month in WS2 28

Figure 8 Seasonal distribution of enterococcal species in impacted streams by cattle contamination 34

Figure 9 *Enterococcus hirae* phylogenetic tree derived from AFLP fingerprints .38

Figure 10 Typical BOX-PCR gel image produced with *E. faecalis* isolates.........40

List of Tables

Table 1 Frequency (±95% CI) of 16S rDNA-based Bacteroides and metagenomic markers in water samples from two watersheds affected by cattle contamination. Watershed 1 (WS1) receives direct impact from cattle, while watershed 2 (WS2) only receives contamination through runoff. Only markers with a frequency between 0.10 and 0.90 were used for the logistic regression analysis 24

Table 2 Enterococcal abundance {CFU/100ml) in Watershed 1 and Watershed 2. Sites were divided based on influence by cattle or type of water resource (streams vs. ponds). 26

Table 3 Description of samples collected in the Rio Añasco Basin, Añasco, Puerto Rico from August 3-14, 2006...29

Table 4 Bacteroidetes 16S rRNA gene marker hits in water samples collected in the Rio Añasco Basin, Añasco, Puerto Rico. The numbers indicate the times the individual primer set was found in each water sample after one amplification round (1x) 32

Table 5 Comparison of the advantages and disadvantages of the BOX-PCR and AFLP methodologies...41

Environmental Issue

The U.S.EPA TMDL 303(d) list fact sheet has indicated that fecal pollution is the #1 cause of impairment in most states, accounting for up to 13% of all reported impairments. Cost-effective development and implementation of TMDLs for impaired aquatic resources requires the rapid and accurate determination of the sources of contamination. Commonly used microbial water quality assessment methods measure densities of fecal indicator bacteria, but do not provide information on the possible sources of contamination producing the elevated indicator concentration.

One tool used to determine the sources of bacterial fecal contamination is to apply a microbial source tracking approach to the system of interest. Microbial source tracking (MST) is based on the assumption that specific strains of bacteria, genetic fingerprints, or DNA-based markers are associated with specific host species. Because accurate source identification of fecal contamination is the objective of MST, more sensitive, selective and reliable molecular markers are required. The two types of genotypic methods that have been applied widely in a variety of environments can be classified as library-independent (LI) and library-dependent (LD). For both types, the temporal and spatial stability of the selected genotypes are aspects that need to be evaluated, and those aspects are often times not well characterize when applying MST to environmental samples. LD-MST methods require the development of large databases comprised of source-specific isolates (Ritter *et al.*, 2003; Wiggins et al., 2003). Once a source-specific fingerprint has been identified, the temporal and spatial

variability of that particular genotype still needs to be characterized. LI-MST is based on the application of culture-independent methods such as amplification of DNA from environmental samples using polymerase chain reaction (PCR). One of the genes that has been widely used for this application is the gene coding for the 16S rRNA, that has been demonstrated to have host specificity (Bernhard & Field, 2000; Layton *et al.*, 2006). However, one drawback of this technique is a degree of cross-reactivity observed with some of the 16S rDNA markers when used in field studies because they target highly conserved 16S regions (Shanks *et al.*, 2006; Shanks et al., 2007). This cross-reactivity prompted the development of alternative PCR assays using metagenomic markers specific for bovine feces (Shanks *et al.*, 2006b). These bovine metagenomic markers were successfully tested in the latter study with little cross reactivity on a large number of bovine feces collected from a variety of locations across the U.S. However, although some spatial variability was covered in that study, a more detailed evaluation of the temporal and spatial variability of the markers was still required to determine their environmental stability and robustness.

This research supports the second long-term goal (LTG 2) established in ORD's Water Quality Multiyear Plan for the protection of watersheds and aquatic communities: "provide the tools to assess and diagnose impairment in aquatic systems and the sources of the associated stressors".

Research Goals

The overall objective of this research was to evaluate the temporal and spatial applicability of DNA-based techniques and markers to identify sources of fecal contamination in a variety of environmental scenarios.

Description of Methods Used in this Research

The methods evaluated were divided between library-dependent (LD) and library-independent (LI) approaches. The library dependent methods were used with a library of enterococci markers isolated from cattle farms (Molina *et al.*, 2007). The two LD methods included amplified fragment length polymorphism (AFLP) and repetitive fragment polymerase chain reaction (PCR) with Box-PCR primers (Box-PCR). Two LI-PCR methods were also compared: 16S rDNA-based Bacteroidales markers and metagenomic markers, both specific to cattle.

Amplified fragment length polymorphism. AFLP consists of selective amplification of restriction fragments resulting from the digestion of total genomic DNA using PCR. The technique has the capability to inspect an entire genome for polymorphisms and is highly reproducible. Molecular genetic polymorphisms are identified by the presence or absence of fragments after restriction and amplification of genomic DNA. AFLP involves four basic steps after DNA extraction from pure cultures: DNA digestion with restriction enzymes; ligation with oligonucleotide adapters; selective amplification with labeled primers; and gel-based analysis of amplified fragments. See Figure 1 for a representation of the procedure. Characteristics of the AFLP procedure include: PCR and

fragment analysis are relatively fast to perform if using automated machines; the entire genome is inspected for polymorphic fragments; uses small amounts of genomic DNA and the DNA concentration does not affect the reproducibility of the assay; provides 50 to 200 fragments per genome assayed allowing for easy identification of polymorphisms; is highly reproducible; and taxon-specific primer sets are not required (commercially available primers work with a large variety of genomes). In addition, the technique can be applied to a large variety of DNA samples including plants, animals, human, and microbial genomes. Some of the most common applications have included generating high resolution genetic maps in plants and animals, analysis of the genetic diversity in plants and animals, characterization of mammalian genotypes, genotypic analysis and epidemiological typing of bacteria, genotypic classification of fungi, and the characterization and classification of pathogens (Blears *et al.*, 1998).

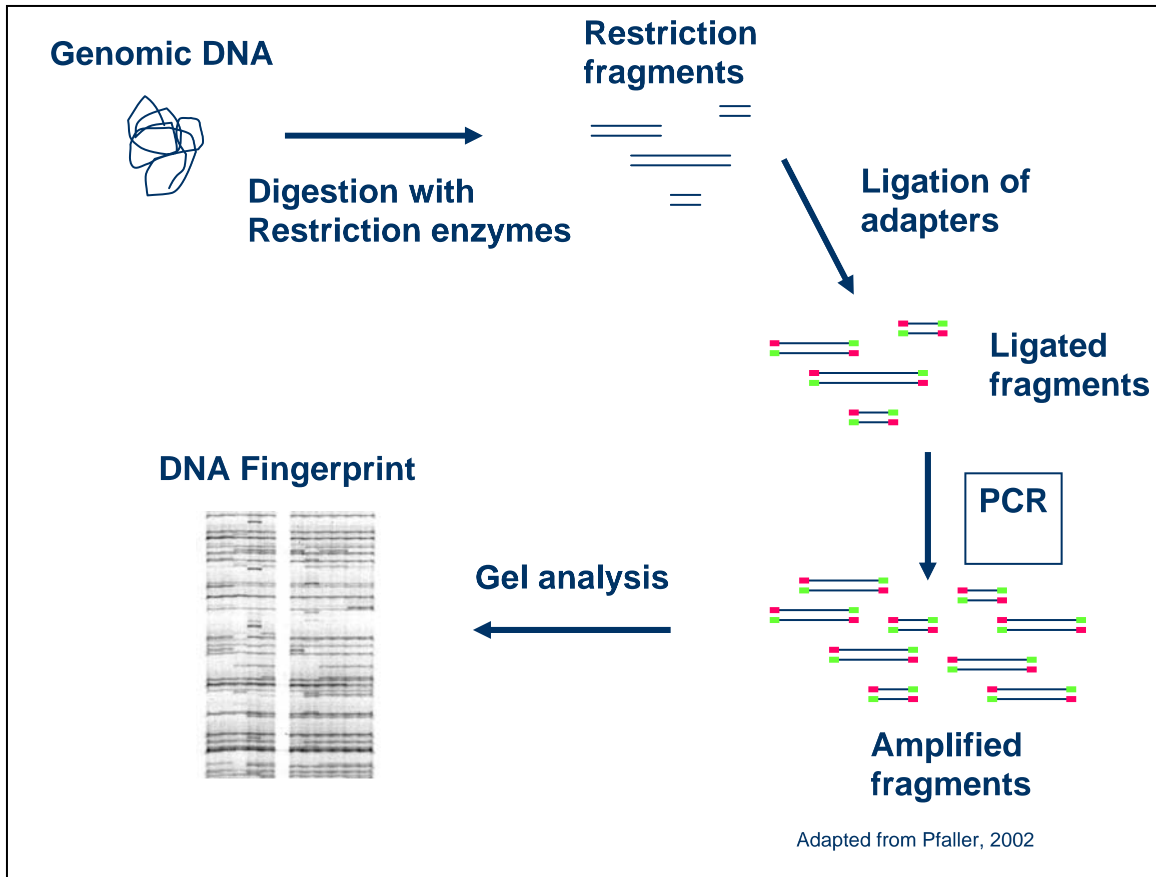


Figure 1 Main steps of the AFLP Procedure

Box-PCR Analysis. BOX-PCR is another PCR-based DNA fingerprinting technique based on amplification of the interspersed repetitive sequences (rep-PCR) found in the DNA of many bacterial species (Koeuth *et al.*, 1995). The BOX element originally described for *Streptococcus pneumoniae* consists of three, highly conserved, interspersed, repetitive sequences: boxA, boxB, and boxC (Martin *et al.*, 1992) that contain 59, 45 and 50 basepairs in length, respectively. BOXA1R and BOXA2R primers are based on the boxA sequence, and have been widely applied for rep-PCR amplification of DNA from a wide variety of bacterial species (Koeuth *et al.*, 1995), including *Enterococcus*. A comparison of BOX-PCR to pulse field gel electrophoresis (PFGE), identified as

the gold standard for *Enterococcus* sp. fingerprinting, indicated that both techniques yield very similar results at the subspecies level for *Enterococcus faecalis* (Malathum *et al.*, 1998).

Host specific 16S-rDNA markers. The majority of molecular tools currently being applied for microbial source tracking rely on the development of an extensive library of cultured isolates to which DNA fingerprints from environmental samples can be compared. The two aforementioned methods fall into this category. LD methods are labor-intensive and limit the target indicator bacteria to those that can be readily grown in a laboratory and can also survive outside the intestine (Simpson *et al.*, 2002). Combining technological advances in molecular biology, such as polymerase chain reaction (PCR) and 16S rDNA gene sequence analysis, has provided powerful tools for characterizing microbial populations without the need for cultivation of the targeted indicators. These combined techniques have become very useful for of MST application. For example, PCR amplification of 16S rRNA gene sequences of the genera *Bacteroides-Prevotella* has proven useful for the identification of specific hosts, such as human, cattle, horses, and pigs (Allsop & Stickler, 1985, Bernhard & Field, 2000b, Dick *et al.*, 2005, Kreader, 1995). These anaerobic bacteria are restricted to the intestinal environment of warm-blooded animals. Unlike some other fecal coliform bacteria, these Bacteroidetes do not survive long in water, and make up 30 to 40% of the total fecal bacteria (Harmsen *et al.*, 1999; Layton *et al.*, 2006), which could account for up to 10% of the fecal mass. Therefore, these anaerobic bacteria could be used as suitable indicators of species-specific

contamination. Prior to the development of culture-independent molecular methods, the use of *Bacteroides* as indicators was limited because of the difficulty to grow them in culture.

Metagenomic Markers In addition to PCR amplification of specific 16S rRNA genes, recent development of another culture-independent technique, genome fragment enrichment, also seems promising for selecting for host-specific metagenomic markers (Shanks *et al.*, 2006a). This technique enriches for genes that are specific in host organisms by subtracting the genes that are common in other organisms. The metagenomic approach not only targets the 16S rRNA gene, but all genes involved in bacterial-host interactions, such as surface proteins (Shanks *et al.*, 2006a). One drawback that the 16S rRNA gene of *Bacteroides*-like species seems to have is its cross-reactivity with non-target fecal sources (Lamendella *et al.*, 2007). This is especially true for the cattle-specific markers. The metagenomic markers developed for bovine sources are a good alternative that could possibly reduce the identification of false positives due to that cross-reactivity. The bovine metagenomic markers developed by Shanks et al. (2006b) were successfully tested with minimal cross reactivity on 148 different bovine feces collected from a variety of locations across the U.S. However, although the latter assays were tested against fecal samples obtained from different regions, more detailed site tests of the temporal and spatial variability of the markers are still required to determine their environmental stability and robustness.

General Research Approach

This research was divided into two general approaches:

- Evaluation and comparison of the presence of 16S-rDNA and metagenomic markers in both water and sediment samples collected from two watersheds associated with cattle farms under different management practices (see Figure 2), and from a rural community serviced by individual household septic wastewater treatment systems.
- Comparison of amplified fragment length polymorphism (AFLP) and repetitive polymerase chain reaction with BOX-primer (BOX-PCR) methodologies to genotype an *Enterococcus sp*. source library, and determine the usability of each methodology for host-specific source identification (see Figure 3).

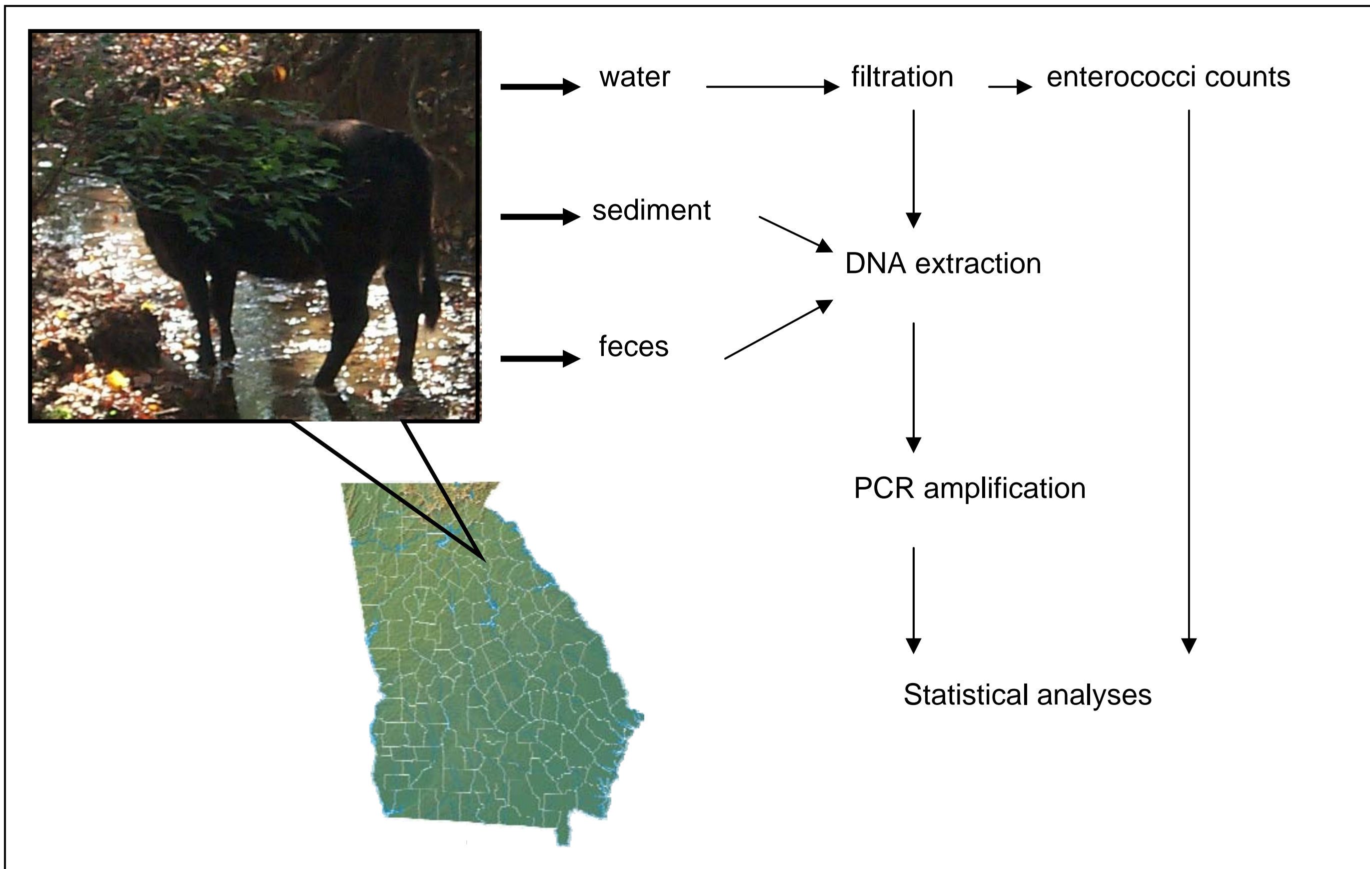


Figure 2 Experimental scheme to perform 16S rDNA and metagenomic marker analyses.

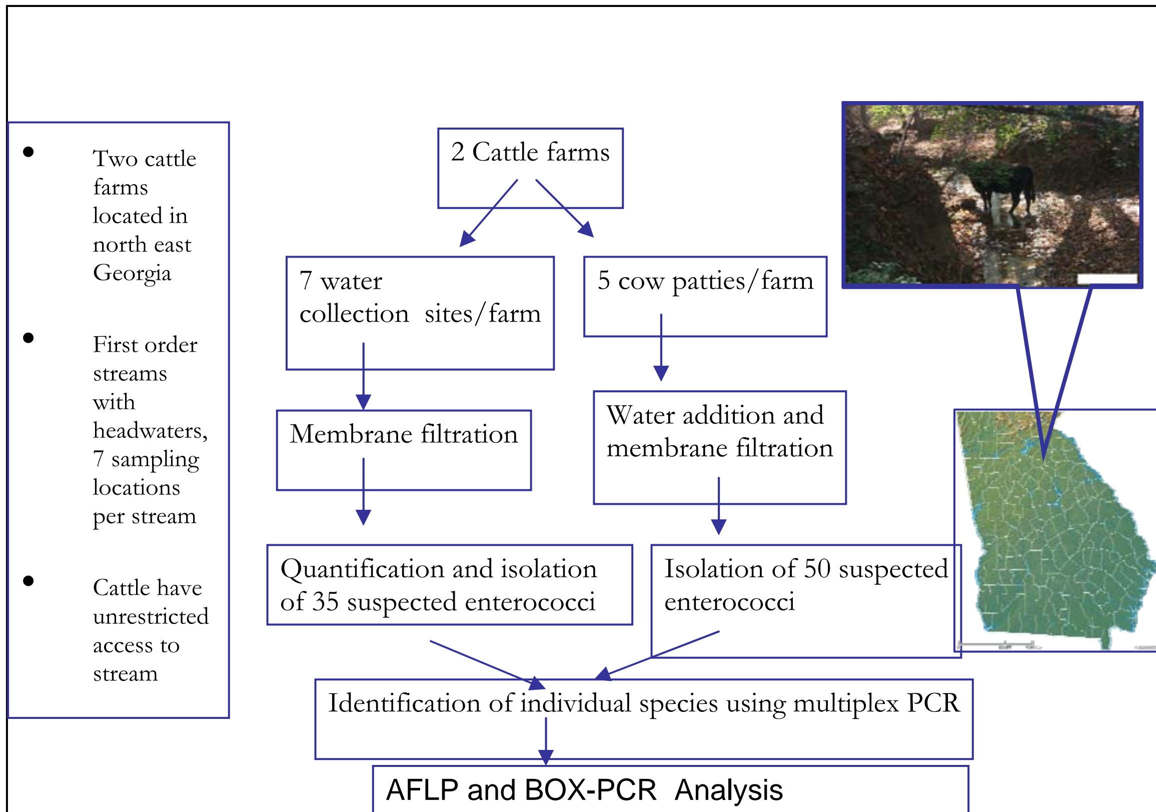


Figure 3 General experimental design to isolate and fingerprint enterococcal species

Methodology

Sampling locations for 16S-rDNA and Metagenomic Markers. The study sites to compare the metagenomic markers and 16S-rDNA primers consisted of two watersheds associated with cattle farms. Watershed 1 (WS1) flows across Farm A located in Madison County, GA. Watershed 2 (WS2) starts in Farm B located in USDA-owned land in Watkinsville, GA. In WS1, samples were collected from 4 sites along a creek and a pond (Figure 4).

Figure 4 Four sites were sampled in WS1 (Farm 1), three sites were located along the stream while site 4 was located in a pond used by the cattle for bathing and drinking. Aerial photo courtesy of GlobeXplorer.com.

Cattle had direct access to all sampling sites except for site 1 that was located upstream from the farm, outside of the property fence approximately 0.13 miles downstream from the origin of the stream. Sites 2 and 3 were located in the middle and end of the stream crossing the farm, respectively. Site 4 was located in a pond used by the cattle for drinking and bathing. On average, 60 head of cattle were present on the farm during the course of this study. Wildlife, such as geese and deer, also could contribute to the fecal sources impacting both water bodies in this farm.

There were 12 sampling sites in WS2; seven of the sites were located along the headwater stream and a pond within Farm B, while five sites were

located in the same creek downstream, outside the farm (Figure 5). Sites 5 and 11 consisted of agricultural and community ponds, respectively. An average of 140 head of cattle were kept and rotated among 16 fenced pastures in Farm B during our study. The cattle had no access to the stream or the pond at this farm. Other possible fecal sources affecting the stream and ponds in this watershed include wildlife such as deer, geese and raccoons. Neither watershed was deemed significantly impacted by human fecal pollution.

Figure 5 Sampling sites related to WS2 (Farm B). Sites 1-7 are located within the farm boundary, 8 and 9 are located in a buffer zone between the farm and a subdivision, and 10-12 are located within a subdivision. Aerial photo courtesy of GlobeXplorer.com.

A separate study site was selected to evaluate the human-specific 16S-rDNA markers. The site consisted of a rural community located in the town of Añasco, Puerto Rico. Water from five sampling locations was collected over a two week period. The sampling locations included one site along an intermittent creek that crossed the community; three sites were located in the Casey River basin (upstream and downstream from the community); and one site consisted of a shallow well (30 feet).

Sample collection. For cattle primers, water and fecal samples were collected on a monthly basis between September 2005 and February 2007. Water samples were collected in sterilized 1-liter bottles, kept on ice for transport to the laboratory and processed for enterococci enumerations and nucleic acid extractions within 6 hours after collection. Two fecal samples per sampling event were collected aseptically from each farm. Fecal samples were stored at -20°C until processed.

The samples collected to test the human-specific Bacteroidetes primers were collected in collaboration with an ongoing study sponsored by the Puerto Rico Water Resources and Environmental Research Institute in an effort to provide information for the development of TMDLs for the Rio Añasco. Water samples (100, 250, and 500 ml) were filtered through polycarbonate filters (0.2 µm). The filters were transferred to microcentrifuge tubes and stored at -20° C, then shipped on ice overnight to our laboratory.

Physico-chemical and microbiological methods. The temperature and pH of water samples were measured on-site using a portable pH meter, Orion

250A plus (Thermo Orion, Beverly, Mass.). Daily precipitation data for WS1 and WS2 were obtained from station ID 092517 of the National Oceanic and Atmospheric Administration (http://hurricane.ncdc.noaa.gov/dly/DLY) and the Georgia Automated Environmental Monitoring Network (http://www.georgiaweather.net/), respectively. Water sample turbidity was measured using a 2020 Turbidimeter (LaMotte Co., Chesterfield, MD) according to the manufacturer's instructions. Enterococcal densities of the water samples were determined using the membrane filtration technique described in EPA method 1600. The colonies were counted twice after 24 and 42 hour incubation at 41°C.

DNA Extraction and PCR amplification. In both WS1 and WS2, approximately 100 ml water samples, and 0.2-0.25 g of cattle feces were used for DNA extractions using an UltraClean Soil DNA Kit (MoBio Inc., California) according to the manufacturer's instructions with some modifications. Specifically, water samples were filtered onto polycarbonate filter membranes (0.22 μm; Millipore Inc., Bedford, MA). Each filter was then transferred to a 6 ml sterile tube containing bead solution and solution S1, and vortexed for 10 min. Inhibitor removal solution (IRS) was added after solution S2, followed by the steps in the manufacturer's instructions. The nucleic acid fraction was eluted to 65 μl of Tris-EDTA buffer. DNA was quantified photometrically using a NanoDrop ND-1000 UV/Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE), and the DNA concentration was adjusted approximately to 10 ng/μl.

The Puerto Rico water samples were also extracted using the MoBio kit. The primers used included a general Bacteroides-Prevotella marker (32F or Gen-Bac), two human-Bacteroides markers (HF183 and HF654), and two cattle-Bacteroides primers (CF128 and CF193). Gels were examined on 1.5% agarose, mostly for 90 min at 90volts, with one or two exceptions at 100 volts for 60min.

PCR assays were performed using GoTaq Green master mix (Promega, Madison, WI) with either 16S rDNA-based *Bacteroidales*-specific primer sets or six cattle-specific metagenomic primer sets. The annealing temperature for each PCR assay was determined using a gradient PCR. The thermal cycling conditions for the 16S rDNA-based markers were an initial denaturation of 2.5 min at 94°C, followed by 30 cycles of denaturation at 94°C for 30 sec each, annealing at an optimized temperature for each primer set for 30 sec plus extension at 72°C for 1 min, and a final extension of 5.5 min at 72°C. The thermal cycling conditions for the metagenomic markers were 3 min of initial denaturation at 94°C, followed by 35 cycles of 1 min each of denaturation (94°C), annealing and extension (72°C), and a final extension step of 5 min at 72°C. Amplification products were visualized on a 2 % agarose gel stained with 0.2X SYBR Safe DNA gel stain (Invitrogen). The limit of detection for each molecular marker set was determined by PCR using serial dilutions of the extracted bovine fecal DNA as templates, starting at 10 ng/µl. Negative controls included DNA extracts from sterilized nanopure water and no DNA template reactions, while DNA extracts

from feces freshly obtained at each sampling event were used as positive controls.

Sources for AFLP and BOX-PCR Analysis. The library of enterococci used for the fingerprinting analyses consisted of 1600 isolates collected over a seasonal cycle at two separate bovine farms where cattle had unrestricted access to the streams at all times (Molina, 2005). Samples were collected from pre-farm (non-impacted, upstream-from-the-farm) stream sites, farm stream sites (impacted), and fecal matter. The library of enterococcal species was developed by isolating colonies from mEI plates and identifying them at the species level using a multiplex PCR procedure (Jackson *et al.*, 2004).

AFLP and BOX-PCR Procedures. Genomic DNA extraction from each *Enterococcus* isolate was performed using a Qiagen DNeasy Tissue Kit. The AFLP procedure was adapted from (Antonishyn *et al.*, 2000). AFLP restriction and ligation was performed using *Hin*dIII, and *Mbo*I. Digested genomic DNA was amplified in parallel reactions using two different selective primer sets, *Mbo*I-AC and *Mbo*I-CTG. The BOX-PCR procedure was an adaptation from (Malathum *et al.*, 1998). BOX-PCR Amplification was performed using Gitschier buffer (pH 8.0) (Kogan *et al.*, 1987), and a BOXA2R primer. To perform fragment analysis, the PCR products were electrophoresed through 6% polyacrylamide denaturing gels with a well-to-read distance of 30 cm for 3 hours on a MJ Research BaseStation 51 DNA Fragment Analyzer. The parallel reactions were run on separate gels with a custom size standard in each lane (BioVentures), allowing accurate sizing of fragments in the 50-600 bp range. For the phylogenetic

analyses, we analyzed gel images using BioNumerics v3.0. Dendrograms were created from the *Mbo*I-AC and *Mbo*I-CTG fingerprints using a curve-based similarity coefficient (Pearson correlation) with the unweighted pair group method (UPGMA).

Results and Discussion

Evaluation of Library-Independent Methods

Comparison of 16S rDNA-based vs. metagenomic marker performance in farm waters impacted by cattle fecal contamination. The general 16S rDNA marker (32F) was detected in all sampling sites at a very high frequency (81%), except in the sites related to the ponds or their effluents (17%) (Table 1). This general marker was followed in frequency order by the cattle-specific 16S rDNA marker (CF128), and then by the metagenomic markers Bac 2, 1, 5, and 3. The metagenomic markers as a whole were found to be 41-60% less frequent than the 16S rDNA cattle marker in stream waters under direct impact (WS1), and between 3-5% less frequent in stream water under indirect impact (WS2), depending on the sampling site. These results suggest that the metagenomic markers are less sensitive than the 16S-rDNA based markers, they are less stable in the environment, or their presence in cattle is more variable. The fact that the metagenomic markers were not found in every single cattle patty sampled at a given time (data not shown) points to the possibility of a higher variability in cattle manure, but it does not discard the other two possibilities. Nevertheless, the CF 128 marker was found in relatively low

frequencies (3-9%) in WS2 versus the directly impacted stream of WS1 (71-94%) (Table 1). This indicates a rather low impact of cattle fecal contamination reaching the stream water in WS2 through run-off compared to the direct inputs in WS1, even though enterococci numbers were rather high in both farm streams.

Table 1. Frequency (±95% CI) of 16S rDNA-based Bacteroides and metagenomic markers in water samples from two watersheds affected by cattle contamination. Watershed 1 (WS1) receives direct impact from cattle, while watershed 2 (WS2) only receives contamination through runoff. Only markers with a frequency between 0.10 and 0.90 were used for the logistic regression analysis

	Marker Frequency ±95% CI					
Site	**Bac32F**	**CF128F**	**Bac1**	**Bac2**	**Bac3**	**Bac5**
WS1-1	**0.76 ±0.09**	**0**	**0**	**0**	**0**	**0**
WS1-2, 3	**1.00 ±0.00**	**0.94 ±0.02**	**0.62 ±0.08**	**0.65 ±0.08**	**0.26 ±0.06**	**0.59 ±0.08**
WS1- 4	**0.88 ±0.05**	**0.71 ±0.10**	**0.12 ± 0.05**	**0.18 ±0.07**	**0 ±0.00**	**0.12 ±0.05**
WS2- 1-4	**0.72 ±0.05**	**0.09 ±0.02**	**0.03 ±0.01**	**0.06 ±0.01**	**0.04 ±0.01**	**0.04 ±0.01**
WS2- 5, 6, 11, 12	**0.17 ±0.04**	**0.03 ±0.01**	**0**	**0.03 ±0.01**	**0**	**0**
WS2- 7-10	**0.70 ±0.06**	**0.04 ±0.01**	**0**	**0.04 ±0.01**	**0**	**0**

Relationship between enterococci enumeration and the occurrence of molecular markers. Enterococci counts were performed for all locations where the markers were tested to establish the relationship with the alternative markers under the two types of farm management. The geometric mean of the enterococcal numbers in the areas with the highest probability of cattle impact

ranged from approximately 24 to 1924 CFU/100ml in WS2 and WS1, respectively (Table 2). The counts taken at the ponds or pond outflows were the lowest (site 4 in WS1 and sites 5, 6, 11, and 12 in WS2), being 93 and 4 CFU/100 ml in WS1 and WS2, respectively. The upstream locations (site 1 in both watersheds), exhibited counts of 74 and 17 in WS1 and WS2, respectively. In general, WS2 exhibited much lower counts than WS1, which was expected due to best management practice implementation in WS2 (fencing cattle out of the stream). When these results were compared to the observed DNA marker frequencies, no significant statistical relationships between the monthly enterococcal counts and the presence of the molecular markers in WS1 (Figure 6) were observed. In WS2, the enterococcal counts were statistically compared only to the general marker 32F due to the absence of the other markers from most sites (Figure 7). In this case also, no significant relationship was identified between the marker and the enterococcal counts. The only marker that indicated a slightly similar trend to that observed with the enterococcal counts was CF 128, and this only during a brief time of the sampling period (Dec 05-Feb 06) in WS1. However, this relationship didn't persist during the warmer months of the year or the following winter season.

In accordance with previous reports, enterococcal counts reported here could not be related to the occurrence of microbial source tracking markers, suggesting that more information is necessary to understand the dynamics of DNA source identifiers in a watershed in relation to the densities of traditional fecal indicators such as *E. coli* (Shanks *et al.*, 2006b) and enterococci. One

possible explanation for the discrepancy could be the differences in the physiological and biochemical features between the two targeted bacterial groups. *Bacteroides* are strict anaerobes and have low environmental persistence, indicating recent contamination (Fiksdal *et al.*, 1985, Kreader, 1998, Oshiro & Fujioka, 1995 1995, Ott *et al.*, 2001). Although enumeration of enterococci provides information on the level of impairment of a system, it does not identify the specific source of contamination (Scott, 2005). Therefore, it is recommended to employ a combination of molecular and traditional methods in field studies to provide more accurate and reliable results in risk assessment and prevention or reduction of contamination.

Table 2 Enterococcal abundance {CFU/100ml) in Watershed 1 and Watershed 2. Sites were divided based on influence by cattle or type of water resource (streams vs. ponds).

		95% Confidence Interval	
Site #	Geometric mean	Lower bound	Upper bound
WS1 site 1	74	9	640
WS1 sites 2, 3	1924	1130	3275
WS1 site 4	94	42	207
WS2 sites 1-4	26*	10	68
WS2 sites 5, 6, 11, 12	4*	2	9
WS2 sites 7-10	123*	59	257

* Zero values in the data were converted to 0.01.

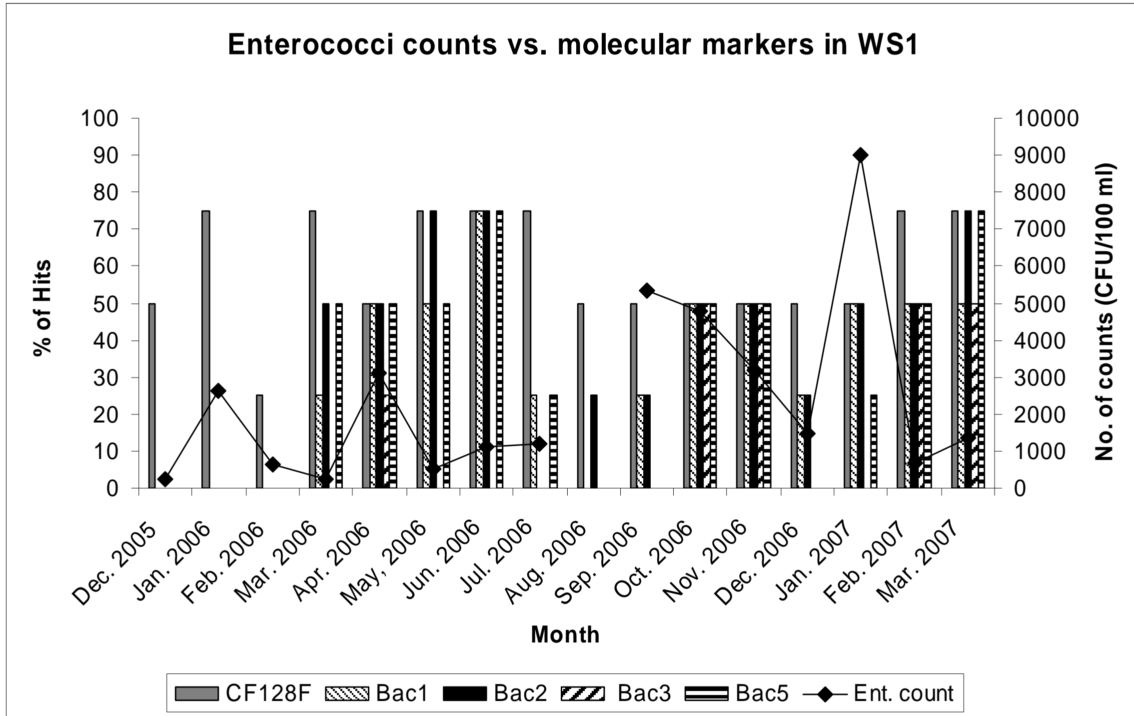


Figure 6 Relationship between the monthly enterococcal counts and the average frequency of the DNA markers per month in WS1

.

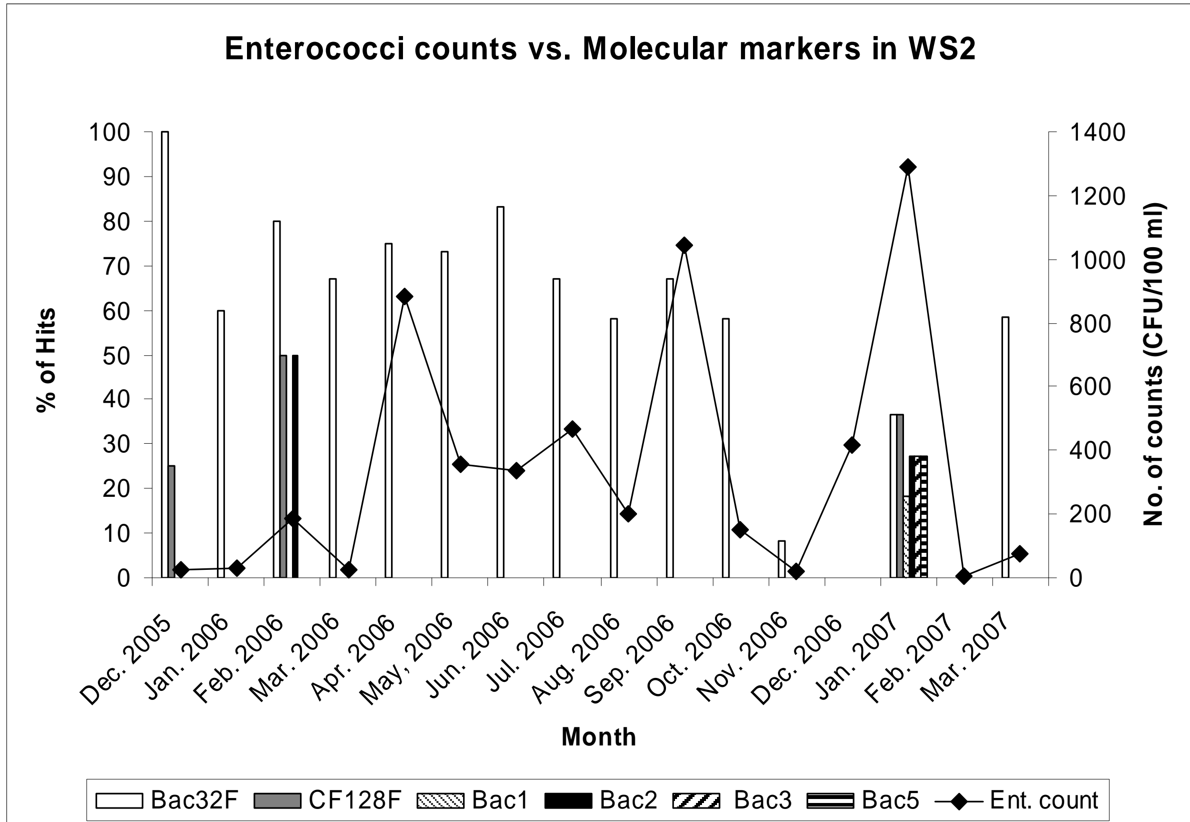


Figure 7 Relationship between monthly enterococcal counts and the average frequency of the DNA markers per month in WS2

Evaluation of human-specific 16S-rDNA markers in freshwater streams impacted by rural non-point sources in Puerto Rico. In this set of samples, each primer group was run as follows: Gen-Bac 32F and HF654 -four times each; primer HF 183 - six times; and primers CF128 and 193 - twice each with the objective to determine whether the locations sampled were impacted by either human (HC) or cattle fecal contamination (CC).

Table 3 Description of samples collected in the Rio Añasco Basin, Añasco, Puerto Rico from August 3-14, 2006

Sample Code	Site Description, volume of sample filtered, or sampling date
Site A	Bridge 406
Site B	Bridge 430
Site C	Shallow well near Bridge 430
Site D	Intermittent stream crossing community
Site E	Casey River
WTC* and WTD	Centrifuged sludge from Athens Water Treatment Plant
Samples 1,4,7,10,13	500 ml of water filtered
Samples 2,5,8,11,14	250 ml of water filtered
Samples 3,6,9,12,15	100 ml of water filtered
Samples 1,2,3	Sampled on 8/3/06
Samples 4,5,6	Sampled on 8/5/07
Samples 7,8,9	Sampled on 8/7/06
Samples 10,11,12	Sampled on 8/9/06
Samples 13,14,15	Sampled on 8/14/07

*When CF primers were tested, WTC was substituted for Cow Fecal DNA as positive control.

When indicated, 2x means that 1ul of PCR product from a first round (x1) was used as template for a second round (x2). Many of the (x2) gels had some non-specific banding, however, when the correct band size for the primer listed was present, the gel was scored with a (+); those gels without the correct band size, but with non-specific bands, were labeled “M” for multiple bands. Results were scored as Clean, Human, Cow, and Human-Cow, based on the PCR results.

Although 2x amplification assays can increase the signal in those cases where low initial concentrations of the target DNA are present, our results were not considered solid enough due to the fact that some gels exhibited a large number of non-specific bands. Further confirmation, for example through sequencing of the bands obtained in the 2x amplification, would be necessary to confirm the presence of the target DNA.

Results for the 1X runs are presented in Table 4. Although the human primer was not amplified in all samples collected from sites A and B, the results indicate that human contamination seems to be present in the system at some level during most sampling dates. The fact that human contamination was not indicated in every single sample could indicate a low level of contamination or presence of inhibitors in some samples. The level of contamination is hard to assess without a real quantification assay. Only once during the sampling period (8/5/06) did the results indicate that cattle contamination was present in one of the sampling locations (site B). The presence of the Gen-Bac in the absence of cattle or human contamination may point to another source of contamination (neither human nor cattle), or cross-amplification with a natural bacterial population.

Samples obtained from the well seem to be free from cattle and human contamination. Only one sample gave a positive human signal, this on the last day of sampling (out of triplicates), and might not be enough evidence to indicate an actual human impact. In a situation like this, inadvertent sample contamination can not be discounted. The stream that crosses the rural

community (site D) is clearly impacted by human contamination, since that assay was positive for every triplicate sample collected throughout the whole sampling period. This community is served mainly by septic systems. The contamination observed points to the fact that these septic systems might not be working properly and are leaking into the intermittent stream traversing the community. The Casey River also seems to be strongly impacted by human contamination, but in this case the contamination might be intermittent, since no contamination was detected in any of the samples collected on the last day of sampling. One possible explanation for this observation is dilution of the assay signal in the river or fecal bacterial decay after the initial contamination episode.

There was no relationship observed between the volume of sample filtered and the presence of a marker, meaning that sometimes a marker was positive in the 100 ml-sample while it was absent in the 500 ml-sample and vice versa. This result could be a function of the amount of inhibitors present in a sample at a given time, or it could just reflect sample randomness. Duplicate and, if possible, triplicate sample collection is recommended to cover sample variability.

Table 4 Bacteroidetes 16S rRNA gene marker hits in water samples collected in the Rio Añasco Basin, Añasco, Puerto Rico. The numbers indicate the times the individual primer set was found in each water sample after one amplification round (1x)

	Number of Primer Set Hits					
	General Bacteroides Marker	Human Bacteroides Marker	Human Bacteroides Marker	Cattle Bacteroides Marker	Cattle Bacteroides Marker	
Sample	**Gen-Bac**	**HF-183**	**HF-654**	**CF-128**	**CF-193**	**Comments:**
Neg Control	0	0	0	0	0	Control
WT C/CowCF*	4	2	0	1	0	Control
WT D	4	3	0	0	0	Control
A-1	1	1	0	0	0	Human
A-2	1	0	0	0	0	General
A-3	0	0	0	0	0	Clean
A-4	1	0	0	0	0	General
A-5	2	1	0	0	0	Human
A-6	0	0	0	0	0	Clean
A-7	0	0	0	0	0	Clean
A-8	0	1	0	0	0	Human
A-9	0	0	0	0	0	Clean
A-10	0	0	0	0	0	Clean
A-11	0	1	0	0	0	Human
A-12	0	2	0	0	0	Human
A-13	0	1	0	0	0	Human
A-14	0	0	0	0	0	Clean
A-15	0	1	0	0	0	Human
B-1	1	0	0	0	0	General
B-2	1	0	0	0	0	General
B-3	1	1	0	0	0	Human
B-5	2	1	0	1	0	Human and cow
B-6	1	0	0	1	0	Cow
B-7	0	0	0	0	0	Clean
B-8	1	0	0	0	0	General
B-9	0	0	0	0	0	Clean
B-10	1	0	0	0	0	General
B-11	0	1	0	0	0	Human
B-12	0	0	1	0	0	Human
B-13	0	0	0	0	0	Clean
B-14	0	0	0	0	0	Clean
B-15	0	1	0	0	0	Human
C-1	0	0	0	0	0	Clean
C-2	0	0	0	0	0	Clean
C-3	0	0	0	0	0	Clean
C-4	0	0	0	0	0	Clean
C-5	0	0	0	0	0	Clean
C-6	0	0	0	0	0	Clean
C-7	0	0	0	0	0	Clean

C-8	0	0	0	0	0	Clean
C-9	0	0	0	0	0	Clean
C-10	0	0	0	0	0	Clean
C-11	0	0	0	0	0	Clean
C-12	0	0	0	0	0	Clean
C-13	0	0	0	0	0	Clean
C-14	0	0	0	0	0	Clean
C-15	0	2	0	0	0	Human
D-1	2	0	0	0	0	General
D-2	0	1	0	0	0	Human
D-3	0	2	0	0	0	Human
D-4	0	1	1	0	0	Human
D-5	1	1	0	0	0	Human
D-6	2	2	0	0	0	Human
D-7	2	2	0	0	0	Human
D-8	3	1	0	0	0	Human
D-9	0	2	0	0	0	Human
D-10	2	0	0	0	0	General
D-11	2	1	0	0	0	Human
D-12	0	2	0	0	0	Human
D-13	1	2	0	0	0	Human
D-14	0	1	1	0	0	Human
D-15	0	1	0	0	0	Human
E-1	0	1	0	0	0	Human
E-2	0	2	0	0	0	Human
E-3	0	2	0	0	0	Human
E-4	2	2	0	0	0	Human
E-5	2	2	0	0	0	Human
E-6	0	1	0	0	0	Human
E-7	2	2	0	0	0	Human
E-8	0	1	0	0	0	Human
E-9	0	0	0	0	0	Clean
E-10	2	2	0	0	0	Human
E-11	0	1	0	0	0	Human
E-12	0	0	0	0	0	Clean
E-13	0	0	0	0	0	Clean
E-14	0	0	0	0	0	Clean
E-15	0	0	0	0	0	Clean

Evaluation of Library-Dependent Methods

Seasonal Distribution of Enterococci Isolates. Application of Pearson's chi-squared statistics to our data indicated that the proportions of the different bacterial species varied seasonally. However, this variability did not stand a monthly statistical distribution analysis, which means that the differences observed were due to sample randomness rather than true seasonal differences. Nevertheless, some general trends were identified with *E. faecalis* and *E. hirae*, although not with *E. casseliflavus*. *E. faecalis* seems to be the only species showing a trend of higher occurrence frequency during the warmer months of the sampling period, April through November 2004 (Figure 8). *E. hirae* was present more commonly during colder months (spring, fall and winter). *E. casseliflavus* indicated no correlation with season, farm sample site or source of sample (water vs. manure). In addition to the behavior of the three former species, *E. faecium* was found in higher abundance only during the winter of 2005. The winter months also reflected the highest diversity in terms of number of species identified and the evenness of the different populations.

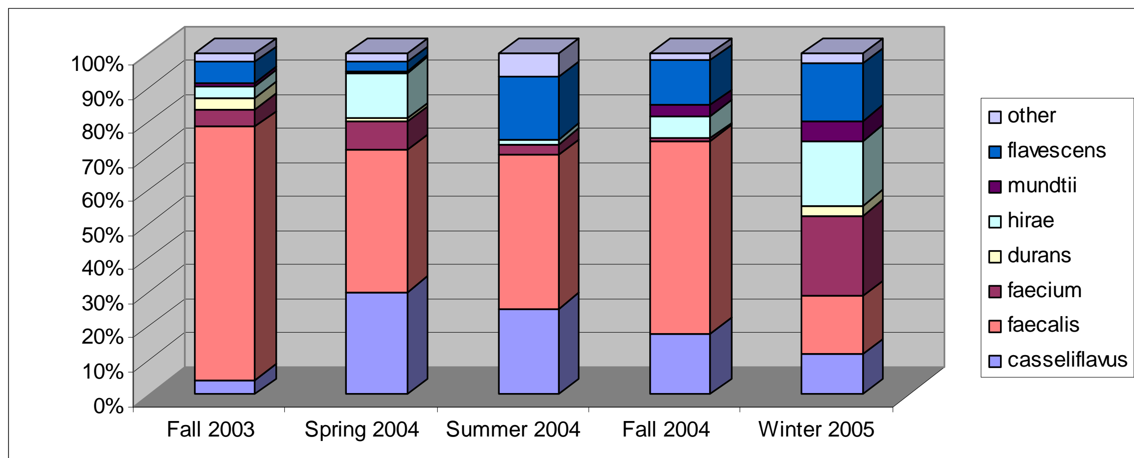


Figure 8 Seasonal distribution of enterococcal species in impacted streams by cattle contamination

Some researchers have suggested that species such as *E. faecalis* can be used as markers for human contamination (Wheeler et al., 2002). However, results from this research suggest that the large seasonal variability exhibited by the different enterococcal populations identified make the use of individual *Enterococcus* species unreliable due to lack of temporal stability. The observed variability, combined with the observed presence of the same E*nterococcus* species in the cattle farm stream water and the water upstream from the farm, highlights the fact that enterococci populations are widespread in nature. This could make their use as markers at the species level undependable. To evaluate the suitability of *Enterococcus* at the subspecies level to serve as markers of bovine contamination, it was necessary to perform fingerprinting analysis of some of those subspecies that were observed to be present in the cattle farm streams throughout the year.

Some methodological considerations developing the phylogeny of *Enterococcus strains* using AFLP. The two primer sets, *MboI*-CTG and *MboI*-AC, for Hex and Fam, respectively, exhibited congruencies of up to 60%. The 40% incongruence can be explained, in part, by the dynamic phylogeny produced due to the high species diversity in the library. Detailed analysis using band matching and maximum parsimony will need to be performed in order to obtain more detailed information. The phylogenetic trees produced by each primer set for *E. faecalis* yielded the greatest incongruence; however, the two primers produced the same basic groupings for both *E. hirae* and *E.*

casseliflavus. These results suggest that *E. faecalis* exhibits the highest species diversity in the environment among the three species studied.

Phylogeny of *E . faecalis*, *E. hirae* ,and *E. casseliflavus* strains using AFLP. Using primer set *Mbo*I-CTG, *E. faecalis* isolates separated into two distinct clusters depending on the farm from which they were isolated. Other than the farm differences, isolates were not found to group by source (manure vs. water), season, or location (stream sites within the farm or stream sites upstream from the farm). One possible explanation for this division is that cattle uptake part of their *E. faecalis* fecal population from their drinking water. Because their drinking water includes the upstream-from-the-farm water, this possibly explains why the *E. faecalis* isolated from manure could not be differentiated from that isolated from the upstream water. This observation also implies that the *E. faecalis* population present in the wildlife inhabiting each farm differs from each other, since no similar fingerprints were identified across farms.

E. hirae also showed two distinct clusters, one containing isolates mainly collected during autumn 2003 in Farm 1 from manure, and the other comprised of isolates from all seasons and sources (Figure 9). The autumn 2003 cluster was not observed at any of the upstream locations, suggesting that it is composed of species mostly present in the feces of the cattle on Farm 1. This cluster is in close phylogenetic relationship to a spring cluster found from both farms, composed of isolates obtained mostly from the water within the farms, but absent in the water collected upstream of the farms. Because the fingerprints in these two clusters are absent in the water upstream from the farms, they could

be developed and tested as MST markers for cattle fecal contamination. However, one drawback observed is the fact that these two clusters only showed-up during autumn and possibly spring, but not during other times of the year. This makes them temporally unstable and unreliable. A good indicator needs to be present throughout all seasons (Simpson *et al.*, 2002).

E. casseliflavus isolates also grouped into two basic clusters, with one cluster accounting for 73% of the library. For this species, no seasonal or source trends were observed, and many fingerprints were found in the water upstream of the farms. In addition, no difference was observed between farms. These results suggest that *E. casseliflavus* fingerprints are widespread in the environment, making it hard to distinguish contributions of cattle vs. wildlife.

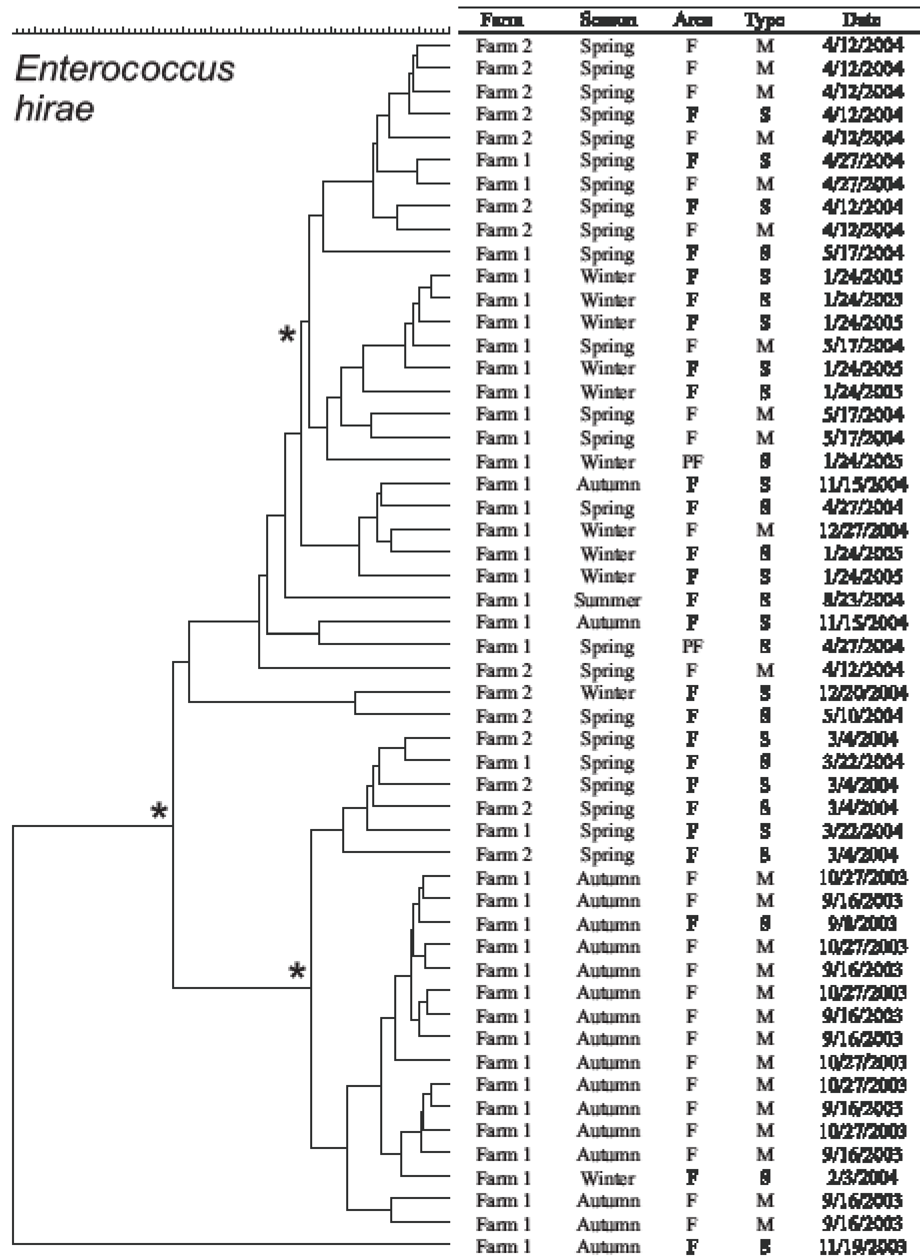


Figure 9 *E. hirae* phylogenetic tree derived from AFLP fingerprints of isolates obtained from water and manure samples collected at two cattle farms with impacted streams.

Comparison of AFLP and BOX-PCR analysis. The genotyping methods of BOX-PCR and AFLP each have distinct advantages and disadvantages (Table 5). Our results showed that, in general, AFLP is far superior at discriminating closely related strains of *Enterococcus.* AFLP produced a greater number of bands per PCR reaction, providing greater discriminatory power; had greater precision in band sizing; and allowed for the use of multiple primer sets. Additionally, the quality of the AFLP gels was very consistent in terms of both band reproducibility and overall gel usability for data analysis. Throughput of samples was also greater with AFLP due to the high sensitivity of the fluorescence based-detection, thereby allowing the use of much smaller band lane widths.

The advantages of BOX-PCR are: the simplicity of the method (fewer steps, technically easier); much lower cost of equipment (only a regular thermal cycler is required) and reagents (BOXA2R primer, enzymes and buffer); and no production of hazardous waste. However, the procedure produced highly variable results in terms of band detection. It also produced lower discriminating power than the AFLP procedure because for most species we were only able to obtain between 9 and 18 different bands (Figure 10). In contrast, the AFLP analysis consistently produced over 100 bands. The BOX-PCR procedure was also highly sensitive to the buffer pH, which can affect band detection. In addition, sensitivity and band brightness was highly affected by gel quality.

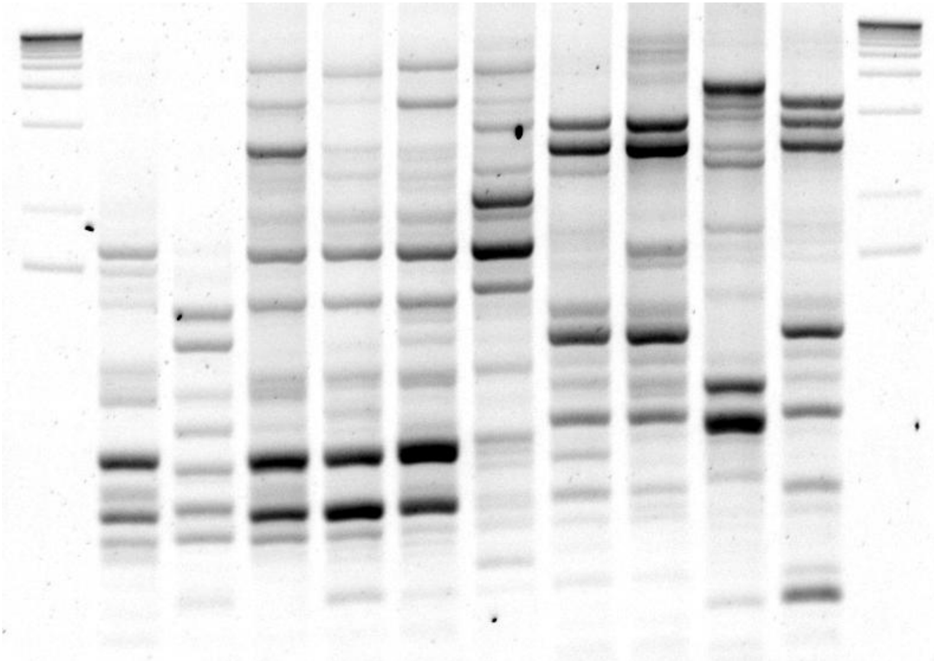


Figure 10 Typical BOX-PCR gel image produced with *E. faecalis* isolates.

Table 5 Comparison of the advantages and disadvantages of the BOX-PCR and AFLP methodologies

Methodology			
BOX-PCR		AFLP	
Disadvantages	Advantages	Disadvantages	Advantages
Poor consistency in gel quality (affects band 'brightness' or sensitivity).	Inexpensive (no expensive equipment, primers, or standards).	Expensive (machine, primers, and standard).	Much greater consistency in gel quality (although not perfect, very sensitive).
Poor reproducibility (pH variability affects band detection; high PCR assay variability).	Technically simple.	Technically more challenging (more steps).	High reproducibility.
Low band sizing precision (inability to discern similarly sized bands).	No hazardous waste.	Hazardous Waste.	Very high band sizing precision.
Requires certain gel lane width for accurate detection (reduces throughput).			High throughput.
Assay produces fewer bands (low discriminatory power).			Many bands (high discriminatory power).
			Option to use different selective primers.

Conclusions

Application of AFLP methodology vs. DNA markers. Studies examining bacterial strain diversity and temporal variability in aquatic and terrestrial habitats using the level of genetic specificity undertaken in this study are uncommon. Our work helps fill this void by providing a genotyping study that

involves hundreds of *Enterococcus* strains from multiple species, seasons, and two aquatic systems, as well as a detailed temporal screening of 16S and metagenomic markers. AFLP genotyping of our *Enterococcus* strain library provided a large and robust data set, that supplied many unique fingerprints. We identified a fingerprint of *E. hirae* that seems to be fairly specific to cattle manure samples; however, the fingerprint showed-up only during two out of the five seasons sampled. This makes the fingerprint unsuitable for MST applications due to the lack of temporal stability and reliability. The fact that *E. faecalis* isolates grouped by farm and showed no correlation to source (upstream-of-the-farms and farm water, or manure) suggests that the cattle in our study may uptake part of their *E. faecalis* population from their drinking water which then gets transferred to their manure. Such environmental uptake masks identification of cattle-specific fingerprints of *E. faecalis*.

Although the AFLP methodology is very reproducible and has high discriminating power, its application as a rapid and resource-efficient methodology is limited because the library production is highly time and resource consuming. Its application is probably most appropriate in very specific scenarios where discrimination among few selected sources is necessary. In contrast, application of DNA, PCR-based markers produced fairly rapid results, and had the capability to screen multiple scenarios in a short period of time. Once stability and cross-amplification aspects have been addressed, it can be a highly efficacious approach to determine sources of contamination in a variety of scenarios.

From our results, we conclude that a combination of the ruminant-specific marker, CF128F, with the metagenomic markers, Bac1, 2 and 5, may provide a solid application package for tracking bovine fecal contamination sources to surface waters. Because enterococcal counts did not show a strong correlation with the occurrence of any of the DNA markers, the dynamics of fecal source tracking markers in a watershed need to be further investigated to be able to determine their correlation with the densities of traditional indicators of fecal contamination.

Significance of Research

This research supports an area of high priority for the Office of Water and has been listed in the Twenty Needs Report as the highest priority for Regions and States. This work supports assessment of aquatic systems impairment under Long Term Goal 2 (LTG 2) of the Office of Research and Development Water Quality Multiyear Plan. LTG2 provides the tools to assess and diagnose the causes and pollutant sources of impairment in aquatic systems. Specifically, the results of this research provide an evaluation of selected LI- and LD-mthods as to their usability for early and rapid assessment of fecal contamination sources. Included is a specific application and comparison of some of the available DNA-based methodologies for discriminating among sources of contamination in impaired surface waters.

Future Directions

During the next phase of this research, we will focus on the application of the LI-approaches to quantify the loadings of agricultural, human and other non-human non-point sources of bacterial contaminants into aquatic resources:

- Determine the loadings, fate and transport of bacterial contaminants from agricultural non-point sources in surface waters using quantitative PCR methods that will provide such information in an accurate, fast and informed way.
- Provide a basis for comparison between traditional fecal indicators, true pathogenic bacteria and DNA-based fecal indicators.
- Develop and validate a technique by which the recovery of an ecosystem from bacterial contamination can be measured, and provide information to watershed managers about the effectiveness of alternative BMP approaches.

References

[1] Allsop, K. and J.D. Stickler (1985) An assessment of *Bacteroides fragils* group organisms as indicators of human faecal pollution. *J Appl Bacteriology* 58: 95-99.

[2] Antonishyn NA, R.R. McDonald, E. L. Chan, G. Horsman, C. E.Woodmansee, P.S. Falk and A.C.G. Mayhall (2000) Evaluation of Fluorescence-Based Amplified Fragment Length Polymorphism Analysis for Molecular Typing in Hospital Epidemiology: Comparison with Pulsed-Field Gel Electrophoresis for Typing Strains of Vancomycin-Resistant *Enterococcus faecium. Journal of Clinical Microbiology* 38: 4058-4065.

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

[4] Blears, M., S.A. De Grandis, H. Lee and J. Trevors (1998) Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology & Biotechnology* 21: 99-114.

[5] Dick, L.K., A.N. 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 Environmental Microbiology* 71: 3184-3191.

[6] Fiksdal L, J.S. Maki, S.J. LaCroix and J.T. Staley (1985) Survival and detection of *Bacteroides* spp., prospective indicator bacteria. *Appl Environmental Microbiology* 49: 148-150.

[7] Harmsen HJM, G.R. Gibson, and P. Elfferich (1999) Comparison of viable cell counts and fluorescence in situ hybridization using
specific rRNA-based probes for the quantification of human fecal bacteria. *FEMS Microbiol. Lett.* 183: 125–129.

[8] Jackson CR, P.J. Fedorka-Cray and J.B. Barrett (2004) Identification of enterococci using a genus and species multiplex PCR. *J Clinical Microbiol* 42: 3558-3565.

[9] Koeuth T., J. Versalovic and J. R. Lupski (1995) Differential Subsequence Conservation of Interspersed Repetitive Streptococcus pneumoniae BOX Elements in Diverse Bacteria. *Genome Research* 5: 408-418.

[10] Kogan S.C., M. Doherrty and J. Gitschier J (1987) An improved method for prenatal diagnosis of genetic disease by analysis of amplified DNA sequences: application to hemophilia. *A. N. Engl. J. Med.* 317: 985-990.

[11] Kreader, C.A. (1995) Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. *Appl Environmental Microbiology* 61: 1171-1179.

[12] Kreader, C.A. (1998) Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Appl Environmental Microbiology* 64: 4103-4105.

[13] Lamendella R, J.W. Santo Domingo, D.B. Oerther, J.R. Vogel & D. M. Stoekel (2007) Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene. *FEMS Microbiol Ecol* 59: 651-660.

[14] Layton A, L. McKay, D. Williams, V. Garrett, R. Gentry and G. Sayler (2006) Development of *Bacteroides* 16S rRNA Gene TaqMan-Based Real-Time PCR Assays for Estimation of Total, Human and Bovine Fecal Pollution in Water. *Appl Environmental Microbiology* 72: 4214-4224.

[15] Malathum K, K.V. Singh, G. M. Weinstock and B.E. Murray (1998) Repetitive sequence-based PCR versus pulsed-field gel electrophoresis for typing of enterococcus faecalis at the subspecies level. *Journal of Clinical Microbiology* 36: 211-215.

[16] Martin B, O. Humbert, M. Camara, E. Guenzi, J. Walker, and T. Mitchell (1992) A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. *Nucleic Acid Research* 20: 3479-3483.

[17] Molina M. (2005) Temporal and Spatial Variability of Fecal Indicator Bacteria: Implications for the Application of MST Methodologies to Differentiate Sources of Fecal Contamination. U. S EPA, Athens

[18] Molina M., M. Cyterski, J. Maimes, J. Fisher and B. Johnson (2007) Comparison of the temporal variability of enterococcal clusters in impacted streams using a multiplex polymerase chain reaction procedure. (Hatcher K, ed.), University of Georgia, Athens Georgia.

[19] Oshiro R. and R. Fujioka (1995) Sand, soil and pigeon droppings-Sources of indicator bacteria in the waters of Hanauma Bay, Oahu, Hawaii. *Water Sci Technol* 31: 251-254.

[20] Ott E-M, T. Mller, M. Mller, C.M.A.P. Franz, A. Ulrich, M. Gabel and W. Seyfarth (2001) Population dynamics and antagonistic potential of enterococci colonizing the phyllosphere of grasses. *J Appl Bacteriology* 91: 54-66.

[21] Ritter Kj, E. Carruthers, and C. A. Carson(2003) Assessment of statistical methods used in library-based approaches to microbial source tracking. *Journal of Water and Health* 1: 209-223.

[22] 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 *Environmental Science Technology* 39: 283-287.

[23] Shanks OC, J.W. Santo Domingo & J.E. Graham (2006a) Use of competitive DNA hybridization to identify differences in the genomes of bacteria. *J Microbiol Methods* 66: 321-330.

[24] Shanks O.C., J.W. Santo Domingo, R. Lamdella, C. A. Kelty & J.E. Graham (2006b) Competitive Metagenomic DNA Hybridization Identifies Host-Specific Microbial Genetic Markers in Cow Fecal Samples. *Appl Environmental Microbiology* 72: 4054-4060.

[25] Simpson J., J.W. Santo Domingo & D. J. Reasoner (2002) Microbial source tracking: state of the science. *Environmental Science and Technology* 36: 5279=5288.

United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

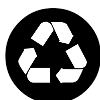
EPA 600/R-07/123
December 2007

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE

☐; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35



Recycled/Recyclable
Printed with vegetable-based ink on paper that contains a minimum of 50% post-consumer fiber content processed chlorine free