Development of host-specific metagenomic markers for microbial source tracking using a novel metagenomic approach

Jorge W. Santo Domingo,[1*] Jingrang Lu, [1] Orin Shanks, [1] Regina Lamendella, [2] Catherine Kelty, [1] and Daniel Oerther [2]

[1]U. S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH.

[2]Department of Civil and Environmental Engineering, University of Cincinnati

[*] Corresponding author: 26 W. Martin Luther King Dr., AWBERC Building, MS-387, Cincinnati, OH; Email: santodomingo.jorge@epa.gov

ABSTRACT

Fecal contamination of source waters is an important issue to the drinking water industry. Improper disposal of animal waste, leaky septic tanks, storm runoff, and wildlife can all be responsible for spreading enteric pathogens into source waters**.** As a result, methods that can pinpoint fecal pollution sources in natural waters are needed to assist in the development and evaluation of adequate management practices targeting pollution control. In the last decade, several methods have been developed to identify fecal sources, collectively known as microbial source tracking (MST) methods. Early studies focused on the use of methods that rely on generating library dependent databases. More recently library independent PCR-based approaches have become more popular among MST practitioners as they do not rely on the development of large culture-based databases. One potential concern associated with the use of library independent approaches relates to the development of host-specific assays using sequencing information from genes not involved in host-microbial interactions. To address this issue, our research group has applied a novel method called genome fragment enrichment (GFE) to select for genomic regions that differ between different fecal metagenomes. We have shown that the vast majority of the selected fragments are indeed specific to the fecal community in question suggesting that this is an efficient way of enriching for community specific DNA regions. Sequences from the enriched fragments were used to develop host-specific PCR assays. Thus far, we have successfully developed several assays specific to human, cow, and chicken fecal communities. Additionally, several assays are capable of detecting multiple avian species, further suggesting that similar gut environments can select for similar host-specific populations. The latter finding is relevant as it shows that is possible to develop assays targeting multiple species. Signals from metagenomic-based assays were detected in water samples demonstrating their potential as source tracking tools. A brief summary of the history and limitations of current MST tools and the discussion of results with metagenomic markers will be the main topics of this manuscript. The need for a risk assessment tool-box that includes assays targeting indicators, source identifiers and pathogens will also be discussed.

KEYWORDS
Microbial source tracking, fecal contamination, metagenomics, source water

INTRODUCTION

Safe drinking water is essential for public health due to the risks associated with consumption of polluted water. In developed countries microbial water quality has greatly improved as a result of the use of modern disinfection treatment technologies. However, the increasing demands associated with population growth are having a negative impact on the microbiological quality of water sources used for drinking water. Due to the high number of polluted watersheds, identifying the type of fecal contamination impacting drinking water sources is becoming a necessary step towards the implementation of adequate management practices targeting source water protection. Adequate watershed management practices can also be beneficial to the water treatment industry from the standpoint of reducing the disinfectant demand required to deliver safe water to the public. In order to decide which management practice is needed, it is first necessary to identify the potential sources responsible for the impairment. Microbial monitoring of source waters is performed using culture based techniques that enumerate bacterial indicators of fecal pollution, like *Escherichia coli* and enterococci. The latter methods are relatively effective at predicting levels of fecal pollution but cannot provide any information on the type of fecal pollution sources impacting surface waters and groundwater.

Recently, a number of microbial source tracking (MST) methods primarily targeting fecal bacteria have been suggested to discriminate fecal pollution associated with different animals (Table 1). Intestinal bacteria within animal groups are expected to be different due to differences in gut conditions. For example, temperature, diet, and type of digestive system can contribute to the conditions specific to each gut type. Age and overall health of the animal are additional contributing factors that might also influence the microbial dynamics and structure of the gut microbiota. Taken all together, it is assumed that each gut type can select for microbial communities that are relatively beneficial to the host, and that particular populations might prefer to inhabit a particular gut type over another.

In general, MST relies on good source identifiers, that is, host-specific populations or genetic targets among host specific populations that are only (or preferentially) present in specific gut systems. Ideal candidates for source identifiers must fit the following criteria: exhibit host-specificity, be abundant in the host, high host distribution (most individual hosts must carry the source identifier), show clonal diversity, show temporal stability, and show geographic continuity. From a user's perspective the ideal MST method should be rapid, simple, sensitive, robust, and inexpensive. Watershed managers dealing with Total Maximum Daily Loads (TMDLs) also require that the method is capable of producing quantitative data (i.e., to calculate fecal loads), while public health officials would prefer that the target/data also correlates with health risks and can be of great value in risk assessment models.

Non-point sources are the primary targets in source tracking as waste discharges from point sources (e.g., wastewater treatment plants) are scrutinized via National Pollution Discharge Elimination Standard (NPDES) permits. The different types of fecal sources can be classified as human (e.g., leaky septic tanks), domesticated animals (e.g., cattle, poultry, swine), and wildlife (e.g., deer, geese, duck). Identifying primary contamination sources in many watersheds is greatly complicated in light of the fact that several sources can contribute simultaneously and by issues related to manure application, handling of biosolids, urban runoff, and the increased usage of reclaimed water, to mention a few.

As a field, MST has a very broad relevance to many areas related to environmental monitoring. For example, fecal source tracking activities are critical to meeting the demands of the Clean Water Act once waters are identified as being in violation of established water quality standards. Specifically, identification of polluted sources can help set the adequate management practices needed to remediate polluted waters and to prevent future pollution events. Several methods used in MST studies can also be used in forensic investigations relevant to food safety, lome land security and ecosystem health.

History of Microbial Source Tracking

The need to discriminate between sources of pollution in water sources stems from the concept that higher risks are associated with human fecal pollution than those associated with feces from other warm-blooded animals. Moreover, risks associated with the feces of other less-studied carriers of fecal indicators, including fish (Geldreich and Clarke, 1966) and insects, are assumed to be relatively less than higher animals, although the contribution of non-mammalian fecal coliforms and enterococci to water fecal pollution and their relevance to public health is poorly understood. As previously established, current culture-based methods used to assess the microbial quality of water do not provide any information regarding the sources of pollution, although when coupled with sanitary surveys the data can identify “hot spots” within watershed systems.

Some of the first attempts to discriminate between the different sources of fecal pollution were performed by Geldreich and colleagues in the mid to late 1960s (Van Donsel et al., 1967; Geldreich and Kenner, 1969). By then several studies had demonstrated that often the levels of fecal bacterial indicators in human feces are different than in other mammals. Recognizing the importance of differentiating between human and non-human sources, Geldreich proposed that fecal coliform to fecal streptococci (FC/FS) ratios above 4 was a sign of human fecal pollution. Similarly, ratios below 4 were primarily associated with non-human pollution sources, but more specifically, ratios between 4.0 and 0.1 were assumed to be associated with domesticated animals while less than 0.1 were thought of being representative of wildlife fecal pollution. The FC/FS ratio was discontinued in the mid 1980’s as studies showed that differences in the survival rates of members within the coliform and streptococci groups may vary depending on the environmental conditions and depending on the time between the contamination event and sampling. Additionally, some fecal streptococci species are associated with soils and vegetation and are difficult to discriminate from those associated with fecal waste (Geldreich et al., 1964; Kibbey et al., 1978; Muller et al., 2001).

More than a dozen MST methods have been developed in the last decade by several laboratories (Table 1). A brief summary of MST methods will follow but additional details on the assumption and limitations for each of these methods have been discussed elsewhere (Scott et al., 2002; Simpson et al., 2002). Fecal source tracking methods can be grouped as library dependent

Table 1. Examples of recent methods targeting fecal microbial populations relevant to source tracking (modified from Microbial Source Tracking Guide Document; USEPA, 2005)

METHOD	**Type of target**	**Reference**
Mitochondrial PCR	Animal DNA	Martellini et al., 2005
Gene Specific PCR	*E. coli* toxin genes	Khatib et al., 2002 Khatib et al., 2003 Scott et al., 2005 Ufnar et al., 2006
rDNA-based PCR	*Bacteroides* *Bifidobacterium* *Enterococcus* *Clostridium* *Cryptosporidium*	Bernhard and Field, 2000 Matsuli et al., 2004 Santo Domingo et al., 2004 Jiang et al., 2005
Host Specific QPCR	*Bacteroides* *Bifidobacteriim* *Clostridium*	Layton et al., 2006 Seurinck et al., 2005 Matsuki et al., 2004
Metagenome-based PCR	unidentified	Shanks et al., 2006
Gene sequencing	*E. coli* *Bacteroides*	Lamendella et al., in press Ram et al., 2004

DGGE	*E. coli*	Buchan et al., 2001;
		Seurinck et al., 2003;
		Lasalde et al., 2005a

Phage Genotyping	F+ coliphage	Stewart et al., 2006
Eukaryotic virus	Enterovirus Adenovirus	Fong et al., 2005 Maluquer de Motes et al., 2004
AFLP	*E. coli*	Leung et al., 2004
rep-PCR	*E. coli*	Dombek et al., 2000 Hassan et al., 2005
PFGE	*E. coli* *Enterococcus*	Parveen et al., 2001 Lasalde et al., 2005b
Ribotyping	*E. coli* *Enterococcus* spp.	Carson et al., 2001 Hartel et al., 2002
ARA/MAR (antibiotic resistance)	*Escherichia coli* *Steptococcus* *Enterococcus spp.*	Wiggins, 1996
Fatty acid methyl esther	*E. coli*	Parveen et al., 2001 Duran et al., 2006
Chemical indicators	Fecal sterols, detergents	Glassmeyer et al., 2005 Blanch et al., 2005

methods (LDMs) and library independent methods (LIMs). Most researchers refer to an MST library as a fingerprint database of bacterial strains isolated from known sources. Researchers have used both phenotypic and genotypic fingerprints to classify bacterial isolates using LDMs. Studies using three phenotypic LDMs are available in the literature, these being: carbon utilization profile (CUP), antibiotic resistance analysis (ARA), and fatty acid metyl esther profile (FAME) analysis. Early on, the ARA approach was used in most studies as it is relatively simple, inexpensive, and "low-tech". Some of the genotypic LDMs are PFGE (pulse field gel electrophoresis), AFLP (amplified fragment length polymorphism), rep-PCR (repetitive extragenic palindromic), ribotyping (restriction fragment length polymorphism; RFLP) using rDNA probes. Additionally, sequencing of specific alleles has been used to discriminate between bacterial strains isolated from different fecal sources (Ram et al., 2004).

Chemicals associated with anthropogenic activities (e.g., caffeine, detergents, fragrances) have also been used to detect sources of pollutions. Glassmeyer et al (2005) showed that 35 different chemicals can be considered potential indicators of human fecal pollution. Wastewater compounds like ethyl citrate, galaxolide, and tonalide, pharmaceuticals like carbamazepine and diphenhydramine, and fecal sterols like coprostanol were the best indicators as determined by the occurrence and concentration levels. Some of the concerns associated with the use of chemical surrogates relate to environmental fate, sensitivity, and cost of the analysis. Additionally, the correlation between their occurrence and health risks has not been established.

While LDMs have been used extensively in MST laboratory exercises the main issue regarding LDM-based studies relates to the need of establishing large databases in order to correctly classify water isolates and determine the level of uncertainty (i.e., error rate) associated with the method of choice. Early studies showed high levels of correct classification but most of them have in common the fact that the libraries were significantly small, which tends to artificially produce high levels of average rate of correct classification. The need for the development of large fingerprint libraries was addressed by Jenkins et al. (2003) who demonstrated the requirement of at least 900 isolates to address the genetic complexity of *E. coli* strains from a small cow herd. The occurrence of microbial populations closely related to fecal counterparts inhabiting vegetation, soils, and sediments can increase dramatically the number of unclassified isolates, complicating even further the statistical analysis and data interpretation. Moreover, statistical analyses of MST data have shown a high degree of variability depending on the method used and sample size (Ritter et al., 2003). Parameters used to assign operational taxonomic units as well as the selection of statistical approaches used to establish the significance of the results have also been subject to criticism (Ritter et al., 2003).

Due to the difficulty of using LDMs in complex watersheds, the use of culture independent LIMs targeting host-specific genes has recently gained significant attention. Most of these methods use the polymerase chain reaction (PCR) to detect host-specific targets in environmental samples. Some of the benefits of using recently developed LIMs include: methods are culture-independent (in most cases), rapid detection, sensitive, defined target, amenable to automated analysis, potential for multiple assays, assay cost. A number of library independent methods (LIMs) have been developed to determine the type of animals implicated in the fecal pollution of natural water systems. In general terms, LIMs do not require the development of extensive fingerprint libraries as most rely on the detection of the host specific marker via a PCR step. In some cases a pre-enrichment step is needed in order to detect the targeted gene (Scott et al., 2005). The primary target used for LIM development has been the 16S ribosomal RNA gene (16S rDNA), which is vital for protein synthesis and therefore present in all bacteria. Sequences from the *Bacteroidetes* class have been particularly useful in the development of host-specific assays. For example, assays targeting ruminant and human fecal pollution have been successfully applied to impaired watersheds (Lamendealla et al., in press). Additional host-specific assays have been developed using this approach (Dick et al., 200) but their validation outside of the laboratory has not been conducted. Quantitative assays based on *Bacteroidetes* 16S rDNA have recently been published, potentially allowing not only for the detection of primary sources of pollution but also the level of pollution (Layton et al., 2006). While 16S rDNA sequences from other fecal bacteria have been used to develop genus and species specific assays, none of these assays are useful in source tracking studies from the standpoint of pinpointing specific pollution sources. Moreover, further evaluation of *Bacteroidetes* 16S rDNA assays against non-target samples have showed that the assays can cross-react with non-specific targets, suggesting that some subpopulations might prefer a more cosmopolitan lifestyle.

Other PCR-based assays have targeted *E. coli* and *E. faecium* toxin genes (Khatib et al., 2002; Scott et al., 2005). Pre-enrichment steps are necessary to increase the sensitivity of the assays as the populations carrying host-specific genes are 2-3 orders of magnitude less abundant than the cosmopolitan close relatives. Alternatively, Martinelli et al (2005) developed several assays targeting human, bovine, ovine and porcine mitochondrial genes. In the latter study, the assays were challenged against DNA extracts from different environmental samples. One problem of using mitochondrial genes is the potential presence of eukaryotic DNA in some of the PCR reagents resulting in higher numbers of false-positives. The fate of animal cells in the environment needs to be studied in greater detail in order to better assess the value of this approach.

Most MST-LIMs target on one bacterial group and one gene to predict the link between sources and pollution (Fig. 1). This approach is limited in terms of the predictive power and the level of uncertainty that each assay inherently possesses and the number of signature sequences that are needed to develop a host-specific assay. Consequently, multiple assays could be used to increase the predictive power of MST tools. These assays can be derived by targeting multiple genomic regions of a host specific population. Isolating such a host specific strain(s) is the first step; however, this is a difficult task as most gut microorganisms (with the exception of enteric viruses) seem to be primarily cosmopolitan and some might be very difficult to isolate in pure culture (as is the case of most *Bacteroides* spp.). Another strategy to develop multiple assays is to explore the molecular diversity of as many different microbial species inhabiting each gut system as possible. In order to select genes unique to a particular gut environment it is necessary to eliminate to some extent the background of genetic markers that are shared by fecal microorganisms. The objective of this manuscript is to provide the reader with a general description of the results obtained in our laboratory using a metagenomic approach to develop MST host specific markers from function specific genes.

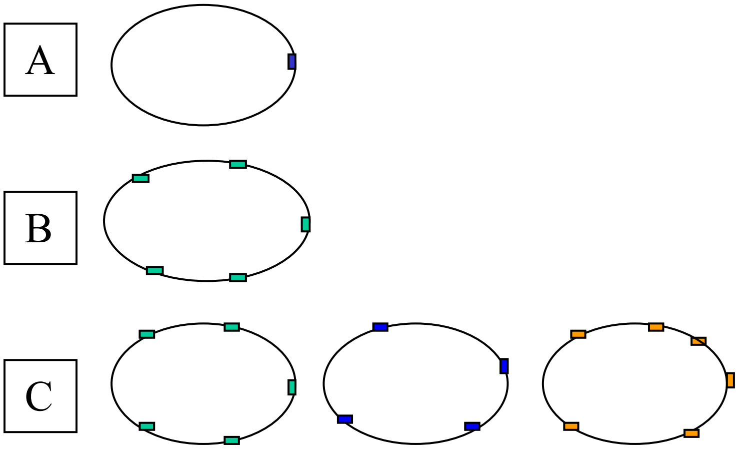

Figure 1. Different approaches used to discover host-specific microbial genes. Each circle represents a microbial genome and the blocks potential host specific genes. The one gene-one microbial group (A) is the most commonly used approach in MST. The best examples are the 16S rDNA *Bacteroides* markers. Expanding from one marker to several host-specific markers per bacterial group (B) will be possible as the genome sequences of fecal bacteria isolated from each host type are available. Metagenomic approaches provide the opportunity of selecting for multiple genes from multiple fecal microbial groups (C) increasing the predictive power of environmental monitoring assays.

MATERIALS AND METHODS

Sample collection and DNA extraction

Fecal samples used in this study were collected from diverse geographic locations in the United States. Feces from pig, goat, sheep, horse, cow, chicken, turkey, Canadian geese, seagull, pigeon, coyotes, squirrel, deer, possum, dog, cat, bobcat, raccoon, hedgehog, vulture, and human were used as negative and positive controls. (Shanks et al., 2006b). These samples are examples of the three classes of potential fecal sources (i.e., human, domesticated animals and wildlife). DNA extracts were used to test the host specificity and host distribution of the potential host-specific markers. Feces were placed into sterile tubes and stored at -80 ˚C until required. Fecal DNA was extracted from the feces using commercially available kits (Mo Bio Fecal kit,Mo Bio Laboratories Inc., CA; the FastDNA Kit, Q-Biogene; Carlsbad, CA) as suggested by the manufactures. Total DNA was eluted in water to achieve a DNA concentration of 1-2 ng/µl.

Water samples were collected in sterile bottles, transported to the laboratory on ice, filtered (100-1000 ml) onto 0.2 µm pore size, 47 mm polycarbonate membranes (Osmonics), and stored in sterile conical tubes at -80ºC. DNA from water samples was extracted using the FastDNA Kit and used to evaluate different host specific assays.

Enrichment of potential host-specific targets

We the GFE method developed by Shanks et al. (2006b) to enrich for host-specific metagenomic fragments for cattle and human marker while a modified GFE was used for avian markers. This method was originally developed to discriminate between genomes of closely related bacteria (Shanks et al., 2006a). Figure 2 shows the basic design for fragment enrichment experiments.

Briefly, genomic DNA extracts from individual fecal samples are mixed to create fecal microbial community DNA composites. The DNA composite used to isolate host-specific fragments is called "tester" while DNA extracts of the non-targeted animal(s) is referred as "blocker". DNA purification steps and washing steps after the competitive hybridization steps were performed as previously described (Shanks et al., 2006b). Thus far, we have performed experiments to isolate DNA fragments for human, cow, chicken, and other avian species using pig metagenomic DNA as blocker. The general steps of GFE will be described in the following paragraph.

To prepare the DNA, capturing surface fecal DNA from the tester is mechanically sheared into approximately 100 to 900 base pair (bp) fragments, and labeled with biotin (Sigma). To prepare target DNA used to enrich for host-specific fragments, oligonucleotide primers having both a common 5' sequence and nine random residues (called K9 primers) are linked to sheared tester fecal DNA using Klenow polymerase extension. The pre-hybridization solution is prepared by mixing genomic DNA capturing surface and blocker DNA solutions overlaid with mineral oil. This mixed solution is briefly heated and then allowed to pre-hybridize at a lower temprature. The DNA with the linkers is incubated separately, before adding an aliquot to the pre-hybridization solution for the final competitive hybridization step. DNA hybrids are then isolated by streptavidin binding and the captured tagged genomic fragments are amplified by lone-linker PCR (Grothus 1993). PCR products from a previous competitive hybridization round are used for the next enrichment round. PCR products from enrichment rounds are pooled and cloned into pCR4.1 TOPO following the manufacturer's instructions (Invitrogen). Individual clones are grown on Luria Broth plus ampicillin and used in M13-PCR assays to screen for inserts. Inserts are confirmed using agarose gel electrophoresis and PCR products are purified using Qiaquick 96 Plate (Qiagen). Clones are sequenced using Big Dye terminator chemistry and capillary gel electrophoresis (Applied Biosystem PRISM 3730XL DNA Analyzer) and using the M13 forward and reverse primers. Sequence editing and alignment are performed using Sequencher software (Gene Codes Corporation, Ann Arbor, MI). BlastX search (Altschul et al. 1997) was conducted for all DNA sequences to assign potential protein homologies. The latter information is used to select DNA fragments to be used in PCR assay development.

For host-specific PCR assays, primers were designed using commercially available software (e.g., Primer Designer, Cary, NC) using the following conditions: no hairpin, no primer dimer formation, and relatively high annealing temperatures (e.g., 60-65˚C). Assays are first optimized using temperature gradients and target fecal DNA. Specificity and sensitivity are determined by challenging the assays against various fecal DNA templates and by varying concentrations of the target DNA, respectively. The primers that showed host-specificity are further challenged against fecal DNA extracts from individual target animals to determine host distribution. The selected host-specific PCR assays are challenged against DNA extracted from water samples presumed to be impacted with fecal contamination. PCR assays specific to *Bacteroides* spp. and *Clostridium coccoides* are used to determine the presence of potential PCR inhibitors in DNA extracts (Bernhard and Field, 2000a; Matsuki et al., 2002). For host specificity studies, DNA for each fecal sample is first extracted and then equal amounts of each DNA extract are mixed to create fecal DNA composites. The presence of PCR products is visualized using 2% agarose gel electrophoresis and GelStar as the nucleic acid stain from FMC Bioproducts (Rockland, ME).

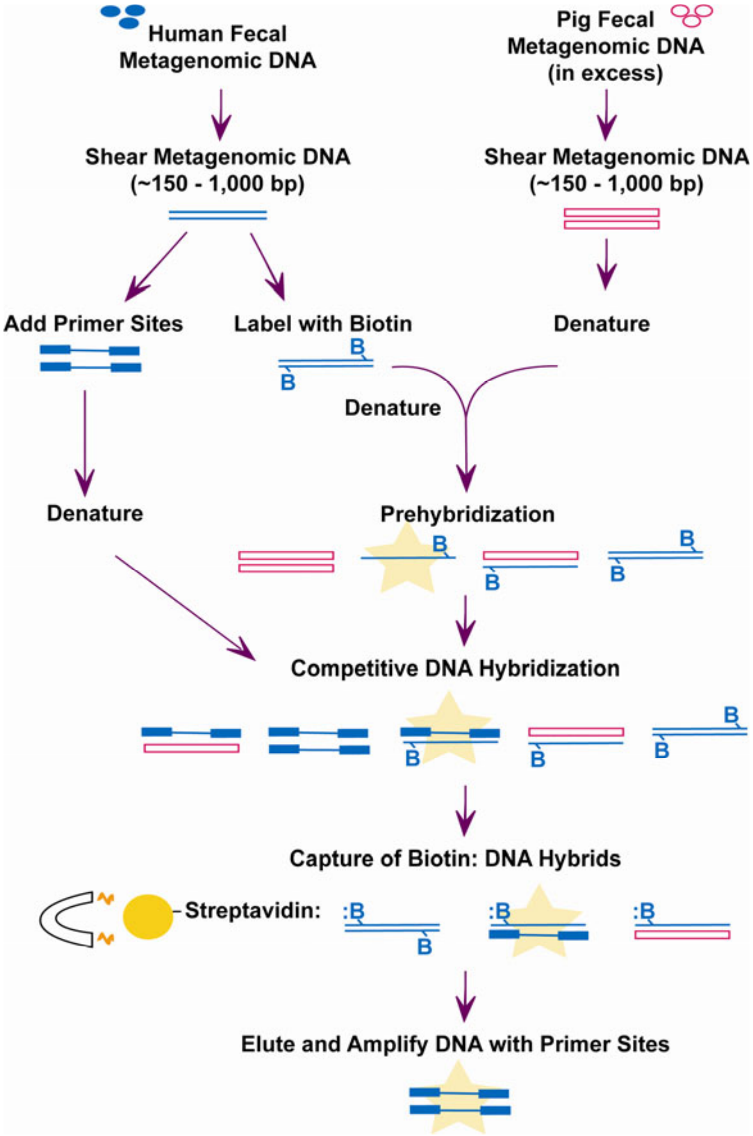
Human Fecal Metagenomic DNA
Pig Fecal Metagenomic DNA (in excess)
Shear Metagenomic DNA (~150 - 1,000 bp)
Shear Metagenomic DNA (~150 - 1,000 bp)
Add Primer Sites
Label with Biotin
Denature
B
B
Denature
Denature
Prehybridization
B
B
B
B
Competitive DNA Hybridization
B
B
B
B
Capture of Biotin: DNA Hybrids
Streptavidin:
:B
B
:B
:B
Elute and Amplify DNA with Primer Sites

Figure 2. Schematic representation of the approach used in our laboratory to generate host specific fragments. In this particular diagram the human fecal metagenomic DNA represents the pool of DNA that fragments are selected for while pig fecal metagenomic DNA was used as the tester. Each pool could be substituted with different metagenomic DNA depending on the type of host-specific assay.

RESULTS and DISCUSSION

Source identifiers must exhibit a high level of preferential host distribution. In most cases the level of host specificity is unknown, and therefore, it is difficult to assess how accurate current source tracking methods are at identifying primary sources. However, it is reasonable to presume that MST methods targeting genes that have little to do with host-microbial interactions will be more prone to cross-reactivity than methods targeting functional genes. We hypothesize that gut microbial populations that develop mechanisms to deal with long-term evolutionary pressures will possess functional genes that confer a selective niche advantage over potential competing populations. From an ecosystem standpoint, ecologically driven processes are responsible for the selection and exclusion of specific microbial populations. These processes in the gut are impacted by dietary, anatomical, and physiological constraints that may lead to three different types of host-microbial interactions: parasitic, commensal, or mutualistic. Gut microbial populations must handle the metabolic competition for the resources available as well as immunological pressure from the host. Evolution between host and microbe will therefore entail a network of genes playing a role in cell-cell recognition and communication (e.g., cell surface proteins, toxins, and adhesin) between different members of the gut microbial community and the host's gut cells. Metabolic pathways that confer advantage to a particular population will likely be selected over those that have a more neutral role in maintaining overall fitness. Overall, these are a good pool of genes to develop host specific markers for source identification. Unlike phylogenetic markers, there is little sequence information for these genes primarily because the identity of these genes in most cases is unknown. In order to address this issue, our laboratory has extracted total DNA from fecal microbial communities (i.e., fecal metagenomes) and challenged these fecal DNA extracts in separate competitive hybridization studies to enrich for DNA sequences that are present in one pool of DNA and absent in another.

Genomic hybridizations were performed to enrich for human, cattle, and chicken specific DNA fragments. Between 350 and 540 enriched metagenomic sequences were generated for each of the three potential fecal sources. *Bacteroidales*- and *Clostridium*-like sequences dominated the pool of potential markers in the human and cattle experiments, while *Clostridium*-like sequences were more predominant in the chicken metagenomic enrichments. Many of the enriched sequences are likely to be part of membrane associated proteins, suggesting that they are involved in host-microbial associations. Other genes include transcription factors, chaperones, and lipoprotein genes. Most sequences shared homology with bacterial genes, with some archaea genes also being represented in the enrichment clone libraries. Additionally, many clones had no significant similarity to sequences in the current databases. This is not surprising as there are few fecal metagenomes currently annotated and due to the enormous diversity of sequences present in complex microbial communities. The latter fact is relevant to source tracking from the standpoint of the large pool of targets that can potentially be used for environmental monitoring.

A limited number of sequences were selected to develop host-specific PCR assays. Some of the assays cross-amplify non-target samples suggesting that these genes are present in multiple host types and therefore are not good markers for source tracking. However, several human-, cattle-, and chicken-specific assays showed a high degree of host specificity, host distribution, and geographic stability. Interestingly, several of the chicken-specific assays generated positive signals when challenged against fecal DNA extracts from chicken, turkey, seagull, and geese suggesting the conservation of some genetic markers among different avian species (Table 2). Positive signals were also obtained when host-specific PCR assays were tested against DNA extracts from fecally contaminated surface waters.

Table 2. Number of clones examined and number of assays developed using metagenomic approach

Host	Number of metagenomic fragments analyzed	Number of host specific assays
Cattle	380	6
Human	351	4
Chicken	541[b]	12
Chicken/Turkey	541 [b]	3
Chicken/Turkey/Geese/Seagull[a]	541 [b]	4

[a] Assays tested positive to more than two avian species.

[b] Fragments obtained from the same study.

Differences in the occurrence of signals were observed for some host-specific assays when challenged against the same DNA extracts. This is probably due to the differences in limit of detection for each different assay or due to differential survival of targeted bacterial populations. Both limit of detection and survival of host-specific populations are important factors when applying source tracking assays. Relying on several markers and/or several host-specific assays might alleviate some of these problems. Since the clones examined in this study represent a small fraction of the total metagenomic sequences present in any given fecal microbial community, it is reasonable to assume that the number of potential host-specific markers is significantly larger than the portion of the markers selected in this study. The results from our studies suggest that competitive hybridization is a rapid and viable approach for the simultaneous identification of multiple genetic markers to identify fecal sources of pollution. This approach does not rely on the availability of sequence information and can potentially generate hundreds of microbial host-specific functional genes.

In the original GFE protocol (Shanks et al., 2006), DNA extracts from only one individual were used as tester and blocker. More recently, we have used DNA composites in an effort to better represent the diversity of host metagenomes of both tester and blocker. The latter was done to decrease the potential for assay cross-reactivity. However, this assumption needs to be validated in future studies. Additionally, we have used smaller amounts of tester and blocker DNA while still maintaining a tester-blocker ratio of 1:15. This is important as it is difficult to extract significant amounts of DNA from some of the potential sources (i.e., birds).

While culture independent methods are becoming popular among source tracking scientists, there are limited number of studies showing the host specificity and geographic stability of the assays. Moreover, most studies addressing these factors have only challenged the assays against a small number of individual feces and a limited number of different animals from relatively few geographically distinct locations. Laboratory cross-validation studies have not been performed for most MST methods and in most cases there are no universal standard operating procedures. In addition, there is little information on the survival of host-specific populations in the environment and the role sediments play as a potential repository of secondary habitat populations. The latter is important as the sediments can promote the survival and growth of bacterial groups relevant to source identification.

There are limitations associated with the current LIMs that are relevant to metagenomic markers as well. For example, most of the published methods depend on the direct amplification of the genetic targets from environmental DNA. DNA extraction procedures are known to co-extract substances that can cause PCR inhibition. This problem is often resolved by diluting the DNA extracts, although the number of false negatives might increase if the target marker is close to the detection limits in that particular sample. Another important limitation relates to the fact that most MST methods rely on targeting only one gene from only one bacterial group and that they target only one marker per host. The metagenomic approach herein described can rapidly generate multiple targets from different fecal microbial groups. However, some of the targets might be found in low numbers and therefore might not be sensitive enough when challenged against environmental samples. In order to increase detection limits it is possible to use an enrichment methods or fluorescent probes (Scott et al., 2005; Layton et al., 2006). Additionally, the success of any PCR assay depends on the presence of conserved regions used for primer annealing. For most assays representative sequence databases are lacking in spite of the fact that sequencing of host-specific targets is necessary to further confirm the specificity of the target in question.

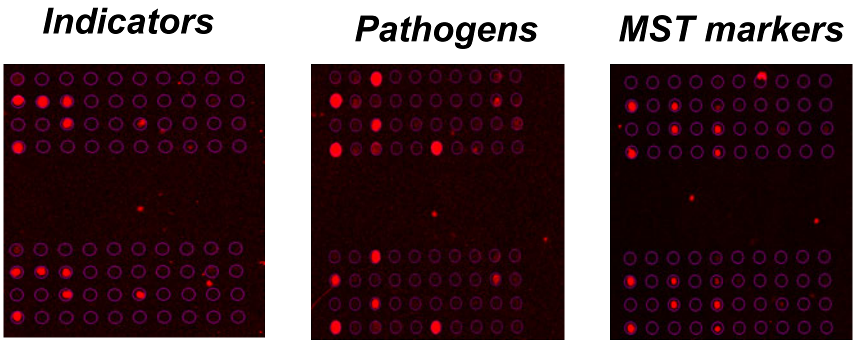

Figure 3. Hypothetical water quality biochip. Each panel could include genetic markers specific to the target in question. This biochip will be useful to develop microbial risk assessment models and to study the ecology of microbial populations relevant to microbial water quality.

For decades, microbial water quality has been performed using environmental monitoring tools that target bacterial indicators of fecal pollution. Problems associated with the poor correlation of pathogen levels with indicator densities are well documented (Harwood et al, 2005). Moreover, the differences between types of pollution sources (e.g., human vs. animal) and health risks are not well understood, although they are the basis of regulatory activities. The possibility for integrating a multi-level approach to environmental monitoring, risk assessment, and risk management is now possible due to the recent progress in biotechnology and genomics. Such approach could entail examining microbial water quality by targeting different classes of markers specific to indicator organisms, pathogens, and source identifiers (Fig. 3). While presence - absence data is useful to those identifying fecal sources, quantitative assays are needed in order to determine daily loads of specific fecal sources, to better assess fate and transport of fecal contaminants, and to establish the correlation between current bacterial indicators and public health risks. Since no single host-specific assay meets all the criteria of a perfect MST method, the use of multiple markers seems to be necessary in environmental monitoring of fecal pollution sources, regardless if the water is used for recreational activities or as a source of drinking water.

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.

Blanch, A. R., L. Belanche-Munoz, X. Bonjoch, J. Ebdon, C. Gantzer, F. Lucena, J. Ottoson, C. Kourtis, A. Iversen, I. Kuhn, L. Moce, M. Muniesa, J. Schwartzbrod, S. Skraber, G. Papageorgiou, H. D. Taylor, J. Wallis, and J. Jofre. 2004. Tracking the origin of faecal pollution in surface water: an ongoing project within the European Union research programme. J. Water Health. 2:249-260.

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

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.

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

Duran, M., B. Z. Haznedaroğlu, and D.H. Zitomer. 2006. Microbial source tracking using host specific FAME profiles of fecal coliforms. Water Res. 40:67-74.

Fong, T. T., D. W. Griffin, and E. K. Lipp. 2005. Molecular assays for targeting human and bovine enteric viruses in coastal waters and their application for library-independent source tracking. Appl. Environ. Microbiol. 71:2070-2078.

Geldreich, E. E., B. A. Kenner, and P. W. Kabler. 1964. Occurrence of coliforms, fecal coliforms, and streptococci on vegetation and insects. Appl Microbiol. 12:63-69.

Geldreich, E. E. and N. A. Clarke. 1966. Bacterial pollution indicators in the intestinal tract of freshwater fish. Appl. Microbiol. 14:429-437.

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

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

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., A. D. Levine, T. M. Scott, V. Chivukula, J. Lukasik, S. R. Farrah, and J. B. Rose. 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl. Environ. Microbiol. 71:3163-3170.

Hassan, W. M., S.Y. Wang, and R. D. Ellender. 2005. Methods to increase fidelity of repetitive extragenic palindromic PCR fingerprint-based bacterial source tracking efforts. Appl. Environ. Microbiol. 71: 512-518.

Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann. 2003. Putative temporal variability of *Escherichia coli* ribotypes from young steers. J. Environ. Qual. 32:305-309.

Jiang, J., K. A. Alderisio, and L. Xiao. 2005. Distribution of *Cryptosporidium* genotypes in storm event water samples from three watersheds in New York. Appl. Environ. Microbiol. 71:4446-4454.

Kibbey, H. J., C Hagedorn, and E. L. McCoy. 1978. Use of fecal streptococci as indicators of pollution in soil. Appl. Environ. Microbiol. 35:711-717.

King, E. L., D. S. Bachoon, and K. W. Gates. 2006. Rapid detection of human fecal contamination in estuarine environments by PCR targeting of *Bifidobacterium adolescentis*. J. Microbiol. Methods. Jul 27; [Epub ahead of print]

Khatib, L. A., Y. L. Tsai, and B. H. Olson. 2002. A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. 59:97-104.

Khatib, L. A., Y. L. Tsai, and B. H. Olson. 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. 63:231-238.

Lamendella, R., J. Santo Domingo, Oerther, D., Vogel, J., Stoeckel, D. 2006. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of *Bacteroidetes* 16S rDNA. FEMS Microbiol. Ecol. In press

Lasalde, C., R. Rodriguez, G. A. Toranzos, and H. H. Smith. 2005a. Heterogeneity of uidA gene in environmental *Escherichia coli* populations. J. Water Health. 3:297-304.

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

Leung, K. L., R Mackereth, Y Tien, 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.

Layton, A., L. McKay, D. Williams, V. Garrett, R. Gentry, and G. Sayler. 2006. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl. Environ. Microbiol. 72:4214-4224.

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

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

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

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

Muller, T., A. Ulrich, E. M. Ott, and M. Muller. 2001. Identification of plant-associated enterococci. J. Appl. Microbiol. 91:268-278.

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

Ram, J. L., R. P. Ritchie, J. Fang, F. S. Gonzales, and J. P. Selegean. 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J. Environ. Qual. 33:1024-1032.

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

Shanks, O., J. W. Santo Domingo, and J. Graham. 2006. Use of Competitive DNA hybridization to identify differences in the genomes of two closely related fecal indicator bacteria. J. Microbiol. Methods. 66:321-330.

Shanks, O., J. W. Santo Domingo, R. Lamendella, C. A. Kelty, and J. Graham. 2006. Competitive metagenomic DNA hybridization identifies host-specific genetic markers in cattle fecal samples. Appl. Environ. Microbiol. 72:4054-4060.

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.

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

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

Stewart, J. R., J. Vinje, S. J. Oudejans, G. I. Scott, and M. D. Sobsey. 2006. Sequence variation among group III F-specific RNA coliphages from water samples and swine lagoons. Appl. Environ. Microbiol. 72:1226-1230.

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.

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

USEPA (U.S. Environmental Protection Agency). 2005. Microbial Source Tracking Guide Document. Office of Research and Development, Washington, DC EPA-600/R-05/064. 131 pp. (http://www.epa.gov/ORD/NRMRL/pubs/600r05064/600r05064.pdf)

Van Donsel, D. J., E. E. Geldreich, and N. A. Clarke. 1967. Seasonal variations in survival of indicator bacteria in soil and their contribution to storm-water pollution. Appl. Environ. Microbiol. 15:1362-1370.

Wiggins, B. A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner. 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. 69:3399-3405.

Wiggins, B.A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman. 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. 65:3483-3486.

Wiggins, B.A. 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. 62:3997-4002.