Evaluation of molecular community analysis methods for discerning fecal sources and human waste

Yiping Cao[1], Laurie C. Van De Werfhorst[2,3], Eric A. Dubinsky[4], Brian D. Badgley[5,], Michael J. Sadowsky[5], Gary L. Andersen[4], John F. Griffith[1] and Patricia A. Holden[2,3]*

Abstract

Molecular microbial community analyses provide information on thousands of microorganisms simultaneously, and integrate biotic and abiotic perturbations caused by fecal contamination entering water bodies. A few studies have explored community methods as emerging approaches for microbial source tracking (MST); however, an evaluation of the current state of this approach is lacking. The present study utilized three types of community-based methods with 64 blind, single- or dual-source, challenge samples generated from 12 sources, including: humans (feces), sewage, septage, dogs, pigs, deer, horses, cows, chickens, gulls, pigeons, and geese. Each source was a composite from multiple donors from four representative geographical regions in California. Methods evaluated included terminal restriction fragment polymorphism (TRFLP), phylogenetic microarray (PhyloChip), and next generation (Illumina) sequencing. These methods correctly identified dominant (or sole) sources in over 90% of the challenge samples, and exhibited excellent specificity regardless of source, rarely detecting a source that was not present in the challenge sample. Sensitivity, however, varied with source and community analysis method. All three methods distinguished septage from human feces and sewage, and identified deer and horse with 100% sensitivity and 100% specificity. Method performance improved if the composition of blind dual-source reference samples were defined by DNA contribution of each single source within the mixture, instead of by *Enterococcus* colony forming units. Data analysis approach also influenced method performance, indicating the need to standardize data interpretation. Overall, results of this study indicate that community analysis methods hold great promise as they may be used to identify any source, and they are particularly useful for sources that currently do not have, and may never have, a source-specific single marker gene.

Introduction

Beach water quality is monitored for microbial contamination through measurements of fecal indicator bacteria (FIB), such as *E. coli* and *Enterococcus*

[1]*Southern California Coastal Water Research Project, Costa Mesa, CA*

[2]*University of California, Bren School of Environmental Science & Management, Santa Barbara, CA*

[3]*University of California, Earth Research Institute, Santa Barbara, CA*

[4]*University of California, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA*

[5]*University of Minnesota, BioTechnology Institute, St. Paul, MN*

[*]*Currently at: Virginia Tech, Department of Crop and Soil Environmental Sciences, Blacksburg, VA*

spp., as surrogates for human pathogens. However, FIB are not specific to any waste or fecal source. Assessment of public health risks and effective remediation of impaired waters therefore require identifying contributing sources through microbial source tracking (MST) studies. Numerous MST methods have been developed, most of which are single marker-based methods designed to discern one particular type of fecal source. Because no single marker gene has been found to be 100% sensitive and specific for its targeted source, a recommended strategy for MST is to simultaneously measure multiple markers (Harwood *et al.* 2005). In line with this strategy, a new class of MST methods based on molecular microbial community analysis is emerging as a useful addition to the MST tool box (Cao *et al.* 2011b).

Microbial communities are each a composite of populations comprised of thousands of microorganisms and markers, whose collective presence and relative abundance directly reflect conditions of the surrounding environment. The foundation of microbiological MST methods is that gut microbial communities of various hosts vary significantly by host species owing to differences in their gut environments, including the types of nutrients introduced by dietary differences (Ley *et al.* 2008, Shanks *et al.* 2011). Microbial communities in feces therefore differ by host animals; similarly, microbial communities in sewage differ greatly from those in pristine waters (McLellan *et al.* 2010). As feces (or sewage) enter ambient water, the microbial community in the receiving water can be altered directly by addition of microbes from the feces and indirectly by addition of chemicals from feces changing water chemistry. Characterization of the overall water microbial community therefore can be used directly in MST for discerning waste and fecal sources (Unno *et al.* 2010, Cao *et al.* 2011a, Dubinsky *et al.* 2012).

In contrast to single marker MST methods where one single marker is measured as one tracer for one type of fecal source, molecular microbial community analysis-based MST methods rely on culture-independent techniques to characterize hundreds or even thousands of markers simultaneously as tracers for all types of contamination sources (Lee *et al.* 2011). Some of these sources may not currently have source-specific single markers available. Motivated by the potential power of this class of MST method, several recent studies have developed the application of community analysis in MST. Cao *et al.* (2011a) demonstrated an integrated community analysis approach, combining terminal restriction fragment length polymorphism (TRFLP) community profiles with multivariate statistical analysis, for determining human waste contamination in a coastal creek in southern California, USA. Dubinsky *et al.* (2012) showed the capacity of a phylogenetic microarray (PhyloChip) for detecting influence from bird, grazer, and human fecal sources in marine waters from coastal California, USA. Unno *et al.* (2010) illustrated how next generation sequencing (454-pyrosequencing) community analysis was used to define sources of fecal contamination in a river basin in South Korea. However, a simultaneous evaluation of these various community analysis methods for differentiating sources is currently lacking.

In this study, three types of community analysis methods were evaluated with 64 blind, single- or dual-source, samples generated from 12 fecal sources, including those from: humans (feces), sewage, septage, dogs, pigs, deer, horses, cows, chickens, gulls, pigeons, and geese. Methods evaluated included community fingerprinting (TRFLP, with two TRFLP assays included), a phylogenetic microarray (PhyloChip), and next generation sequencing (Illumina). The goal of this study was to assess the general performance of these methods for discerning various sources in unknowns when the sources were provided as references, and to deliberate factors affecting the performance metrics. This study did not intend to characterize geographic and population variations of microbial communities for the same host fecal sources; thus also did not evaluate if reference samples from different geographic regions or host populations could generally serve for source identification by community analysis methods.

Methods

Study Design

Sixty-four blind challenge and 12 reference samples, created from freshly collected fecal material from the 12 sources described above, were used for the evaluation. The 64 challenge samples (i.e., a blind duplicate set of 32 blind samples) contained either a single fecal source (38 singletons) or two fecal sources (26 doubletons). Each fecal source was a composite of at least 12 individuals (or 9 sewage treatment, or 6 septage collection, facilities) with equal contribution from 4 representative California geographies: central CA, Los Angeles county,

Orange county, and San Diego county. A singleton slurry was made for each composite fecal source via blending to mix the 6-12 individual fecal samples in the appropriate volume of 0.2 µm pore size filtered artificial freshwater. The 38 singleton challenge samples included 24 full strength and the fourteen 1:10 strength singletons, which were created by filtering 200 ml and 20 ml of the corresponding singleton slurry, respectively, through polycarbonate membrane filters (Isopore Millilpore, 47 mm dia., 0.4 µm pore size). Each of the 26 doubleton samples was created by filtering 200 ml of a corresponding doubleton slurry created by mixing 90 and 10% (by volume) of the corresponding singleton slurries. An additional set of 12 full strength singleton samples, one for each of the 12 fecal sources, was created from the same singleton slurries and provided as known reference samples. More details on the field fecal material collection and laboratory sample preparation are described elsewhere (Boehm *et al.* 2013).

All 76 samples (64 unknown + 12 reference samples) were analyzed by each of the following methods (see Supplemental Information (SI) Figure SI-1; ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_461_474SI.pdf): TRFLP targeting all bacteria (Univ-TRFLP; Cao *et al.* 2011a) or the order of Bacteroidales (Bac-TRFLP; Cao *et al.* 2013), PhyloChip targeting bacteria (Dubinsky *et al.* 2012), and Illumina next generation sequencing targeting bacteria. Potential source(s) in the 64 blind challenge samples were determined through comparing microbial communities in the challenge samples to those in the 12 reference samples (according to procedures described in the following corresponding subsections).

TRFLP Assays

Laboratory procedures for the two TRFLP assays, Univ-TRFLP and Bac-TRFLP, were the same except for the PCR step where genes encoding 16S rRNA were amplified. Briefly, following DNA extraction and quantification, duplicate PCR was performed to amplify genes encoding 16S rRNA from either all Bacteria (Univ-TRFLP) or Bacteroidales (Bac-TRFLP); the pooled PCR products were then purified and digested with each of the two restriction enzymes *Hha*I and *Msp*I separately; the digested products were analyzed on a capillary gel to provide TRFLP community profiles in the form of electropherograms. DNA was extracted with the DNA-EZ kit following the manufacturer's protocol (GeneRite, North Brunswick, NJ), quantified by a fluorometric assay for total DNA concentration (Quant-iT™; Invitrogen, Carlsbad, CA), then stored at -20°C until use. For Univ-TRFLP, the 50 µl PCR reactions used 0.525 µM each of universal bacterial primers 8F hex (fluorescently labeled forward primer; 5'-AGAGTTTGATCCTGGCTCAG) and 1389R (5'-ACGGGCGGTGTGTACAAG) and 25 ng (or 5 µl maximum) sample DNA. For Bac-TRFLP, the 50 µl PCR reactions used 0.525 µM fD1-Hex (fluorescently labeled forward primer fD1; 5'-AGAGTTTGATCCTGGCTCAG) and 0.5 µM rBacPre (5'-TCACCGTTGCCGGCGTACTC, (Wood *et al.* 1998)) and 16 ng of sample DNA. The PCR thermal programs and other details are described elsewhere (Bac-TRFLP in Cao *et al.* 2013; Univ-TRFLP in Cao *et al.* 2006). The same DNA extracts were used for both TRFLP assays. Two laboratories performed both assays following the same corresponding standard operating procedures including the entire process from DNA extraction to data analysis for each TRFLP assay. Archived DNA from an activated sludge sample (Montecito Sanitary District, Santa Barbara, CA) was analyzed by both laboratories, and the electropherograms were compared for quality assurance.

Data analysis for determining sources in the unknown challenge samples was based on the similarity of the overall community between unknown and reference samples and is described in detail elsewhere (Cao *et al.* 2011a, 2013). Briefly, first, raw TRFLP data were processed to provide two multivariate datasets (one for each of the two restriction enzymes) with samples as rows and relative abundance of operational taxonomic unit (OTUs) as columns. Here the OTUs are terminal restriction fragments (Liu *et al.* 1997). Then, two multivariate analysis techniques (detrended correspondence analysis and Bray-Curtis similarity analysis) were performed on each dataset to identify the reference sample(s) to which an unknown challenge sample was most similar. The source represented by the identified reference sample(s) was then deemed to be present in the unknown sample. Lastly, source identification results from analyzing TRFLP data for each of the two enzymes with two multivariate techniques were combined to provide one final source identification answer for each of the 64 blind samples. The Univ- and Bac-TRFLP data were analyzed separately to provide two separate sets of source identification

answers. Additionally, the pair of final answers from Univ- and Bac-TRFLP for each sample was considered together to report a final answer from combining both TRFLP assays (Univ- and Bac-TRFLP).

PhyloChip Analysis

Laboratory procedures for PhyloChip analysis are described in detail elsewhere (Dubinsky *et al.* 2012). Briefly, following DNA extraction and quantification, replicate PCR was performed to amplify genes encoding 16S rRNA from Bacteria; pooled PCR products were purified then fragmented with DNAaseI; the fragmented products were then labeled with biotin followed by hybridization overnight onto the microarray; the microarray was then stained and scanned to provide raw PhyloChip data in the form of fluorescent image files. DNA was extracted as in TRFLP, quantified by a fluorometric assay for total DNA concentration (QuBit; Invitrogen), then stored at -20°C until use. The bacterial 16S rRNA gene was amplified from each sample using PCR with primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). Each 25 µl PCR reaction contained 1× Ex Taq buffer (Takara Bio Inc., Japan), 0.025 units/µl Ex Taq polymerase, 0.8 mM dNTP mixture, 1.0 µg/µl BSA, and 200 pM each primer and 1 ng DNA (gDNA) as template for the 12 known reference samples and 10 ng gDNA for the 64 unknown challenge samples. For the PhyloChip assay each sample was amplified in 8 replicate 25 µl reactions spanning a range of annealing temperatures. PCR conditions were 95°C (3 minutes), followed by 30 cycles 95°C (30 seconds), 48-58°C (25 seconds), 72°C (2 minutes), followed by a final extension 72°C (10 minutes).

Two approaches were used to analyze the fluorescent image files following array scanning. The first used the standard OTU approach described in Dubinsky *et al.* (2012). In this approach the presence of 59,316 different bacterial OTUs was determined by positive hybridization of multiple probes that correspond to distinguishing 16S rRNA gene polymorphisms (average of 37 probes/OTU). The 12 reference samples were used to define identifier OTUs for each source. For each source, OTUs unique to that source were identified. For this analysis, some sources that are known to have similar bacterial communities were grouped based on Dubinsky *et al*. (2012). Grouped sources were human wastes (human feces, sewage, septage), wild birds (gull, goose, pigeon) and domestic grazing mammals (cow, horse). Identifier OTUs for individual sources in each of these three groupings could be shared with other sources in the same group but not with sources outside the group. For all other sources (dog, pig, deer, chicken) an OTU needed to be exclusive to the individual source to be recruited as an identifier OTU. For source determination of blind samples, the OTU composition of each sample was determined by PhyloChip analysis as described above and screened for identifier OTUs for each source. A source was considered a match if >20% of its identifier OTUs were found in the blind sample. If two or more sources met these criteria, and those sources were all found in one of the three source animal groupings (human wastes, wild birds or domestic grazing mammals), then the source with the highest percentage of matches was considered the true match. The 20% cutoff was chosen because previous work (Dubinsky *et al.* 2012) showed that 20% was the minimum percentage of identifier taxa matching a known fecal source in waters exceeding FIB limits. This cutoff was shown effective in field tests of marine waters that were contaminated with sewage or bird feces.

The second analysis approach considered each of the PhyloChip's 1,016,064 individual oligonucleotide probe features individually. Each reference sample was screened for probes that exceeded 100 and 1000 fluorescence intensity units upon hybridization. Source identifier probes were defined as probes that exceeded 1000 intensity units in the source sample but never exceeded 100 intensity units in any other reference sample, unless the other samples were in the same source animal grouping (human wastes, wild birds or domestic grazing mammals). For source determination of blind samples using the probe-based approach, the probes that exceeded 100 intensity units in each blind sample were determined and screened for identifier probes of each source. A source was considered a match if >20% of its identifier probes were found in the blind sample. If two or more sources met these criteria, and those sources were all found in one of the three source animal groupings (human wastes, wild birds or domestic grazing mammals), then the source with the highest percentage of matches was considered the true match.

Illumina Sequencing

Briefly, following DNA extraction and quantification, triplicate PCR was performed to amplify the V6 hypervariable regions (Huber *et al.* 2007) of the

16S rRNA gene from Bacteria; PCR products were purified then pooled for Illumina next generation sequencing. DNA was extracted from filters using MO BIO PowerSoil DNA extraction kits (MO BIO, Carlsbad, CA), according to the manufacturer's instructions. DNA was quantified using a QuBit DNA quantification system (Invitrogen) with Qubit high sensitivity assay reagents, then stored at -20°C until use. All PCR reactions used 25 ng (or 10 µl maximum) of DNA as template and were performed in triplicate. Primer sets were designed with a 6 bp ID tag on the 5' end of the reverse primer, which was unique to each DNA sample. This allowed for multiplexed sequencing. PCR amplicons were visualized using gel electrophoresis to confirm amplification of properly sized products. Reactions were each purified using the Qiaquick PCR purification kit (Qiagen, Valencia, CA), eluted in 30 µl of 10 mM Tris-Cl buffer, pH 8.0, and pooled. Purified PCR products were quantified as with the DNA extracts above, then stored at -20°C before pooling for sequencing.

Equimolar aliquots of each PCR product (12 reference and 64 blind samples) were pooled to give ~1 µg of DNA in a 100 µl total volume. Final pooled DNA concentrations were measured as with the DNA extracts above. Amplicon size analysis was done using an Agilent DNA 1000 chip and a 2100 BioAnalyzer (Agilent, Santa Clara, CA). The pooled samples were sequenced, as paired end reads, at the University of Minnesota Biomedical Genomics Center (St. Paul, MN) using Illumina Miseq technology, following the manufacturer's protocols (Illumina, Hayward, CA).

Sequence data were processed and analyzed using the Fastq-Join program (http://code.google.com/p/ea-utils/wiki/FastqJoin) and the MOTHUR program (Schloss *et al.* 2009). Because amplicon sizes were small enough that reads in each pair overlapped, paired ends were merged using the Fastq-Join program. Merged sequences were binned according to barcode sequence, and barcode and amplicon primer sequences were trimmed using the MOTHUR program (Schloss *et al.* 2009). To ensure high quality data for analysis, sequence reads containing ambiguous bases, homopolymers >7 bp, more than one mismatch in the primer sequence, or an average per base quality score below 35 were removed. Sequences that only appeared once in the total set were assumed to be a result of sequencing error and removed from the analysis. Chimeric sequences were also removed from the data set using the UCHIME algorithm within the MOTHUR program (Edgar *et al.* 2011). Using these criteria, 18.5 million initial sequences were filtered down to a total of 12.9 million quality sequences, ranging from 20,000 to 320,000 per sample.

For determining sources in challenge samples, the unique sequence reads were directly analyzed, or clustered into OTUs and then analyzed, by multivariate analysis techniques. For either approach, a random subset of 90,000 sequences was chosen from each sample to balance read numbers, except for five samples that returned less than 90,000 sequences. For these five samples, all available reads were used. The subsampling size was also varied from 20,000 to 90,000 per sample to evaluate the potential effect of subsampling size (or depth of sequencing) on source identification. For the OTU approach, this subset of sequence reads was aligned to the RDP7 16S rRNA database and clustered into OTUs at a cutoff value of >90% or >97% (Cole *et al.* 2009). Taxonomy was assigned to OTU consensus sequences using the RDP7 taxonomy database using the Bayesian method with a bootstrap algorithm (100 iterations) and a probability cutoff of 0.60. The overall microbial communities from unknown challenge samples were compared to those from reference samples to determine their sources using either the OTUs or unique sequence reads with various subsampling sizes. Dendrograms, based on Bray-Curtis distances, were used to cluster samples with similar communities together. An unknown blind sample that clustered with a reference sample was reported to have the source represented by the reference as dominant source. When an unknown sample did not indicate clear clustering with any reference samples on the dendrogram, the raw Bray-Curtis distances were used to determine the most similar source in the unknown sample. Bray-Curtis distances and dendrograms were generated using the MOTHUR program (Schloss *et al.* 2009). Multiple distance measures (UniFrac and BC) and multivariate analysis techniques (PCoA, NMDS, and dendrogram) were used in the exploratory data analysis stage on selected samples to link sources with unknowns. These method variations yielded similar results. Therefore, for the formal data analysis on all samples, dendrograms based on BC distances and the raw BC coefficient were used for source identification, as they were easy to perform and sufficient for the source identification tasks. Additionally, analyzing the data with various subsampling size (20,000 to 90,000 reads) at the

unique read or OTU levels (with 97 or 90% similarity cutoff for clustering reads into OTUs) led to identical source identification answers, thus only one set of Illumina results was reported.

Performance Evaluation

Two sets of sample keys were used to define source(s) present in the 64 blind challenge samples. The ENT key was based on percentage *Enterococcus* contribution (by EPA Method 1600) from each source to each blind sample, and the DNA key was based on percentage DNA contribution (determined by NanoDrop method). Briefly, concentrations of *Enterococcus* measured in the source slurries (or total DNA measured in the single-source samples) were used to approximate the proportion of enterococci (or total DNA) contributed by each source to the dual-source samples based on a 90 and 10% (by volume as during the dual-source challenge sample preparation) *in silico* mixing. The ENT and DNA keys do not differ for single-source challenge samples but do affect the interpretation of dominant source in some of the dual-source challenge samples (Table SI-1). Dominance was defined as when the contribution from one source was at least two times the contribution from the other source. For a few samples (three by the DNA key and two by the ENT key, Table SI-1), the contribution from one source was higher but less than or equal to two times the contribution from the other source. For these samples, either source could be considered as the dominant source for performance evaluation.

The reported source identification results from TRFLP, PhyloChip, and Illumina sequencing were compared to both keys for the performance evaluation. Each result was classified into one of seven categories depending on how it compared to the key (Table 1). For singletons, the percentage of correct identification was calculated as the number of samples where the source was correctly identified (i.e., category "correct") divided by the number of samples where an answer was reported. For doubletons, the percentage of correct identification was calculated as the number of samples where the dominant source was correctly identified and no incorrect source was listed (i.e., sum of categories "correct" and "correct dominant & similar minor") divided by the number of samples where an answer was reported.

Additionally, sensitivity and specificity were calculated for all 12 sources separately. The three human sources were considered either together as one category of source (as done with single human-associated indicator assays; Boehm *et al.* 2013) or separately as three different sources. For each particular source A, sensitivity was calculated as the number of challenge samples correctly identified as containing source A divided by the total number of samples that contained source A; specificity was calculated as the number of challenge samples that was not falsely reported as containing source A divided by the total number of samples that did not contain source A. Note that sensitivity and specificity metrics of TRFLP (Univ), TRFLP (Bac), and PhyloChip (OTU) for combined human (feces, septage, sewage together), and the other non-human sources (except pigeon) were also reported elsewhere in comparison with single indicator assays (Boehm *et al.* 2013). Also note that the three community analysis methods as evaluated in the current study were each a complete entity including the whole process from DNA extraction to amplicon detection, as depicted in Figure SI-1. The performance evaluation

Table 1. Seven categories of how reported results for challenge samples compared to the key.

Category	For Singletons	For Doubletons
Correct	Source correctly identified	Dominant source (or both sources) correctly identified
Correct dominant & similar minor*	n/a	Correct dominant source but similar minor source identified *
Minor source instead	n/a	Minor source correctly identified but dominant source not identified
Similar minor source*	n/a	Similar minor source identified *; dominant source not identified
Partially wrong	Two sources listed but only one correct	Correct dominant source listed with one or more incorrect source(s), or correct minor source listed with one or more incorrect source(s)
Wrong	Source(s) listed not present in sample	
No answer	No answer provided	

* Similar sources refer to human feces, sewage, and septage which were all considered human waste

was therefore on the complete methods as opposed to on the detection platforms alone. As such, potential target and PCR differences were organic parts of the differences between methods, and investigation on these detailed elements was thus beyond the scope of the study.

Results

Regardless of the molecular methods and data analysis approach used, a high percentage of correct identification was achieved by the community analysis methods when tested with 12 sources in the form of single- and dual-source challenge samples (Table 2). The greatest percentages of correct identification of singletons were 100, 95, and 92% for TRFLP, Illumina, and PhyloChip, respectively. The greatest percentages of correct identification of the dominant source in doubletons were 100, 96, and 92% for PhyloChip, Illumina, and TRFLP, respectively. While PhyloChip provided answers to all challenge samples, there were two to eight samples for which TRFLP data did not have sufficient evidence for source identification, and there was one sample where the Illumina method could not provide an answer. As performance from the two laboratories using the TRLP method were largely similar, only one lab's results (with slightly better performance metrics) are presented in this manuscript. Detailed across-laboratory evaluation of the TRFLP method is presented elsewhere (Cao *et al.* 2013).

Community analysis method performance with doubletons greatly improved when performance was evaluated against the DNA key instead of the ENT key. Both the number (an increase of 2 to 9 samples; Figure 1) and overall percentage (a jump of 8 to 36%; Table 2) of correct identifications increased from evaluation based on the ENT key to evaluation based on the DNA key. Using the ENT key, the community analysis methods reported the minor source as the dominant source for 8 to 35% of doubletons. Indeed, 92% of the reported answers that were categorized as "minor source instead" occurred when the ENT key was used. However, when the DNA key was used, the dominant source in those doubletons was correctly identified by most of the community analysis methods (Figure 1).

The data analysis approaches greatly impacted the performance of TRFLP and PhyloChip, but did not change source identification by Illumina (Table 2; Figure 1). For TRFLP, combining Univ- and Bac-TRFLP information together reduced the number of challenge samples with no results and increased the overall percentage of correct identification. For PhyloChip, analyzing the data at the probe level vs. at the OTU level increased percentages of correct identification by 16 and 11% for singleton and doubleton, respectively. Similarly, the number of incorrect identifications (i.e., partially wrong and wrong; Figure 1) was also reduced when combining Univ & Bac (vs. considering Univ and Bac separately, for TRFLP) or analyzing data at the probe level

Table 2. Summary of overall performance.

Molecular Method	Data Analysis	Singleton (n = 38)[a]		Doubleton (n = 26)[b]			All (n = 64)
		% Correct	No Answer[c]	% Correct (DNA Key)	% Correct (ENT Key)	No Answer[c]	% Correct[d]
TRFLP[e]	Univ	91%	4	86%	50%	4	-
	Bac	100%	4	84%	52%	1	-
	Univ & Bac	97%	1	92%	56%	1	95%
PhyloChip	OTU	76%	0	77%	46%	0	-
	Probe	92%	0	100%	92%	0	95%
Illumina[f]	Reads/OTU	95%	0	96%	60%	1	95%

[a] For singletons, the percentage of correct identification was calculated as the number of samples where the source was correctly identified (i.e., category "correct", Table 1) divided by the number of samples where an answer was reported.

[b] For doubletons, the percentage of correct identification was calculated as the number of samples where the dominant source was correctly identified and no incorrect source was listed (i.e., sum of categories "correct" and "correct dominant & similar minor", Table 1) divided by the number of samples where an answer was reported.

[c] The no answer column lists the number of challenge samples where no source identification answer was provided by the community analysis method.

[d] Overall % correct were based on DNA key for doubletons, and "-" indicates not calculated.

[e] As TRFLP performance from the two laboratories was largely similar, only one lab's results (with slightly better performance metrics) are presented. Detailed cross-laboratory evaluation of the TRFLP method was presented elsewhere (Cao et al. 2013).

[f] All data analysis approaches (Reads, OTU at 90%, or OTU at 97%) for Illumina provided the same source identification answers for all 64 samples. Therefore, only one set of Illumina results is presented and the data analysis approach is denoted by reads or OTU (i.e. reads/OTU).

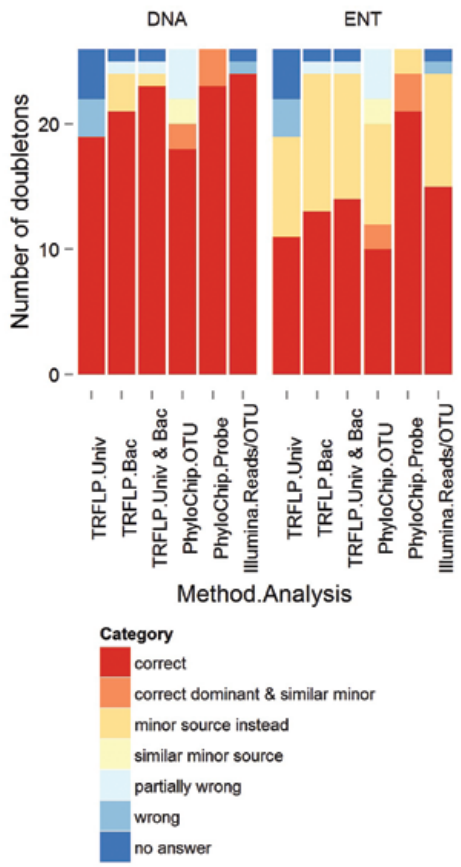

Figure 1. Performance evaluation of doubletons based on DNA and ENT keys.

(vs. analyzing data at the OTU level, for PhyloChip). However, for Illumina, analyzing the data at either the unique read or OTU level provided identical source identification results for all 64 samples.

The degree of "correctness", as defined in Table 1 and based on the DNA key, indicated trends for certain sources and/or community analysis methods. All challenge samples containing septage (n = 6), deer (n = 2), and horse (n = 4) were correctly identified with highest degree of correctness (i.e., category of "correct" in Table 1), regardless of community analysis method and data analysis approach. Challenge samples containing goose (n = 4) were rarely correctly identified by TRFLP regardless of data analysis approach. Yet, for challenge samples containing cow, pig, and dog, Bac-TRFLP and Univ&Bac-TRFLP performed better than Phylochip with OTU analysis, and better than or at least similar to PhyloChip with probe analysis or Illumina. Both Bac- and Univ&Bac-TRFLP provided "correct" (as defined in Table 1) answers for all challenge samples containing cow (n = 8), pig (n = 10), and dog (n = 8), while PhyloChip and Illumina methods did not achieve the "correct" category for up to 4 samples for each source.

Regardless of the source(s), community analysis methods exhibited excellent specificity, rarely reporting a source that was not present (Table 3). Further, all methods were able to distinguish the three different types of human waste sources (i.e., septage, sewage, and human feces), except where PhyloChip reported sewage as a minor source for six doubletons containing human feces as the minor source (OTU approach), and reported septage as a minor source for three doubletons containing sewage as the minor source (probe approach). The higher resolution PhyloChip (probe analysis approach) and Illumina methods were even able to distinguish pigeon and gull, while TRFLP could not regardless of the data analysis approach (Tables SI-2 and SI-3). Nevertheless, PhyloChip with a lower resolution data analysis approach (i.e., OTU) did not distinguish cow and horse (Tables SI-2 and SI-3).

Sensitivity varied by target source and community analysis method, and by the type of challenge sample (Tables 3 and 4). Sensitivity to a few sources was particularly low: Bac-TRFLP was insensitive to gull (Tables SI-2 and SI-3); all community analysis methods were more sensitive to human feces and septage than to sewage, which by nature is a mixture of multiple fecal and non-fecal sources (Table 3). However, the type of challenge sample containing each target source appeared to have a more prevalent effect on sensitivity in that sources that were represented in the 64 challenge samples mostly as singleton or dominant in doubletons were identified with higher sensitivity. The majority of the false negatives (19 out of 22) occurred in doubletons, among which most (14 out of 19) contained the target source as a minor source (Table 4). For example, all human source (sewage, septage, and human feces) false negatives occurred in doubletons; all challenge samples where both replicates were false negative contained the target source as a minor source. While sensitivity to sewage was lower across all methods than to human feces, 40% of the challenge samples containing sewage contained sewage as a minor source whereas only 20% of the challenge samples containing human feces contained human as a minor source (Table 3). Additionally, the highest resolution community analysis approach in this study (i.e., PhyloChip with data analysis based on >1million individual probes) was generally the most sensitive. Within a given molecular community analysis

Table 3. Sensitivity and specificity of community analysis methods for all sources, calculated based on all 64 blind samples.

Source	TRFLP[c] (Univ& Bac)		PhyloChip (Probe)		Illumina[c]		Number of Source-Containing Samples for Evaluation[e]					
	sen	spe	sen	spe	sen	spe	total n	full.S	diluted.S	dom.D	mix.D	minor.D
septage	0.67	1	1	0.95	0.67	1	6	2	2	-	-	2
sewage	0.43	1	0.71	1	0.43	1	14	2	2	2	2	6
human[a]	0.89	1	1	1	0.72	1	18	2	2	6	4	4
HUMAN[b]	0.68	1	0.97	1	0.61	1	38	8	8	8	6	12
chicken	0.75	1	1	1	1	1	4	2	-	2	-	-
goose	0.25	1	1	0.97	1	0.97	4	2	-	2	-	-
gull[d]	0.5	0.94	0.83	1	0.5	1	12	2	2	2	2	4
pigeon[d]	1	0.89	1	1	1	1	2	2	-	-	-	-
cow	0.75	1	0.75	0.98	0.38	1	8	2	2	2	-	2
deer	1	1	1	1	1	1	2	2	-	-	-	-
dog	0.63	1	1	1	0.5	1	8	2	2	-	2	2
horse	1	1	1	1	1	1	4	2	-	2	-	-
pig	1	0.98	1	1	1	0.98	8	2	2	2	2	-

[a] Lower case human refers to human feces.

[b] Capital HUMAN refers to all three human sources (human feces, sewage, septage) combined.

[c] Only one lab's TRFLP results and one set of Illumina results are presented as in Table 2.

[d] Note that only PhyloChip (probe) and Illumina could distinguish between gull and pigeon. For TRFLP and PhyloChip (OTU), gull and/or pigeon were reported when either source was identified. A "gull and/or pigeon" answer was considered either a true positive or false positive, depending on the source being evaluated and whether or not that source was present in the unknown challenge sample. All other occasional "and/or" answers were evaluated similarly.

[e] For each source, total n refers to the total number of challenge samples that contained the target source. Full.S, diluted.S, dom.D, mix.D, minor.D refer to, respectively, the number of full strength singletons, 1:10 strength singletons, doubletons where the target source is the dominant source, doubletons where dominance could not be established, and doubletons where the target source is a minor source. Dominance determination was based on % DNA contribution from each source that makes up the doubleton. Dominance could not be established if contribution from one source was higher but less than two times higher than the contribution from the other source.

method, high resolution data analysis approaches also reduced the number of false negative for human feces, e.g., for TRFLP (Univ- and Bac- combined vs. separate) and for PhyloChip (probe vs. OUT; Table 4). Moreover, sensitivity was sometimes lower for the Illumina method as it aimed at identifying the dominant source only, while TRFLP and PhyloChip reported one, or, when sufficient evidence existed in data, two sources. TRFLP and PhyloChip reported two sources for 1 to 4 and 12 to 24 challenge samples, respectively, depending on the data analysis approach. However, the Illumina method can be used to partition more than one source in a sample by incorporating shared OTUs into the data analysis (e.g., Unno *et al.* 2010, Knights *et al.* 2011). However, the software development for such a tool is yet to be completed.

For the different sources tested in this study, all community analysis methods had 100% sensitivity and specificity for deer and horse (Table 3). At least one community analysis method had 100% sensitivity and specificity for human feces (PhyloChip), chicken (PhyloChip, Illumina), pigeon (PhyloChip, Illumina), dog (PhyloChip), and pig (PhyloChip). At least one community analysis method had >80% sensitivity and >80% specificity for each of the 12 sources except sewage and cow (Table 3). Reported answers for all 64 challenge samples and a detailed performance evaluation by each source are presented in Tables SI-2 and SI-3.

Discussion

Microbial community analysis methods demonstrated great promise to become universal MST tools for identifying any source, or even dual sources simultaneously. Because community analysis methods identified sources by characterizing the microbial communities in the suspected sources then comparing with unknown sample communities, their source identification capacity was not restricted by the single marker genes that have been developed, or by the need to perform multiple single marker assays for multiple sources. Theoretically, the suspected source can be any source. This unrestricted and comparative nature of this class of MST methods enabled the three community methods to correctly identify the dominant sources in 95% of the unknown samples created from 12 different sources and to successfully identify sources such as deer that has no single marker MST assay. While this evaluation study focused on identifying dominant sources, the

Table 4. Number of false negatives by targeted source, unknown challenge sample, and community analysis method. False negatives were determined against each targeted source, i.e., a doubleton challenge sample could be false negative to either contributing source. Note that n = 2 for each unknown challenge sample, and a "2" in the table indicates both replicates was false negative for the target source.

Source [a]	Unknown Challenge Sample [b]	TRFLP			PhyloChip		Illumina
		Univ	Bac	Univ & Bac	OTU	Probe	Reads/OTU
septage	horse:septage 82:18	2	2	2	2		2
sewage	chicken:sewage 92:8		2	2	2	2	2
	gull:sewage 92:8	1		2	2		2
	pig:sewage 99:1	2	2	2	2	2	2
	sewage:pig 54:46	2	2	2			2
human	cow:human 92:8	2	2	2	2		2
	dog:human 62:38	1			2		
	goose:human 91:9	2			2		2
	human:cow 88:12	2					
	human:dog 98:2						1
	human:gull 98:2	1					
gull	gull 1:10		2				
	gull:human 56:44	2	2	2	2		2
	gull:sewage 92:8		2				
	human:gull 98:2	2	2	2	2	2	2
	sewage:gull 88:12	2	2	2	2		2
cow	cow						1
	cow 1:10	1			1		1
	cow:human 92:8	1			2		1
	human:cow 88:12	2	2	2	2	2	2
dog	dog:human 62:38	2		1			2
	human:dog 98:2	2	2	2	2		2

[a] Only the three human sources and all animal sources that were present in at least eight challenge samples (i.e. sample size >8, Table 3) were presented here. Animal sources with a small sample size were not subjected to this detailed examination of false negatives.

[b] Cow, cow 1:10, and gull 1:10 denote cow full strength singleton, 1:10 strength singleton, and gull 1:10 strength singleton. The doubletons are denoted by the two contributing sources followed by numeric proportions representing the percent DNA contributions from each source to the doubleton. A source is said to be dominant in the doubleton if its contribution is at least two times the contribution from the other source.

capability of community analysis methods to identify minor sources or multiple sources simultaneously has been realized through either a superior data analysis approach (i.e., analyzing PhyloChip data at the probe level, this study) or a more focused target microbial community (Bac-TRFLP targeting the order of Bacteroidales in this study, or pyrosequencing targeting the phylum of Bacteroidetes in Unno *et al.* 2011).

Community analysis methods inherently use multiple lines of evidence for identifying a source, instead of relying on detection of one host-specific DNA marker as in single marker PCR or qPCR assays. The multiple lines of evidence are reflected in the fact that overall community similarities (TRFLP, Illumina) or multiple source-specific identifier OTUs or identifier probes (PhyloChip; 23 - 466 identifier OTUs per source, 50 - 7703 identifier probes per source) were used to identify source(s) in the blind samples. This characteristic of generating multiple lines of evidence likely contributed to the excellent specificity observed for all sources tested, and to the capability to distinguish closely-related, within-group sources. For example, while no PCR or qPCR assay could differently trace human fecal contamination to septic systems versus leaking sewer versus transient populations, all three community analysis methods were capable of distinguishing each of the three types of human waste (septage, sewage, and raw feces) within the human waste group. Additionally,

while many gull-specific single indicator PCR or qPCR assays cross-reacted with pigeon and/or goose sources (Boehm *et al.* 2013), PhyloChip (probe) and Illumina were able to distinguish all three birds. Other benefits of using community analysis methods include more tolerance to temporal variability and less susceptibility to geographic variations because such methods do not entirely depend on the fate of one single marker that is developed in specific geographic regions (Cao *et al.* 2011a, Dubinsky *et al.* 2012).

Despite their advantages, community analysis methods usually have lower sensitivity than single indicator PCR or qPCR assays (Cao *et al.* 2011a, Boehm *et al.* 2013, the present study). Because community analysis methods measure all indicators and target all sources simultaneously, signals from the less abundant (or rare) sources can be low and overwhelmed by signals from dominant contributing sources. This explains why false negatives in this study occurred mostly with doubletons and even more frequently with the minor sources in the doubletons. This may also partially explain the lower sensitivity with sewage, naturally a multiple-source mixture, compared to that with pure human feces. It is reasonable that it would be easier to match an unknown doubleton (containing human feces or sewage and another animal source) to a "pure reference source" (i.e., human feces) than to a "mixed reference source" (i.e., sewage which may itself contain other animal sources), particularly when sewage was not the dominant source in the unknown doubleton. Nevertheless, detection of minor contributing sources can be improved through utilizing higher resolution data analysis approaches (i.e., probe-based data analysis for PhyloChip) or higher resolution molecular techniques (i.e., Illumina next generation deep sequencing instead of TRFLP community fingerprinting). Another possible reason for the observed low sensitivity of community analysis methods is that they mostly focused on identifying dominant sources in this study. For example, although high resolution data were obtained (20K to 100K unique sequence reads per sample), Illumina data were only analyzed to the extent sufficient for identifying dominant sources. Advancement in bioinformatics will continue improving the sensitivity of microbial community analysis methods for source identification (Unno *et al.* 2011).

While it might be optimal to identify every contributing source, in practice, dominant source detection is still very useful for management to prioritize remediation efforts. Dominant source detection, however, may not be achieved by one single marker method that aims to detect its target source without providing information on other contributing sources. It is important to recognize that dominance determination by community analysis methods was better when dominance was defined by DNA contribution vs. by *Enterococcus* contribution. This is expected as all three community analysis methods were DNA-based molecular methods. As the relative abundance of other microbial community members vs. members of the *Enterococcus* genus may not be the same across different sources (i.e., community composition differs among sources), one would not expect total bacterial DNA to correlate well with *Enterococcus* concentrations, particularly when the latter was determined by a culture-based method (USEPA 2002). Although there may be a strong desire to perform *Enterococcus* source allocation since *Enterococcus* is often specified for compliance monitoring and TMDL development, source allocation based on total DNA contribution from each source provides an alternative that is relevant to public health protection (Field *et al.* 2003, Ervin *et al.* 2013).

Besides source dominance definition (DNA key vs. ENT key), other factors also influenced community analysis method performance. Data analysis approaches greatly improved TRFLP and PhyloChip performance, likely because more information was utilized when combining Univ & Bac (vs. considering Univ and Bac separately, for TRFLP) or analyzing data at the probe level (vs. analyzing data at the OTU level, for PhyloChip). This is consistent with a previous study where an integrated data analysis approach using the overall community TRFLP profiles helped identify human sources, while using a few isolated signature OTUs from the overall TRFLP profile might not have been successful (Cao *et al.* 2011a). However, increasing the amount of input information per sample (through increasing subsampling size from 20K to 90K reads, or through analyzing the data at the unique reads level vs. at the OTU level), did not change Illumina performance. That is, Illumina performance was not affected by sequencing depth or clustering. This is likely because the input information was very large at the base level and, while further increases may have included more rare sequences or OTUs, they did not alter the overall community composition that

was used for source identification by Illumina in this study. Nevertheless, deeper sequence analysis may be required to detect non-dominant sources, particularly against an environmental microbial community background. Further study is required to determine the amount of community sequence information that is needed to resolve sources in real monitoring situations. Regardless, the influence and complexity of data analysis approaches for community analysis methods indicates the need for developing standardized and automated data analysis approaches for wider application of this class of methods in MST (Unno *et al.* 2011, Cao *et al.* 2013).

Although this study provides a promising overall assessment of source differentiation by community analysis methods, it is important to recognize certain limitations of this evaluation. First, challenge samples were prepared in sterile filtered artificial freshwater, free of ambient bacterial communities that can dilute or confound signals from fecal sources, therefore potentially lower the sensitivity of community analysis methods (Unno *et al.* 2010, Cao *et al.* 2011a, Dubinsky *et al.* 2012). Second, fresh fecal material was used in both reference and challenge samples. It is unknown how differential degradation of microbial community constituents in ambient samples (Walters *et al.* 2009) would change source identification by community analysis and other MST methods. Third, the number of challenge samples per source was relatively low, particularly for animal sources, which could contribute to substantial variability in estimating sensitivity. Having drastically more negative than positive challenge samples for a given source inevitably created a stringent study design for assessing specificity, but an inadequate setting for assessing sensitivity. Lastly, the types of challenge samples (defined by the relative concentration of a target source in the challenge sample, i.e., singleton vs. doubleton, doubleton with minor target source vs. doubleton with dominant target source) varied by target source (Table 3), which makes it less meaningful to compare the reported sensitivity across sources for a given community analysis method. This is because the relative concentration of a target (and non-target) source in challenge samples greatly affects method performance metrics in evaluation studies. However, comparison of performance for the same source across methods is not affected.

Overall, at the current stage, microbial community analysis may not be a replacement but could be a very useful complementary tool to single marker qPCR assays. The relative low sensitivity makes this class of methods inappropriate for management applications where high analytical sensitivity is preferred, e.g., for detecting very low levels of human waste input. Source identification results by the community-based methods are currently qualitative (dominant vs. minor), which may not be sufficient for comparing the extent of contamination by one particular source across sites. However, community analysis methods can be most useful for identifying sources that currently have no developed single qPCR marker and for confirming source identification answers by a single marker that lacks the certainty provided by multiple lines of evidence in community analysis. For example, a manager may wish to pin point whether the septic or sewer systems is the source of human fecal contamination so that appropriate management action may be taken. In addition, community analysis may be used to compare microbial communities originating from non-fecal sources such as sand and kelp to that in the receiving waters, in order to determine the impact of non-fecal sources on water quality at a site.

Nevertheless, community analysis methods are currently more expensive, and require a higher level of expertise for analysis and data interpretation than an individual qPCR assay. Among the three community analysis methods evaluated in this study, TRFLP is currently the least expensive and technically most accessible by common molecular laboratories; PhyloChip is the most expensive and can be performed by several facilities with microarray capacity; Illumina is currently intermediately priced on a per sample basis and can be performed by specialized facilities with next generation sequencing (NGS) capacity. However, community fingerprinting methods such as TRFLP provide much less information than comprehensive microarray and NGS such as PhyloChip and Illumina. NGS is a dynamic field with rapid technology advancement in sequencing and bioinformatics that may dramatically reduce the cost and time required for analysis and improve technology accessibility in the future.

Literature Cited

Boehm, A.B., L.C. Van De Werfhorst, J.F. Griffith, P.A. Holden, J.A. Jay, O.C. Shanks, D. Wang and S.B. Weisberg. 2013. Performance of fourty-three microbial source tracking methods: A twenty-seven lab evaluation study. *Water Research* 47:6812-6828.

Cao, Y., G.N. Cherr, A.L. Córdova-Kreylos, T.W.-M. Fan, P.G. Green, R.M. Higashi, M.G. LaMontagne, K.M. Scow, C.A. Vines, J. Yuan and P.A. Holden. 2006. Relationships between sediment microbial communities and pollutants in two California salt marshes. *Microbial Ecology* 52:619-633.

Cao, Y., L.C. Van De Werfhorst, E. Scott, M. Raith, P.A. Holden and J.F. Griffith. 2013. Bacteroidales terminal restriction fragment length polymorphism (TRFLP) for fecal source differentiation in comparison to and in combination with universal bacteria TRFLP. *Water Research* 47:6944-6955.

Cao, Y., L.C. Van De Werfhorst, B. Sercu, J.L.S. Murray and P.A. Holden. 2011a. Application of an integrated community analysis approach for microbial source tracking in a coastal creek. *Environmental Science & Technology* 45:7195-7201.

Cao, Y., C.H. Wu, G.L. Andersen and P.A. Holden. 2011b. Community analysis-based methods. pp. 251-282 *in*: C. Hagedorn, A. Blanch and V. Harwood (eds.), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer. New York, NY.

Cole, J.R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity and J.M. Tiedje. 2009. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* 37:D141-D145.

Dubinsky, E.A., L. Esmaili, J.R. Hulls, Y. Cao, J.F. Griffith and G.L. Andersen. 2012. Application of phylogenetic microarray analysis to discriminate sources of fecal pollution. *Environmental Science & Technology* 46:4340-4347.

Edgar, R.C., B.J. Haas, J.C. Clemente, C. Quince and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194-2200.

Ervin, J.S., T.L. Russell, B.A. Layton, K.M. Yamahara, D. Wang, L.M. Sassoubre, Y. Cao, C.A. Kelty, M. Sivaganesan, A.B. Boehm, P.A. Holden, S.B. Weisberg and O.C. Shanks. 2013. Characterization of fecal concentrations in human and other animal sources by physical, culture, and quantitative real-time PCR methods. *Water Research* 47:6873-6882.

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

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. *Applied and Environmental Microbiology* 71:3163-3170.

Huber, J.A., D.B. Mark Welch, H.G. Morrison, S.M. Huse, P.R. Neal, D.A. Butterfield and M.L. Sogin. 2007. Microbial population structures in the deep marine biosphere. *Science* 318:97-100.

Knights, D., J. Kuczynski, E.S. Charlson, J. Zaneveld, M.C. Mozer, R.G. Collman, F.D. Bushman, R. Knight and S.T. Kelley. 2011. Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* 8:761-763.

Lee, J.E., S. Lee, J. Sung and G. Ko. 2011. Analysis of human and animal fecal microbiota for microbial source tracking. *ISME Journal* 5:362-365.

Ley, R.E., M. Hamady, C. Lozupone, P.J. Turnbaugh, R.R. Ramey, J.S. Bircher, M.L. Schlegel, T.A. Tucker, M.D. Schrenzel, R. Knight and J.I. Gordon. 2008. Evolution of mammals and their gut microbes. *Science* 32:1647-1651.

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

McLellan, S.L., S.M. Huse, S.R. Mueller-Spitz, E.N. Andreishcheva and M.L. Sogin. 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology* 12:378-392.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn and C.F. Weber. 2009. Introducing mothur: Open-source, platform-independent, community-supported

software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75:7537-7541.

Shanks, O.C., C.A. Kelty, S. Archibeque, M. Jenkins, R.J. Newton, S.L. McLellan, S.M. Huse and M.L. Sogin. 2011. Community structures of fecal bacteria in cattle from different animal feeding operations. *Applied and Environmental Microbiology* 77:2992-3001.

US Environmental Protection Agency (USEPA). 2002. Method 1600: Enterococci in Water by Membrane Filtration Using membrane-*Enterococcus* Indoxyl-β-D-Glucoside Agar (mEI). EPA-821-R-02-022. USEPA Office of Water. Washington, DC.

Unno, T., D.Y.W. Di, J. Jang, Y.S. Suh, M.J. Sadowsky and H.-G. Hur. 2011. Integrated online system for a pyrosequencing-based microbial source tracking method that targets Bacteroidetes 16S rDNA. *Environmental Science & Technology* 46:93-98.

Unno, T., J. Jang, D. Han, J.H. Kim, M.J. Sadowsky, O.-S. Kim, J. Chun and H.-G. Hur. 2010. Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environmental Science & Technology* 44:7777-7782.

Walters, S.P., K.M. Yamahara and A.B. Boehm. 2009. Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: Implications for their use in assessing risk in recreational waters. *Water Research* 43:4929-4939.

Wood, J., K.P. Scott, G. Avguštin, C.J. Newbold and H.J. Flint. 1998. Estimation of the relative abundance of different bacteroides and prevotella ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. *Applied and Environmental Microbiology* 64:3683-3689.

Acknowledgements

Funding for this project has been provided in part through an agreement with the California State Water Resources Control Board. The contents of this document do not necessarily reflect the views and policies of the California State Water Resources Control Board, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. A portion of this work was performed under the auspices of the US Department of Energy under contract DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory. The authors would like to thank staff from the Southern California Coastal Water Research Project, Stanford University, University of California, Los Angeles, and University of California, Santa Barbara for collecting the fecal sources and preparing the blind samples, all the staff members at the community analysis labs for assisting in laboratory work.

Supplemental Information

Supplemental Information is available at ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_461_474SI.pdf.