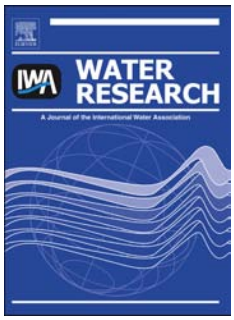
Accepted Manuscript



Evaluation of Molecular Community Analysis Methods for Discerning Fecal Sources and Human Waste

Yiping Cao, Laurie C. Van De Werfhorst, Eric A. Dubinsky, Brian D. Badgley, Michael J. Sadowsky, Gary L. Andersen, John F. Griffith, Patricia A. Holden

PII: S0043-1354(13)00554-X

DOI: 10.1016/j.watres.2013.02.061

Reference: WR 10063

To appear in: *Water Research*

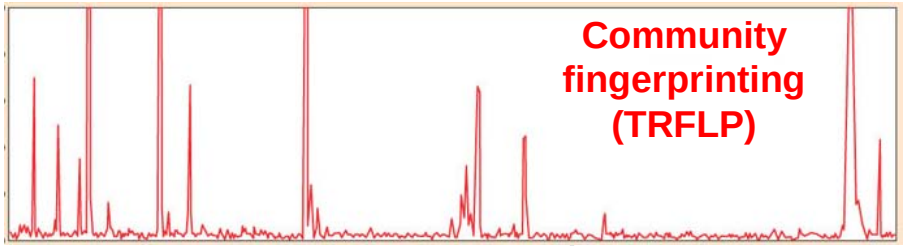
Received Date: 26 September 2012

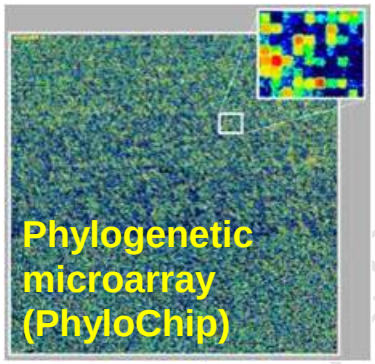
Revised Date: 2 February 2013

Accepted Date: 7 February 2013

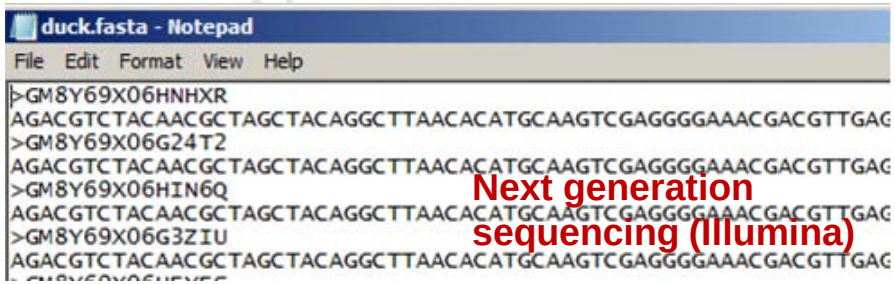
Please cite this article as: Cao, Y., Van De Werfhorst, L.C., Dubinsky, E.A., Badgley, B.D., Sadowsky, M.J., Andersen, G.L., Griffith, J.F., Holden, P.A., Evaluation of Molecular Community Analysis Methods for Discerning Fecal Sources and Human Waste, *Water Research* (2013), doi: 10.1016/j.watres.2013.02.061.

Fecal source identification by three community analysis technology





Each correctly identifies dominant sources in 95% of the 64 unknown samples created from 12 sources



Evaluation of Molecular Community Analysis Methods for Discerning Fecal Sources and Human Waste

Yiping Cao[a], Laurie C. Van De Werfhorst[b,c], Eric A. Dubinsky[d], Brian D. Badgley[e, 1], Michael J. Sadowsky[e], Gary L. Andersen[d], John F. Griffith[a], Patricia A. Holden[b,c]

[a] Southern California Coastal Water Research Project Authority, Costa Mesa, CA 92626;

[b] Bren School of Environmental Science & Management, University of California, Santa Barbara, CA 93106-5131;

[c] Earth Research Institute, University of California, Santa Barbara, CA 93106-3060;

[d] Lawrence Berkeley National Laboratory, Earth Sciences Division, Berkeley, CA;

[e] BioTechnology Institute, University of Minnesota, St. Paul, MN 55108

[1]Present address: Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061

Correspondence: Yiping Cao, Ph.D., Southern California Coastal Water Research Project Authority, 3535 Harbor Blvd, Suite 110, Costa Mesa, CA 92626. Phone: (714) 755-3241, Fax: (714) 755-3299, Email: yipingc@sccwrp.org

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. Here, we 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.

Key words: microbial source tracking; microbial community analysis; TRFLP; PhyloChip; Next generation sequencing

1. INTRODUCTION

Beach water quality is monitored for microbial contamination through measurements of fecal indicator bacteria (FIB), such as *E. coli* and *Enterococcus* 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, i.e. 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 (Cao et al. 2011a, Dubinsky et al. 2012, Unno et al. 2010).

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.

2. METHODS

2.1 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 200ml 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. In press).

All 76 samples (64 unknown + 12 reference samples) were analyzed by each of the following methods (Figure S1): TRFLP targeting all Bacteria (Univ-TRFLP, (Cao et al. 2011a)) or the order of Bacteroidales (Bac-TRFLP, (Cao et al. submitted)), 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).

2.2 TRFLP

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^{TM}; 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 5ul 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 were described elsewhere (Bac-TRFLP (Cao et al. submitted); Univ-TRFLP (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. submitted, Cao et al. 2011a). 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&Bac-TRFLP).

2.3 PhyloChip

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 min), followed by 30 cycles 95°C (30 s), 48-58°C (25 s), 72°C (2 min), followed by a final extension 72°C (10 min).

Two approaches were used to analyze the fluorescent image files following array scanning. The first used the standard operational taxonomic unit (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 we determined which OTUs were unique to that source. For this analysis we grouped some sources that are known to have similar bacterial communities 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.

2.4 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^{o}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 specific 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.

213 Purified PCR products were quantified as with the DNA extracts above, then stored at -20 ^{o}C before
214 pooling for sequencing.

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

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

232 For determining sources in challenge samples, the unique sequence reads were directly analyzed, or
233 clustered into OTUs and then analyzed, by multivariate analysis techniques. For either approach, a
234 random subset of 90,000 sequences was chosen from each sample to balance read numbers, except for
235 five samples that returned less than 90,000 sequences. For these five samples, all available reads were
236 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, produced 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 dentrograms 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.

2.5 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 Method EPA 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 S1). 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 S1), 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. In press)) 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. In press). 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 S1. The performance evaluation 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.

3. 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 2 to 8 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. submitted).

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% - 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 (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", i.e. 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 (Table S2, S3). Nevertheless, PhyloChip with a lower resolution data analysis approach (i.e. OTU) did not distinguish cow and horse (Table S2, S3).

Sensitivity varied by target source and community analysis method, and by the type of challenge sample (Tables 3, 4). Sensitivity to a few sources was particularly low: Bac-TRFLP was insensitive to gull (Table S2, S3); 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

359 analysis based on >1million individual probes, was generally the most sensitive . Within a given
360 molecular community analysis method, high resolution data analysis approaches also reduced the number
361 of false negative for human feces, e.g. for TRFLP (Univ- and Bac- combined vs. separate) and for
362 PhyloChip (probe vs. OTU) (Table 4). Moreover, sensitivity was sometimes lower for the Illumina
363 method as it aimed at identifying the dominant source only, while TRFLP and PhyloChip reported one, or,
364 when sufficient evidence existed in data, two sources. TRFLP and PhyloChip reported two sources for 1-
365 4 and 12-24 challenge samples, respectively, depending on the data analysis approach. However, the
366 Illumina method can be used to partition more than one source in a sample by incorporating shared OTUs
367 into the data analysis (e.g., (Knights et al. 2011, Unno et al. 2010)). However, the software development
368 for such a tool is yet to be completed.

369 For the different sources tested in this study, all community analysis methods had 100% sensitivity and
370 specificity for deer and horse (Table 3). At least one community analysis method had 100% sensitivity
371 and specificity for human feces (PhyloChip), chicken (PhyloChip, Illumina), pigeon (PhyloChip,
372 Illumina), dog (PhyloChip), and pig (PhyloChip). At least one community analysis method had >80%
373 sensitivity and >80% specificity for each of the 12 sources except sewage and cow (Table 3). Reported
374 answers for all 64 challenge samples and a detailed performance evaluation by each source are presented
375 in the Supporting Information (Tables S2, S3).

376

377 **4. DISCUSSION**

378 Microbial community analysis methods demonstrated great promise to become universal MST tools for
379 identifying any source, or even dual sources simultaneously. Because community analysis methods
380 identified sources by characterizing the microbial communities in the suspected sources then comparing
381 with unknown sample communities, their source identification capacity was not restricted by the single
382 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 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 (this study), or pyrosequencing targeting the phylum of Bacteroidetes (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 , i.e. of generating multiple lines of evidence, likely contributed to the excellent specificity observed for all sources tested (Table 3), 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 (Table 3, Tables S2, S3). Additionally, while many gull-specific single indicator PCR or qPCR assays cross-reacted with pigeon and/or goose sources (Boehm et al. In press), PhyloChip (probe) and Illumina were able to distinguish all three birds (Table 3, Tables S2, S3). 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 (this study, (Boehm et al. In press, Cao et al. 2011a)). 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 (Table 2, Figure 1). This is expected as all three community analysis methods

433 were DNA-based molecular methods. As the relative abundance of other microbial community members
434 vs. members of the *Enterococcus* genus may not be the same across different sources (i.e., community
435 composition differs among sources), one would not expect total bacterial DNA to correlate well with
436 *Enterococcus* concentrations, particularly when the latter was determined by a culture-based method (U.S.
437 EPA 2002). Although there may be a strong desire to perform *Enterococcus* source allocation since
438 *Enterococcus* is often specified for compliance monitoring and TMDL development, source allocation
439 based on total DNA contribution from each source provides an alternative that is relevant to public health
440 protection (Ervin et al. in revision, 2013, Field et al. 2003).

441 Besides source dominance definition (DNA key vs. ENT key), other factors also influenced community
442 analysis method performance. Data analysis approaches greatly improved TRFLP and PhyloChip
443 performance, likely because more information was utilized when combining Univ & Bac (vs. considering
444 Univ and Bac separately, for TRFLP) or analyzing data at the probe level (vs. analyzing data at the OTU
445 level, for PhyloChip). This is consistent with a previous study where an integrated data analysis approach
446 using the overall community TRFLP profiles helped identify human sources, while using a few isolated
447 signature OTUs from the overall TRFLP profile might not have been successful (Cao et al. 2011a).
448 However, increasing the amount of input information per sample (through increasing subsampling size
449 from 20K to 90K reads, or through analyzing the data at the unique reads level vs. at the OTU level) did
450 not change Illumina performance. That is, Illumina performance was not affected by sequencing depth or
451 clustering. This is likely because the input information was very large at the base level, and while further
452 increases may have included more rare sequences or OTUs, they did not alter the overall community
453 composition that was used for source identification by Illumina in this study. Nevertheless, deeper
454 sequence analysis may be required to detect non-dominant sources, particularly against an environmental
455 microbial community background. Further study is required to determine the amount of community
456 sequence information that is needed to resolve sources in real monitoring situations. Regardless, the
457 influence and complexity of data analysis approaches for community analysis methods indicates the need

458 for developing standardized and automated data analysis approaches for wider application of this class of
459 methods in MST (Cao et al. submitted, Unno et al. 2011).

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

478 Overall, at the current stage, microbial community analysis may not be a replacement but could be a very
479 useful complementary tool to single marker qPCR assays. The relative low sensitivity makes this class of
480 methods inappropriate for management applications where high analytical sensitivity is preferred, e.g. for
481 detecting very low levels of human waste input. Source identification results by the community-based
482 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.

5. CONCLUSIONS

- The TRFLP, PhyloChip, and Illumina community analysis methods correctly identified the dominant source in 95% of the unknown samples created from 12 difference sources, demonstrating the potential of this class of methods to become a universal MST tool for identifying any source.

- Community analysis methods are particularly useful for distinguishing sources that currently do not have source-specific single marker genes, e.g. for identifying deer fecal material, and for distinguishing the different types of human fecal sources (human feces, sewage, and septage), and closely-related birds: gulls, pigeons and geese.
- Dominant source definition (of the dual-source challenge samples) greatly affected the perceived method performance, indicating the need to resolve the current disconnect between regulatory water quality standards (based on culturable FIB) and molecular MST methods (based on DNA).
- Data analysis approach greatly affected TRFLP and PhyloChip performance, stressing the need to standardize data interpretation for wider application of these methods in MST.
- The community analysis methods were most effective at identifying dominant sources as their current sensitivity was limited by molecular method resolution or available data analysis tools for high resolution data. Method refinement and further evaluation is warranted to improve minor source identification particularly in ambient water samples.

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 U.S. Department of Energy under contract DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory. We 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.

REFERENCES

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

Cao Y., Cherr G.N., Córdova-Kreylos A.L., Fan T.W.-M., Green P.G., Higashi R.M., LaMontagne M.G., Scow K.M., Vines C.A., Yuan J. and Holden P.A., Relationships between sediment microbial communities and pollutants in two California salt marshes, *Microbial Ecology* 52 (4), 2006, 619-633.

Cao Y., Van De Werfhorst L.C., Scott E., Raith M., Holden P.A. and Griffith J.F., Bacteriodales terminal restriction fragment length polymorphism (TRFLP) for Fecal Source Differentiation In Comparison to and In Combination with Universal Bacteria TRFLP, *Water Research*, submitted.

Cao Y., Van De Werfhorst L.C., Sercu B., Murray J.L.S. and Holden P.A., Application of an integrated community analysis approach for microbial source tracking in a coastal creek, *Environmental Science & Technology* 45 (17), 2011a, 7195-7201.

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

Cole J.R., Wang Q., Cardenas E., Fish J., Chai B., Farris R.J., Kulam-Syed-Mohideen A.S., McGarrell D.M., Marsh T., Garrity G.M. and Tiedje J.M., The Ribosomal Database Project: improved alignments and new tools for rRNA analysis, *Nucleic Acids Research* 37 (suppl 1), 2009, D141-D145.

Dubinsky E.A., Esmaili L., Hulls J.R., Cao Y., Griffith J.F. and Andersen G.L., Application of phylogenetic microarray analysis to discriminate sources of fecal pollution, *Environmental Science & Technology* 46 (8), 2012, 4340-4347.

Edgar R.C., Haas B.J., Clemente J.C., Quince C. and Knight R., UCHIME improves sensitivity and speed of chimera detection, *Bioinformatics* 27 (16), 2011, 2194-2200.

Ervin J.S., Russell T.L., Layton B.A., Yamahara K.M., Wang D., Sassoubre L.M., Cao Y., Kelty C.A., Sivaganesan M., Boehm A.B., Holden P.A., Weisberg S.B. and Shanks O.C., Characterization of Fecal Concentrations in Human and Other Animal Sources by Physical, Culture, and Quantitative Real-Time PCR Methods, *Water Research*, in revision, 2013.

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

Harwood V.J., Levine A.D., Scott T.M., Chivukula V., Lukasik J., Farrah S.R. and Rose J.B., Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection, *Applied and Environmental Microbiology* 71 (6), 2005, 3163-3170.

Huber J.A., Mark Welch D.B., Morrison H.G., Huse S.M., Neal P.R., Butterfield D.A. and Sogin M.L., Microbial Population Structures in the Deep Marine Biosphere, *Science* 318 (5847), 2007, 97-100.

Knights D., Kuczynski J., Charlson E.S., Zaneveld J., Mozer M.C., Collman R.G., Bushman F.D., Knight R. and Kelley S.T., Bayesian community-wide culture-independent microbial source tracking, *Nat Methods* 8 (9), 2011, 761-763.

Lee J.E., Lee S., Sung J. and Ko G., Analysis of human and animal fecal microbiota for microbial source tracking, *ISME Journal* 5 (2), 2011, 362-365.

576 Ley R.E., Hamady M., Lozupone C., Turnbaugh P.J., Ramey R.R., Bircher J.S., Schlegel M.L., Tucker
577 T.A., Schrenzel M.D., Knight R. and Gordon J.I., Evolution of mammals and their gut microbes,
578 *Science* 32, 2008, 1647-1651.
579 Liu W.-T., Marsh T.L., Cheng H. and Forney L.J., Characterization of microbial diversity by determining
580 terminal restriction fragment length polymorphisms of genes encoding 16S rRNA, *Applied and*
581 *Environmental Microbiology* 63 (11), 1997, 4516-4522.
582 McLellan S.L., Huse S.M., Mueller-Spitz S.R., Andreishcheva E.N. and Sogin M.L., Diversity and
583 population structure of sewage-derived microorganisms in wastewater treatment plant influent,
584 *Environmental Microbiology* 12 (2), 2010, 378-392.
585 Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley
586 B.B., Parks D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J. and Weber
587 C.F., Introducing mothur: Open-source, platform-independent, community-supported software for
588 describing and comparing microbial communities, *Applied and Environmental Microbiology* 75
589 (23), 2009, 7537-7541.
590 Shanks O.C., Kelty C.A., Archibeque S., Jenkins M., Newton R.J., McLellan S.L., Huse S.M. and Sogin
591 M.L., Community structures of fecal bacteria in cattle from different animal feeding operations,
592 *Applied and Environmental Microbiology* 77 (9), 2011, 2992-3001.
593 U.S. EPA, Method 1600: Enterococci in Water by Membrane Filtration Using membrane-*Enterococcus*
594 Indoxyl-β-D-Glucoside Agar (mEI). EPA-821-R-02-022., 2002, p.^pp., Washington, D.C.
595 Unno T., Di D.Y.W., Jang J., Suh Y.S., Sadowsky M.J. and Hur H.-G., Integrated Online System for a
596 Pyrosequencing-Based Microbial Source Tracking Method that Targets Bacteroidetes 16S rDNA,
597 *Environmental Science & Technology* 46 (1), 2011, 93-98.
598 Unno T., Jang J., Han D., Kim J.H., Sadowsky M.J., Kim O.-S., Chun J. and Hur H.-G., Use of barcoded
599 pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds,
600 *Environmental Science & Technology* 44 (20), 2010, 7777-7782.
601 Walters S.P., Yamahara K.M. and Boehm A.B., Persistence of nucleic acid markers of health-relevant
602 organisms in seawater microcosms: Implications for their use in assessing risk in recreational
603 waters, *Water Research* 43 (19), 2009, 4929-4939.
604 Wood J., Scott K.P., Avguštin G., Newbold C.J. and Flint H.J., Estimation of the relative abundance of
605 different bacteroides and prevotella ribotypes in gut samples by restriction enzyme profiling of
606 PCR-amplified 16S rRNA gene sequences, *Applied and Environmental Microbiology* 64 (10),
607 1998, 3683-3689.

608

609

610 **Figure Captions**

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

612

613 **Tables**

614 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	

615

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

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. submitted).

[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).

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]	0.50	0.94	0.83	1	0.50	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	0.50	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.

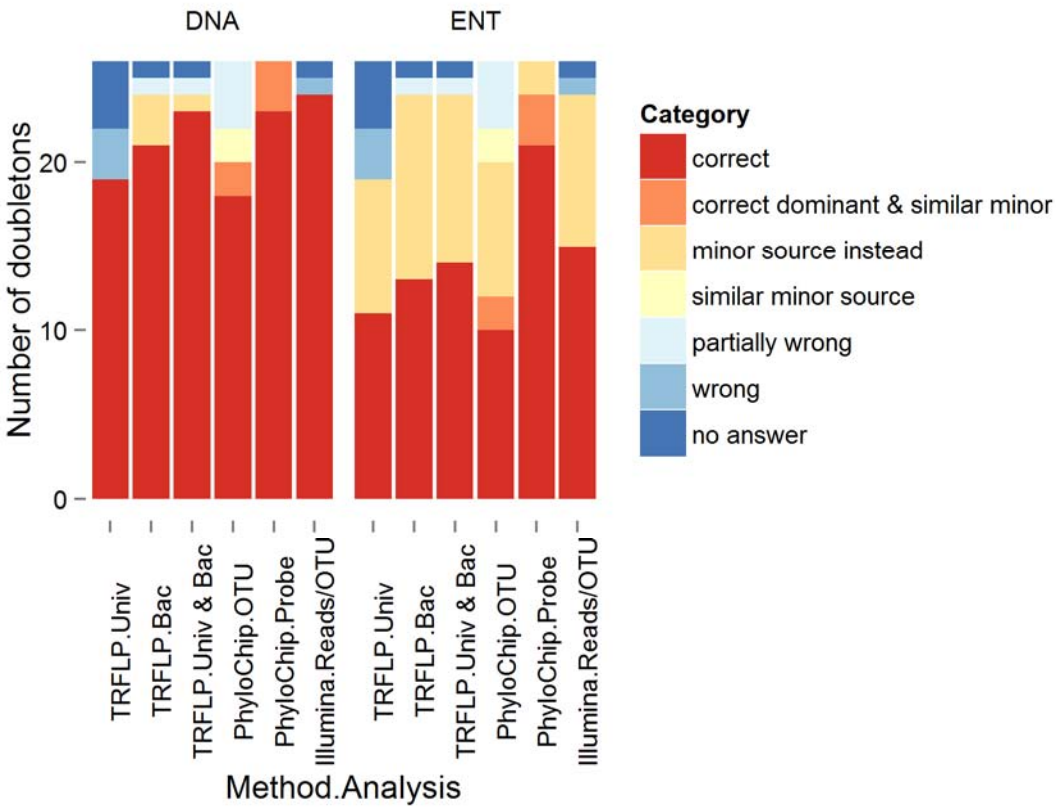
Table 4. Number of false negatives by targeted source, unknown challenge sample, and community analysis method. False negatives were determined against each targeted source, i.e. a doubleton challenge sample could be false negative to either contributing source. Note that n=2 for each unknown challenge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

Figure 1.



- Community analysis (TRFLP, PhyloChip, Illumina) were tested with 64 blind samples from 12 sources
- All three methods exhibited 95% correct identification and excellent specificity (89-100%)
- Sensitivity varied (25-100%) by community analysis methods and by challenge sources
- Data analysis approach and challenge source definition greatly affected method performance
- All methods distinguished septage, sewage and human feces

Supporting Information for:

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], Patricia A. Holden[2,3]

[1]Southern California Coastal Water Research Project Authority, Costa Mesa, CA 92626; [2]Bren School of Environmental Science & Management, University of California, Santa Barbara, CA 93106-5131; [3]Earth Research Institute, University of California, Santa Barbara, CA 93106-3060; [4]Lawrence Berkeley National Laboratory, Earth Sciences Division, Berkeley, CA; [5]BioTechnology Institute, University of Minnesota, St. Paul, MN 55108

[§]Present address: Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061

Contains:

12 pages

1 Figure

3 Tables

RESULTS AND DISCUSSION

Detailed assessment by each source (Table S2, S3)

The various combinations of community analysis methods and data analysis approaches are denoted as TRFLP (Univ), TRFLP (Bac), TRFLP (Univ&Bac), PhyloChip (OTU), and PhyloChip (probe) when necessary. As TRFLP performance from the two laboratories was largely similar, therefore 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, submitted). For Illumina, analyzing the data with various subsampling size (20K to 90K reads) at the unique read or OTU levels (with 97% or 90% similarity cutoff for clustering reads into OTUs) led to identical source identification results. Therefore, only one set of Illumina results is reported.

Sources that were incorrectly listed as present in each challenge sample are indicated in red font (Table S2, S3). Also, although results are presented in much detail here, this evaluation study aimed at providing an overall performance assessment of these types of community analysis-based methods (as presented in the main manuscript). This evaluation study was not designed to comprehensively evaluate community analysis methods against each of the 12 sources. The numbers and types of challenge samples for each source are presented below prior to the method results for each source. Evaluations presented here on the dominant source in doubleton samples are based on the DNA key.

Human feces: Two full and two 1:10 strength singletons, 14 doubletons containing human feces (six as dominant and four as minor sources, and four where dominance could not be established) were included in the evaluation study. All four singletons were correctly identified by all community-based methods used here. However, PhyloChip (OTU) also listed pig in addition to human feces for the two full-strength singletons, and dog for the two 1:10 strength singletons. For the six doubletons where human feces were the dominant sources, all community analysis methods correctly identified human feces as the dominant

source with a few exceptions. The exceptions were that TRFLP (Univ) mistakenly reported goose as the dominant source for one doubleton, TRFLP (Bac) could not distinguish human feces from sewage for one doubleton, and Illumina mistakenly reported pig as the dominant source for one doubleton. For the four human feces-containing doubletons where no dominance between sources could be established based on DNA contribution, all community analysis methods correctly identified the presence of human waste except PhyloChip (OTU) which reported sewage instead of human feces for two doubletons. TRFLP (Bac) and PhyloChip (probe) also identified the animal source in two and four, respectively, of these doubletons containing human feces. For the four doubletons where human feces were the minor source, PhyloChip (probe) identified both sources in all four doubletons, while TRFLP (Bac, Univ & Bac) identified human feces in the two doubletons where goose was the dominant source, likely benefiting from the fact that TRFLP (Bac) did not detect any bird signal due to low abundance of Bacteroidales in seabirds (Fogarty and Voytek 2005, Lu et al. 2009, Lu et al. 2008). Moreover, TRFLP (Univ) and Illumina did not report an answer for four and one doubleton, respectively.

Sewage: Two full and two 1:10 strength singletons, ten doubletons containing sewage (two as dominant and six as minor source, and two where dominance could not be established) were included in the evaluation study. All four singletons were correctly identified by all methods, except TRFLP (Bac) which did not distinguish sewage from human feces for the two full strength sewage singletons. For the two doubletons where sewage was the dominant source (sewage:gull 88:12), all community analysis methods correctly identified sewage as dominant sources except TRFLP (Bac) which did not distinguish sewage from human feces for one doubleton. Additionally, PhyloChip (probe) correctly identified both sewage and gull in these doubletons. For the two sewage-containing doubletons where no dominance between sources could be established based on DNA contribution (sewage:pig 54:46), only the PhyloChip (OTU, probe) identified sewage, along with pig. For the six doubletons where sewage was the minor source, only PhyloChip (probe) and TRFLP (Univ) correctly reported sewage for two and three (out of six)

doubletons, respectively. The TRFLP (Bac) and TRFLP (Univ&Bac) methods each did not report an answer for one doubleton.

Septage: Two full and two 1:10 strength septage singletons, and two doubletons containing septage as a minor source (horse:septage 82:18) were included in the evaluation study. All four singletons were correctly identified by all methods regardless of data analysis approach. These four septage singletons were clustered tightly together during the multivariate data analysis (detrended correspondence analysis or non-metric multidimensional scaling) with other sources (TRFLP, Illumina), indicating potentially distinctly different microbial communities in septage than in other sources included in this study. For the two doubletons (horse:septage 82:18), only the PhyloChip (probe) identified the minor source septage while all other methods only reported the dominant source horse.

Gull: Two full and two 1:10 strength singletons, and eight doubletons containing gull (two as dominant source, four as minor sources, and two where dominance could not be established) were included in the evaluation study. Generally, TRFLP and PhyloChip (OTU) did not distinguish between gull and pigeon and therefore listed gull and/or pigeon when either source was reported. This is consistent with the finding that all qPCR methods targeting gull cross-reacted with pigeon (Boehm et al. In press). However, PhyloChip with data analysis at the probe level and Illumina did separate these two bird species. Without distinguishing gull and pigeon, all four gull singletons were correctly identified. The only exception was that the PhyloChip (probe) results also listed cow as a source in addition to gull for one of the 1:10 strength gull singletons.

For the two doubletons where gull was the dominant source (gull:sewage 92:8), all community analysis methods correctly identified gull as the dominant source with three exceptions: TRFLP (Bac) reported sewage as dominant for both doubletons, and PhyloChip (OTU) also mistakenly listed dog as a contributing source to these two doubletons. Because the abundance of Bacteroidales in seabirds is generally low (Fogarty and Voytek 2005), it is not surprising that TRFLP (Bac) reported sewage instead

of gull as the dominant source for the gull:sewage 92:8 doubleton. For the two gull-containing doubletons where no dominance between sources could be established based on DNA contribution (gull:human 56:44) and for the four doubletons where gull was the minor source, only the PhyloChip (probe) method identified the presence of gull feces.

Pigeon: Only two full strength pigeon singletons were included in this study. As presented above for gull, TRFLP and PhyloChip (OTU) could not distinguish gull and pigeon and therefore listed gull and/or pigeon when either source was reported. However, PhyloChip (with data analysis at the probe level) and Illumina correctly identified the two pigeon singletons.

Goose: Two full strength goose singletons and two doubletons where goose was the dominant source (goose:human 91:9) were included in this study. In general, TRFLP had difficulty in distinguishing goose from other sources and therefore was not able to provide an answer for the two goose singletons, which were correctly identified by PhyloChip and Illumina. For the two doubletons where goose was dominant source, only PhyloChip (probe), Illumina, and TRFLP (Univ) were able to identify goose.

Chicken: Two full strength chicken singletons and two doubletons where chicken was the dominant source (chicken:sewage 92:8) were included in this study. In general, chicken was correctly identified by all community analysis methods regardless of data analysis approaches. However, TRFLP (Bac, Univ&Bac) did not provide an answer for one of the doubletons.

Dog: Two full and two 1:10 strength dog singletons, and four doubletons containing dog feces (two as minor sources and two where dominance could not be established) were included in this evaluation study. All four dog singletons were correctly identified by all community analysis methods regardless of data analysis approaches. For the two doubletons where dominance could not be established (dog:human 62:38), only PhyloChip (OTU, probe) and TRFLP (Bac) identified the presence of dog feces. For the two doubletons where dog was the minor source, only PhyloChip (probe) identified the presence of dog feces.

Cow: Two full and two 1:10 strength cow singletons, and four doubletons containing cow feces (two as minor, and two as the dominant, source) were included in this evaluation study. Most cow singletons were correctly identified, except that PhyloChip (OTU) could not distinguish cow and horse, and therefore reported cow and/or horse when either source was reported. Additionally, one replicate of the full strength singleton and one replicate of the 1:10 strength singleton were mistakenly identified as goose by one or all three community analysis methods. This indicates either high similarity between goose and cow fecal microbial communities, or potential cross contamination during challenge sample preparation. Goose has previously also been found difficult to classify with a 454-based community analysis method (Unno et al. 2010). For the two doubletons where cow was the dominant source, all methods correctly identified the presence of cow except PhyloChip (OTU) which reported a deer and sewage mixture. Interestingly, deer is commonly the source that causes false positives for qPCR assays that target cow (Boehm et al. In press). For the two doubletons where cow was the minor source, no community analysis method reported cow as a contributing source.

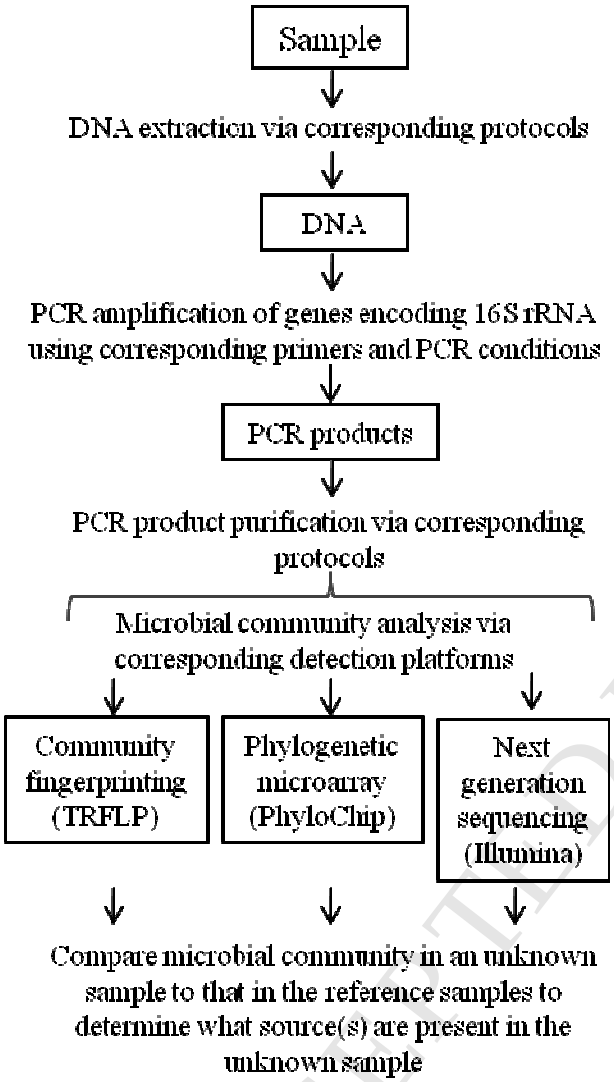
Pig: Two full and two 1:10 strength pig singletons, and four doubletons containing pig feces (two as a dominant source and two where dominance could not be established) were included in this evaluation study. The four pig singletons were correctly identified by TRFLP (Bac, Univ&Bac), PhyloChip (probe) and Illumina. However, the PhyloChip (OTU) method reported septage along with pig for all four pig singletons and TRFLP (Univ) only correctly identified one pig singleton and did not report an answer for the other three pig singletons. For the two doubletons where pig was the dominant source (pig:sewage 99:1), pig was correctly identified by all community analysis methods as the dominant source. For the two doubletons where dominance could not be established (sewage:pig 54:46), pig was identified by all community analysis methods except TRFLP (Univ) where only goose was listed as the contributing source.

Horse: Two full strength horse singletons, and two doubletons containing horse feces as the dominant source (horse:septage 82:18) were included in this evaluation study. Horse was correctly identified for

the two horse singletons and correctly identified as the dominant source in the two doubletons by all community analysis methods. The PhyloChip (OTU) could not distinguish cow and horse, and therefore listed cow and/or horse when either source was identified.

Deer: Only two deer full strength singletons were included in this evaluation study. Both singletons were correctly identified by all community analysis methods, regardless of data analysis approach.

Figure S1. Flow diagram for source identification by community analysis methods.



TABLES

Table S1. Comparison of doubleton challenge sample keys as defined by total DNA contribution (NanoDrop) or by culturable *Enterococcus* (Method EPA 1600) from each of the two sources. Numeric ratios following the two sources indicate percent contributions from each source. Dominance was defined as when contribution from one source was at least two times the contribution from the other source. The bold font indicates no dominance between the two sources, and italic and bold font indicates reversal of dominance based on ENT key.

DNA Key	**ENT Key**
chicken:sewage 92:8	chicken:sewage 96:4
cow:human 92:8	cow:human 99:1
dog:human 62:38	dog:human 99.9:0.1
goose:human 91:9	goose:human 71:29
gull:human 56:44	gull:human 99:1
gull:sewage 92:8	gull:sewage 99.98:0.02
horse:septage 82:18	horse:septage 98:2
human:cow 88:12	**human:cow 50:50**
human:dog 98:2	***human:dog 6:94***
human:gull 98:2	***human:gull 33:67***
pig:sewage 99:1	pig:sewage 98:2
sewage:gull 88:12	***sewage:gull 2:98***
sewage:pig 54:46	**sewage:pig 60:40**

Table S2. Reported source identification results[§] for singletons. Red fonts indicate reported sources not present in the challenge sample.

Singletons[a]	**TRFLP (Univ)**[b]	**TRFLP (Bac)**[b]	**TRFLP (Univ & Bac)**[b]	**PhyloChip (OTU)**[c]	**PhylocChip (Probe)**[c]	**Illumina**[d]
chicken	chicken	chicken	chicken	BIRD (chicken)	chicken	chicken
chicken	chicken & sewage	chicken	chicken	BIRD (chicken)	chicken	chicken
cow	cow	cow	cow	cow/horse	cow	cow
cow	cow	cow	cow	cow/horse	cow	goose
deer	deer	deer	deer	deer	deer	deer
deer	deer	deer	deer	deer	deer	deer
dog	dog	dog	dog	dog	dog	dog
dog	dog	dog	dog	dog	dog	dog
goose	pig	-	-	BIRD (goose)	goose	goose
goose	-	-	gull/pigeon	BIRD (goose)	goose	goose
gull	gull/pigeon	gull/pigeon	gull/pigeon	BIRD (gull/pigeon)	gull	gull
gull	gull/pigeon	gull/pigeon	gull/pigeon	BIRD (gull/pigeon)	gull	gull
horse	horse	horse	horse	cow/horse	horse	horse
horse	horse	horse	horse	cow/horse	horse	horse
human	human	human	human	HUMAN (feces) & pig	human	human
human	human	human	human	HUMAN (feces) & pig	human	human
pig	-	pig	pig	pig & HUMAN (septage)	pig	pig
pig	pig	pig	pig	pig & HUMAN (septage)	pig	pig
pigeon	pigeon/gull	gull/pigeon	gull/pigeon	BIRD (gull/pigeon)	pigeon	pigeon
pigeon	gull/pigeon	gull/pigeon	gull/pigeon	BIRD (gull/pigeon)	pigeon	pigeon
septage	septage	septage	septage	HUMAN (septage)	septage	septage
septage	septage	septage	septage	HUMAN (septage)	septage	septage
sewage	sewage	human/sewage	sewage	HUMAN (sewage)	sewage	sewage
sewage	sewage	human/sewage	sewage	HUMAN (sewage)	sewage	sewage
cow 1:10	cow	cow	cow	cow/horse	cow	cow
cow 1:10	goose	cow	cow	BIRD (goose)	goose; cow	goose
dog 1:10	dog	dog	dog	dog	dog	dog

dog 1:10	dog	dog	dog	dog	dog	dog
gull 1:10	gull/pigeon	-	gull/pigeon	BIRD (gull/pigeon)	gull	gull
gull 1:10	gull	-	gull/pigeon	BIRD (gull/pigeon)	gull; cow	gull
human 1:10	human	human	human	HUMAN (feces) & dog	human	human
human 1:10	human	human	human	HUMAN (feces) & dog	human	human
pig 1:10	-	pig	pig	pig & HUMAN (septage)	pig; goose	pig
pig 1:10	-	pig	pig	pig & HUMAN (septage)	pig	pig
septage 1:10	septage	septage	septage	HUMAN (septage)	septage	septage
septage 1:10	septage	septage	septage	HUMAN (septage)	septage	septage
sewage 1:10	sewage	sewage	sewage	HUMAN (sewage)	sewage	sewage
sewage 1:10	sewage	sewage	sewage	HUMAN (sewage)	sewage	sewage

[§] The "/" in the source identification answer denotes "and/or", "&" denotes "and", and "-" denotes no answer was reported.

[a] Samples included blind duplicates of 19 singletons (full strength or 1:10 strength).

[b] TRFLP generally reported a dominant source unless sufficient evidence suggested presence of two sources. As TRFLP performance from the two laboratories was largely similar, only one lab's results (with slightly better performance metrics) is presented in this manuscript. Detailed cross-laboratory evaluation of the TRFLP method was presented elsewhere (Cao et al. submitted).

[c] PhyloChip reported one or, when sufficient evidence suggested so, two sources. For PhyloChip (with OTU data analysis approach), human waste and bird sources were reported as the broad category in capital letters followed by the actual source in parenthesis.

[d] Illumina only reported a dominant source. All data analysis approaches (Reads, OTU at 90% and OTU at 97%) for Illumina provided the same source identification answers. Therefore, only one set of Illumina results is presented in the manuscript.

Table S3. Reported source identification results[§] for doubletons. Red fonts indicate reported sources not present in the challenge sample.

Doubletons[a]	**TRFLP (Univ)**[b]	**TRFLP (Bac)**[b]	**TRFLP (Univ&Bac)**[b]	**PhyloChip (OTU)**[c]	**PhyloChip (probe)**[c]	**Illumina**
chicken/sewage 92/8	sewage & chicken	chicken	chicken	BIRD (chicken)	chicken; septage	chicken
chicken/sewage 92/8	sewage & chicken	-	-	BIRD (chicken)	chicken; septage	chicken
cow/human 92/8	cow	cow	cow	deer & HUMAN (sewage)	cow; human	-
cow/human 92/8	-	cow	cow	deer & HUMAN (sewage)	human; cow	cow
dog/human 62/38	human	human & dog	human	dog & HUMAN (sewage)	dog; human	human
dog/human 62/38	-	human & dog	human & dog	dog & HUMAN (sewage)	dog; human	human
goose/human 91/9	goose	human	human	HUMAN (sewage)	goose; human	goose
goose/human 91/9	goose	human & pig	goose/human/pig	HUMAN (sewage)	goose; human	goose
gull/human 56/44	human	human	human	HUMAN (feces)	gull; human	human
gull/human 56/44	human	human	human	HUMAN (feces)	gull; human	human
gull/sewage 92/8	gull	sewage	gull/pigeon	BIRD (gull/pigeon) & dog	gull; sewage	gull
gull/sewage 92/8	gull & sewage	sewage	gull/pigeon	BIRD (gull/pigeon) & dog	gull; sewage	gull
horse/septage 82/18	horse	horse	horse	cow/horse	horse; septage	horse
horse/septage 82/18	horse	horse	horse	cow/horse	horse; septage	horse
human/cow 88/12	goose	human	human	HUMAN (feces)	human	human
human/cow 88/12	-	human	human	HUMAN (feces)	human	human
human/dog 98/2	human	human	human	HUMAN (feces)	human; dog	human
human/dog 98/2	human	human	human	HUMAN (feces)	human; dog	pig
human/gull 98/2	-	human/sewage	human	HUMAN (feces)	human	human
human/gull 98/2	human	human	human	HUMAN (feces)	human	human
pig/sewage 99/1	pig	pig	pig	pig	pig	pig
pig/sewage 99/1	pig	pig	pig	pig	pig; septage	pig
sewage/gull 88/12	sewage	human/sewage	sewage	HUMAN (sewage)	gull; sewage	sewage
sewage/gull 88/12	sewage	sewage	sewage	HUMAN (sewage)	gull; sewage	sewage
sewage/pig 54/46	goose	pig	pig	pig & HUMAN (sewage)	pig; sewage	pig
sewage/pig 54/46	goose	pig	pig	pig & HUMAN (sewage)	pig; sewage	pig

[§] The "/" in the source identification answer denotes "and/or", "&" denotes "and", and "-" denotes no answer was reported.

[a] Samples included blind duplicates of 13 doubletons. Numeric ratios following the two sources indicate relative DNA contribution from each source.

[b] TRFLP generally reported a dominant source unless sufficient evidence suggested presence of two sources. As TRFLP performance from the two laboratories was largely similar, only one lab's results (with slightly better performance metrics) is presented in this manuscript. Detailed cross-laboratory evaluation of the TRFLP method was presented elsewhere (Cao et al. submitted).

[c] PhyloChip reported one or, when sufficient evidence suggested so, two sources. For PhyloChip (with OTU data analysis approach), human waste and bird sources were reported as the broad category in capital letters followed by the actual source in parenthesis.

[d] Illumina only reported a dominant source. All data analysis approaches (Reads, OTU at 90% and OTU at 97%) for Illumina provided the same source identification answers. Therefore, only one set of Illumina results is presented in the manuscript.

REFERENCES

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

Cao Y., Van De Werfhorst L.C., Scott E., Raith M., Holden P.A. and Griffith J.F., Bacteriodales terminal restriction fragment length polymorphism (TRFLP) for Fecal Source Differentiation In Comparison to and In Combination with Universal Bacteria TRFLP, *Water Research*, submitted.

Fogarty L.R. and Voytek M.A., Comparison of *Bacteroides-Prevotella* 16S rRNA genetic markers for fecal samples from different animal species, *Applied and Environmental Microbiology* 71 (10), 2005, 5999-6007.

Lu J., Santo Domingo J.W., Hill S. and Edge T.A., Microbial Diversity and Host-Specific Sequences of Canada Goose Feces, *Applied and Environmental Microbiology* 75 (18), 2009, 5919-5926.

Lu J.R., Santo Domingo J.W., Lamendella R., Edge T. and Hill S., Phylogenetic diversity and molecular detection of bacteria in gull feces, *Applied and Environmental Microbiology* 74 (13), 2008, 3969-3976.

Unno T., Jang J., Han D., Kim J.H., Sadowsky M.J., Kim O.-S., Chun J. and Hur H.-G., Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds, *Environmental Science & Technology* 44 (20), 2010, 7777-7782.