Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Application of phylogenetic microarray analysis to discriminate sources of fecal pollution

Permalink

<https://escholarship.org/uc/item/6qn8j88f>

Author

Dubinsky, E.A.

Publication Date

2012-03-15

DOI

[dx.doi.org/10.1021/es2040366](https://doi.org/10.1021/es2040366)

Peer reviewed

Application of Phylogenetic Microarray Analysis to Discriminate Sources of Fecal Pollution

Eric A. Dubinsky,[†] Laleh Esmaili,[†] John R. Hulls,[†] Yiping Cao,[‡] John F. Griffith,[‡] and Gary L. Andersen^{*,†}

[†]Lawrence Berkeley National Laboratory, Earth Sciences Division, Berkeley, CA

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

ABSTRACT: Conventional methods for fecal source tracking typically use single biomarkers to systematically identify or exclude sources. High-throughput DNA sequence analysis can potentially identify all sources of microbial contaminants in a single test by measuring the total diversity of fecal microbial communities. In this study, we used phylogenetic microarray analysis to determine the comprehensive suite of bacteria that define major sources of fecal contamination in coastal California. Fecal wastes were collected from 42 different populations of Humans, birds, cows, horses, elk, and pinnipeds. We characterized bacterial community composition using a DNA microarray that probes for 16S rRNA genes of 59 316 different bacterial taxa. Cluster analysis revealed strong differences in community composition among fecal wastes from human, birds, pinnipeds, and grazers. Actinobacteria, Bacilli, and many Gammaproteobacteria taxa discriminated birds from mammalian sources. Diverse families within the Clostridia and Bacteroidetes taxa discriminated human wastes, grazers, and pinnipeds from each other. We found 1058 different bacterial taxa that were unique to either human, grazing mammal, or bird fecal wastes. These OTUs can serve as specific identifier taxa for these sources in environmental waters. Two field tests in marine waters demonstrate the capacity of phylogenetic microarray analysis to track multiple sources with one test.

INTRODUCTION

Beach closures and public health advisories have a major economic impact on coastal communities whose economies are based largely on tourism from beach recreation. Most closings and advisories are triggered by water samples that exceed microbial water quality standards for fecal indicator bacteria (FIB), usually culturable coliforms, *E. coli*, or enterococci that are considered a proxy for human health risk in recreational waters. Because the direct measurement of all human pathogens is often impractical and unreliable under field conditions, water monitoring relies on the detection of bacterial indicators that have some demonstrated correlation with human illness in areas mostly impacted by human sewage.^{1,2} Sewage, however, is one of many potential sources of FIB, and monitoring results are often confounded by inputs from a variety of wildlife and nonfecal sources.^{1,3-5} FIB are common in most warm-blooded animals, and many studies demonstrate that FIB occur in several environmental sources aside from feces, including soils and sediments, algal wrack, and beach sands.³⁻⁵ Thus, water bodies often contain measurable amounts of FIB even where anthropogenic inputs are absent, and the presence of FIB provides an insufficient indication of health risk without additional source tracking data.

Shortcomings of the current FIB monitoring approach combined with widespread development and implementation of total maximum daily load (TMDL) requirements for microbiological pollution are fueling interest in microbial source tracking (MST) methods.^{6,7} Many approaches to source tracking are under development, most of which rely on single phenotypic or genotypic biomarkers to measure sources.^{1,6} A limitation of single targets is that no single assay is known to be 100% specific for any one type of waste,⁶ and MST based on single targets is entirely dependent on the fate of one biomarker once it enters receiving waters.⁸⁻¹⁰

A huge diversity of microorganisms is resident in human and animal guts. Approximately 1000 different microbial taxa are now known to reside in the human gut alone, but the potential for this diversity to be used as a means for identifying sources remains largely unexplored and there have been few comparative surveys of microbial community composition among important sources of fecal contamination.¹¹⁻¹³ New

techniques for high-throughput DNA sequence analysis such as high-density microarrays and next-generation sequencing (NGS) technologies like pyrosequencing are enabling comprehensive surveys of diverse microbial communities that occur in a sample. Targeting the whole microbial community for source identification is a fundamentally different approach than traditional molecular methods that are dependent on the detection on one gene sequence under complex environmental conditions.^{11,14–16} Sequence analysis of entire microbial communities creates an opportunity to discover a multitude of different bacterial species that are unique to fecal and environmental sources of FIB in recreational waters.

In this study, we used a high-density oligonucleotide microarray to census the 16S rRNA gene diversity in different sources of fecal contamination. The microarray targets 59 316 different 16S rRNA gene polymorphisms that represent most known phyla of bacteria. We test the assumption that different avian and mammalian fecal sources can be distinguished on the basis of their bacterial community composition. We screened a variety of fecal sources of concern in coastal California to identify the microbial groups that are source-specific and then used these unique taxa to detect influence from these sources in marine samples that exceeded water quality limits for fecal indicator bacteria.

METHODS

Feces Sampling and DNA Extraction. Human fecal wastes and freshly deposited droppings from animals were collected at numerous locations throughout California (Supporting Information). Human fecal wastes included primary influent or effluent from eight different municipal wastewater treatment plants, leachate samples from two community septic tanks serving more than 30 households each, and one composite sample of 10 holding tanks from individual households. Sampled animal populations included cows (4), horses (4), tule elk (4), western and California gulls (9), Canada geese (4), pelican (3), pigeon (2), cormorant (1), sea lion (3), elephant seal (1). Each animal sample was a composite of droppings from at least five different individuals from one location and every replicate sample was from a unique population. Individual fecal samples were homogenized and immediately frozen upon collection. Samples were stored at -80°C until DNA extraction.

Each fecal sample was extracted in triplicate to obtain genomic DNA from the microbial community. Two extraction methods were employed: a CTAB extraction method and a kit extraction. In the CTAB method, 0.5 g of homogenized fecal sample were added to a Lysing Matrix E tube (MP Biomedicals, Solon, OH), and 650 μL TE buffer, 250 μL 10 $\times$ phosphate buffered saline, and 100 μL 10% SDS were added. The tubes were bead-beat at 5.5 m/s for 25 s in a FastPrep-24 instrument (MP Biomedicals, Solon, OH), 5.5 m/s for 25 s, 10 μL of 20 mg/mL Proteinase K was added and tubes were incubated at 37°C for 30 min. Following centrifugation at 10 000g for 5 min, supernatant was transferred to 2 mL microcentrifuge tube and amended with 80 μL 5 M NaCl and 80 μL 10% CTAB buffer solution. Tubes were heated at 65°C for 10 min, 700 μL 24:1 chloroform/isopropanol added and then centrifuged at 6000g for 5 min. Supernatant was transferred to a clean microcentrifuge tube, amended with 0.8 volumes of isopropanol, gently mixed, and incubated at -20°C for 1 h. Following centrifugation at 16 000g 4°C for 15 min the supernatant was discarded and the remaining extract washed with ice-cold 70%

ethanol. Following centrifugation at 16 000g at 4°C for 5 min, the supernatant was discarded, and the remaining DNA pellet was air-dried and suspended in 30 μL elution buffer. The second DNA extraction method was conducted with the DNA EZ extraction kit (GeneRite, North Brunswick, NJ) per manufacturer's instructions. We extracted a subset of samples with both methods and saw little difference in the measured community profiles, and we saw no similarity patterns in our final results that were explained by extraction method.

Water Sampling and DNA Extraction. Water monitoring samples were collected from the field at sites with known sources of contamination. One set of samples was collected during a 10-day period following a 720 000 gallon spill of primary effluent from the Sausalito-Marín City Sanitary District treatment plant in Sausalito, California, that occurred in February 2009. Samples were collected daily for three days following the onset of the spill and then once more 10 days after the initial spill during an accidental rupture that occurred during the repair. Sample locations included eight onshore and offshore sites ranging from directly adjacent to the ruptured pipe at the plant to up to 1 km away from the spill origin. Water samples were collected in 1 L bottles and stored at 4°C until laboratory processing (within 6 h of collection). For FIB tests, 20 mL of water was subsampled for total coliforms and *E. coli* (Colilert, IDEXX Laboratories, Westbrook, ME) and enterococcus (Enterolert, IDEXX Laboratories). For DNA extraction, 250 mL $\times$ 3 of each sample was vacuum filtered through Whatman Anodisc membrane filters (47 mm dia., 0.2 μm pore size) and immediately frozen and stored at -80°C until DNA extraction. DNA was extracted from filters using the DNA EZ kit per manufacturer's instructions.

Water samples were also collected in conjunction with the County of Sonoma as part of the State of California AB411 monitoring program at Campbell Cove, Bodega Bay. Samples were collected weekly from September 2008 to October 2009 in knee-deep water with 1 L sampling bottles and processed in the laboratory as described above. A subset of samples was analyzed based on FIB counts. The analysis focused on nine samples that exceeded 1-day concentration limits and an additional 10 samples scattered throughout the sampling period that fell below FIB limits.

Polymerase Chain Reaction. The bacterial 16S rRNA gene was amplified from each sample using PCR with primers 27F (5'-AGAGTTTGTATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTACGACTT-3') for bacteria. Each PCR reaction contained 1 $\times$ Ex Taq buffer (Takara Bio Inc., Japan), 0.025 units/ μL Ex Taq polymerase, 0.8 mM dNTP mixture, 1.0 $\mu\text{g}/\mu\text{L}$ BSA, and 200 pM each primer, and 1 ng genomic DNA (gDNA) as template for fecal samples and 10 ng gDNA for water samples. 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), $48-58^{\circ}\text{C}$ (25 s), 72°C (2 min), followed by a final extension 72°C (10 min). Amplicons from each reaction were pooled for each sample, purified with the QIAquick PCR purification kit (Qiagen, Valencia, CA), and eluted in 50 μL elution buffer.

PhyloChip Assay Description and Analysis. A complete description of the PhyloChip design and analysis is available in the supplementary methods of Hazen et al.¹⁷ The PhyloChip (Second Genome, San Bruno, CA) was designed to detect most 16S rRNA gene sequences that identify bacteria and archaea. The PhyloChip probes for 59 959 different bacterial and

archaeal taxa that represent 147 phyla, 1123 classes, 1219 orders, and 1464 families according to the placement of its member organisms in the taxonomic outline as maintained by Philip Hugenholtz.¹⁸ The microarray includes 1 016 064 probe features, the majority of which target 16S rRNA gene sequences that are useful for differentiating taxa. Additional probes are for quality management, processing controls, image orientation, and normalization controls.¹⁷

PhyloChip Assay Analysis. For PhyloChip hybridization, we used 500 ng of bacterial PCR product for each microarray. PCR products were fragmented with DNase I to a range of 50–200 bp as verified by agarose gels. Commercial kits were utilized for DNA preparation: Affymetrix (Santa Clara, CA) WT Double Stranded DNA Terminal Labeling, and Affymetrix GeneChip Hybridization, Wash, and Stain kits were used for PhyloChip analysis. Briefly, fragmented 16S amplicons and non-16S quantitative amplicon reference controls were labeled with biotin in 40 μ L reactions containing: 8 μ L of 5X TDF buffer, 40 units of TDF, 3.32 nanomoles of GeneChip labeling reagent. After incubating at 37 °C for 60 min, 2 μ L of 0.5 M EDTA was added to terminate the reaction. Labeled DNA was combined with 65 μ L of 2X MES hybridization buffer, 20.4 μ L of DMSO, 2 μ L of Affymetrix control oligo B2, and 0.4 μ L nuclease free water. Each reaction mixture was injected into the hybridization chamber of an array cartridge and incubated for 16 h in an Affymetrix hybridization oven at 48 °C and 60 rpm. Hybridization solution was removed and the microarrays were stained and scanned according to the manufacturer's instructions.

Analysis procedures for fluorescent image files are described in detail in the Supporting Information of Hazen et al.¹⁷ Briefly, each individual array feature occupied approximately 8×8 pixels in the image file corresponding to a single probe 25mer on the surface. Probe intensities were background-subtracted and scaled to quantitative standards (non-16S rRNA gene spike-ins) as previously described.¹⁹ Presence/absence calling of each microbial taxon (operational taxonomic unit – OTU) was based on positive hybridization of multiple probes that correspond to an OTU (average of 37 probes/OTU). Differences in mean hybridization intensity (fluorescence) of an OTU probe set among different PhyloChips reflected differences in the relative abundance of the OTU.¹⁹ The PhyloChip data used in this study are available for download at http://greengenes.lbl.gov/Download/Microarray_Data/.

PhyloChip results are output as lists of detected OTUs and their hybridization scores with associated taxonomic information and references to represented sequences in public 16S rRNA gene repositories (greengenes.lbl.gov). Hybridization results were reduced to a community profile from each PhyloChip assay to a format useful for multivariate statistics consisting of log transformed hybridization intensity values for all detected OTUs. Interprofile dissimilarity was calculated with the Bray–Curtis metric, and the resulting distance matrix was analyzed with hierarchical cluster analysis and nonmetric multidimensional scaling (NMDS) ordination using the Primer v.6.1.13 statistical package. Analysis of Similarity (ANOSIM) was used to test the significance of differences in community composition among sample groups.

Determination of Source Identifier Taxa. Source identifier taxa were defined as individual OTUs that were detected in a single source type but never detected in any samples from other sources. The criteria for identifier bacteria selection were as follows: Identifier bacteria for animal sources

needed to be unique to a single animal type and present in at least three distinct populations. Identifier bacteria for human sources needed to be present in at least 7 of 8 samples. More stringent requirements were selected for human sources because a greater number of bacterial taxa were detected in human sources than animal sources likely due to over-representation of human-specific bacteria in the 16S rRNA gene database from human microbiome sequencing projects.

Source Identification in Environmental Water Samples. Application of PhyloChip for source identification in marine waters was tested in two field scenarios with known sources of human and avian contamination. The first was a monitoring study of a 720 000 gallon spill that occurred in Richardson Bay, an arm of San Francisco Bay, off the coast of Sausalito, CA, in February 2009. The PhyloChip was used to determine which bacterial taxa significantly increased in relative abundance in samples with high FIB counts and whether these enriched bacteria included the expected human identifier bacteria described above. To determine which taxa were specifically associated with high FIB counts, water samples with FIB concentrations that exceeded any 30-day geometric mean concentration limit were compared to samples that fell below all FIB concentration limits. Baseline microbial communities were defined by mean abundance of taxa in low FIB samples. Taxa whose relative abundance significantly exceeded baseline ($>\text{mean} + 2\sigma$) were determined in high FIB samples. The presence of source identifier bacteria in this enriched subset was used to determine the association between fecal sources and FIB exceedances. Results are reported as the percent of expected identifier taxa that were detected in each sample. The expected number of identifier taxa for a given source was defined as the average number of source-specific identifier taxa detected in individual populations of that source. A positive signal for source detection was defined as $>20\%$ enrichment of expected identifier taxa in a sample. Significant association between the detection of each source type and high FIB exceedances (enterococcus above regulatory limit) was tested with contingency analysis (JMP 7.0.1).

The second field test occurred at Campbell Cove in Bodega Bay, California, a recreational beach that frequently exceeds FIB water quality limits. Contamination at this beach is not from human sources and is likely due to gull feces.²⁰ We collected weekly monitoring samples over a 1-year period at this beach in conjunction with the county as part of the California Clean Beaches Initiative. Samples were split for both routine FIB testing and filtration for subsequent PhyloChip analysis. The analysis approach was similar to the sewage spill monitoring described above in which low FIB samples were analyzed at each site to establish the mean abundance and variance of each OTU under baseline (nonexceedance) conditions, and association between the enrichment of source identifier taxa and high FIB counts was tested with contingency analysis.

■ RESULTS AND DISCUSSION

Fecal Source Microbial Communities. A total of 20 368 bacterial OTUs were detected across all fecal samples. Samples clustered by source type indicating fecal bacterial communities of the same type of source animal were more similar to each other than to those of other sources (Figure 1). The deepest branching clusters separated all mammalian sources from avian sources indicating that microbial community composition is a distinctive characteristic of these two classes of vertebrates. Within the mammals, samples clustered into three distinct

Figure 1. Cluster analysis dendrogram of 16S rRNA gene composition showing similarity among microbial communities in fecal sources. Each sample represents a distinct animal population or sewage source and is a composite of individual fecal samples from the population.

groups comprised of grazing animals (cows, elk, horses), human wastes, and pinnipeds (Figure 1). Grazing mammals were further partitioned into two clusters comprising ruminants (cow, elk) and horses. Geese formed a distinct cluster within the birds. There was no obvious clustering among the other bird types (gulls, pelicans, pigeons, cormorants), and clustering patterns among these birds were not related to geography. All sources contained taxonomic groups that encompass *E. coli* and enterococcus that are used as regulatory fecal indicators.

Clostridia dominated the taxonomic (OTU) richness of the fecal bacteria in mammalian fecal sources (Figure 2). The

Figure 2. Composition of OTUs detected in each fecal source. OTUs are shown that were detected in at least half of the samples for each source animal. Total number of OTUs is shown in parentheses and lists with taxonomic descriptions found in the Supporting Information, Table 2.

remainder of taxonomic diversity in mammals was comprised of mainly Bacteroidales, Gammaproteobacteria, and Bacilli. These results are consistent with previous surveys of other mammalian gut microbial communities.^{12,13,21} In contrast to mammals,

avian feces contained far less taxa in the Clostridia and Bacteroidales and instead were dominated by Gammaproteobacteria and Bacilli (Figure 2).

Analysis of avian fecal samples revealed that seabirds and pigeons had similar composition of bacteria and were dominated by Gammaproteobacteria (mostly Enterobacteria)

and Bacilli (mostly Lactobacillales) (Figure 2). Fecal communities in these birds were also characterized by the presence of Fusobacteriaceae OTUs that were generally absent from mammalian and geese communities. Taxonomic composition in geese was distinct from other types of birds and contained greater numbers of taxa in the Actinobacteria, Alphaproteobacteria, and Clostridia. Geese differ from other birds in this study because of their unique diet and digestive system. Geese consume high-fiber feed, such as grass, and contain a well-developed cecum that facilitates their breakdown in the large intestine.²² This more rumenlike digestive system facilitates the activity of fermentative Clostridia.²³ Canada geese often forage for plants and insects in the soil, and consequently ingest bacteria that are resident in soil or on plant surfaces, which may explain the prominence of Actinobacteria and Alphaproteobacteria in their feces. Despite these differences between geese and other birds, geese bacterial communities were more similar to other types of birds than they were to any mammalian fecal sources (Figure 1).

Within the mammals, a variety of Clostridia, Bacilli, and Bacteroidetes dominated taxonomic (OTU) richness of grazing mammals (Figure 2). These bacteria are known to digest cellulose and other plant polysaccharides in the ruminant gut.²⁸ Clostridia, Gammaproteobacteria (mostly coliforms), and Bacteroidetes dominated taxonomic richness of human wastes (Figure 2). Human wastes were further distinguished by the presence of several Betaproteobacteria and Verrucomicrobia. Clostridia and Gammaproteobacteria dominated the taxonomic richness of pinnipeds.

On the basis of similarities in community composition (Figure 1), the data were partitioned into four major groups for identifier bacteria analysis: human wastes, birds, grazers, and pinnipeds. Source identifier taxa were defined as individual OTUs that were detected in a single source type but never detected in any samples from other sources. The number of OTUs that met criteria for selection as source identifier taxa was 304 for birds, 213 for grazers, 0 for pinnipeds, and 541 for human wastes (Figure 3).

Human identifier bacteria were primarily Bacteroidaceae and Clostridiales OTUs that matched known human fecal bacteria 16S rRNA gene sequences (Figure 3). Human Clostridiales OTUs were mainly found in the families Eubacterium, Faecalibacterium and Ruminococcus. Verrucomicrobia in the family Akkermansia were also indicative of human wastes, and are known to be mucin degraders in the human GI tract.²⁹

Bird identifier taxa included several different groups of Bacilli, mainly Lactobacillales, and Staphylococcaceae (Figure 3). In addition, bird identifier bacteria included one unclassified family in the Clostridiales, as well as Enterobacteriaceae and Fusobacteriaceae. Bacteroidetes are a minor component in avian microbial communities.²⁴ We found several Lactobacilli OTUs that are included in the same subfamily as *Catelicoccus marimammalium* and that are closely related to *Enterococcaceae*. Lu et al.²⁴ found gull feces were dominated by Bacilli (37% sequences), most of which were closely related to *Catelicoccus marimammalium*.

Grazer identifier taxa included a variety of Clostridia, many of which are known from cattle rumen, consisting of Clostridium, Ruminococcus, unclassified Clostridiales, RF6, RF30, RF39, and SHA-32 (Figure 3). In addition, grazer identifiers included several Bacilli taxa found in the Planococcaceae, and Bacteroidales taxa that were distinct from those found in human wastes (Figure 3).

Figure 3. Phylogenetic tree (a) and taxonomic composition (b) of source identifier OTUs for human wastes, birds, and grazers. The phylogenetic tree was constructed from using full-length 16S rRNA gene sequences of representative taxa in each OTU using the approximately maximum likelihood algorithm implemented in FastTree,²⁶ and the tree was displayed using the Interactive Tree of Life tool.²⁷ The outer bar of the phylogenetic tree (a) represents major bacterial phyla, the next bar is proportional to the number of OTUs for each of the 1053 source identifier OTUs, and the inner circle represents the phylogenetic placement for each OTU with branch lengths proportional to change in 16S rRNA gene sequence. Detailed taxonomic description and reference sequence information for each source identifier OTU is provided in Tables 3–5 of the Supporting Information.

Pinniped microbial communities were distinct from other fecal sources, but all OTUs found in at least three pinniped samples were also found in at least one other human or animal sample. For this reason, this study did not generate identifier taxa for pinnipeds due to the potential for cross-reactivity.

Source identification field tests. Application of PhyloChip for source identification in marine waters was tested in two field scenarios with defined sources of human and avian contamination. The first test looked investigated microbial communities in Richardson Bay waters adjacent to a large sewage spill. Out of 26 water samples collected during the spill, two exceeded the 1-day enterococcus concentration limit for marine water (104 MPN/mL) and an additional five exceeded the 30-day geometric mean limit (35 MPN/mL). These exceedance samples had significantly different microbial community compositions than samples that fell below FIB limits (Figure 4). All samples that were above enterococcus limits contained most (78–96%) of the expected fecal identifier bacteria for human fecal wastes (Figure 5). In contrast, there was little enrichment of bird or grazer identifier bacteria (0–10%) in samples with high enterococcus counts. Contingency analysis showed greater than expected numbers of samples with enrichment in human identifier taxa (>20% identifiers enriched) in high enterococcus samples ($P < 0.001$) but insignificant enrichment of grazer or bird identifier taxa ($P > 0.05$). The results show the PhyloChip analysis is sensitive to human fecal signal in marine waters.

The second field test was conducted along the beach of Campbell Cove in Bodega Bay, a site where a previous source tracking investigation found no evidence of human fecal contamination.²⁰ We analyzed a total of nine samples with high enterococcus counts (>35 MPN/mL) and eleven nonexceedance samples collected over 13 months. An average of 1093

Figure 4. Microbial community analysis of 26 water and two sewage samples collected during the 2009 sewage spill from the Sausalito Marin-City Sanitary District treatment plant. Ordination was conducted using nonmetric multidimensional scaling with the Bray–Curtis distance metric (stress = 0.03). Numerical values are enterococcus counts (MPN/mL) of individual samples. Microbial community composition in water samples that exceeded regulatory limits for enterococcus (>35 MPN/mL) was significantly different than in samples with low enterococcus counts (ANOSIM, Global $R = 0.927$, $P = 0.001$).

out of 6046 detected OTUs were significantly enriched over baseline relative abundances in high enterococcus samples. Several samples with high enterococcus had significant enrichment of identifier bacteria associated with bird feces (Figure 6). Contingency analysis showed greater than expected numbers of samples with enrichment in bird identifier taxa (>20% identifiers enriched) in high enterococcus samples ($P = 0.033$). Neither human nor grazer fecal identifiers were

Figure 5. Source identification in San Francisco Bay water samples collected during sewage spill monitoring ($N = 26$). Results from enterococcus FIB tests are plotted against the percent of source-identifier taxa that were significantly enriched above background (low FIB) conditions. High and low dashed lines show the single-day enterococcus concentration limit and 30-day geometric mean limits, respectively. Higher than expected numbers of high enterococcus samples were enriched in human identifier taxa ($P < 0.001$) but not bird or grazer identifier taxa ($P > 0.05$).

significantly enriched when enterococcus counts were high ($P > 0.05$). From these results, we conclude that birds and not human or grazer fecal inputs were associated with high enterococcus counts at Campbell Cove. These results are consistent with the findings of the previous source tracking investigation at this site that used *E. coli* ribotyping to determine gulls and not humans were a fecal source.²⁰ We also found four high enterococcus samples at Campbell Cove with negligible enrichment in any source identifier taxa indicating additional sources of FIB that were not tested. Further investigation of the phylogenetic inventory of all

Figure 6. Source identification at Campbell Cove, Bodega Bay ($N = 19$). Results from enterococcus FIB tests are plotted against the percent of source-identifier taxa that were significantly enriched above background (low FIB) conditions. High and low dashed lines show the single-day enterococcus concentration limit and 30-day geometric mean limits, respectively. Higher than expected numbers of high enterococcus samples were enriched in bird identifier taxa ($P = 0.033$) but not human or grazer identifier taxa ($P > 0.05$).

bacterial taxa from this site, not just fecal identifiers, could help reveal additional fecal or environmental sources of FIB.

The phylogenetic microarray approach to source identification uses simultaneous occurrence of many diverse taxa to determine fecal sources. Future work needs to address how fate and transport influences detection rates of these different taxa once they enter the environment. In the sewage spill example presented in this study, almost all human source identifier taxa were detected in water samples with high FIB. These fecal bacteria were input from a large release of sewage directly into the tested waters, and subject to little aging and decay. This situation is in contrast to the nonpoint source situation at

Campbell Cove where high FIB samples contained around 20% of the identifier taxa from a known fecal source (gull feces). Inputs of fecal bacteria at Campbell Cove were not necessarily direct into receiving waters but also from shoreline runoff and leaching through beach sands and sediments.²⁰ As a result, fecal microbial communities were subject to more modification before entering receiving waters compared to direct inputs by the sewage spill. Application of the community identifier approach to source tracking will benefit from adjusting the analysis based on the expected persistence of different taxa.

There are potential advantages and limitations to using a phylogenetic microarray for source identification. An advantage is sensitive detection of taxa with low abundance in the community.¹⁹ As fecal sources are diluted in receiving waters, taxa that are critical for source identification will decrease in relative abundance as they mix with the complex microbial background of the environment. The microarray probes for target sequences from the entire sample of PCR amplicons, and this amplicon pool consists of many billions of 16S rRNA gene sequences. Hybridization the entire pool amplified sequences may offer an advantage over pyrosequencing or other types of next generation sequencing because these methods randomly sequence a relatively small fraction of the amplified PCR product and consequently are not always reliable for detecting less abundant members of the community that may be critical for source identification.²⁵

A limitation in applying phylogenetic microarrays to MST may be the insufficient number of probes for sources that are underrepresented in 16S rRNA gene databases. For example, few studies have surveyed microbial diversity in pinnipeds, and as a likely consequence we found no unique taxa in pinnipeds using the PhyloChip. More thorough assessments of sequence composition in some source types will be needed to find additional host-specific targets. In addition, cost and complexity can be barriers to widespread adaptation of this technology in its current form. Measuring the full range of 16S rRNA gene sequences in the microbial community is not necessary, however, and a down-selected microarray that targets only the subset of microorganisms that is useful for source identification would simplify analysis and reduce cost.

The results of this study show that 16S rRNA gene composition of the bacterial community can be used to discriminate sources of fecal contamination. Differences in the diversity among fecal sources reveal hundreds of unique taxa that are specific to human, bird and grazer feces. Several different phylogenetic lineages, most of which are not considered in existing MST assays, differentiate these sources and are mainly found in the Clostridia, Bacilli, and Bacteroidetes. Comprehensive interrogation of microbial communities for these diverse identifier taxa has great potential to improve the reliability of source detection in the environment. Phylogenetic microarrays are an effective tool for rapidly measuring the full assortment of microbial taxa that distinguish fecal contaminants and deserve serious consideration for source tracking.

■ ASSOCIATED CONTENT

● Supporting Information

List of fecal types and source locations, list of human waste identifier taxa, list of bird identifier taxa, and list of grazer identifier taxa. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: glandersen@lbl.gov.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful for funding and support provided by the California State Water Resources Control Board Clean Beaches Initiative (07-576-550-0), Rathmann Family Foundation, County of Marin, National Institute of Health (R01-ES013515), US Environmental Protection Agency and the City of Dana Point, and U.S. Department of Energy under Contract No. DE-AC02-05CH11231 with the Lawrence Berkeley National Laboratory. We thank Phil Smith for his assistance with project administration, and Shariff Osman, Fran Reid, Donna Ferguson, Darcy Ebentier, and Meredith Raith for their assistance in the field and laboratory. The following individuals and organizations facilitated sample collection: County of Sonoma, Omar Arias (Sausalito-Marín City Sewage District), California Dept. of Fish & Game, National Park Service, U.S. Fish & Wildlife Service, Carly Schachter (San Francisco Bay Bird Observatory), Marine Mammal Center, Pier 39 Marina, Oakland Zoo, Gale Ranch, Lunny Ranch, Strauss Family Creamery, Sasha Gennet (Nature Conservancy), Melinda Fowler, Daniela Maldini (OKEANIS), Richard Bailey (Lake Merritt Institute), and Stinson Beach County Water District.

■ REFERENCES

- (1) Field, K. G.; Samadpour, M. Fecal Source Tracking, the Indicator Paradigm, and Managing Water Quality. *Water Res.* **2007**, *41* (16), 3517–3538.
- (2) Wade, T. J.; Pai, N.; Eisenberg, J. N. S.; Colford, J. M.; Do, U. S. Environmental Protection Agency Water Quality Guidelines for Recreational Waters Prevent Gastrointestinal Illness? A Systematic Review and Meta-Analysis. *Environ. Health Perspect.* **2003**, *111* (8), 1102–1109.
- (3) Boehm, A. B. Enterococci Concentrations in Diverse Coastal Environments Exhibit Extreme Variability. *Environ. Sci. Technol.* **2007**, *41* (24), 8227–8232.
- (4) Boehm, A. B.; Yamahara, K. M.; Love, D. C.; Peterson, B. M.; McNeill, K.; Nelson, K. L. Covariation and Photoinactivation of Traditional and Novel Indicator Organisms and Human Viruses at a Sewage-Impacted Marine Beach. *Environ. Sci. Technol.* **2009**, *43* (21), 8046–8052.
- (5) Yamahara, K. M.; Layton, B. A.; Santoro, A. E.; Boehm, A. B. Beach Sands along the California Coast Are Diffuse Sources of Fecal Bacteria to Coastal Waters. *Environ. Sci. Technol.* **2007**, *41* (13), 4515–4521.
- (6) Santo Domingo, J. W.; Bambic, D. G.; Edge, T. A.; Wuertz, S. Quo Vadis Source Tracking? Towards a Strategic Framework for Environmental Monitoring of Fecal Pollution. *Water Res.* **2007**, *41* (16), 3539–52.
- (7) USEPA, *Microbial Source Tracking Guide Document*, EPA-600/R-05/064; U.S. Environmental Protection Agency: Washington, DC, 2005.
- (8) Bae, S.; Wuertz, S. Rapid Decay of Host-Specific Fecal Bacteroidales Cells in Seawater As Measured by Quantitative PCR with Propidium Monoazide. *Water Res.* **2009**, *43* (19), 4850–4859.
- (9) Balleste, E.; Blanch, A. R. Persistence of Bacteroides Species Populations in a River as Measured by Molecular and Culture Techniques. *Appl. Environ. Microbiol.* **2010**, *76* (22), 7608–7616.

- (10) Walters, S. P.; Field, K. G. Survival and Persistence of Human and Ruminant-Specific Faecal Bacteroidales in Freshwater Microcosms. *Environmental Microbiology* **2009**, *11* (6), 1410–1421.
- (11) Cao, Y.; Wu, C. H.; Andersen, G. L.; Holden, P. A., Community Analysis-Based Methods. In *Microbial Source Tracking: Methods, Applications, and Case Studies*, Hagedorn, C.; Blanch, A. R.; Harwood, V. J., Eds. Springer: New York, NY, 2011; pp 251–282.
- (12) Lee, J. E.; Lee, S.; Sung, J.; Ko, G. Analysis of Human and Animal Fecal Microbiota for Microbial Source Tracking. *The ISME Journal* **2011**, *5* (2), 362–5.
- (13) Unno, T.; Jang, J.; Han, D.; Kim, J. H.; Sadowsky, M. J.; Kim, O. S.; Chun, J.; Hur, H. G. Use of Barcoded Pyrosequencing and Shared OTUs To Determine Sources of Fecal Bacteria in Watersheds. *Environ. Sci. Technol.* **2010**, *44* (20), 7777–7782.
- (14) Wu, C. H.; Sercu, B.; Van de Werfhorst, L. C.; Wong, J.; DeSantis, T. Z.; Brodie, E. L.; Hazen, T. C.; Holden, P. A.; Andersen, G. L. Characterization of Coastal Urban Watershed Bacterial Communities Leads to Alternative Community-Based Indicators. *PLoS One* **2010**, *5* (6), e11285.
- (15) Cao, Y. P.; Van De Werfhorst, L. C.; Sercu, B.; Murray, J. L. S.; Holden, P. A. Application of an Integrated Community Analysis Approach for Microbial Source Tracking in a Coastal Creek. *Environ. Sci. Technol.* **2011**, *45* (17), 7195–7201.
- (16) Jeong, J. Y.; Park, H. D.; Lee, K. H.; Weon, H. Y.; Ka, J. O. Microbial Community Analysis and Identification of Alternative Host-Specific Fecal Indicators in Fecal and River Water Samples Using Pyrosequencing. *J. Microbiol.* **2011**, *49* (4), 585–594.
- (17) Hazen, T. C.; Dubinsky, E. A.; DeSantis, T. Z.; Andersen, G. L.; Piceno, Y. M.; Singh, N.; Jansson, J. K.; Probst, A.; Borglin, S. E.; Fortney, J. L.; Stringfellow, W. T.; Bill, M.; Conrad, M. E.; Tom, L. M.; Chavarria, K. L.; Alusi, T. R.; Lamendella, R.; Joyner, D. C.; Spier, C.; Baelum, J.; Auer, M.; Zemla, M. L.; Chakraborty, R.; Sonnenthal, E. L.; D'Haeseleer, P.; Holman, H. Y. N.; Osman, S.; Lu, Z. M.; Van Nostrand, J. D.; Deng, Y.; Zhou, J. Z.; Mason, O. U. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science* **2010**, *330* (6001), 204–208.
- (18) Hugenholtz, P. Exploring Prokaryotic Diversity in the Genomic Era. *Genome Biology* **2002**, *3*, 1–8.
- (19) DeSantis, T. Z.; Brodie, E. L.; Moberg, J. P.; Zubieta, I. X.; Piceno, Y. M.; Andersen, G. L. High-Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library When Sampling the Environment. *Microbial Ecology* **2007**, *53*, 371–383.
- (20) *Final Interim Report for Bodega Bay—Campbell Cove Tidal Circulation Study, Water Quality Testing and Source Abatement Measures Project*; County of Sonoma Department of Health Services: Santa Rosa, CA, 2004.
- (21) Ley, R. E.; Hamady, M.; Lozupone, C.; Turnbaugh, P. J.; Ramey, R. R.; Bircher, J. S.; Schlegel, M. L.; Tucker, T. A.; Schrenzel, M. D.; Knight, R.; Gordon, J. I. Evolution of mammals and Their Gut Microbes. *Science* **2008**, *320* (5883), 1647–1651.
- (22) Clench, M. H.; Mathias, J. R. The Avian Cecum—A Review. *Wilson Bulletin* **1995**, *107* (1), 93–121.
- (23) Wang, Z. Y.; Shi, S. R.; Xu, M. J.; Yang, H. M. 16S rRNA-Based Analysis of Bacterial Diversity in the Microbial Flora of the Goose Intestinal Tract. *Journal of Animal and Feed Sciences* **2009**, *18* (3), 531–540.
- (24) Lu, J. R.; Santo Domingo, J. W.; Lamendella, R.; Edge, T.; Hill, S. Phylogenetic Diversity and Molecular Detection of Bacteria in Gull Feces. *Appl. Environ. Microbiol.* **2008**, *74* (13), 3969–3976.
- (25) Zhou, J. Z.; Wu, L. Y.; Deng, Y.; Zhi, X. Y.; Jiang, Y. H.; Tu, Q. C.; Xie, J. P.; Van Nostrand, J. D.; He, Z. L.; Yang, Y. F. Reproducibility and Quantitation of Amplicon Sequencing-Based Detection. *ISME Journal* **2011**, *5* (8), 1303–1313.
- (26) Price, M. N.; Dehal, P. S.; Arkin, A. P., FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **2010**, *5* (3), Article No.: e9490.
- (27) Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **2007**, *23* (1), 127–128.
- (28) Flint, H. J.; Bayer, E. A.; Rincon, M. T.; Lamed, R.; White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **2008**, *6*, 121–131.
- (29) Derrien, M.; Collado, M. C.; Ben-Amor, K.; Salminen, S.; de Vos, W. M. The Mucin Degrader *Akkermansia muciniphila* Is an Abundant Resident of the Human Intestinal Tract. *Appl. Environ. Microbiol.* **2008**, *74*, 1646–1648.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.