A Novel Microbial Source Tracking Microarray for Pathogen Detection and Fecal Source Identification in Environmental Systems

ARTICLE *in* ENVIRONMENTAL SCIENCE & TECHNOLOGY · MAY 2015

Impact Factor: 5.33 · DOI: 10.1021/acs.est.5b00980 · Source: PubMed

6 AUTHORS, INCLUDING:

Xiang Li
West Virginia University
6 PUBLICATIONS **24** CITATIONS
SEE PROFILE

Valerie J Harwood
University of South Florida
115 PUBLICATIONS **3,217** CITATIONS
SEE PROFILE

Christopher Staley
University of Minnesota Twin Cities
34 PUBLICATIONS **345** CITATIONS
SEE PROFILE

Michael J Sadowsky
University of Minnesota Twin Cities
195 PUBLICATIONS **4,540** CITATIONS
SEE PROFILE

Environmental Science & Technology

Article

pubs.acs.org/est

A Novel Microbial Source Tracking Microarray for Pathogen Detection and Fecal Source Identification in Environmental Systems

Xiang Li,[†] Valerie J. Harwood,[‡] Bina Nayak,[‡] Christopher Staley,[§] Michael J. Sadowsky,[∥] and Jennifer Weidhaas*,[†]

[†]Department of Civil and Environmental Engineering, West Virginia University, P.O. Box 6103, Morgantown, West Virginia 26506, United States

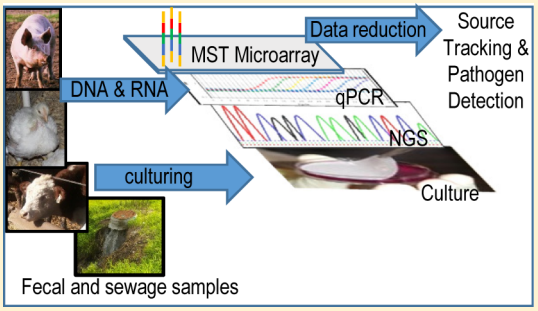
[‡]Department of Integrative Biology, University of South Florida, Tampa, Florida 33620, United States

[§]BioTechnology Institute, University of Minnesota, St. Paul, Minnesota 55108, United States

[∥]Department of Soil, Water, and Climate, BioTechnology Institute, University of Minnesota, St. Paul, Minnesota 55108, United States

Supporting Information

ABSTRACT: Pathogen detection and the identification of fecal contamination sources are challenging in environmental waters. Factors including pathogen diversity and ubiquity of fecal indicator bacteria hamper risk assessment and remediation of contamination sources. A custom microarray targeting pathogens (viruses, bacteria, protozoa), microbial source tracking (MST) markers, and antibiotic resistance genes was tested against DNA obtained from whole genome amplification (WGA) of RNA and DNA from sewage and animal (avian, cattle, poultry, and swine) feces. Perfect and mismatch probes established the specificity of the microarray in sewage, and fluorescence decrease of positive probes over a 1:10 dilution series demonstrated semiquantitative measurement. Pathogens, including norovirus, *Campylobacter fetus, Helicobacter pylori*, *Salmonella enterica*, and *Giardia lamblia* were detected in sewage, as well as MST markers and resistance genes to aminoglycosides, beta-lactams, and tetracycline. Sensitivity (percentage true positives) of MST results in sewage and animal waste samples (21−33%) was lower than specificity (83−90%, percentage of true negatives). Next generation DNA sequencing revealed two dominant bacterial families that were common to all sample types: *Ruminococcaceae* and *Lachnospiraceae*. Five dominant phyla and 15 dominant families comprised 97% and 74%, respectively, of sequences from all fecal sources. Phyla and families not represented on the microarray are possible candidates for inclusion in subsequent array designs.



INTRODUCTION

Approximately 166 000 river and stream miles in the U.S. are impaired due to the presence of waterborne pathogens.[1] These microorganisms, which originate from fecal contamination, can result in disease outbreaks, particularly in recreational water near lake and marine shores.[2] Impairment, or chronic contamination of waters by fecal contamination, is typically evaluated using analyses of fecal indicator bacteria (FIB). However, the presence of FIB does not necessarily indicate that pathogens are present in the environmental system, as the morphology, physiology, and ecology (including distribution among host species and fate outside of the host) of microbial pathogens is widely variable. Therefore, the expectation that one indicator (e.g., *Escherichia coli* or enterococci) can predict the risk from all pathogens is likely to be problematic.

Fecal contamination of environmental systems can originate from a variety of sources, including agriculture[3] and runoff from urban surfaces[4] or direct release of untreated sewage to waterbodies. Additionally, nonpoint sources such as wild animals (deer, raccoon), wild avian (ducks, geese), and domesticated animals (cats, dogs, cattle) can contribute fecal contamination to water bodies.[5] To remediate fecal contamination in environmental systems, it would be valuable to be able to identify the source(s) of fecal material contributing pathogens to waterbodies, in addition to quantifying pathogens in these systems to assess risk to human health. The FIB have been shown to be poor indicators of sources of fecal contamination in water bodies due to their broad distribution in the gastrointestinal tracts of most warm-blooded animals,[6] and their ability to survive outside the host environment. Various microbial source tracking (MST) methods have been proposed for identification of different fecal pollution sources in environmental systems,[7] and offer more rapid turnaround times and specificity of the tests to particular pathogens or host associated fecal microbes. However, these MST methods suffer from low throughput as a limited number of targets can be assayed at one time.

Received: February 24, 2015
Revised: May 2, 2015
Accepted: May 13, 2015
Published: May 13, 2015

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

Microarrays, wherein thousands to hundreds of thousands of targets can be assayed at one time, overcome the limitations of current culture and qPCR based assays and are of increasing interest in the research and regulatory realms.[8−10] While microarrays have been developed previously for MST[11,12] and detection of pathogens in sewage[13,14] or other media,[15−17] there are numerous drawbacks to use of these microarrays. First, previously developed arrays use PCR based amplification, wherein only a handful of genes are targeted for amplification. While multiplex PCR (e.g., targeting up to 10 different organisms[18]) has been done previously, there are still unmanageable costs and complexity associated with combining gene specific PCR with microarrays. Second, the previous arrays have only targeted specific groups of organisms such as pathogens (*E. coli*,[13] *Cryptosporidium*[19]), indicator organisms (*Enterococcus* spp.,[11] *Bacteroidales*[12]) or viruses in clinical applications.[17] For example, a recently reported array focuses on detection of key virulence genes of pathogens (i.e, the PathoChip),[20] and Ishii and colleagues[21] reported the use of microfluidic PCR to detect pathogens in food and water. Other community based microarrays targeting the 16S rRNA gene of bacteria (e.g., the Phylochip) have been used for microbial source tracking.[22−24] Drawbacks to these community based microarrays include decreased sensitivity (38−80%) for detecting fecal sources in environmental samples compared to specific MST gene targets tested via qPCR[23] and an inability to simultaneously target Bacteria, Eukaryotes and viruses with the microarray.

The purpose of this study was to design and validate an MST microarray for detecting fecal contamination and pathogens in environmental samples, while simultaneously indicating the source of fecal contamination. Specifically, the MST microarray was designed to detect pathogens (bacteria, protozoa, and viruses), FIB, and MST markers. Further studies were conducted to evaluate the correlation between the microarray fluorescence and more established methods, namely qPCR and culture based methods. Finally, next generation sequencing was conducted to determine the relative coverage of the fecal microbiome represented by the microarray-based probes.

MATERIALS AND METHODS

Fecal Sample Collection and Handling. Swine feces, cattle feces, and soiled poultry litter were collected at the West Virginia University Animal Sciences Farm, while duck and goose feces were collected at the Monongahela River (Star City, WV). Wild avian DNA was obtained from composite samples of gull feces collected on Delaware beaches. The wild avian DNA was combined with the duck and goose nucleic acids after extraction. Raw municipal sewage was collected from the Star City Wastewater Treatment Plant (Star City, WV) immediately after bar screening in the headworks building. To evaluate the potential for food safety monitoring, spinach was contaminated by dipping a leaf into raw sewage, then microorganisms were removed from the surface of the leaf, concentrated and nucleic acids were extracted. Additional details regarding sample collection, homogenization and compositing are provided in the Supporting Information.

Culture Based Enumeration. *E. coli* and *Enterococcus faecalis*, *Salmonella enterica*, and *Staphylococcus aureus* were quantified using EPA or Standard Methods[25−28] from 1 g of cattle feces, swine feces, or poultry litter and 100 mL of raw sewage. A 1 g aliquot of solid fecal or litter samples were mixed with 100 mL of sterile 1× PBS and stirred for 3 h at room temperature. Samples were allowed to settle for 5 min and the concentration of microorganisms in the supernatant was quantified. Confirmation of the pathogens *S. enterica* and *S. aureus* were determined via qPCR detection of the *invA* and *sec* genes, respectively (Table S1, Supporting Information).

Co-Extraction of Nucleic Acids. Nucleic acids were extracted from approximately 0.5 g fecal or litter sample, or 600 mL of raw sewage, using a previously published DNA and RNA coextraction method.[29] The complementary DNA (cDNA) was synthesized by the Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA) following the manufacture's protocol.

qPCR Based Enumeration. Primers and probes used in this study, qPCR reaction efficiency and R^2 values are presented in Table S1 (Supporting Information). The 2× Master Mix and TaqMan probes were purchased from Life Technology (Woburn, MA), and all qPCR experiments were performed using the Applied Biosystem 7300 Real-Time PCR system (Foster city, CA). TaqMan-based qPCR detection of *E. faecalis*, *E. coli*, *S. enterica*, *S. aureus*, and *Bacteroidales* (GenBac) was carried out as previously reported.[30] The thermocycler conditions for SYBR green based detection of *Bacteroides*, human polyomavirus and human norovirus was 50 °C, 2 min; 95 °C, 10 min; 45 cycles of 95 °C for 30 s, then 60 °C (or 57 °C for norovirus) for 1 min; followed by dissociation curve. Each qPCR run included a positive control (plasmid containing the template DNA) and a negative control (PCR grade water). Triplicate analyses were done for each sample. Details regarding the plasmid construction are provided in the supplemental data.

Microarray Design. Microorganisms and gene targeted on the microarray were included only if they could aid in answering the following three questions. Are common waterborne pathogens present in the sample? Are pathogens that originate from fecal contamination present in the sample? Which fecal source contributed these pathogens? As such the following types of microorganisms and genes were included in the microarray design: (1) common waterborne pathogens including Bacteria, Eukaryotes and viruses, (2) previously published MST marker genes and organisms, (3) antibiotic resistance genes, (4) fecal indicator bacteria, and (5) universal bacterial probes. All probes on the custom microarray (Custom CGH, 8 X15K platform, Agilent, Santa Clara, CA) were 60 mers and had a targeted melt temperature of 65−82 °C. Probes were selected for the array according to the following preference: (1) previously published and validated microarray probes, (2) previously published qPCR probes that could be lengthened to 60 mers without cross hybridizing with known environmental microbes, and (3) probes designed during this study, targeting microorganisms and genes of interest listed above. CommOligo2.0[31] was used to design probes for this study, targeting GC contents from 40 to 65% within the target melting temperatures. Each custom designed probe sequence was validated for specificity against the NCBI Blastn database. The targeted organisms, genes, probe sequences and references are listed in Supporting Information Table S2. The distribution of target genes or microorganisms for the probes on the microarray included 40% rRNA (5S, 16S, 18S), 16% virus (17 different viruses), 14% mitochondrial DNA (28 different organisms), 11% pathogen virulence genes (77 different genes), and 6% antibiotic resistant genes (beta-lactams, tetracycline resistance, aminoglycoside resistance). The remaining genes (13%) were of unknown function. Three positive

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

control probes and either three (on microarray (1) or six (on microarray (2) negative control probes were included in the design.

The perfect match (PM) and mismatch (MM) method previously proposed[32] was used to minimize the potential for false positive detections. A total of 27 MM probes were designed for each PM probe by replacing 1, 2, or 3 nucleotides at the following three probe regions along the 60 mer probes: the 14, 15, 16 positions; the 29, 30, 31 positions; and/or the 44, 45, 46 positions. For generation of the MM probes, nucleotides targeted for substitution containing cytosine or guanine in the PM probe were replaced with an adenine nucleotide, while probe nucleotides containing an adenine or thymine in the PM probe were replaced with a guanine. In total 15,699 probes were added to the microarray design including: 13,860 PM probes (each replicated 5 times), MM probes (each replicated 1−3 times), positive control or negative control probes (each replicated five times); 1262 normalization probes (Agilent designed); and 577 Agilent control probes.

Sample Whole Genome Amplification, Digestion, Labeling, Hybridization and Scanning. Samples assayed on microarray 1 included: cattle and swine feces, poultry litter, raw sewage, wild avian feces, and a spiked positive control. Samples assayed on microarray 2 included: cattle and swine feces, poultry litter, raw sewage and diluted raw sewage (1:5, 1:10, and 1:50 dilution) and spinach dipped in raw sewage. The positive control sample consisted of PCR grade water containing 80 000 gene copies $\times 5\ \mu L^{-1}$ each of plasmids or genomic DNA from (1) plasmids of *Bacteroides* sp. HF-183, polyomavirus, and norovirus (constructed for this study using regions targeted by primers in Supporting Information Table S1), (2) plasmids of *Brevibacterium* sp. LA35 (NCBI FJ462358), *E. faecalis*[30] and general *Bacteroidales* (GenBac)[33] and (3) cultures of *S. aureus* (ATCC 25923), *S. enterica* (ATCC 14028), and *E. coli* (ATCC 9637).

The general sample handing protocol as shown in Figure 1 was as follows: (1) total DNA or cDNA is amplified separately by whole genome amplification (WGA) then combined, (2) restriction enzyme digestion reduces the length of the nucleic acids to <3000 bp for optimum hybridization to the array, (3) positive controls at known concentrations are added to the digested nucleic acids, (4) nucleic acids are labeled with Cy3, (5) samples are hybridized to the array, then unbound nucleic acids are washed from the array, and (6) the array is scanned and the data normalized for analysis. Specifically, one microliter of DNA or cDNA at 100 ng $\times\ \mu L^{-1}$ from the samples were amplified by WGA separately using the Illustra Genomiphi V2 DNA Amplification Kit (GE Healthcare, Pittsburgh, PA) and then 5 μL of both WGA DNA and cDNA were combined for further processing.

A combination of four restriction enzymes, *PvuII*, *RsrII*, *SgrAI*, and *Nb.BbvcI* (New England Biolabs, Ipswich, MA) were selected to achieve a nucleic acid fragment length less than 3000 bp that were shown not to digest within the 60 mer PM probe sequences based on in silico analysis. All the samples were digested with $1U \times \mu L^{-1}$ each of *PvuII*, *RsrII*, *SgrAI*, and *Nb.BbvcI*, combined with 3 μL of 10× buffer, 13 μL water and 10 μL of WGA sample. The samples were digested at 37 °C for 4 h and the reaction was terminated by heating to 80 °C for 20 min. Resulting DNA fragments were analyzed by a 2100 Bioanalyzer (Agilent, Santa Clara, CA) to determine the length of the nucleic acids.

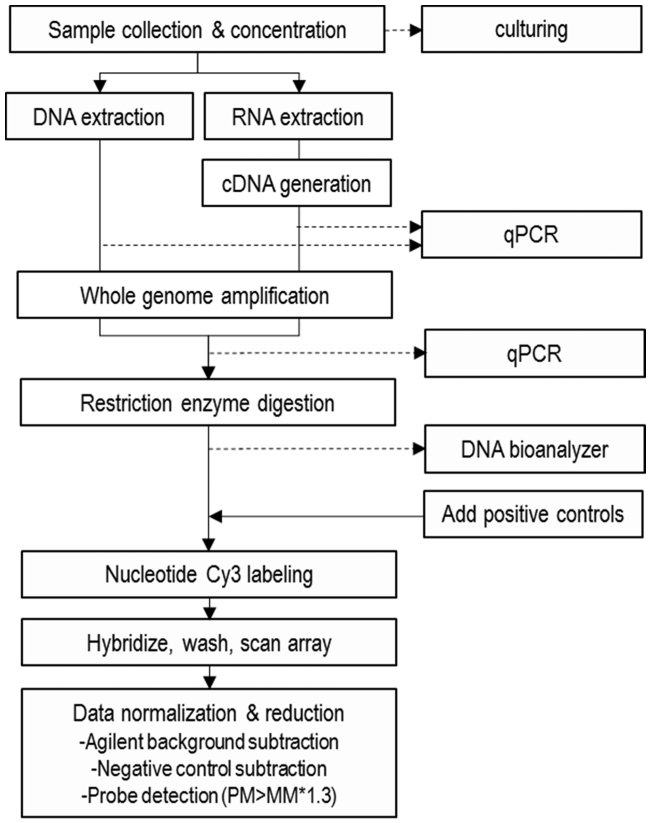


Figure 1. Microarray sample handling and analysis procedure (solid lines). Dashed lines indicate periodic quality control and method validation testing.

Two positive controls, namely PCR products containing a 220 bp fragment of rainbow trout 5.8S rRNA gene and a 74 bp fragment of *Dehalococcoides mccartyii* 16S rRNA,[34] were added to all samples at a final concentration of 11 ng $\times\ \mu L^{-1}$ and 5.5 ng $\times\ \mu L^{-1}$, respectively. Details regarding the plasmid construction for these positive controls are provided in the Supporting Information. Enzyme-digested DNA fragments were labeled with Cy3 according to the manufacture's protocol (SureTaq DNA labeling kit, Agilent, Santa Clara, CA). The labeled samples and the microarrays were analyzed by the Duke University Microarray Facility using an Agilent C Scanner.

Microarray Data Normalization and Analysis. All fluorescence signals were log transformed and background subtracted according to Agilent data normalization protocols.[35] Then the average of the negative control and nonsense probe fluorescence for each microarray was subtracted from PM and MM probe log fluorescence values. Target genes were called a detect on the microarray only if (1) all five PM probes had log fluorescence greater than the average log fluorescence of the negative control probes, and (2) the average log fluorescence of the five PM probes were greater than 1.3 times the average log fluorescence of MM probes for that gene target.[36] The 1.3 log fluorescence difference between the PM and MM probes was suggested previously for the Phylochip.[37]

Next Generation Sequencing. Samples of the raw sewage, swine feces, poultry litter, and cattle feces collected for microarray 2 were sent to the University of Minnesota Genomics Center (Minneapolis) for next generation sequencing. The V5−V6 hypervariable regions of the 16S rRNA[38] were amplified using the BSF784/R1064 primer set. Amplicons were

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

Table 1. Increase in DNA Concentrations during Whole Genome Amplification (WGA) from Fecal and Wastewater Samples

		concentration, log gene copies × L^{-1} or g^{-1}				
fecal source	sample	enterococcus 23S rRNA	*E. coli uidA*	*Bacteroidales* 16S rRNA	*Salmonella* sp. *invA*	average (standard deviation)
wastewater influent	pre-WGA	2.8	0.9	6.5	1.8	2.7 (0.8)
	post-WGA	5.6	3.7	10.0	3.3	
	log increase	2.9	2.8	3.5	1.5	
beef and dairy cattle manure	pre-WGA	2.0	2.9	9.0	3.0	1.9 (0.2)
	post-WGA	4.1	4.8	11.0	4.6	
	log increase	2.0	1.9	2.1	1.6	
swine manure	pre-WGA	1.3	1.8	7.5	2.8	0.7 (1.2)
	post-WGA	1.1	2.3	9.9	2.8	
	log increase	−0.1	0.6	2.4	0.1	
poultry litter	pre-WGA	3.3	3.2	2.9	2.8	1.4 (0.7)
	post-WGA	5.5	4.2	4.5	3.5	
	log increase	2.3	1.0	1.6	0.7	

gel purified and pooled in equal amounts for sequencing at a read length of 2 × 150 nt on a HiSeq2500. Raw data was received as fastq files and is deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) under accession number SRP054964.

Sequences were processed using mothur ver 1.29.2.[39] Details of sequence processing are provided in the Supporting Information. Briefly, sequences were paired-end aligned using fastq-join,[40] trimmed for quality, and binned into operational taxonomic units (OTUs) at 97% using the furthest-neighbor algorithm. Taxonomic assignments were made against the Ribosomal Database Project database ver. nine.[41] For comparison among samples, sequence read numbers were normalized to 234 281 reads per sample by random subsampling.

Statistical Methods. Nonmetric multidimensional scaling (NMDS) of the microarray data was used to visually discriminate between closely related fecal samples based on evolutionary distance. Genes detected on the microarray were assigned a value of "1", and nondetects a "0" for all 423 PM probes (excluding controls), and a matrix of 13 samples was generated. The two-dimensional NMDS plots were generated using PROC MDS of a Bray−Curtis distance matrix (Gower method) on the binomial data set using SAS (ver. 9.4, SAS Institute, Inc., Cary, NC). A total of 20 replicate plots were generated and the plots with the least stress were selected. A stress measure of less than 0.1 corresponds to ideal ordination with little likelihood of misinterpretation.[42]

Using only the probes with known host specificity (e.g., MST markers or host specific pathogens) the sensitivity and specificity of the microarray was determined as follows. The positive predictive value or the percentage of reported probe detections that are true positives (TP), was estimated as TP/(TP+FP), where FP is false positives. Conversely, the negative predictive value or the percentage of probes that were not detected on an array that are true negatives (TN) was calculated at TN/(FN+TN), where FN is false negatives. Percentage true positive (sensitivity) was calculated as the number of correctly identified positive probes detected on the microarray in a sample divided by the number of probes associated with a particular host on the microarray, TP/(TP+FN). The percentage of false positives was calculated as the number of false positive probes detected on a microarray (e.g., cattle marker detected in swine feces) divided by the total number of probes on the microarray, FP/(TP+TN+FP+FN). Specificity was calculated as the percentage of true negatives TN/(FP+TN).[43,44]

Plots were generated in Sigma Plot (Systat Software, Inc., Chicago, IL). Correlations were calculated using SigmaStat and were considered significant at the $P < 0.05$ level. Similarity percentages (SIMPER) were calculated with PAST.[45] Diversity indices including number of OTUs, Shannon index and abundance based coverage estimators (ACE) for the NGS data were performed using mother ver 1.29.2.[39]

RESULTS

Microarray Sample Processing and Method Validation. Whole genome amplification increased the nucleic acid concentration in samples an average of 1.7 ± 1.0 log gene copies × L^{-1} or g^{-1} for all microbial targets in the samples tested (Table 1). After restriction enzyme digestion of WGA products from raw sewage, the average length of the digested products was 2,406 bp (Figure S1, Supporting Information). None of the PM probes contained complementary restriction enzyme digestion sites, however, 13 of the MM probes did. Location of the restriction enzyme digestion site on the MM probes was positively correlated (Pearson Product Moment, $r = 0.38$, $P = 0.007$) with log fluorescence detected on the microarrays from raw sewage, swine feces, cattle feces, and poultry litter (Figure S2, Supporting Information). The longer the WGA product after restriction enzyme digestion (i.e., cut site at 40 bp rather than 11 bp when reading 5′-3′), the larger the relative fluorescence of the digested WGA product.

Eight of nine genes or organisms in the spiked positive control sample were detected in 22 individual probes, for an 89% detection rate. Thirty-two false positive probes were detected on microarray 1. The average false positives fluorescence was 1.08 ± 0.54 log fluorescence units (FU) while the true positive detections were 1.17 ± 0.62 log FU or 8% higher on average. For microarray 2 these false positive probes were removed or redesigned (Table S2, Supporting Information).

The relative FU of PM probes on the microarray may be correlated with (1) the number of adenosine nucleotides on the

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

Table 2. Summary of Viruses and Pathogens Detected via the Microarray in Various Fecal Samples Percentage of Samples Tested Containing the Pathogen or Virus (Number of Samples Tested)

pathogen or virus detected	raw and diluted sewage	cattle feces	swine feces	poultry litter	wild avian feces
Viruses					
adenovirus	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
bacteriophages	80 (5)	50 (2)	50 (2)	100 (2)	100 (1)
bocavirus	80 (5)	100 (2)	50 (2)	100 (2)	100 (1)
influenza C	20 (5)	0 (2)	0 (2)	0 (2)	0 (1)
norovirus	40 (5)	50 (2)	50 (2)	100 (2)	0 (1)
polyomavirus	40 (5)	0 (2)	0 (2)	0 (2)	0 (1)
pepper mild mottle virus	0 (4)	0 (1)	0 (1)	100 (1)	NA*
torque teno virus	0 (4)	0 (1)	100 (1)	0 (1)	NA
Pathogens (Bacterial or Eukaryotic)					
Campylobacter fetus	100 (4)	100 (1)	100 (1)	100 (1)	NA
C. jejuni	0 (5)	0 (2)	0 (2)	100 (2)	100 (1)
C. lari	0 (5)	50 (2)	0 (2)	0 (2)	0 (1)
Clostridium botulinum	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
C. clostridiiforme	50 (4)	100 (1)	100 (1)	100 (1)	NA
C. difficile	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
C. perfringens	100 (5)	100 (2)	50 (2)	50 (2)	100 (1)
C. tetani	100 (5)	100 (2)	100 (2)	50 (2)	100 (1)
Escherichia coli	80 (5)	50 (2)	50 (2)	100 (2)	100 (1)
Giardia lamblia	60 (5)	100 (2)	50 (2)	50 (2)	100 (1)
Helicobacter pylori	100 (4)	100 (1)	100 (1)	100 (1)	NA
Listeria monocytogenes	20 (5)	0 (2)	0 (2)	50 (2)	0 (1)
Leptospirillum ferriphilum	0 (4)	0 (1)	100 (1)	100 (1)	NA
Mycobacterium tuberculosis	100 (4)	100 (1)	100 (1)	100 (1)	NA
Naegleria gruberi	50 (4)	0 (1)	100 (1)	100 (1)	NA
Salmonella enterica	60 (5)	50 (2)	50 (2)	100 (2)	100 (1)
S. enterica typhimurium	50 (4)	0 (1)	0 (1)	100 (1)	NA
Schistosoma incognitum	0 (5)	50 (2)	50 (2)	0 (2)	0 (1)
Shigella flexneri	50 (4)	100 (1)	100 (1)	100 (1)	NA
Staphylococcus aureus	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
β-Staphylococcus hemolyticus	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
Vibrio cholerae	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
V. parahemolyticus	60 (5)	100 (2)	50 (2)	100 (2)	100 (1)
Yersinia enterocolitica	100 (5)	100 (2)	100 (2)	100 (2)	100 (1)
Y. ruckeri	40 (5)	0 (2)	0 (2)	50 (2)	100 (1)

*NA = not applicable since the probe was not included on microarray for the sample tested.

oligonucleotides hybridized, (2) concentration of PM oligonucleotides added to the array, or (3) the average length of the oligonucleotides hybridized to the array probes (e.g., if >60 mer). To test the first hypothesis we evaluated the number of adenosine nucleotides in the PM probes compared to the relative FU on a microarray. In all cases, the PM probe sequence relative FU was not correlated with the number of adenosine nucleotides (Pearson's product moment, $P > 0.05$) (Figure S3, Supporting Information) expected from the probe sequence. The average relative fluorescence of the negative control probes on microarray 1 and 2 was 3.1 ± 0.1 and 2.2 ± 0.7 log FU, respectively. In 9 of 13 samples (69%) the fluorescence of the negative controls was not correlated (Pearson's Product Moment, $P > 0.05$) with the number of adenosine nucleotides in the probe sequence, which would be hybridized with Cy3 labeled uracil in the sample. Only in the diluted raw sewage samples and the produce dipped in raw sewage was the relative FU of the negative controls positively correlated ($P < 0.05$) with the number of adenosine nucleotides.

Fecal Sample Testing Results via Microarray. The microorganisms or targeted genes detected in each sample are presented in Table S3 (Supporting Information). There were a total of 70 and 132 probes detected in the raw sewage on microarray 1 and 2, respectively. Similarly, 73 and 111 probes were detected in the cattle feces, 68 and 110 probes detected in the swine feces, 104 and 130 probes detected in the poultry litter on microarray 1 and 2, respectively. Ninety-one probes were detected from the mixed avian fecal sample, and 100 probes were detected in the spinach sample dipped in raw sewage on microarray 1.

Eight different viruses and 25 different pathogens were detected in the sewage, feces, and poultry litter samples (Table 2). In most cases, several different virulence or housekeeping genes were detected. For example, six of the 26 different *E. coli* virulence genes on the microarray were detected in various samples (Supporting Information Table S3), including the *flic* gene of *E. coli* 0157:H7 and O55:H7, *argY* gene of *E. coli* K12, *sfaD* gene of *E. coli* CFTO73, *itpb* gene of ETEC *E. coli* and the *uidA* gene of *E. coli*. In some cases, microorganisms were not detected on microarray 1 (e.g., *E. coli*), but after redesign of the probes and inclusion of additional probes these microbes were detected on microarray 2. Similarly, multiple virus genes were detected for many targets. The cattle and sheep associated

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

capsid protein gene of norovirus was detected in cattle feces, swine feces, and poultry litter, while the human associated RNA dependent, RNA polymerase gene of norovirus was detected in raw human sewage and poultry litter. Nineteen antibiotic resistance genes for aminoglycosides, tetracycline, and beta-lactams, were detected in raw sewage, cattle and swine feces, and poultry litter (Table 3).

Table 3. Summary of the Number of Samples in Which Antibiotic Resistance Genes Were Detected via the Microarray in Various Fecal Samples

antibiotic---resistance gene	raw and diluted sewage (n = 4)[a]	cattle feces (n = 1)	swine feces (n = 1)	poultry litter (n = 1)
Aminoglycosides---*aac*(3)-III	2	0	1	1
Aminoglycosides---*aacC1*	1	0	0	0
Aminoglycosides---*aadE*	3	1	1	0
Beta-lactams---*bla* CMY-2	2	0	0	1
Beta-lactams---*bla* FOX-2	1	1	1	0
Beta-lactams---*bla* IMP-2	4	1	1	1
Beta-lactams---in wastewater	4	1	1	1
Beta-lactams---in wastewater	4	1	0	0
Tetracycline---*tetA-Aeromonas*	2	0	0	0
Tetracycline---*tetA-E. coli*	2	0	0	0
Tetracycline---*tetA-Shigella*	2	0	0	0
Tetracycline---*tetA*	4	1	1	1
Tetracycline---*tetB-Salmonella*	0	0	0	1
Tetracycline---*tetB*	2	0	1	1
Tetracycline---*tetC*	2	0	0	0
Tetracycline---*tetM*	0	0	1	1
Tetracycline---*tetO*	1	1	1	1
Tetracycline---*tetQ*	2	1	1	0
Tetracycline---*tetW*	4	0	0	0

[a] n indicates the number of samples tested for a particular class of ARG.

Correlations of Microarray Results with qPCR and Culture Based Methods. The microarray log FU for probes targeting *Enterococcus* spp. and *Bacteroidales* in sewage were linearly correlated with the microbial abundance estimated by qPCR, and to culture based densities of *Enterococcus* spp. (Figure 2). Microarray fluorescence for the *uidA* gene of *E. coli* followed the trend with microbial abundance estimated by qPCR, but was not significantly correlated. There was a statistically significant decline (linear regression, P = 0.023) in the microarray log fluorescence (Figure 3) detected in the dilution series of sewage samples (e.g., no dilution, 1:5, 1:10, and 1:50 dilution). Culture based counts of *E. faecalis* in the 1:50 dilution were 196 MPN (100 mL)$^{-1}$, which is below the discharge limits for fecal coliforms from the sewage treatment plant sampled.

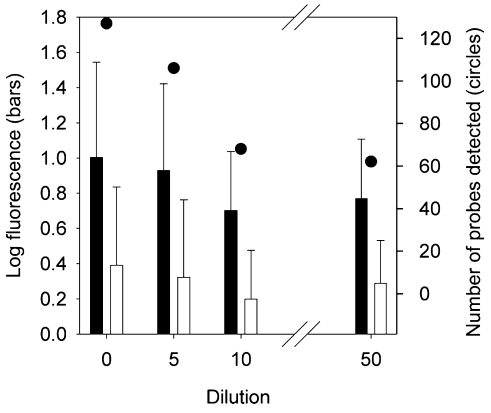


Figure 3. Mean ± SD of relative fluorescence of perfect match (■) and mismatch (□) probes in a dilution series of raw wastewater (0, 5, 10, and 50 to 1 dilution) and number of probes detected (●) on each microarray.

Comparison of Microarray and NGS Results. All fecal samples characterized by next-generation sequencing had sequence coverage estimations of ≥99.9%. The number of operational taxonomic units (OTUs) observed in the sewage, poultry litter, and cattle and swine feces were 404, 497, 1063, and 1102, respectively. Classified OTUs belonged to 27 phyla, which were dominated by members of the *Firmicutes*,

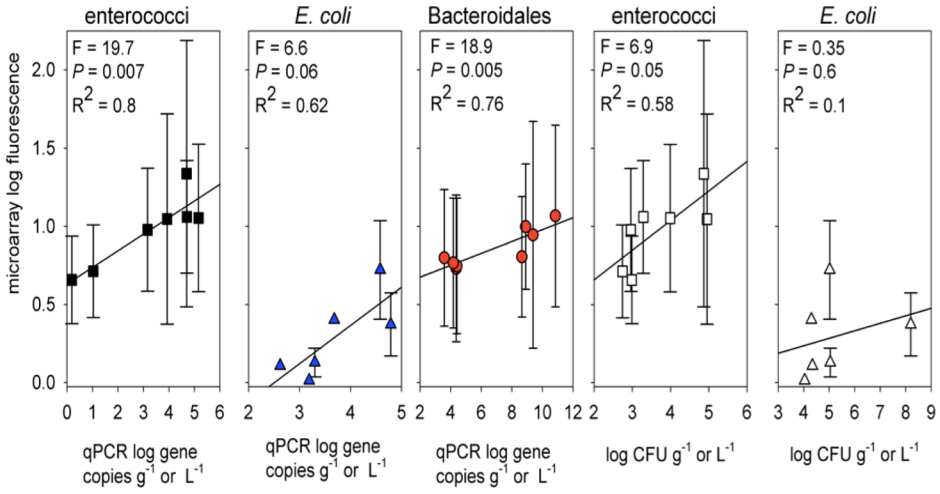


Figure 2. Correlation between microarray relative log fluorescence units (n = 6 perfect match probes) and microbial abundance determined via qPCR and culture methods. Average and standard deviation of microarray log fluorescence for multiple probes on the microarray targeting *Enterococcus spp.*, *E. coli*, and *Bacteroidales* versus the qPCR log gene copies per L or g based on *E. coli uidA* gene, *Bacteriodales* (Genbac) 16S rRNA, and *Enterococcus* 23S rRNA.

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

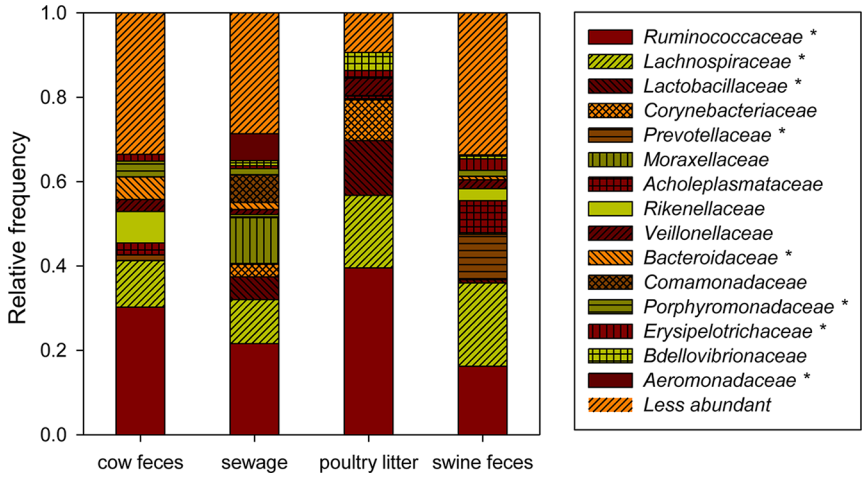


Figure 4. Relative abundance of families based on NGS of fecal, litter and wastewater samples. Families indicated with an asterisk are represented on the microarray.

Table 4. Predictive Accuracy of Microarray for the Microbial Source Tracking Markers (Average ± Standard Deviation)

sample type (number of samples)	percentage true-positive, sensitivity	positive predictive value	percentage false-positive	specificity	negative predictive value	host associated probes detected	host associated probes on array
raw wastewater (5)	21.1 ± 6.7	47.7 ± 7.7	9.0 ± 1.6	85.1 ± 2.5	62.1 ± 2.0	16 ± 5[a]	76
cattle feces (2)	26.7 ± 11.5	35.4 ± 2.9	10.5 ± 2.9	86.5 ± 3.6	80.7 ± 2.5	12 ± 5[b]	43
swine feces (2)	21.4 ± 6.1	26.8 ± 2.5	10.7 ± 3.9	87.0 ± 4.6	83.5 ± 1.1	8 ± 2[c]	35
poultry litter (2)	33.3 ± 2.1	49.5 ± 4.0	11.5 ± 1.0	82.8 ± 1.3	71.1 ± 0.7	22 ± 1[d]	66
avian feces (1)	30.3	62.5	6.5	90.0	70.1	20[e]	66

[a]Human or animal feces associated markers including: hexon gene of human adenovirus and *RdRp* gene of human associated norovirus[44]; [b]Cattle, animal or ruminant feces associated markers including: capsid protein gene of cattle associated norovirus,[44] ruminant associated *nihf* gene of *Methanobrevibacter ruminantium*,[45] and 16S rRNA gene of the order *Bacteroidales* (i.e., Rum-2- Bac associated)[46]; [c]Swine or animal feces associated markers including: two hexon genes of swine associated adenoviruses[44]; [d]Poultry litter, avian feces or animal feces associated markers including: 16S rRNA genes of *Catellicoccus marimammalium*,[47,48] *Brevibacterium* sp. LA35,[49] and *Helicobacter* spp.[47] and mitochondrial DNA associated with chickens, gulls, ostriches and pigeons[50]; [e]Poultry litter, avian feces or animal feces associated markers including: 16S rRNA genes of *Catellicoccus marimammalium*,[47,48] *Brevibacterium* sp. LA35,[49] and *Helicobacter* spp.[47] and mitochondrial DNA associated with cormorants, duck, geese, and quail[50].

Proteobacteria, *Bacteroidetes*, *Tenericutes*, and *Actinobacteria*. The result is consistent with findings from a recent PhyloChip study,[24] although *Tenericutes* was also found in this study which was not identified previously. These phyla accounted for 97 ± 2% of sequences among fecal types, and four are represented by probes on the microarray. Similarly, the 15 most highly represented families accounted for 74 ± 12% of the sequences among sample types, although only eight of these families are represented by probes on the microarray (Figure 4). Diversity was highest in the swine feces and lowest in the poultry litter based on Shannon indices of 4.14 (raw sewage), 3.34 (poultry litter), 4.53 (cattle feces), and 5.10 (swine feces). The abundance based coverage estimators, (i.e., a nonparametric estimator of the number of OTUs in the original sample) were 1190.2 (raw sewage), 1254.4 (poultry litter), 1962.0 (cattle feces), and 1921.5 (swine feces).

Potential for MST with the Microarray. The sensitivity and specificity of the microarray for microbial source tracking is shown in Table 4, along with the positive and negative predictive values. The sensitivity of the microarray was fairly low, ranging from 21 to 33% for all sample types. In contrast, the microarray MST markers were found to be quite specific with calculated specificities ranging from 85 to 90%. Additionally, the false positive percentage of the microarray was low, ranging from 6.5 to 11.5%. The number of MST probes detected associated with various target organisms[46−52] compared to the total number of MST probes by host are also indicated in Table 4. Commonly detected genes among all samples included animals feces associated bacteriophages of *Bacteroides fragilis*,[53] and the 16S rRNA genes of *Bacteroides thetaiotaomicron*, *Bacteroides vulgatus*, *Bacteroides distasonis*, *Bifidobacterium adolescentis*, *Rhodococcus coprophilus*,[12] all enteric bacteria, all Gram-negative and Gram-positive bacteria, and *Bacteria*.

The percent similarity (SIMPER) between replicate sample types on the microarray were estimated to be 78.1, 75.7, 73.9, and 73.4 for the raw sewage, cattle feces, swine feces and poultry litter samples, respectively. In total, the number of probes detected on replicate samples taken months apart that were similar (and differed) were 98 similar (34 differed) for raw sewage; 76 similar (35 differed) for the cattle feces; 126 similar (46 differed) in the poultry litter; and 74 similar (36 differed) in the swine fecal sample. Further, NMDS analysis was used to assess clustering of like samples between microarrays using both a subset of the probes (just 16S rRNA corresponding to the same families detected via NGS) and all the probes on the microarrays (Figure 5). Replicate samples clustered together regardless of whether all probes on the microarray (n = 422, Figure 5B) were used for clustering or the rRNA genes alone were used (n = 116, Figure 5D), suggesting a similarity between like samples analyzed months apart and reproducibility of the method. Only the cattle and swine feces were found to cluster

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

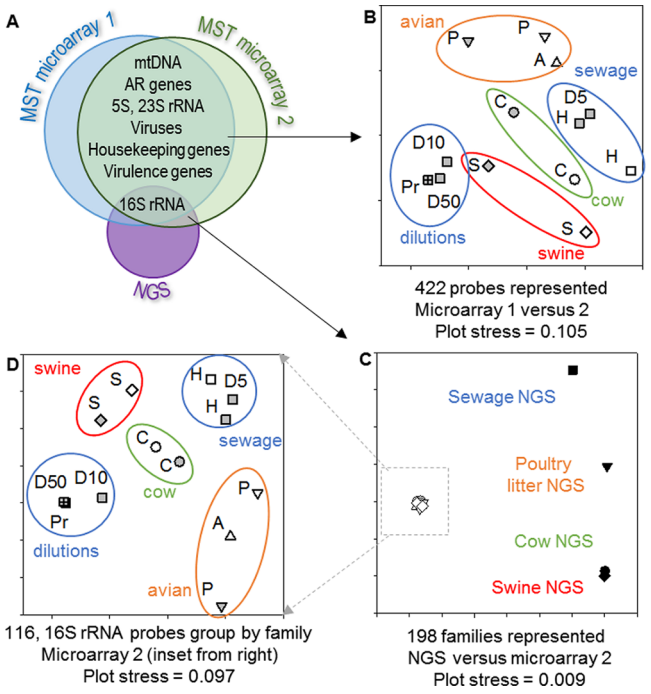


Figure 5. A. Overview of separation of fecal sources by NMDS of microarray (B and D) and NGS (C) data. Venn diagram shows gene categories represented on the microarrays and NGS. Microarray 1 and 2 results are in white and gray symbols, respectively, whereas NGS results are in black symbols. P = poultry litter; Pr = spinach contaminated with sewage; C = cattle; S = swine; H = human sewage; A = avian; D5, D10, and D50 = human sewage diluted 5:1, 10:1, and 50:1.

together based on NGS of the fecal samples (Figure 5C), suggesting very dissimilar microbial communities in the human and poultry litter samples as compared to the cattle and swine feces.

DISCUSSION

Microarrays have emerged as viable platforms for pathogen detection in clinical settings,[20,52] in food[47,49] and in environmental applications.[11,14,54,55] Microarrays represent a more effective method for detection of pathogens in samples as compared to multiplex qPCR or culture based methods. This microarray design draws on the extensive work conducted over the past 100 years[7] to develop culture based and molecular methods for detection of waterborne pathogens, fecal indicator bacteria and MST marker genes. Currently microarrays can be designed to rapidly query an environmental sample for only those pathogens most likely to cause risk to human health while simultaneously identifying potential sources of the fecal contamination. This offers a significant advantage over NGS or other rRNA based microarrays, which is currently more useful for exploring the diversity of organisms in fecal or water samples[56,57] rather than identifying human health risk and microbiological water quality.

While NGS methods are achieving longer reads with each generation of equipment and show great promise for detection of pathogens in the future,[58] microarrays currently offer a significant advantage over the phylogenetic resolution currently possible with relatively short NGS reads (e.g, 150 to 200 bp).[59] For example our microarray provided significant coverage of the major phyla found in the fecal and sewage samples. Furthermore, while NGS sequencing generally targets a single gene, we targeted different genes with the microarray (e.g, pathogen virulence genes, 16S rRNA, mtDNA), many of which are explicitly associated with virulence or particular pathogens. We were therefore able to detect the presence of multiple pathogens as well as the potential source of those fecal pathogens. Finally, the MST microarray is capable of targeting multiple viruses that do not have common gene targets such as 16S rRNA genes of the domain *Bacteria* which is the target of the commercially available Phylochip.

Previously it has been suggested that drawbacks to the use of microarrays for routine monitoring for pathogens include the lack of quantitative data regarding pathogen concentrations and the inability to offer low detection levels available by qPCR.[54] The results presented here and by others[60] show the validity of using WGA for uniformly amplifying nucleic acids of relatively scarce pathogens from environmental samples to levels detectable via the microarray. Further, a reproducible decrease in relative fluorescence of probes in serially diluted samples was observed here, suggesting that microarrays may be useful for semiquantitative analysis of gene concentrations in samples. In our studies we found a correlation between the length of the product hybridized to the array and relative fluorescence, which further supports our method for using 6, 7, and 8 bp restriction enzymes that did not digest within the 60 mer probes. Commonly recommended methods such as digestion with 4 or 5 bp restriction enzymes or heating to 95 °C, as recommended by the vendor,[61] could result in significantly shorter nucleic acids and greater potential for false negative results. Others have shown that the use of internal probes to adjust chip-to-chip fluorescence allows for correction of raw fluorescence intensities, and gives a linear correlation between spot intensities and target concentrations.[62]

The detection of free-living, nonpathogenic organisms such as cyanobacteria, *Naegleria gruberi* and *Nitrosococcus halophilus* in sewage and animal feces was unexpected. The amoeboflagellate, *Naegleria gruberi*, is widely present in freshwater and soil. *Nitrosococcus halophilus* is an ammonia-oxidizing bacterium found exclusively in saline environments. *Vibrio cholerae* and *V. parahemolyticus* are opportunistic human pathogens that are autochthonous in estuarine and marine environments. Potential reasons for false positive detection of these organisms in fecal samples include (1) errors in sampling methods (collection of soil during cattle feces sampling), (2) errors in the probe design, (3) insufficient hybridization and washing stringency, and (4) incomplete knowledge of the microbial ecology of these and closely related pathogens. To resolve issues with false positive and negatives, additional probes targeting these organisms will be added to future designs of the array. Further confirmation of the organisms in samples can be conducted by PCR, qPCR and culture based methods.

One limitation to the use of the MST microarray is the current knowledge of microbial ecology and scientific understanding of the specificity of microorganisms or genes to a particular host. The estimated sensitivity of the microarray is relatively low compared to previously published studies for qPCR and library dependent methods[6] although the results presented herein are comparable to those presented for the Phylochip.[22,23] However, the sensitivity calculation for the microarray are estimated differently than qPCR based methods. Specifically, sensitivity for the microarray is estimated as the number of correctly detected positive probes on the array divided by the number of probes on the array for that particular

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

host. Therefore, we find that on average 1 out of every 3 or 4 MST probes on the array associated with a particular host is correctly detected. It is likely that optimization of sample concentration methods prior to nucleic acid extraction can improve the sensitivity of the analysis.

Ideally, microbiological water quality assays would be able to not only detect pathogen presence in an environmental sample, but also determine the origin of the fecal pollution for guiding remediation efforts. As treated sewage is increasingly used to augment declining freshwater reserves, methods identifying pathogens and potential sources of fecal contamination will be highly valued. The MST microarray presented herein overcomes many of the drawbacks associated with culture and qPCR based methods currently used for water borne microbiological quality monitoring, and unlike NGS, is specialized to detect relatively scarce targets like pathogens and MST markers. Future work will evaluate the utility of the microarray for detecting pathogens in surface water samples and in reclaimed wastewater. Additionally the NGS results will be used to determine potential phyla not currently represented on the microarray that are potential candidates for inclusion as new probes in next microarray versions. For example, NGS showed that all four fecal samples contained sequences corresponding to the genus *Acholeplasma* in the phylum *Tenericutes*, but no corresponding probes in this phylum were used on the microarray.

Overall the results of this study show that

- Common waterborne pathogens (bacterial, eukaryotic and viral) can be detected via microarrays in fecal samples, raw sewage, poultry litter and contaminated produce,
- Microarray fluorescence may be correlated with qPCR and culture based enumeration of FIB,
- Microarray data reproducibly separated fecal material from different sources into different clusters.
- The current MST microarray contains probes accounting for four of the five most dominant phyla and eight of the 15 most dominant families found in fecal samples as determined via NGS.

■ ASSOCIATED CONTENT

Supporting Information

Table S1. qPCR primers and probes used in this study, Figure S1. Electrophoresis image of restriction enzyme digested WGA treated, DNA and cDNA isolated from sewage influent, Figure S2. Correlation between the location of restriction enzyme cut sites and log relative fluorescence of MM probes cut, Figure S3. Relative log fluorescence of PM probes detected on the microarrays compared to the number of adenosine nucleotides in the 60-mer probe sequences, Table S2. Microarray probe sequences, targeted microorganisms and genes, Table S3. Organisms and genes detected (1) and not detected (0) in the fecal samples on the microarray, "—"indicates the probes were not included on that microarray. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b00980.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 304-293-9952; e-mail: jennifer.weidhaas@mail.wvu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Duke Microarray Core facility (a Duke NCI Cancer Institute and a Duke Genomic and Computational Biology shared resource facility) for their technical support, microarray data management and feedback on the generation of the microarray data reported in this manuscript. Next-generation sequencing analysis was performed using the resources of the Minnesota Supercomputing Institute. Partial funding for this project was provided by a National Science Foundation grant to Jennifer Weidhaas (CBET 1234366) and the NSF ADVANCE IT Program (Award HRD-1007978). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

■ REFERENCES

(1) USEPA Water Quality Reporting Database (ATTAINS), US Environmental Protection Agency, http://www.epa.gov/waters/ir/ (accessed January 2015).

(2) Hlavsa, M. C.; Roberts, V. A.; Kahler, A. M.; Hilborn, E. D.; Wade, T. J.; Backer, L. C.; Yoder, J. S. Recreational water−associated disease outbreaks—United States, 2009−2010. *Morb. Mortal. Weekly Rep.* **2014**, *63* (1), 6−10.

(3) Jay-Russell, M. T.; Madigan, J. E.; Bengson, Y.; Madigan, S.; Hake, A. F.; Foley, J. E.; Byrne, B. A. Salmonella Oranienburg isolated from horses, wild turkeys and an edible home garden fertilized with raw horse manure. *Zoonoses Public Health* **2014**, *61* (1), 64−71.

(4) Zhao, H.; Li, X. Understanding the relationship between heavy metals in road-deposited sediments and washoff particles in urban stormwater using simulated rainfall. *J. Hazard Mater.* **2013**, *246*−*247*, 267−76.

(5) Lu, J.; Ryu, H.; Hill, S.; Schoen, M.; Ashbolt, N.; Edge, T. A.; Domingo, J. S. Distribution and potential significance of a gull fecal marker in urban coastal and riverine areas of southern Ontario, Canada. *Water Res.* **2011**, *45* (13), 3960−8.

(6) Stoeckel, D.; Harwood, V. Performance, design and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* **2007**, *73* (8), 2405−2415.

(7) Harwood, V. J.; Staley, C.; Badgley, B. D.; Borges, K.; Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **2014**, *38* (1), 1−40.

(8) 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), 3539−3552.

(9) Field, K.; Samadpour, M. Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res.* **2007**, *41* (16), 3517−3538.

(10) Berthet, N.; Dickinson, P.; Filliol, I.; Reinhardt, A.; Batejat, C.; Vallaeys, T.; Kong, K.; Davies, C.; Lee, W.; Zhang, S.; Turpaz, Y.; Heym, B.; Coralie, G.; Dacheux, L.; Burguiere, A.; Bourhy, H.; Old, I.; Manuguerra, J.; Cole, S.; Kennedy, G. Massively parallel pathogen identification using high-density microarrays. *Microbial Biotechnol.* **2008**, *1* (1), 79−86.

(11) Indest, K.; Betts, K.; Furey, J. Application of oligonucleotide microarrays for bacterial source tracking of environmental *Enterococcus* sp. Isolates. *Int. J. Env. Res. Public Health* **2005**, *2* (1), 175−185.

(12) Wang, R.-F.; Beggs, M. L.; Robertson, L. H.; Cerniglia, C. E. Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. *FEMS Microbiol. Lett.* **2002**, *213* (2), 175−182.

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

(13) Lee, D.-Y.; Shannon, K.; Beaudette, L. A. Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. *J. Microbiol. Methods* **2006**, *65* (3), 453−467.
(14) Lee, D.-Y.; Lauder, H.; Cruwys, H.; Falletta, P.; Beaudette, L. A. Development and application of an oligonucleotide microarray and real-time quantitative PCR for detection of wastewater bacterial pathogens. *Sci. Total Environ.* **2008**, *398* (1−3), 203−211.
(15) Rasooly, A.; Herold, K. Food microbial pathogen detection and analysis using DNA microarray technologies. *Foodborne Pathog. Dis.* **2008**, *5* (4), 531−550.
(16) Wu, C.; Valdes, J.; Bentley, W.; Sekowski, J. DNA microarray for discrimination between pathogenic O157:H7 EDL933 and non-pathogenic *Escherichia coli* strains. *Biosensors Bioelectron* **2003**, *19*, 1−8.
(17) Erlandsson, L.; Rosenstierne, M. W.; McLoughlin, K.; Jaing, C.; Fomsgaard, A. The microbial detection array combined with random Phi29-amplification used as a diagnostic tool for virus detection in clinical samples. *PLoS One* **2011**, *6* (8), e22631.
(18) Panicker, G.; Call, D. R.; Krug, M. J.; Bej, A. K. Detection of Pathogenic *Vibrio* spp. in shellfish by using multiplex PCR and DNA microarrays. *Appl. Environ. Microbiol.* **2004**, *70* (12), 7436−7444.
(19) Straub, T.; Daly, D.; Wunschel, S.; Rochelle, P.; DeLeon, R.; Chandler, D. Genotyping *Cryptosporidium parvum* with an *hsp70* single-nucleotide polymorphism microarray. *Appl. Environ. Microbiol.* **2002**, *67* (4), 1817−1826.
(20) Lee, Y. J.; van Nostrand, J. D.; Tu, Q.; Lu, Z.; Cheng, L.; Yuan, T.; Deng, Y.; Carter, M. Q.; He, Z.; Wu, L.; Yang, F.; Xu, J.; Zhou, J. The PathoChip, a functional gene array for assessing pathogenic properties of diverse microbial communities. *ISME J.* **2013**, *7* (10), 1974−84.
(21) Ishii, S.; Segawa, T.; Okabe, S. Simultaneous quantification of multiple food and waterborne pathogens by use of microfluidic quantitative PCR. *Appl. Environ. Microbiol.* **2013**, *79* (9), 2891−2898.
(22) Dubinsky, E. A.; Esmaili, L.; Hulls, J.; Cao, Y.; Griffith, J.; Andersen, G. L. Application of phylogenetic microarray analysis to discriminate sources of fecal pollution. *Environ. Sci. Technol.* **2012**, *46*, 4340−4347.
(23) Boehm, A. B.; Van De Werfhorst, L. C.; Griffith, J. F.; Holden, P. A.; Jay, J. A.; Shanks, O. C.; Wang, D.; Weisberg, S. B. Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res.* **2013**, *47* (18), 6812−6828.
(24) 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.
(25) *Method 1200: Analytical Protocol for Non-Typhoidal Salmonella in Drinking Water and Surface Water*; U.S. Environmental Protection Agency, 2012.
(26) *Method 1103.1: Escherichia coli (E. coli) in Water by Membrane Filtration Using membrane-Thermotolerant Escherichia coli Agar (mTEC)*; U.S. Environmental Protection Agency, 2010.
(27) *Method 1611: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay*, EPA-821-R-12-004; US Environmental Protection Agency, 2012.
(28) *Standard Methods for the Examination of Water and Wastewater*, 21st ed.; American Public Health Association, Inc.: Washington, DC, 2005.
(29) Griffiths, R. I.; Whiteley, A. S.; O'Donnell, A. G.; Bailey, M. J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **2000**, *66* (12), 5488−91.
(30) Weidhaas, J.; Garner, E.; Basden, T.; Harwood, V. J. Runoff studies demonstrate parallel transport behavior for a marker of poultry fecal contamination and *Staphylococcus aureus*. *J. Appl. Microbiol.* **2014**, *117* (2), 417−429.
(31) He, Z.; Gentry, T. J.; Schadt, C. W.; Wu, L.; Liebich, J.; Chong, S. C.; Huang, Z.; Wu, W.; Gu, B.; Jardine, P.; Criddle, C.; Zhou, J. GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* **2007**, *1* (1), 67−77.
(32) Deng, Y.; He, Z.; Van Nostrand, J. D.; Zhou, J. Design and analysis of mismatch probes for long oligonucleotide microarrays. *BMC Genomics* **2008**, *9*, 491.
(33) Weidhaas, J.; Mantha, S.; Hair, E.; Nayak, B.; Harwood, V. J. Evidence for extra-intestinal growth of *Bacteroidales* originating from poultry litter. *Appl. Environ. Microbiol.* **2015**, *81* (1), 196−202.
(34) Loffler, F. E.; Yan, J.; Ritalahti, K. M.; Adrian, L.; Edwards, E. A.; Konstantinidis, K. T.; Muller, J. A.; Fullerton, H.; Zinder, S. H.; Spormann, A. M. Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov. and family Dehalococcoidaceae fam. nov., within the phylum Chloroflexi. *Int. J. Syst. Evol Microbiol* **2013**, *63* (Pt 2), 625−35.
(35) *Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Enzymatic Labeling for Blood, Cells, or Tissues (with a High Throughput option) Protocol, Version 7.2*; Agilent, July 2012.
(36) Severgnini, M.; Cremonesi, P.; Consolandi, C.; Bellis, G. D.; Castiglioni, B. Advances in DNA microarray technology for the detection of foodborne pathogens. *Food Bioprocess Technol.* **2011**, *4* (6), 17.
(37) Brodie, E. L.; DeSantis, T. Z.; Joyner, D. C.; Baek, S. M.; Larsen, J. T.; Andersen, G. L.; Hazen, T. C.; Richardson, P. M.; Herman, D. J.; Tokunaga, T. K.; Wan, J. M.; Firestone, M. K. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl. Environ. Microbiol.* **2006**, *72* (9), 6288−6298.
(38) Claesson, M. J.; Wang, Q.; O'Sullivan, O.; Greene-Diniz, R.; Cole, J. R.; Ross, R. P.; O'Toole, P. W., Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **2010**, *38*, e200. DOI: 10.1093/nar/gkq873.
(39) Schloss, P. D.; Westcott, S. L.; Ryabin, T.; Hall, J. R.; Hartmann, M.; Hollister, E. B.; Lesniewski, R. A.; Oakley, B. B.; Parks, D. H.; Robinson, C. J.; Sahl, J. W.; Stres, B.; Thallinger, G. G.; Van Horn, D. J.; Weber, C. F. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75* (23), 7537−7541.
(40) Aronesty, E. Comparison of sequencing utility programs. *Open Bioinf. J.* **2013**, *7*, 1−8.
(41) 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.; Tiedje, J. M. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **2009**, *37* (suppl 1), D141−D145.
(42) Clarke, K. Non-parametric multivariate analysis of changes in community structure. *Aust. J. Ecol.* **1993**, *18*, 117−143.
(43) Harwood, V.; Wiggins, B.; Hagedorn, C.; Ellender, R.; Gooch, J.; Kern, J.; Samadpour, M.; Chapman, A.; Robinson, B.; Thompson, B. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *J. Water Health* **2003**, *1*, 153−166.
(44) Myoda, S. P.; Carson, C. A.; Fuhrmann, J. J.; Hahm, B.-K.; Hartel, P. G.; Yampara-Iquise, H.; Johnson, L.; Kuntz, R. L.; Nakatsu, C. H.; Sadowsky, M. J.; Samadpour, M. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. *J. Water Health* **2003**, *01.4*, 167−180.
(45) Hammer, O.; Harper, D.; Ryan, P. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **2001**, *4*(1), 9.
(46) Green, H.; Dick, L.; Gilpin, B.; Samadpour, M.; Field, K. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Appl. Environ. Microbiol.* **2012**, *78* (2), 503−510.

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329

(47) Kostić, T.; Stessl, B.; Wagner, M.; Sessitsch, A.; Bodrossy, L. Microbial diagnostic microarray for food- and water-borne pathogens. *Microb. Biotechnol.* **2010**, *3* (4), 444−454.
(48) Lu, J.; Santo Domingo, J.; Lamendella, R.; Edge, T.; Hill, S. Phylogenetic Diversity and Molecular Detection of Bacteria in Gull Feces. *Appl. Environ. Microbiol.* **2008**, *74* (13), 3969−3976.
(49) Sergeev, N.; Distler, M.; Courtney, S.; Al-Khaldi, S. F.; Volokhov, D.; Chizhikov, V.; Rasooly, A. Multipathogen oligonucleotide microarray for environmental and biodefense applications. *Biosens. Bioelectron.* **2004**, *20* (4), 684−698.
(50) Vuong, N.-M.; Villemur, R.; Payment, P.; Brousseau, R.; Topp, E.; Masson, L. Fecal source tracking in water using a mitochondrial DNA microarray. *Water Res.* **2013**, *47* (1), 16−30.
(51) Weidhaas, J.; Macbeth, T.; Olsen, R.; Sadowsky, M.; Norat, D.; Harwood, V. Identification of a *Brevibacterium* marker gene specific to poultry litter and development of a quantitative PCR assay. *J. Appl. Microbiol.* **2010**, *309* (1), 334−347.
(52) Yoo, S. M.; Choi, J.; Lee, S.; Yoo, N. Applications of DNA microarray in disease diagnostics. *J. Ind. Microbiol. Biotechnol.* **2009**, *19* (7), 635−646.
(53) Wolf, S.; Hewitt, J.; Greening, G. E. Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. *Appl. Environ. Microbiol.* **2010**, *76* (5), 1388−1394.
(54) Aw, T. G.; Rose, J. B. Detection of pathogens in water: From phylochips to qPCR to pyrosequencing. *Curr. Opin. Biotechnol.* **2012**, *23* (3), 422−430.
(55) 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 Res.* **2013**, *47* (18), 6862−6872.
(56) Cai, L.; Zhang, T. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environ. Sci. Technol.* **2013**, *47* (10), 5433−5441.
(57) Staley, C.; Unno, T.; Gould, T. J.; Jarvis, B.; Phillips, J.; Cotner, J. B.; Sadowsky, M. J. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J. Appl. Microbiol.* **2013**, *115* (5), 1147−1158.
(58) Gomi, R.; Matsuda, T.; Matsui, Y.; Yoneda, M. Fecal source tracking in water by next-generation sequencing technologies using host-specific *Escherichia coli* genetic markers. *Environ. Sci. Technol.* **2014**, *48* (16), 9616−9623.
(59) Mizrahi-Man, O.; Davenport, E. R.; Gilad, Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: Evaluation of effective study designs. *PLoS One* **2013**, *8* (1), e53608.
(60) Wu, L.; Liu, X.; Schadt, C. W.; Zhou, J. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl. Environ. Microbiol.* **2006**, *72* (7), 4931−4941.
(61) *Oligonucleotide Array-Based CGH for Genomic DNA Analysis, Protocol ver 7.3*; Agilent Technologies, Santa Clara, CA, 2014.
(62) Hwang, B. H.; Cha, H. J. Quantitative oligonucleotide microarray data analysis with an artificial standard probe strategy. *Biosensors Bioelectron.* **2008**, *23* (11), 1738−1744.

DOI: 10.1021/acs.est.5b00980
Environ. Sci. Technol. 2015, 49, 7319−7329