Performance of forty-one microbial source tracking methods: A twenty-seven laboratory evaluation study

Alexandria B. Boehm[1], Laurie C. Van De Werfhorst[3,4], John F. Griffith[2], Patricia A. Holden[3,4], Jenny A. Jay[5], Orin C. Shanks[6], Dan Wang[1] and Stephen B. Weisberg[2]

Abstract

The last decade has seen development of numerous new microbial source tracking (MST) methodologies, but many of these have been tested in just a few laboratories with a limited number of fecal samples. This method evaluation study examined the specificity and sensitivity of 41 MST methodologies by analyzing data generated in 27 laboratories. MST methodologies targeting human, cow, ruminant, dog, gull, pig, horse, and sheep were tested against sewage, septage, human, cow, dog, deer, pig, chicken, pigeon, gull, horse, and goose fecal samples. Each laboratory received 64 blind samples containing a single source (singletons) or two sources (doubletons), as well as diluted singleton samples to assess method sensitivity. Laboratories utilized their own protocols when performing the methods and data were deposited in a central database before samples were unblinded. Between one and seven laboratories tested each method. The most sensitive and specific assays, based on an analysis of presence/absence of each marker in target and non-target fecal samples, were HF183 endpoint and HF183 SYBR (human), CF193 and Rum2Bac (ruminant), CowM2 and CowM3 (cow), BacCan (dog), Gull2SYBR and LeeSeaGull (gull), PF163 and pigmtDNA (pig), HoF597 (horse), PhyloChip (pig, horse, chicken, deer), Universal 16S TRFLP (deer), and Bacteroidales 16S TRFLP (pig, horse, chicken, deer); all had sensitivity and specificity higher than 80% in all or the majority of laboratories. When the abundance of MST markers in target and non-target fecal samples was examined, some assays that performed well in the binary analysis were found to not be sensitive enough as median concentrations fell below a minimum abundance criterion (set at 50 copies per colony forming units of enterococci) in target fecal samples. Similarly, some assays that cross-reacted with non-target fecal sources in the binary analysis were found to perform well in a quantitative analysis because the cross-reaction occurred at very low levels. Based on a quantitative analysis, the best performing methods were HF183Taqman and BacH (human), Rum2Bac and BacR (ruminant), LeeSeaGull (gull), and Pig2Bac (pig); no cow or dog-specific assay met the quantitative specificity and sensitivity criteria. Some of the best performing

[1]*Stanford University, Department of Civil and Environmental Engineering, Environmental and Water Studies, Stanford, CA*

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

[3]*Unviersity of California, Bren School of Environmental Science and Management, Santa Barbara, CA*

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

[5]*University of California, Department of Civil and Environmental Engineering, Los Angeles, CA*

[6]*US Environmental Protection Agency, Cincinnati, OH*

assays in the study were run by just one laboratory so further testing of assay portability are needed. While this study evaluated the marker performance in defined samples, further field testing as well as development of frameworks for fecal source allocation and risk assessment are needed.

Introduction

Beach water quality monitoring is based on measurement of fecal indicator bacteria (FIB), which are used as surrogates for human pathogens because they are relatively easy to measure and have been found to correlate with human health outcomes (Pruss 1998, Wade *et al.* 2003). However, FIB can originate from numerous pollution sources, such as human sewage, manure from livestock operations, wildlife, and urban runoff. In addition, non-fecal FIB sources have been well documented (Hardina and Fujioka 1991, Byappanahalli *et al.* 2003, Yamahara *et al.* 2007). Effective beach management requires knowledge of the dominant FIB sources and their potential influences on water quality. Source identification also allows prioritization of watersheds for remediation based on predicted human health risks; risks will differ depending on the host source from which the FIB originated (Soller *et al.* 2010).

Numerous microbiological source tracking (MST) methods intended to discriminate between human and non-human fecal sources have been developed, with some methods designed to differentiate among animal sources. The field was historically dominated by library-dependent methods that match genetic or phenotypic patterns of FIB isolates from a known source to that of isolates in an ambient sample. More recently, genetic markers associated with particular animal feces have gained favor because they do not require building costly isolate libraries, which have been found to be geographically (Wiggins *et al.* 2003, Ebdon and Taylor 2006) and temporally (Jenkins *et al.* 2003, Hansen *et al.* 2009) specific. Several other classes of methods, including viruses specific to human fecal material (Noble *et al.* 2003, Noble and McQuaig 2011), chemical (Hagedorn and Weisberg 2009), community-based (Cao *et al.* 2011a), and metagenomic methods (Unno *et al.* 2010), are also used.

A few large studies to assess efficacy of these methods have been conducted (Griffith *et al.* 2003, Stoeckel *et al.* 2004), but they were conducted prior to development of many library-independent methods. Methods developed since that time have been mostly evaluated within the research laboratories in which they were developed, making it difficult to assess their geographical stability (Stoeckel and Harwood 2007). Moreover, most evaluations have focused on a small number of candidate sources, limiting the ability to assess cross-reactivity that has been reported in some studies (Layton *et al.* 2006, Kildare *et al.* 2007, McLain *et al.* 2009, Van De Werfhorst *et al.* 2011). Studies that have investigated a large number of candidate sources (Shanks *et al*. 2010a,b) were performed in a single laboratory, resulting in no information on potential influence of inter-laboratory variability.

The following provides the study design and broad overview of results from a recent study (the Source Identification Protocol Project, SIPP) in which 41 MST methods implemented by 27 laboratories (Table 1) were challenged with 12 possible fecal sources in 64 blind samples.

Methods

A global call for participating laboratories was distributed by email. All laboratories that indicated they would like to participate were accommodated; this included those who wished to test newly developed assays as well as older assays gaining popularity in the MST field. The methods whereby 64 challenge samples were created and distributed to the 27 participating laboratories are outlined below. Participating laboratories were aware of the types of feces and challenge samples (singletons and doubletons) included in the study, but challenge samples received by the laboratories were labeled with only a number; thus they were blinded. Depending on the needs of the laboratories, challenge samples were delivered as filters or in liquid form. Laboratories that performed viral analyses required unique processing of the sample and filter. Participating laboratories assayed the samples using a variety of methodologies including culture-dependent assays for bacteriophage, culture-independent assays for specific genes, and molecular community profiling methodologies. The number of laboratories testing each assay varied from one to seven; all assays, regardless of the number of laboratories running them, were included in the analysis although it is acknowledged that assays performed by just one laboratory should be further tested. In many cases, assay developers participated in the study. While comparison of the developer's results with the other laboratories' results might be

Table 1. List of participating laboratories.

Principle Investigator	Affiliation
C. Sinigalliano	National Oceanic and Atmospheric Administration
J. Lee	Ohio State Univ.
W. Meijer	Univ. College Dublin
J. Rose	Michigan State Univ.
M. Byappanahalli	U. S. Geological Survey
J. Stewart	Univ. North Carolina
M. Sadowsky	Univ. Minnesota
J. Ebdon & H. Taylor	Univ. Brighton
S. Wuertz	Univ. California Davis
J. Jay	Univ. California Los Angeles
R. Noble	Univ. North Carolina
S. Reynolds	Environmental Canine Services LLC
V. Kannappan & D. Kashian	Wayne State Univ. / Ottowa County
J. Griffith	Southern California Coastal Water Research Project
M. Gourmelon	French Research Institute for Exploration of the Sea
T. Fong	TetraTech
K. Goodwin	National Oceanic and Atmospheric Administration
A. Farnleitner	Vienna University of Technology
J. Santo Domingo	U.S. Environmental Protection Agency
D. Diston & M. Wicki	Federal Office of Public Health, Switzerland
J. Fuhrman	Univ. Southern California
A. Boehm	Stanford
O. Shanks	U.S. Environmental Protection Agency
P. Holden	Univ. California Santa Barbara
R. Rodrigues & J. Brandão	National Institute of Health, Portugal
T. Madi	Source Molecular
G. Andersen	Lawrence Berkeley National Laboratory

informative, such comparisons were not an aim of the present analysis.

Composite challenge samples were created from freshly collected fecal material from 12 sources relevant to California beaches: individual humans, sewage, septage, horse, cattle, deer, pigs, geese, chicken, pigeon, sea gull, and dogs. A minimum of 12 individual samples were collected for each animal type, except for sewage (untreated influent from 9 treatment facilities) and septage (6 septage collection trucks or community systems). An approximately equal number of each fecal sample type was collected from four California (CA) geographies: central CA, Los Angeles County, Orange County, and San Diego County. In all cases, samples were collected fresh (i.e., wet and recently deposited), except in the case of feral deer in which case feces were collected even if they appeared aged.

All individual fecal samples were collected in 50 ml polypropylene culture tubes using sterile technique and stored at 4°C until used to create composite challenge samples (between 1 and 3 days). First, composite slurries were created for each of the 12 source types as follows. Between 1 and 55 g of feces, 75 ml (septage) or 100 ml (sewage) from each individual fecal sample were added to between 1 and 5 L of 0.2 µm pore size filtered (Polycap 36 AS, Whatman) artificial freshwater (distilled water with 0.3 mM $MgCl_2$, 0.6 mM $CaCl_2$, and 1.4 mM $NaHCO_3$), and was blended using a Waring (Torrington, CT) 700S blender run at 12000 RPM and screened through 300 µm nylon mesh to remove large debris. The total amount of fecal material added to the artificial freshwater was that estimated to yield concentrations of approximately 2000 colony forming units (CFU) enterococci per 100 ml based on information gleaned from the literature and pilot studies on fecal enterococci concentrations.

Composite challenge samples contained either a single fecal type (singletons) or two fecal types

(doubletons; Table 2). To create the doubleton composite slurries, 90 and 10% (by volume) of the appropriate singleton composite slurries were mixed in a sterile carboy.

Once all the composite slurries were made, they were assigned a number by personnel not involved in the study and all other identifying markers were removed from the carboys. The newly labeled carboys were returned to researchers who further processed the samples for study participants.

Fifty ml of the blinded, composite challenge slurries were filtered to create the composite challenge samples for distribution, except for the few methods that required a liquid sample. Some of the challenge samples included singletons at 1:10 strength to assess method sensitivity, which were created by filtering only 5 ml of the slurries.

All filtrations used sterile disposable filtration devices (Thermo Scientific, Logan, UT) and 47 mm diameter, 0.4 µm pore size polycarbonate membrane filters (Isopore, Millipore, Billerica, MA) with the exception of filtrations for mammalian virus analyses, which used 47 mm diameter, 0.45-µm pore size mixed cellulose ester membrane filters (HA, Millipore) adjusted to pH 3.5 with 20% HCl or the addition of $MgCl_2$ to 0.1 M in some cases (see Supplemental Information (SI) Table SI-1; ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_409_431SI.pdf; Harwood *et al*. 2013). Filters were placed in cryotubes, flash frozen in liquid nitrogen, and stored at -80°C until shipment to participating labs. Bacteriophage were enumerated in unfiltered liquid challenge samples to which 10% glycerol (IBI Scientific, Peosta, IA) was added. Canine scent tracking was carried out with 50 ml unfiltered liquid samples. All filters were shipped on dry ice; liquid samples were shipped on blue ice. All challenge samples were created in duplicate for a total of 64 challenge samples.

Table 2. Challenge samples. Each sample was provided to participating laboratories in duplicate. A sample with '1:10' indicates a 1:10 dilution of the full strength singleton. The numbers after the doubleton name indicates the percent by volume, respectively, combined to create the sample. Further details on sample creation can be found in the method section.

Singletons	Doubletons
chicken	sewage/chicken 10/90
deer	sewage/gull 10/90
dog	sewage/gull 90/10
dog 1:10	sewage/pig 10/90
goose	sewage/pig 90/10
gull	human/cow 10/90
gull 1:10	human/cow 90/10
horse	human/dog 10/90
pig	human/dog 90/10
pig 1:10	human/goose 10/90
pigeon	human/gull 10/90
cow	human/gull 90/10
cow 1:10	septage/horse 10/90
human	
human 1:10	
septage	
septage 1:10	
sewage	
sewage 1:10	

MST Analysis

Analyses for 41 MST methods were carried out by 27 laboratories (Table SI-2). The following assays were tested: CowM2 and CowM3 (Shanks *et al.* 2008), BacCow (Kildare *et al.* 2007), BacR (Reischer *et al.* 2006), Rum2Bac (Mieszkin *et al.* 2010), CF128 and CF193 (Bernhard and Field 2000), Gull2SYBR (Lu *et al.* 2008), Gull2Taqman (Shibata *et al.* 2010, Sinigalliano *et al.* 2010), LeeSeaGullSpecific (Lee *et al*. 2013), Gull2Endpt (Lu *et al.* 2008), Pig2Bac (Mieszkin *et al.* 2010), PF163 (Dick *et al.* 2005), PigmtDNA (Martellini *et al.* 2005), DogBac (Shibata *et al.* 2010), BacCan (Kildare *et al.* 2007), HoF597 (Dick *et al.* 2005), Omvito (Martellini *et al.* 2005), HF183SYBR (Seurinck *et al.* 2005), HF183Taqman (Haugland *et al.* 2010), BacHum (Kildare *et al.* 2007), HumM2 (Shanks *et al.* 2009), BsteriF1 (Haugland *et al.* 2010), *nifH* (Johnston *et al.* 2010), BacH (Reischer *et al.* 2007), gyrB (Lee and Lee 2010), Btheta (Yampara-Iquise *et al.* 2008), GB124 phage (Ebdon *et al.* 2007), HB73 phage (Vijayavel *et al.* 2010), F-specific coliphage (Gourmelon *et al.* 2007), HF183Endpt (Bernhard and Field 2000), enterovirus (DeLeon *et al.* 1990, Donaldson *et al.* 2002, Fuhrman *et al.* 2005, Gregory *et al.* 2006, Walters *et al.* 2009), adenovirus (Xu *et al.* 2000, Jothikumar *et al.* 2005), norovirus I and II (Jothikumar *et al.* 2005, Da Silva *et al.* 2007), polyomavirus (Aksamit 1993, McQuaig *et al.* 2009), MB55 phage (Vijayavel *et al.* 2010), canine scent tracking (Murray 2011), Phylochip (Dubinsky *et al.* 2012), and Bacteroidales

and general terminal restriction fragment length polymorphism (Cao *et al*. 2013 b).

The vast majority of the analyses measured molecular markers by PCR or quantitative real-time PCR (QPCR), and the general methods used are described below (full details are provided in Table SI-2). DNA or RNA was extracted using commercial kits and then quantified (by most laboratories). Conventional PCR or QPCR was carried out using previously published protocols. Standard curves were created with genomic DNA or RNA, plasmid DNA, or *in vitro* transcribed RNA, as indicated in Table SI-2. Quantification thresholds (C_q) were translated into copy numbers using single or master standard calibration models (Sivaganesan *et al.* 2010). Copy numbers per filter were reported to a central data depository where results were compiled. Samples were unblinded once all data were received. For QPCR assays, most laboratories reported a lower limit of quantification (LLOQ). The C_q values for challenge samples that were above that of the LLOQ were either quantified as concentrations below the LLOQ, reported as 'not detected' (ND), or 'detected but not quantified' (DNQ), depending on laboratory (Table SI-2).

Community-Based Analysis

Three community-based methods (PhyloChip, Universal 16S terminal restriction fragment length polymorphism (TRFLP), Bacteroidales 16S TRFLP) were tested (Table S2). PhyloChip was run by one laboratory, and the two TRFLP methods by two laboratories. The method details are described elsewhere, along with additional methods and data analysis approaches not reported here, including next generation (Illumina) sequencing, combined Universal & Bacteroidales TRFLP, and PhyloChip using a probe-based analysis approach (Cao *et al*. 2013 a,b). In brief, a set of 12 reference samples, 1 for each fecal source type, were created at the same time as the 64 blinded challenge samples using the same fecal material. Two hundred ml of full strength singleton challenge slurries were filtered through membrane filters to create the samples. These reference samples were processed and analyzed at the same time as the 64 challenge samples for all of the community-based methods. Using multivariate techniques, the results from the unknown samples were compared to the reference samples to determine the sources likely present (Cao *et al.* 2011b, Dubinsky *et al.* 2012).

F-Specific Bacteriophage

F-specific bacteriophage were cultured from samples (International Organization for Standardization 1995). Plaques were picked and cultivated in media with and without RNase (Gourmelon *et al.* 2007). FRNA phage isolates where then subsequently genotyped into genogroup I or II using reverse transcription-PCR (Ogorzaly *et al.* 2009). Samples were designated as 'inconclusive', 'human', 'animal', or 'both' as follows. If less than five plaques could be genotyped from a sample, the result was deemed inconclusive. Otherwise, a sample was classified as 'human', 'animal', or 'both' if more than 20% of genotyped isolates corresponded to genogroup II and less than 20% corresponded to genogroup I, if more than 20% of genotyped isolates corresponded to genogroup I and less than 20% to group II, or if more than 20% of genotyped isolates corresponded to genogroup II and more than 20% of genotyped isolates corresponded to genogroup I, respectively.

Amount of Feces per Filter

The amount of feces in each sample was estimated using four different metrics: enterococci measured by membrane filtration (ENT-MF), enterococci measured by Enterolert (ENT-MPN), *Bacteroidales* as measured by the general Bacteroidales marker (genbac3), and mass of DNA (DNA).

The concentration of ENT-MF in each challenge singleton slurry (C_n) was determined using EPA Method 1600 (USEPA 2006). Between two and three dilutions were filtered for enumeration. All countable filters (filters with between 1 and 250 CFU) were assigned a number of CFU. If more than one dilution was countable, the counts were averaged to estimate C_n in units of CFU per 100 ml. In the case of the horse singleton, a single plate with an estimated 600 CFU was used to estimate the concentration.

The concentration of ENT-MPN in each singleton slurry was determined using Enterolert (IDEXX, Westbrook, ME) defined substrate assays. For each challenge singleton slurry, between two and three volumes (between 10 and 0.0001 ml) were added to molecular grade water (Hardy Diagnostics, Santa Maria, CA) to a final volume of 100 ml and processed according to manufacturer's directions using Quanti-tray 2000s (IDEXX). Concentrations from all trays that yielded more than one well and less than all 97 wells positive were averaged to estimate C_n in units of MPN per100 ml.

The concentration of genbac3 was estimated by three laboratories using EPA method B (USEPA 2010), with a geometric mean of values among laboratories calculated to estimate C_n in units of copies per 100 ml. The concentration of DNA was estimated by two laboratories using Quant-iT kits (dsDNA High-Sensitivity or dsDNA Broad-Range; Invitrogen-Molecular Probes, Grand Island, NY) applied to 50 ml of slurry filtered through a membrane filter and using Generite DNA extraction kits. This concentration was multiplied by the exact volume of extract and normalized to 100 ml of slurry.

The concentrations were used to calculate ENT-MF, ENT-MPN, genbac3, and ng DNA per filter for full strength singleton (C_n *0.5), for 1:10 strength singleton (C_n *0.05), and doubleton (C_n *0.45 + C_m *0.05), where C_n and C_m are the concentrations of the full strength singletons used to make up 90 and 10% of the doubleton by volume, respectively) challenge samples.

Data Analysis

MST data submitted by the laboratories were formatted for data consistency. For PCR analyses, results reported as 'absence' were considered 'negative', and all other results were considered 'positive'. For quantitative (QPCR and phage) data, results reported as 'negative', 'No C_q', '0' or less than a detection limit were replaced with 'ND', and results reported as 'detected', 'presence', 'positive', 'DNQ' (detected but not quantified), and 'BLOQ' (below limit of quantification) were replaced with 'DNQ'.

Data from each laboratory and assay were translated into positive/negative for each of 64 challenge samples to create a binary data set, with QPCR values reported as DNQ treated as positive. Using the binary data set, the sensitivity and specificity of each MST assay was determined using all challenge samples (singletons, 1:10 singletons, and doubletons). The MST assay result was considered correct if it detected its target when the intended host feces were present in the challenge sample, regardless of the amount of feces present, or it did not detect its target in samples that did not contain the host feces. The MST method result was considered incorrect if it detected its target when the intended host feces were not present in the challenge sample, or it did not detect its intended host feces when the feces were present in the challenge sample. Sewage and septage were treated as human sources even though they potentially could contain other animal sources.

Sensitivity (reported as a percent) was calculated as the number of challenge samples correctly identified as positive for the host feces divided by the total number of samples that contained the host feces:

$$\text{Sensitivity} = TP/(FN+TP) \qquad \textbf{Eq. 1}$$

where TP and FN are true positives and false negatives, respectively.

Specificity (%) was calculated as the number of challenge samples correctly identified as negative for the host feces divided by the total number of samples that did not contain the host feces:

$$\text{Specificity} = TN/(FP+TN) \qquad \textbf{Eq. 2}$$

where TN and FP are true negatives and false positives, respectively.

An assay was considered sensitive and specific if both these metrics exceeded 80% in all the laboratories that ran the assay (if number of labs (n) $\leq$3) or for more than half the laboratories that ran them (if n >3).

The community-based assays were also evaluated in the binary analysis. While the single indicator MST methods target only one particular source and report 'positive' or 'negative' for that particular source, community-based assays target all sources simultaneously and report the most probable source(s) for each unknown sample. At a minimum, the methods identified one dominant source for each unknown. To provide evaluation of the community-based assays parallel to the single indicator methods, the community-based results were translated into multiple binary data sets, one for each of the fecal source types (except for pigeon which could not be distinguished from gull feces in the calibration exercise (Cao *et al*. 2013 b).

The FRNA bacteriophage results were treated similarly to the community-based assays because the final results were not expressed as presence/absence or quantity of a single source, but instead listed 'human', 'animal', or 'both' as being present. To be included in the binary analysis, these results were translated into one binary data set based on the presence/absence of 'human' in the results.

Method performance was also assessed based on quantitative target abundance, but this analysis was limited to the singleton samples (full and 1:10 strength). MST genetic marker abundance was first

normalized by the amount of feces on the filter to give copy number or plaque forming unit (PFU) per CFU ENT-MF. ND values were assigned a value of 10^{-6} copies per CFU; DNQ values were assigned a value of 10^{-5} copies per CFU, values lower than all measured normalized abundances. A method was considered sensitive if the median number of copies per CFU exceeded 50 copies per CFU, which was selected as the lowest number that could be detected in an ambient water sample containing 100 CFU ENT per 100 ml (~ a recreation water quality standard) if 100 ml were filtered for MST analysis, nucleic acid extracted into 100 µl of eluant, and 2 µl of this is run as template in a QPCR assay with a lower limit of quantification (LLOQ) of 100 copies per reaction. This LLOQ value was conservatively and arbitrarily chosen as one that might be achieved by a local laboratory. Additionally, quantitative MST methods were deemed specific based on their level of cross-reactivity with non-target samples. For each MST method, the number of non-target fecal samples that had marker abundances on a per CFU ENT basis greater than or equal to the abundances measured in target fecal samples was determined; only measurements within the range of quantification were considered (i.e., samples yielding ND and DNQ in non-target and target samples were not considered in the comparisons). A method was deemed specific if this number was zero. This approach allows assays to be deemed specific even if low levels of cross reactivity occur.

Review of the data indicated evidence that one challenge sample filter received by a single laboratory (a chicken singleton filter) was contaminated with human feces. Each human MST genetic marker the laboratory measured (n = 5) was found in very high abundance on the filter; and similar results were not reported by any other laboratory or for the replicate chicken singleton filter. Thus, data from this filter were removed from the MST analysis, but not the community-based analysis as a different set of filters was used.

Blanks

Filtration blanks (FB) were created to assess the potential for cross-contamination during the creation of the challenge samples. These consisted of filters placed into filtration devices, doused with phosphate buffered saline or artificial fresh water, and carried through subsequent QPCR processing. Filtration blanks were run in replicate using ten different assays by five cooperating laboratories (Laboratories 1 through 5; Table SI-2). The 10 QPCR methods included: BacHum, BsteriF1, HF183Taqman, HumM2, BacCow, CowM2, DogBact, Gull2Taqman, Pig2Bac, and genbac3.

Individual laboratories ran their own processing controls. For PCR and QPCR methods, these consisted of no template controls and extraction blanks (sterile filters run through the nucleic-acid extraction process and then subjected to analysis). Processing control data were not submitted by the majority of participating laboratories. It was assumed that laboratories conducted their own quality control assessments as per standard practice prior to submitting the data, and thus processing control data will not be presented herein.

Inhibition

As for the process controls, it was assumed that laboratories ran their own inhibition tests and results from these were used during data QA/QC. One participating laboratory (Laboratory 3; Table SI-2) assessed inhibition for QPCR assays HF183Taqman, BacHum, BsteriF1, HumM2, CowM2, DogBact, and genbac3 using a 'spike and dilute' method (Cao *et al*. 2012). In brief, between 10^2 and 10^4 copies of target were spiked into DNA extracts from the 64 challenge samples, and undiluted and 1:10 diluted extracts were run in each QPCR assay in triplicate. If the difference in C_q values between the diluted and undiluted sample was less than 2.3, the samples were deemed 'inhibited'.

Results

Amount of Fecal Material on Challenge Filters

The amount of fecal material on the filters varied over several orders of magnitude regardless of how fecal material was measured (Figure 1). Concentrations of enterococci measured by membrane filtration and defined substrate methods (ENT-MF and ENT-MPN) ranged over five orders of magnitude and were well correlated (Pearson's r between $\log_{10}$-transformed concentrations = 0.83, p = 0.001, n = 12). Genbac3 concentrations ranged over seven orders of magnitude; genbac3 concentrations in full strength singletons did not covary significantly with any other measures. DNA concentrations on the filters ranged over two orders of magnitude and

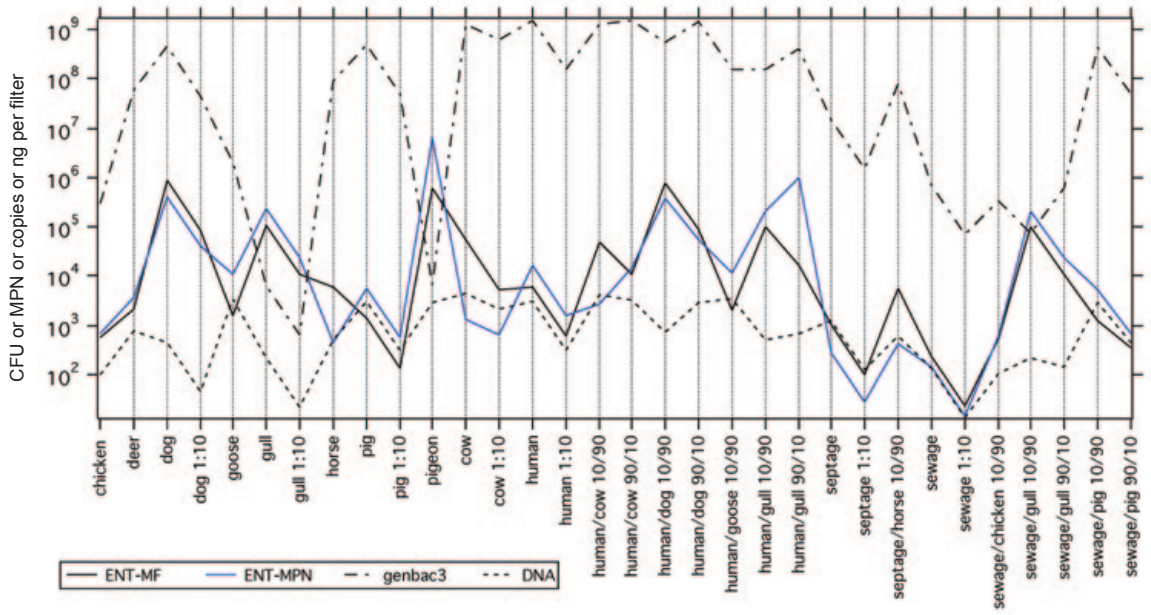

Figure 1. The amount of fecal material on each challenge filter as measured by enterococci measured by membrane filtration (ENT-MF), defined substrate (ENT-MPN), genbac3, and mass of DNA. See the Methods section and Table 2 for definitions of challenge filters shown on x-axis.

also did not correlate with any other metric of fecal material.

Based on ENT-MF, full strength samples (full strength singletons and doubletons) contained between 230 and 8.7 x 10^5 CFU per filter with a median of 6.0 x 10^3 CFU per filter. One-tenth strength singletons contained between 23 and 8.7x10^4 CFU per filter with a median of 600 CFU per filter. All filters, except for the 1:10 sewage and 1:10 septage had over 104 CFU per filter, the amount of enterococci that would remain on a filter if 100 ml of water out-of-compliance with enterococci standards had been filtered for routine monitoring.

Blanks

Filtration blanks (FB) provide checks on whether there was cross-contamination between challenge filters during their creation. Four-hundred fifty-five (455) FB reactions were run across the 10 assays (this counts each replicate QPCR as one reaction). Both a general indicator marker (genbac3) and nine MST markers were tested. A total of 99 reactions (22%) amplified (all C_q >33, median C_q = 38; Figure SI-1). Seventy four (74) of 112 genbac3 FB control reactions were positive, and 0 of 58 DogBact, 8 of 19 Gull2Taqman, 7 of 36 BacCow, 6 of 46 Pig2Bac, 3 of 63 HF183Taqman, 3 of 33 BsteriF1, 3 of 37 BacHum, and 2 of 40 CowM2 assays were positive. For the MST QPCR methods that targeted host-associated markers, C_q values were below the LLOQ for the laboratory running the sample with a few exceptions, and replicate reactions for the same FB were not positive except in one case. For the general indicator marker genbac3 which is generally present in high concentrations in all feces (Figure 1), some FB test reactions yielded concentrations near those measured for gull and pigeon singletons (Figure 1); these species are known to contain low levels of fecal *Bacteroidales* (Lu *et al.* 2008). Overall, the FB results indicated very limited cross-contamination with respect to MST genetic markers, and cross-contamination with genbac3 at low levels relative to the target samples.

Inhibition

To assess inhibition, every blinded sample extract was spiked with targets for the HF183Taqman, BacHum, BsteriF1, HumM2, CowM2, DogBac, and genbac3 QPCR assays and run undiluted and 1:10 diluted to assess for inhibition. In total, 64 samples were run for inhibition with 7 assays for 448 inhibition tests. Of these, only three inhibition tests suggested potential inhibition; for each of these tests, only one of three replicate reactions showed a change in C_q <2.3. Although nucleic-acid extraction methods varied across laboratories (Table SI-2), and not all assays were tested for inhibition, the results suggest that there was not a gross problem of inhibition in the study.

Binary Analysis

The binary analysis calculated the specificity and sensitivity of each MST method using Equations 1 and 2 while considering DNQ as a positive result. A method was considered sensitive and specific if both these metrics exceeded 80% in all the laboratories that ran the method (if number of labs (n) ≤3) or for more than half the laboratories that ran them (if n >3). Sensitive and specific methods for identifying each type of fecal sample are summarized in Table 3, along with an indication of whether the assay was tested by just a single laboratory.

Human Detection

Only two of the human assays were classified as both sensitive and specific using a cutoff of 80% for all or most laboratories: HF183 end point and the HF183SYBR assay (Figure 2; Table 3). Two of the seven laboratories running the HF183 endpoint method reported low sensitivity (50%) while the remaining five laboratories reported sensitivities upward of 80%. One of the laboratories reporting low sensitivity did not detect the genetic marker in many of the samples containing sewage and septage, while the other laboratory with low sensitivity did not detect the genetic marker in samples containing all three types of 'human' waste. All seven laboratories reported specificities from 92 to 100%. The HF183 SYBR method was performed by four laboratories, and all reported high sensitivity (>87%). One laboratory reported low specificity (28%) and detected the genetic marker in all types of non-target host fecal material, while the other three laboratories reported specificity >85%.

Table 3. Summary of sensitive and specific assays based on binary analysis. * indicates run by one lab, ** indicates run by one laboratory which was the developer. These assays will require further testing.

Human	HF183 endpoint, HF183SYBR
Cow	CowM2, CowM3**
Ruminant	Rum2Bac, CF193*
Dog	BacCan**
Gull	Gull2SYBR, LeeSeaGull**
Pig	PF163, mtPigDNA**, PhyloChip**, Bac TRFLP
Horse	HoF597, PhyloChip**, Bac TRFLP
Chicken	PhyloChip**
Deer	PhyloChip**, Bac TRFLP, Univ TRFLP
Goose	-

Other Bacteroidales genetic markers that target the 16S rRNA operon (BacHum, BsteriF1, HF183Taqman, Btheta, BacH) or functional genes (HumM2) performed at high sensitivities, but these methods also exhibited cross-reactivity, which prevented their classification as specific (Figure 2). This cross-reactivity was not limited to one particular type of non-human fecal source. The gyrB method that targets a functional gene in *Bacteroidales* was specific, but not sensitive because it was not detected in some samples with sewage and septage.

The *nifH* method targets a functional gene in an enteric archaeon and was run by five laboratories. This assay did not perform with both specificity and sensitivity over 80% in any single laboratory; however, some individual laboratories reported either good sensitivity or specificity (Figure 2).

Several human viruses, Bacteroidales phage, and *Enterococcus faecium* phage were included in the study. The human virus methods were highly specific, but did not perform at high sensitivities (Figure 2); likely a result of the viruses being at low abundances in feces. Interestingly, enterovirus genomic RNA was detected in several pig samples by multiple laboratories. *Bacteroides* phage and F-specific phage typing did not exhibit good sensitivity and specificity for human fecal material, although specificity was greater than 80% for bacteriophage assays in all but one laboratory (Figure 2). Further analysis and discussion of these results can be found in Harwood *et al.* (2013).

Canine scent tracking was performed by two trained dogs. In this method, trained dogs are told to identify samples containing human waste. The dogs' ability to detect human waste over other types of waste met neither specific nor sensitive criteria of >80% (Figure 2).

None of the community-based assays met the 80% criteria for specific and sensitive detection of human fecal samples. Of the three microbial community profiling methods, the PhyloChip method performed the best with a sensitivity of 79% and specificity of 85% (Figure 2). Both TRFLP methods showed lower sensitivity than the PhyloChip, but showed good specificity (96 - 100%) for the detection of human fecal material in the challenge samples (Figure 2). Further analysis for each of the fecal source types, including additional community-based methods and data analysis approaches can be found in companion papers (Cao *et al*. 2013 a,b).

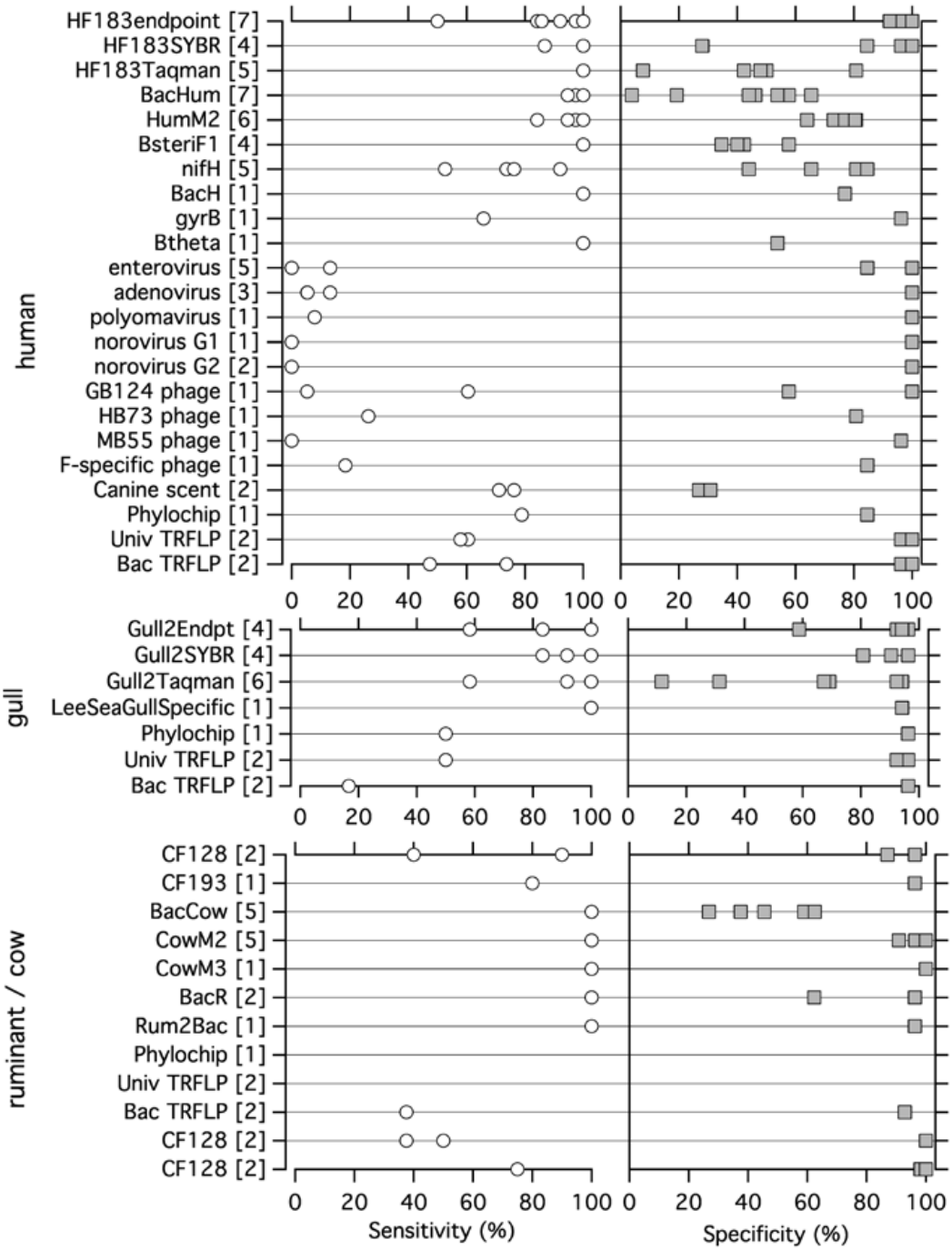

Figure 2. Sensitivity and specificity of human, gull, and ruminant/cow assays calculated from binary presence/absence data. Open circles are used for sensitivity, shaded squares are used for specificity. The number of laboratories that ran each assay is listed next to the assay name (left axis). Note that for some assays, results from two or more laboratories were identical so symbols may overlap.

Cow/Ruminant Detection

Of the methods deemed by their developers to be cow-associated (BacCow, CowM2, and CowM3) CowM2 and CowM3 performed with sensitivity and specificities meeting our criteria across all performing laboratories (Figure 2; Table 3). BacCow had high sensitivity, but poor specificity and exhibited cross-reactivity with all types of non-bovine feces (Figure 2). Of the community profiling methods, the Bacteroidales TRFLP performed the best with sensitivities and specificities by the two performing laboratories of 75 and >98%, respectively (Figure 2). However, no community method met the benchmark of 80% for sensitivity and specificity.

Of the methods described by their developers as ruminant-associated, CF193 and Rum2Bac were sensitive and specific as ruminant markers (Figure 2; Table 3). CF128 performed well in the hands of one laboratory (specificity and sensitivity >80%) but performed with low sensitivity in the other laboratory where it was not detected in a fraction of samples with cow and deer feces (Figure 2). Similarly, BacR met the specificity and sensitivity benchmarks in one laboratory, but in another, showed cross-reactivity with a range of non-target fecal samples (Figure 2).

Gull Detection

Of the gull methods, Gull2SYBR and LeeSeaGull performed with >80% sensitivity and specificity in all laboratories (Figure 2; Table 3). The Gull2 method generally showed cross-reactivity with the goose and pigeon fecal samples, while the LeeSeaGull method cross-reacted with pigeon. Implications of the cross reactivity with other birds is discussed by Sinigalliano *et al.* (2013). While the community profiling assays did not report good sensitivities to gull feces, all 3 assays were specific (>92%; Figure 2).

Pig Detection

PF163, PigmtDNA, PhyloChip, and Bacteroidales TRFLP all performed well with both sensitivities and specificities greater than 80% for all performing laboratories (Figure 3; Table 3). The QPCR method Pig2Bac showed high sensitivity, but reduced specificity (Figure 3), cross-reacting with dog and human feces, as well as septage.

Dog Detection

BacCan was performed by a single laboratory and was the only dog method that performed with sensitivity and specificity >80% (Figure 3; Table 3). DogBact had good sensitivity (>80%), but poor specificity in all but one laboratory; one laboratory reported specificity as low as 2% (Figure 3). The method showed cross-reactivity with all types of non-dog fecal samples. None of the three community profiling methods performed with sensitivity >80% for dog, but all had high specificity (>93%; Figure 3). Dog assay performance is further explored in a companion paper (Schriewer *et al*. 2013).

Horse Detection

HoF597 was performed by four laboratories and performed well in all but one laboratory, where the sensitivity was 50% because the genetic marker was not detected in one of two horse singleton and one of two horse/septage doubleton samples (Figure 3; Table 3). The PhyloChip and Bacteroidales TRFLP community methods also performed well, with sensitivities and specificities over 80%. The Universal TRFLP performed above the 80% in one of the two participating laboratories (Figure 3).

Sheep Detection

No sheep feces were included in our study. However, one laboratory performed the sheep-associated method, Omvito. This method did not detect its target in any of the challenge samples, indicating it potentially could be highly specific (Figure 3). Its sensitivity could not be evaluated.

Other Fecal Hosts

The ability of the three community profiling methods to detect chicken, goose, and deer feces were evaluated in the challenge samples; there were no individual gene-based assays included in the study that are specific for these feces. PhyloChip performed with sensitivities and specificities exceeding 80% for both chicken and deer feces (Figure 3). The two TRFLP methods (Universal 16S and Bacteroidales 16S) were sensitive and specific against deer feces (Figure 3; Table 3). None of the community methods detected goose feces based at the 80% sensitivity metric, but all reported high specificity (>93%; Figure 3).

Quantitative Analysis

The binary analysis does not take advantage of the quantitative nature of some of the methods. Methods with low levels of cross-reactivity are penalized in the same manner as methods with high levels of cross-reactivity in a binary analysis; however such methods may be useful for MST applications. Similarly, methods that detect highly abundant genes in target feces may be preferable over those that detect less abundant genes; yet these methods also cannot be identified from the results of a binary analysis.

The abundance of MST genetic markers in both target and non-target singleton challenge samples were examined to identify sensitive and specific

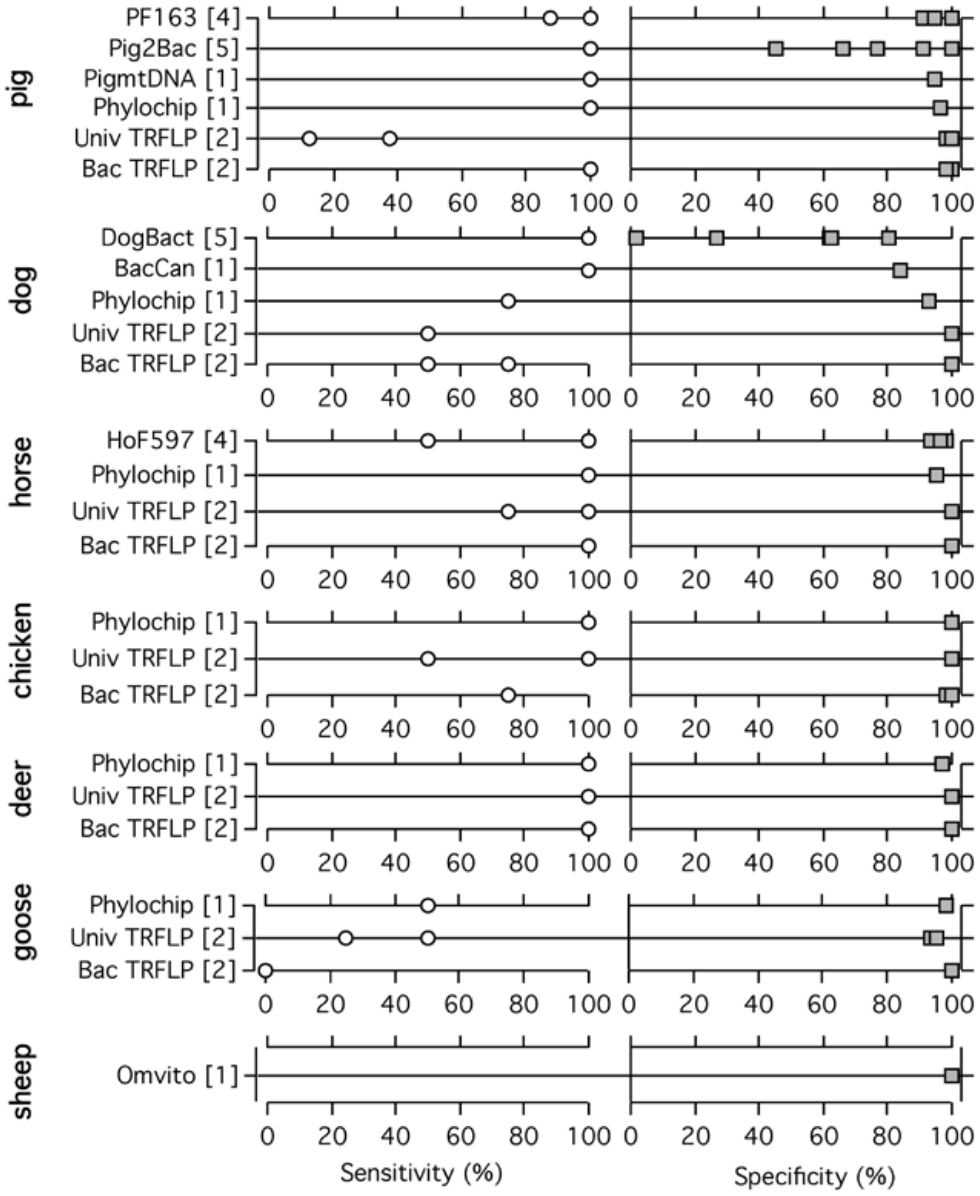

Figure 3. Sensitivity and specificity of pig, horse, chicken, deer, goose, and sheep assays calculated from binary presence/absence data. Open circles are used for sensitivity, shaded squares are used for specificity. The number of laboratories that ran each assay is listed in square brackets next to the assay name (left axis). Note that for some assays, results from two or more laboratories were identical and symbols may overlap.

quantitative MST methods. Because a different amount of feces was filtered onto each challenge filter (Figure 1), the MST method results were normalized to the amount of feces as measured by the number of CFU ENT. For completeness, some summary results are also presented using mass of DNA and copies of genbac3 as normalization factors. Even though some virus and phage methods were performed using quantitative techniques (Table SI-2), those results were not included in this analysis due to poor performance in the binary analysis (Figure 2).

Human Quantification

The abundance of human MST genes in target (i.e., human) challenge samples tended to be higher than in non-target (i.e., other animal) challenge samples (Figure 4; Table SI-3), but the difference in gene abundance between human and non-target samples varied by method. In the case of HF183Taqman, HumM2, BacH, gyrB, and Btheta, genetic marker abundance on a per CFU ENT basis was lower in non-target relative to the human challenge samples (Figure 4). For these methods,

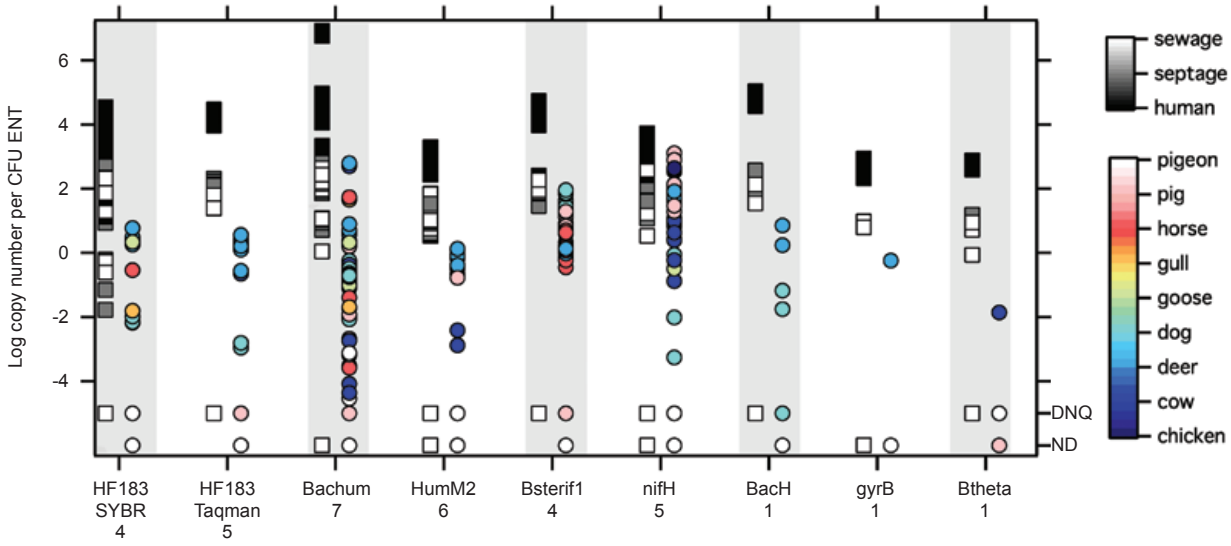

Figure 4. Human assays normalized by CFU ENT. The number below each assay is the number of performing laboratories. The number of copies per CFU ENT for quantitative human assay for the assay's target (squares) and nontarget (circles) fecal samples. Symbol color denotes specific fecal source. Results within the range of quantification are plotted against the left axes. Samples where the assay target was detected but not quantified (DNQ) and not detected (ND) are plotted against the right axes.

non-target samples did not amplify at levels higher than the human samples (Table 4). However, these human-associated methods gave ND or DNQ for some human challenge samples (Table SI-3) suggesting reduced sensitivity at times. The ND and DNQ in human samples occurred exclusively in sewage and septage samples and were most common for the 1:10 diluted sewage and septage samples. As shown in Figure 1, the 1:10 diluted septage and sewage samples had relatively small amounts of fecal material. In general, MST gene abundance was higher in the human fecal challenge samples, than in the septage and sewage samples (Figure 4).

HF183SYBR, BacHum, BsteriF1, and *nifH* measured genetic marker abundances at similar levels in target and some non-target fecal samples (Figure 4; Table 4) suggesting reduced specificity for these assays. BacHum had the greatest cross-reactivity with deer, BsteriF1 with dog, HF183SYBR with deer, and *nifH* with pig. However, these assays also gave many ND or DNQ for non-target samples (Table SI-3); over 50% of the non-target challenge samples tested returned results in these categories.

Most human-associated methods amplified less than 10% of the tested non-human fecal samples at levels on par with the human samples (Table 4). The exceptions were *nifH* and BacHum which cross-reacted with 19 and 11% of the non-target samples they were challenged with, respectively. The human methods that showed no cross-reactivity within the concentration range measured in human samples include BacH, Btheta, gyrB, HF183Taqman, and HumM2.

All of the methods showed good sensitivity for some or all of the target challenge samples in that the genetic markers were present at greater than 50 copies per CFU ENT (Figure 4). However, the lower genetic marker abundance sewage and septage forced the medians of genetic marker abundances in target samples to be lower than 50 copies per CFU ENT for *nifH*, Btheta, gyrB, and HumM2 (Table 5). Following the study criterion that a sensitive method should provide a median of 50 copies per CFU ENT for target samples, the following methods were sensitive: BacH, BacHum, BsteriF1, HF183SYBR, and HF183Taqman.

Cow/Ruminant Quantification

All cow/ruminant methods show limited cross-reactivity with the exception of BacCow (Figure 5c). The abundance of BacCow per CFU ENT was as high in all deer and some pig samples, as it was in cow challenge samples (Table 4). This suggests that this assay may be better represented as a ruminant method (Raith *et al*. 2013).

Table 4. The number (and percent in parentheses) of non-target challenge samples for which source-specific quantitative assays measured gene abundances (normalized per CFU ENT) at levels higher than those measured in the target samples in the range of quantification (ROQ). Results are broken down by fecal source, with the total and percent of non-target challenge samples that amplified in the ROQ shown in the bottom row. Top panel shows the human assays, bottom panel shows the remaining assays. Shaded cells indicate that these are the target challenge samples for the particular assay.

Sample	BacH	BacHum	BsteriF1	Btheta	gyrB	HF183SYBR	HF183Taqman	HumM2	*nif*H
chicken	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
cow	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	6 (30)
deer	0 (0)	12 (86)	0 (0)	0 (0)	0 (0)	5 (63)	0 (0)	0 (0)	4 (40)
dog	0 (0)	3 (11)	10 (63)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
goose	0 (0)	1 (7)	0 (0)	0 (0)	0 (0)	1 (13)	0 (0)	0 (0)	0 (0)
gull	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
horse	0 (0)	1 (7)	0 (0)	0 (0)	0 (0)	1 (13)	0 (0)	0 (0)	0 (0)
pig	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (70)
pigeon	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
human									
septage									
sewage									
Total	**0 (0)**	**20 (11)**	**10 (10)**	**0 (0)**	**0 (0)**	**7 (7)**	**0 (0)**	**0 (0)**	**24 (19)**

Sample	BacCow	CowM2	CowM3	BacR	Rum2Bac	Gull2SYBR	Gull2Taqman	LeeSeaGull	DogBact	BacCan	Pig2Bac
chicken	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (9)	0 (0)	4 (40)	0 (0)	0 (0)
cow						1 (6)	0 (0)	0 (0)	2 (10)	0 (0)	0 (0)
deer	10 (100)	0 (0)	0 (0)			1 (13)	0 (0)	0 (0)	1 (10)	0 (0)	0 (0)
dog	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			0 (0)
goose	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (38)	1 (8)	0 (0)	0 (0)	1 (50)	0 (0)
gull	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)				0 (0)	0 (0)	0 (0)
horse	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
pig	2 (10)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	3 (13)	0 (0)	6 (30)	0 (0)	
pigeon	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (63)	12 (100)	0 (0)	2 (20)	0 (0)	0 (0)
human	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	0 (0)	4 (20)	0 (0)	0 (0)
septage	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (13)	3 (13)	0 (0)	2 (10)	1 (25)	0 (0)
sewage	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	4 (20)	0 (0)	0 (0)
Total	**12 (7)**	**0 (0)**	**0 (0)**	**0 (0)**	**0 (0)**	**14 (10)**	**21 (10)**	**0 (0)**	**25 (15)**	**2 (6)**	**0 (0)**

All methods with the exception of BacCow showed good specificity (Table 4) in that no non-target samples had abundances on par with the target samples. Even for BacCow, only 7% of the tested non-targets cross-reacted at concentrations similar to those in the cow fecal samples and most of these non-target samples were deer feces.

In terms of the sensitivity metric, BacCow was the only cow-associated method with a median abundance greater than 50 copies per CFU ENT in target samples (Table 5). Both ruminant methods measured median abundance greater than 50 copies per CFU ENT for target challenge samples. Thus, BacCow, Rum2Bac, and BacR were deemed sensitive. CowM2 and CowM3 did not meet the sensitivity benchmark. They were each not detected in one cow sample (a 1:10 diluted challenge sample), and were generally found in relatively lower abundance on a per CFU basis (Figure 5c).

Gull Quantification

Gull2SYBR and Gull2Taqman showed cross-reactivity with non-target fecal samples, while the LeeSeaGull showed more limited cross-reactivity (Figure 5b; Table 4). Cross-reactivity, defined as abundance of marker on a per CFU basis in non-target fecal samples on par with that in gull feces, for the Gull2 assays occurred with all types of challenge samples, but was greatest for the pigeon, goose, septage, cow, pig, human, and deer feces, where abundance on a per CFU ENT basis was similar to the gull samples. The only samples that

Table 5. Performance of quantitative MST assays normalized by CFU ENT. The number and percent of non-target samples measuring copies per CFU in the range measured in target samples is provided in column labeled 'non-target in target range'; assays for which this is 0 are deemed specific. The median copies/CFU measured in target challenge samples is provided. The second number for the human assays (the number following the comma) shows the median if DNQs and NDs are not included in the median calculation. Assays with a median greater than 50 copies/CFU are deemed sensitive. Both sensitive and specific assays are indicated. An '*' indicates that the N would change to Y if the median that does not include DNQs and NDs is used. The number of laboratories (N) that ran the assay is provided. Assays run by a single laboratory will require further testing.

Assay	Host	N	# (%) Non-Target in Target Range	Median in Target (Copies/CFU)	Specific	Sensitive	Spec & Sens
BacH	human	1	0 (0)	87, 375	Y	Y	Y
BacHum	human	7	20 (11)	331, 374	N	Y	N
BsteriF1	human	4	10 (10)	123, 130	N	Y	N
Btheta	human	1	0 (0)	11, 13	Y	N	N
gyrB	human	1	0 (0)	0.003, 279	Y	N*	N*
HF183SYBR	human	4	7 (7)	52, 71	N	Y	N
HF183Taqman	human	5	0 (0)	138, 140	Y	Y	Y
HumM2	human	6	0 (0)	7, 48	Y	N	N
nifH	human	5	24 (19)	33, 167	N	N*	N
BacCow	cow	5	12 (7)	13490	N	Y	N
CowM2	cow	5	0 (0)	15	Y	N	N
CowM3	cow	1	0 (0)	1	Y	N	N
BacR	ruminant	2	0 (0)	955	Y	Y	Y
Rum2Bac	ruminant	1	0 (0)	832	Y	Y	Y
Gull2SYBR	gull	4	14 (10)	0.4	N	N	N
Gull2Taqman	gull	6	21 (10)	7	N	N	N
LeeSeaGull	gull	1	0 (0)	55	Y	Y	Y
DogBact	dog	5	25 (15)	145	N	Y	N
BacCan	dog	1	2 (6)	5495	N	Y	N
Pig2Bac	pig	5	0 (0)	107152	Y	Y	Y

LeeSeaGull cross-reacted with were pigeon samples, but the abundance of LeeSeaGull in pigeon samples was lower than that of the gull samples. Thus, the LeeSeaGull method met the study specificity criterion. However, it should be noted that the LeeSeaGull assay was tested by just one laboratory; a companion paper (Sinigalliano *et al*. 2013) tested it further. Cross reactivity between gull assays and non-gull avian feces suggest these assays may be better described as bird assays (Sinigalliano *et al*. 2013).

The LeeSeaGull method quantified genes in all gull challenge samples. The median abundance of the LeeSeaGull marker in gull feces was 55 copies per CFU ENT and thus met the study sensitivity criterion (Table 5). The median abundance measured using the Gull2 methods was less than 50 copies per CFU ENT; they did not quantify genes in some of the gull 1:10 challenge samples.

Dog Quantification

Both DogBact and BacCan showed cross-reactivity with non-target samples (Figure 5a; Table 4). BacCan cross-reacted with septage, goose, and cow challenge samples, but only the goose and septage amplified at levels on par with levels in dog feces. DogBact cross-reacted with every type of sample, measuring gene abundance per CFU ENT in many non-targets on par with dog fecal samples. BacCan measured gene abundances in two (6%) non-target samples on par with the dog fecal samples. While none of the dog assays met the specificity criterion, BacCan performed better than DogBact. BacCan was tested by a single laboratory in the present study, but further testing of the assay is described in a companion paper (Schriewer *et al*. 2013).

Both dog methods measured median gene abundances greater than 50 copies per CFU ENT

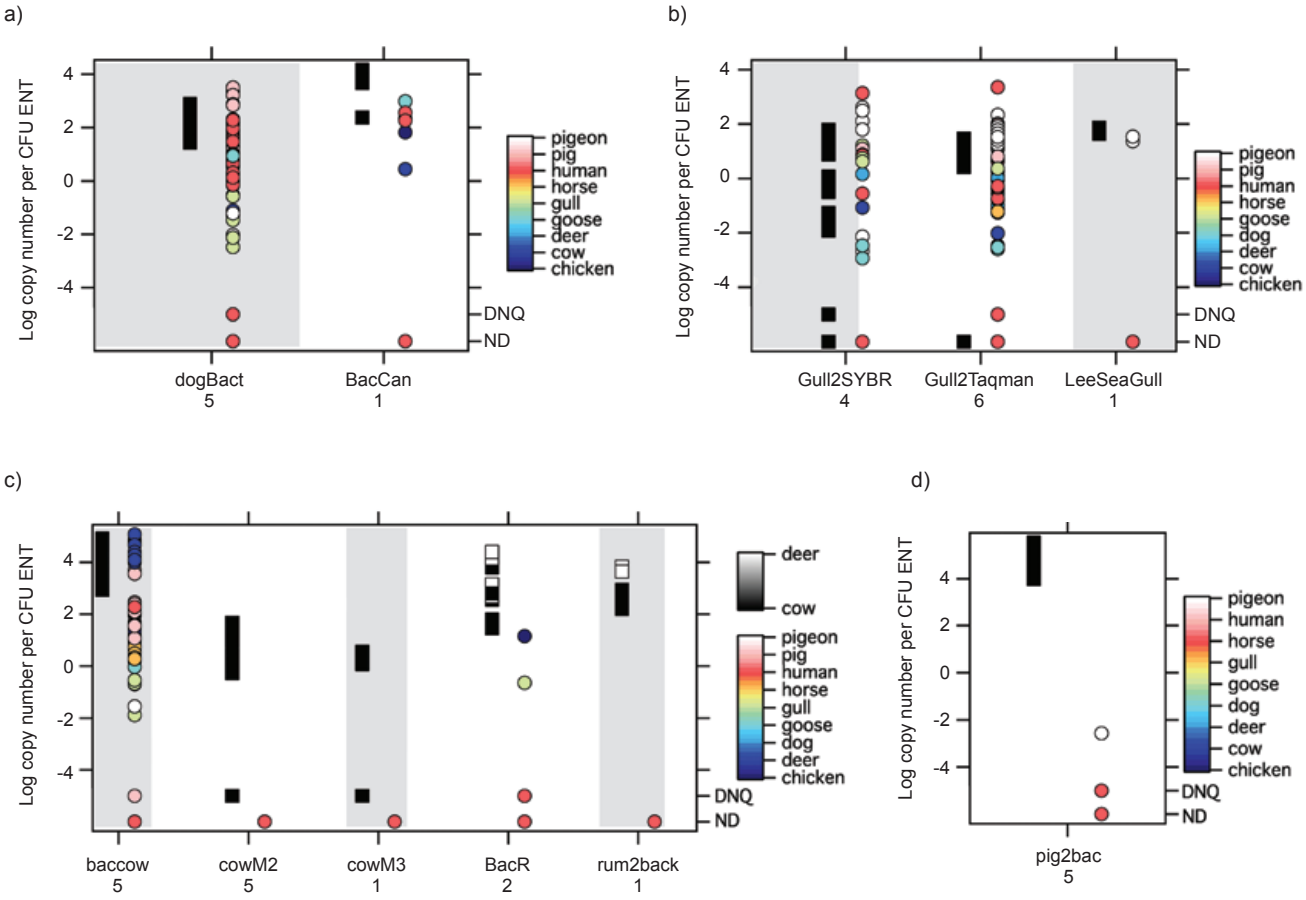

Figure 5. The number of copies per CFU ENT for each dog (panel a), gull (panel b), cow/ ruminant (panel c), and pig (panel d) assay for the assays' target (squares) and non-target (circles) fecal samples. The number below each assay is the number of performing laboratories. Symbol color denotes the specific fecal course (see color legends for each panel). Results within the range of quantification are plotted against the left axes. Samples where the assay target was detected but not quantified (DNQ) and not detected (ND) are plotted against the right axes.

in target challenge samples (Table 5); they also measured genes in all target samples.

Pig Quantification

Only one quantitative pig method was tested and it was both sensitive and specific (Figure 5d; Table 4). The Pig2Bac method measured gene abundances well above 50 copies per CFU ENT in all target challenge samples (Table 5). The assay cross-reacted with a single non-target sample (pigeon) in one laboratory at a low level, below those observed in the pig challenge samples.

Assay Specificity under Different Normalizations

Measured gene abundance in each challenge sample was normalized by the copies of genbac3 and DNA mass in the challenge samples (Figure 1). The specificity analyses were repeated (Tables SI-4, SI-5, and SI-6) for each assay, the number of non-target samples that had gene abundance per genbac3 and DNA mass in the same range as those measured in target fecal samples was determined. Specific methods identified using the ENT-MF normalization were also specific using genbac3 and DNA mass for normalization. For a few methods, changing the normalization scheme resulted in one or two more non-target fecal samples having abundances on par with those found in the assays' target fecal samples. All of the methods that were identified as 'not-specific' using ENT-MF normalization were also 'not-specific' using the alternative normalization schemes, with two exceptions. Both dog-associated methods were specific when their results were normalized to mass of DNA (Schriewer *et al*. 2013).

Discussion

The study findings indicate there has been a substantial advance in MST methodology during the last decade. Similar blind-sample method evaluation studies conducted in the early 2000s found that cultivation-based library-dependent methods that were state-of-the-art at the time performed poorly (Griffith *et al.* 2003, Stoeckel *et al.* 2004). The present study found that numerous library-independent methods developed in subsequent years, including those for detecting human, cow, gull, ruminant, pig, and horse feces, are both sensitive and specific. Additionally, it found that fecal material for which source-associated genetic markers, such as chicken and deer, that were not included in this study could be identified reliably by several community-based methods.

A few methods evaluated in the present study produced only presence-absence information (conventional PCR, community analysis, canine scent tracking approaches), but most also produced quantitative information. When examined from both a presence-absence and quantitative perspective, some methods (HF183Taqman, Pig2Bac, and BacR in particular) performed better when evaluated quantitatively. This is because when these assays cross-reacted with non-target feces, it was at a low level. In the presence-absence evaluation, low level cross reactivity is classified as a false positive; however, when considered in the quantitative evaluation, cross reactivity at low levels (i.e., levels lower than those observed for the target fecal samples) did not count as false-positives. There were, however, a few cases in which quantitative assays performed better in the presence-absence analysis and this is because they did not meet the quantitative sensitivity criterion that required 50 copies per ENT-CFU. One reason methods may have not met this criterion is the low levels of the fecal material in some target samples; this was particularly a concern for sewage and septage samples. Inconsistencies between performance in the presence-absence and quantitative assessments highlight the importance of selecting an appropriate lower threshold for classifying a response, which is further explored in Layton *et al*. (2013), Raith *et al*. (2013), Schriewer *et al*. (2013), and Sinigalliano *et al*. (2013).

The concern about how to interpret low concentration responses highlights the diverse approaches used to interpret high QPCR C_q values observed by numerous laboratories participating in this study. While some laboratories designated a lower limit of quantification (LLOQ) and deemed samples with a C_q higher than this limit as detected but not quantified (DNQ), other laboratories assigned concentrations to samples with C_q values higher than that of the LLOQ by extending the standard curve beyond the lowest standard tested; still others ignored high C_q values and assigned not detected (ND) to such samples. Additionally, laboratories reported vastly different LLOQ values. In this overview analysis, results submitted by the laboratories without modification and assigned DNQ as positive were used. The impact of variable LLOQ assignments is further explored in Layton *et al*. (2013), Raith *et al*. (2013), Schriewer *et al*. (2013), and Sinigalliano *et al*. (2013). These authors found assigning DNQ as negative rather than positive resulted in increased specificity and reduced sensitivity for some assays.

Assessment of quantitative method success depends in part on the approach used to assign the amount of fecal material in the challenge samples. Data were collected that allow for fecal material to be measured using fecal mass, DNA mass, general *Bacteroidales* genetic markers, and culture-based *Enterococcus* measured by membrane filtration and by defined substrate assays. Samples were defined based on *Enterococcus* using membrane filtration because the most common application of MST methodologies is identification of the sources of fecal contamination that have led to enterococci standards exceedances at a beach. However, alternative definitions could be more appropriate for other applications. For these reasons, specificity analyses using DNA mass and general Bacteroidales (genbac3) were repeated to quantify the amount of feces in the challenge samples. Performance of a few methods improved, as further detailed in Layton *et al*. (2013), Raith *et al*. (2013), Schriewer *et al*. (2013), and Sinigalliano *et al*. (2013), but overall conclusions about performance of most methods based on singleton and 1:10 diluted single source filters were unaffected by the different sample definitions. However, further work will need to verify if method performance is affected by source definitions when more than one animal source is present, i.e., doubleton filters (Cao *et al*. 2013, Ervin *et al*. 2013, Layton *et al*. 2013).

Another factor affecting our assessment was treatment of sewage and septage samples as human fecal sources. These sources were included because they are routinely monitored for management action.

Inclusion of these sources was also based on interest in evaluating MST methods that measure a target, such as a pathogen, phage, or rare gene sequence that is more likely to be found in samples from a group of humans, rather than in individual human fecal samples. Sewage and septage, though, can also contain pet fecal material or wildlife feces that infiltrates through leaks in the system or is consciously added to the system by users. In addition, it was observed that septage and sewage generally contained fewer human MST genetic markers on a per CFU ENT basis than human feces, suggesting differential decay of cultivated *Enterococcus* and host-associated genetic markers while aging during their storage and transport in conveyance systems (i.e., sewer lines, storage tanks). Removing septage and sewage samples from our analysis did not materially affect the assessment for most methods, but the lower ratio of MST genetic markers to ENT-MF in these sources did contribute to reduced sensitivity of some human methods, such as gyrB and HumM2.

A subset of the best performing assays were tested by just one laboratory and in many cases, that laboratory was the developer. Those assays (BacH, gyrB, LeeSeaGull, Rum2Bac, CF193, CowM3, BacCan, pigmtDNA, and PhyloChip) need to be further tested to document their portability across laboratories. Participating laboratories began this process; LeeSeaGull and BacCan assays were subjected to further testing on unblinded samples in Schriewer *et al*. (2013) and Sinigalliano *et al*. (2013).

While the results from this study are encouraging, it is important to recognize that they represent a best-case scenario for method performance. First, only 12 fecal sources were assessed, leaving many other sources that can potentially cross-react yet to be assessed. Second, the combinations were limited to two sources, whereas field samples may contain more sources that can potentially interfere with the methods. Third, fresh fecal material was used, and MST targets may degrade at a different rate compared to cultivated *Enterococcus* cells (Walters *et al.* 2009, Schulz and Childers 2011, Jeanneau *et al.* 2012), which would affect interpretation when aged fecal material is discharged into ambient waters. Fourth, the feces were suspended in artificial freshwater free of background microorganisms or other constituents that can interfere with nucleic acid isolation and detection. Fifth, it is important to recognize that the challenge sample fecal material was collected from California, potentially limiting applicability to other areas of the country or overseas. While multiple specimens were collected from each of four diverse geographic areas of the state, repeating a method assessment in other locales is warranted. Sixth, while this study evaluated the performance of MST markers in defined samples, field testing at beaches is needed as some of the assays were developed for use other types of environments (e.g., BacH and BacR were developed for use in alpine watersheds in Europe (Reischer *et al*. 2006, 2007). Finally, work is also needed to develop frameworks that use marker concentrations in ambient waters for fecal source allocation (Kinzelman *et al*. 2011, Wang *et al*. 2013), risk assessment, and identification of human or other animal-dominated polluted beaches.

Literature Cited

Aksamit, A.J. 1993. PCR detection of JC virus. pp. 364 *in*: D.H. Persing, T.F. Smith, F.C. Tenover and T.J. White (eds.), Diagnostic Molecular Microbiology Principles and Applications. Mayo Foundation. Rochester, MN.

Bernhard, A.E. and K.G. Field. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Applied and Environmental Microbiology* 66:4571-4574.

Byappanahalli, M.N., D.A. Shively, M.B. Nevers, M.J. Sadowsky and R.L. Whitman. 2003. Growth and survival of *Escherichia coli* and enterococci populations in the macro-alga Cladophora (Chlorophyta). *FEMS Microbiology Ecology* 46:203-211.

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

Cao, Y., J.F. Griffith, S. Dorevitch and S.B. Weisberg. 2012. Effectiveness of qPCR permutations, internal controls and dilution as means for minimizing the impact of inhibition while measuring Enterococcus in environmental waters. *Journal of Applied Microbiology* 113:66-75.

Cao, Y., L.C. Van De Werfhorst, E.A. Dubinsky, B.D. Badgley, M.J. Sadowsky, G.L. Andersen, J.F. Griffith and P.A. Holden. 2013a. Evaluation of molecular community analysis methods for

discerning fecal sources and human waste. *Water Research* 47:6862-6872.

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

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

Da Silva, A.K., J.C. Le Saux, S. Parnaudeau, M. Pommepuy, M. Elimelech and F.S. Le Guyader. 2007. Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: different behaviors of genogroups I and II. *Applied and Environmental Microbiology* 73:7891-7897.

DeLeon, R., C. Shieh, R.S. Baric and M.D. Sobsey. 1990. Detection of enteroviruses and hepatitis A virus in environmental samples by gene probes and polymerase chain reaction. pp. 833-853 *in*: Proceedings of the Water Quality Technology Conference, Advances in Cater Analysis and Treatment, San Diego, CA. American Water Works Association. Denver, CO.

Dick, L.K., A.E. Bernhard, T.J. Brodeur, J.W. Santo Domingo, J.M. Simpson, S.P. Walters and K.G. Field. 2005. Host distributions of uncultivated fecal *Bacteroidales* reveal genetic markers for fecal source identification. *Applied and Environmental Microbiology* 71:3184-3191.

Donaldson, K.A., D.W. Griffin and J.H. Paul. 2002. Detection, quantitation and identification of enteroviruses from surface waters and sponge tissue from the Florida Keys using real-time RT-PCR. *Water Research* 36:2505-2514.

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

Ebdon, J.E. and H.D. Taylor. 2006. Geographical stability of enterococcal antibiotic resistance profiles in Europe and its implications for the identification of fecal sources. *Environmental Science & Technology* 40:5327-5332.

Ebdon, J.E., M. Muniesa and H.D. Taylor. 2007. The application of a recently isolated strain of *Bacteroides* (GB124) to identify human sources of faecal pollution in a temperate river catchment. *Water Research* 41:3683-3690.

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

Fuhrman, J.A., X. Liang and R.T. Noble. 2005. Rapid detection of enteroviruses in small volumes of natural waters by real-time quantitative reverse transcriptase PCR. *Applied and Environmental Microbiology* 71:4523-4530.

Gourmelon, M., M.P. Caprais, R. Segura, C. Le Mennec, S. Lozach, J.Y. Piriou and A. Rince. 2007. Evaluation of two library-independent microbial source tracking methods to identify sources of fecal contamination in French estuaries. *Applied and Environmental Microbiology* 73:4857-4866.

Gregory, J.B., R.W. Litaker and R.T. Noble. 2006. Rapid one-step quantitative reverse transcriptase PCR Assay with competitive internal positive control for detection of enteroviruses in environmental samples. *Applied and Environmental Microbiology* 72:3960-3967.

Griffith, G.F., S.B. Weisberg and C.D. McGee. 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *Journal of Water and Health* 141-151.

Hagedorn, C. and S.B. Weisberg. 2009. Chemical-based fecal source tracking methods: current status and guidelines for evaluation. *Reviews in Environmental Science and Biotechnology* 8:275-287.

Hansen, D.L., S. Ishii, M.J. Sadowsky and R.E. Hicks. 2009. *Escherichia coli* populations in Great Lakes waterfowl exhibit spatial stability and temporal shifting. *Applied and Environmental Microbiology* 75:1546-1551.

Hardina, C.M. and R.S. Fujioka. 1991. Soil: The environmental source of *Escherichia coli* and enterococci in Hawaii's streams. *Environmental Toxicology and Water Quality: An International Journal* 6:185-195.

Harwood, V.J., A. B. Boehm, L.M. Sassoubre, V. Kannappan, J.R. Stewart, T.-T. Fong, M.-P. Caprais, R.R. Converse, D. Diston, J. Ebdon, J.A. Fuhrman, M. Gourmelon, J. Gentry-Shields, J.F. Griffith, D. Kashian, R.T. Noble, H. Taylor and M. Wicki. 2013. Performance of viruses and bacteriophages for fecal source determination in a multi-laboratory, comparative study. *Water Research* 47:6929-6943.

Haugland, R.A., M. Varma, M. Sivaganesan, C. Kelty, L. Peed and O.C. Shanks. 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Systematic and Applied Microbiology* 33:348-357.

International Organization for Standardization. 1995. Detection and Enumeration of Bacteriophages. Part 1: Enumeration of F-specific RNA Bacteriophages. ISO 10705-1. International Organization for Standardization. Geneva, Switzerland.

Jeanneau, L., O. Solecki, N. Wery, E. Jarde, M. Gourmelon, P.Y. Communal, A. Jadas-Hecart, M.P. Caprais, G. Gruau and A.M. Pourcher. 2012. Relative decay of fecal indicator bacteria and human-associated markers: A microcosm study simulating wastewater input into seawater and freshwater. *Environmental Science & Technology* 46:2375-2382.

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

Johnston, C., J.A. Ufnar, J.F. Griffith, J.A. Gooch and J.R. Stewart. 2010. A real-time qPCR assay for the detection of the *nifH* gene of *Methanobrevibacter smithii*, a potential indicator of sewage pollution. *Journal of Applied Microbiology* 109:1946-1956.

Jothikumar, N., T.L. Cromeans, V.R. Hill, X. Lu, M.D. Sobsey and D.D. Erdman. 2005. Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Applied and Environmental Microbiology* 71:3131-3136.

Kildare, B.J., C.M. Leutenegger, B.S. McSwain, D.G. Bambic, V.B. Rajal and S. Wuertz. 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: A Bayesian approach. *Water Research* 41:3701-3715.

Kinzelman, J., D. Kay and K. Pond. 2011. Relating MST results to fecal indicator bacteria, pathogens, and standards. pp. 337-359 *in*: C. Hagedorn, A.R. Blanch and V.J. Harwood (eds.), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer. New York, NY.

Layton, B., Y. Cao, D. Ebentier, K. Hanley, L. Van De Werfhorst, D. Wang, T. Madi, R. Whitman, M. Byappanahalli, E. Ballesté, W. Meijer, A. Schriewer, S. Wuertz, R. Converse, R. Noble, S. Srinivasan, J. Rose, C.S. Lee, J. Lee, J. Shields, J. Stewart, G. Reischer, A. Farnleitner, C. Sinigalliano, R. Rodrigues, S. Lozach, M. Gourmelon, L. Peed, O. Shanks, J. Jay, T. Holden, A. Boehm and J. Griffith. 2013. Performance of human fecal-associated PCR-based assays in the California Source Identification Protocol Project. *Water Research* 47:6897-6908.

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

Lee, C.S. and J. Lee. 2010. Evaluation of new gyrB-based real-time PCR system for the detection of *B. fragilis* as an indicator of human-specific fecal contamination. *Journal of Microbiology Methods* 82:311-318.

Lee, C., J.W. Marion and J. Lee. 2013. Development and application of a quantitative PCR assay targeting *Catellicoccus marimammalium* for assessing gull-associated fecal contamination at Lake Erie beaches. *Science of the Total Environment* 454-455:1-8.

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

Martellini, A., P. Payment and R. Villemur. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in

fecally contaminated surface water. *Water Research* 39:541-548.

McLain, J.E.T., H. Ryu, L. Kabiri-Badr, C.M. Rock and M. Abbaszadegan. 2009. Lack of specificity for PCR assays targeting human *Bacteroides* 16S rRNA gene: Cross amplification with fish feces. *FEMS Microbiology Letters* 299:38-43.

McQuaig, S.M., T.M. Scott, J.O. Lukasik, J.H. Paul and V.J. Harwood. 2009. Quantification of human polyomaviruses JC virus and BK virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples. *Applied and Environmental Microbiology* 75: 3379-3388.

Mieszkin, S., J.F. Yala, R. Joubrel and M. Gourmelon. 2010. Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *Journal of Applied Microbiology* 108:974-984.

Murray, J. 2011. Canine Scent and Microbial Source Tracking in Santa Barbara, CA. U2R09. Water Environment Research Foundation. Alexandria, VA.

Noble, R.T. and S.M. McQuaig. 2011. Viruses as tracers of fecal pollution. pp. *in*: C. Hagedorn, A.R. Blanch and V.J. Harwood (eds.), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer. New York.

Noble, R.T., S.M. Allen, D. Blackwood, W. Chu, S.C. Jiang, G.L. Lovelace, M.D. Sobsey, J.R. Stewart and D.A. Wait. 2003. Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study. *Journal of Water and Health* 1:195-207.

Ogorzaly, L., A. Tissier, I. Bertrand, A. Maul and G. C. 2009. Relationship between F-specific RNA phage genogroups, faecal pollution indicators and human adenoviruses in river water. *Water Research* 43:1257-1264.

Pruss, A. 1998. Review of epidemiological studies on health effects from exposure to recreational water. *International Journal of Epidemiology* 27:1-9.

Raith, M.R., C.A. Kelty, J.F. Griffith, A. Schriewer, S. Wuertz, S. Mieszkin, M. Gourmelon, G. Reischer, A.H. Farnleitner, J. Ervin, P. Holden, D. Ebentier, K. Hanley, J.A. Jay, D. Wang, A.B. Boehm, T. Gim, J. Rose, M. Sivaganesan and O.C. Shanks. 2013. Comparison of PCR-based assays on the characterization of cattle fecal pollution in California. *Water Research* 47:6921-6928.

Reischer, G., D. Kasper, R. Steinborn, A. Farnleitner and R. Mach. 2007. A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Letters in Applied Microbiology* 44:351-356.

Reischer, G., D. Kasper, R. Steinborn, R. Mach and A. Farnleitner. 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Applied and Environmental Microbiology* 72:5610-5614.

Schriewer, A., C.D. Sinigalliano, K.D. Goodwin, C. A., D. Wanless, J. Bartkowiak, D.L. Ebentier, K.T. Hanley, J. Ervin, L.A. Deering, O.C. Shanks, L.A. Peed, W.G. Meijer, J.F. Griffith, J. SantoDomingo, J.A. Jay, H. P.A. and S. Wuertz. 2013. Performance evaluation of canine associated Bacteroidales assays in a multi-laboratory comparison study. *Water Research* 47:6909-6920.

Schulz, C.J. and G.W. Childers. 2011. Fecal *Bacteroidales* diversity and decay in response to variations in temperature and salinity. *Applied and Environmental Microbiology* 77:2563-2572.

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

Shanks, O.C., E. Atikovic, A.D. Blackwood, J. Lu, R.T. Noble, J. Santo Domingo, S. Siefring, M. Sivaganesan and R.P. Haugland. 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Applied and Environmental Microbiology* 74:745-752.

Shanks, O.C., C.A. Kelty, M. Sivaganesan, M. Varma and R.A. Haugland. 2009. Quantitative PCR for genetic markers of human fecal pollution. *Applied and Environmental Microbiology* 75:5507-5513.

Shanks, O.C., K. White, C.A. Kelty, S. Hayes, M. Sivaganesan, M. Jenkins, M. Varma and R.A.

Haugland. 2010a. Performance assessment PCR-based assays targeting Bacteroidales genetic markers of bovine fecal pollution. *Applied and Environmental Microbiology* 76:1359-1366.

Shanks, O.C., K. White, C.A. Kelty, M. Sivaganesan, J. Blannon, M. Meckes, M. Varma and R.A. Haugland. 2010b. Performance of PCR-based assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples. *Environmental Science & Technology* 44:6281-6288.

Shibata, T., H.M. Solo-Gabriele, C.D. Sinigalliano, M.L. Gidley, L.R. Plano, J.M. Fleisher, J.D. Wang, S.M. Elmir, G. He, M.E. Wright, A.M. Abdelzaher, C. Ortega, D. Wanless, A.C. Garza, J. Kish, T. Scott, J. Hollenbeck, L.C. Backer and L.E. Fleming. 2010. Evaluation of conventional and alternative monitoring methods for a recreational marine beach with nonpoint source of fecal contamination. *Environmental Science & Technology* 44:8175-8181.

Sinigalliano, C., J. Ervin, L. Van De Werfhorst, J. Lee, K. Goodwin, D. Wanless, J. Bartkowiak, S. Wuertz, A. Schriewer, T. Holden, J. Jay, A.B. Boehm, B. Layton, M. Raith, W. Meijer, M. Sadowsky, M. Gourmelon, R. Noble, J. Griffith, M. Byappanahalli, D. Wang and J. SantoDomingo. 2013. A multi-laboratory comparative assessment on the performance of seabird-host-specific microbial source tracking PCR assays targeting *Catellicoccus marimammalium*. *Water Research* 47:6883-6896.

Sinigalliano, C.D., J.M. Fleisher, M.L. Gidley, H.M. Solo-Gabriele, T. Shibata, L.R. Plano, S.M. Elmir, D. Wanless, J. Bartkowiak, R. Boiteau, K. Withum, A.M. Abdelzaher, G. He, C. Ortega, X. Zhu, M.E. Wright, J. Kish, J. Hollenbeck, T. Scott, L.C. Backer and L.E. Fleming. 2010. Traditional and molecular analyses for fecal indicator bacteria in non-point source subtropical recreational marine waters. *Water Research* 44:3763-3772.

Sivaganesan, M., R.A. Haugland, E.C. Chern and O.C. Shanks. 2010. Improved strategies and optimization of calibration models for real-time PCR absolute quantification. *Water Research* 44:4726-4735.

Soller, J.A., M.E. Schoen, T. Bartrand, J.E. Ravenscroft and N.J. Ashbolt. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Research* 44:4674-4691.

Stoeckel, D.M. and V.J. Harwood. 2007. Performance, design, and analysis in microbial source tracking studies. *Applied and Environmental Microbiology* 73:2405-2415.

Stoeckel, D.M., M.V. Mathes, K.E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T.L. O'Brien, T.W. Fenger, M. Samadpour, K.M. Strickler and B.A. Wiggins. 2004. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. *Environmental Science & Technology* 38:6109-6117.

United States Environmental Protection Agency (USEPA). 2006. Method 1600: Enterococci in Water by Membrane Filtration Using membrane-Enterococcus Indoxyl-beta-D-Glucoside Agar (mEI). EPA 821-R-06-009. USEPA Office of Water. Washington, DC.

United States Environmental Protection Agency (USEPA). 2010. Method B: Bacteroidales in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay. USEPA Office of Water. Washington, DC.

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

Van De Werfhorst, L.C., B. Sercu and P.A. Holden. 2011. Comparison of the host specificities of two Bacteroidales quantitative PCR assays used for tracking human fecal contamination. *Applied and Environmental Microbiology* 77:6258-6260.

Vijayavel, K., R. Fujioka, J. Ebdon and H. Taylor. 2010. Isolation and characterization of *Bacteroides* host strain HB-73 used to detect sewage specific phages in Hawaii. *Water Research* 44:3714-3724.

Wade, T.J., N. Pai, J. Eisenberg and J.M. Colford. 2003. Do US EPA water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environmental Health Perspectives* 111:1102-1109.

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

Wang, D., A.H. Farnleitner, K.G. Field, H.C. Green, O.C. Shanks and A.B. Boehm. In press. *E. coli* and *Enterococcus* apportionment at recreational beaches using fecal source tracking markers – Is it feasible? *Environmental Science & Technology Letters*.

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

Xu, W., M.C. McDonough and D.D. Erdman. 2000. Species-specific identification of human adenoviruses by a multiplex PCR assay. *Journal of Clinical Microbiology* 38:4114-4129.

Yamahara, K.M., B.A. Layton, A.E. Santoro and A.B. Boehm. 2007. Beach sands along the California coast are diffuse sources of fecal bacteria to coastal waters. *Environmental Science and Technology* 41:4515-4521.

Yampara-Iquise, H., G. Zheng, J.E. Jones and C.A. Carson. 2008. Use of a *Bacteroides thetaiotaomicron*-specific α-1-6, mannanase quantitative PCR to detect human faecal pollution in water. *Journal of Applied Microbiology* 105:1686-1693.

Acknowledgements

The authors acknowledge all participants in the Source Identification Protocol Project (Table 1) and Chuck Hagedorn for comments on the manuscript. Blythe Layton, Yiping Cao, Todd Russell, Lauren Sassoubre, Meredith Raith, Darcy Ebentier, Ngoc Hoang, and Ning Jiang were instrumental in creating the challenge samples, assisted by other volunteers associated with the authors' labs. Kevan Yamahara created the graphical abstract. This manuscript has been subjected to the US Environmental Protection Agency (USEPA) peer and administrative review and has been approved for external publication. Funding for this project has been provided in full or in part through an agreement with the California State Water Resources Control Board. Any opinions expressed are those of the author(s). The contents of this manuscript do not necessarily reflect the views and policies of the State Water Resources Control Board or the USEPA, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Supplemental Information

Supplemental Information is available at ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_409_431SI.pdf.