Evaluation of the repeatability and reproducibility of a suite of qPCR-based microbial source tracking methods

Darcy L. Ebentier[1], Kaitlyn T. Hanley[1,2], Yiping Cao[3], Brian D. Badgley[4], Alexandria B. Boehm[5], Jared S. Ervin[6,7], Kelly D. Goodwin[8], Michèle Gourmelon[9], John F. Griffith[3], Patricia A. Holden[6,7], Catherine A. Kelty[10], Solen Lozach[9], Charles McGee[11], Lindsay A. Peed[10], Meredith Raith[3], Hodon Ryu[10], Michael J. Sadowsky[4], Elizabeth A. Scott[3], Jorge Santo Domingo[10], Alexander Schriewer[2], Christopher D. Sinigalliano[8], Orin C. Shanks[10], Laurie C. Van De Werfhorst[6,7], Dan Wang[5], Stefan Wuertz[2,12] and Jennifer A. Jay[1]

Abstract

Many PCR-based methods for microbial source tracking (MST) have been developed and validated within individual research laboratories. Inter-laboratory validation of these methods, however, has been minimal, and the effects of protocol standardization regimes have not been thoroughly evaluated. Knowledge of factors influencing PCR in different laboratories is vital to future technology transfer for use of MST methods as a tool for water quality management. In this study, a blinded set of 64 filters (containing 32 duplicate samples generated from 12 composite fecal sources) were analyzed by three to five core laboratories with a suite of PCR-based methods utilizing the standardized reagents and protocols. Repeatability (intra-laboratory variability) and reproducibility

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

[2]*University of California, Department of Civil and Environmental Engineering, Davis, CA*

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

[4]*University of Minnesota, BioTechnology Institute and Department for Soil, Water and Climate, St. Paul, MN*

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

[6]*University of California, Bren School of Environmental Science and Management, Santa Barbara, CA*

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

[8]*National Oceanic and Atmospheric Administration, Atlantic Oceanographic & Meteorological Laboratory, Miami, FL (stationed at SWFSC, La Jolla, CA)*

[9]*Ifremer, Laboratoire de Microbiologie, MIC/LNR, Département Ressources Biologiques et Environnement, Unité Environnement, Microbiologie et Phycotoxines, Plouzané, France*

[10]*US Envrionmental Protection Agency, National Risk Management Research Laboratory, Cincinnati, OH*

[11]*Orange County Sanitation District, 10844 Ellis Ave, Fountain Valley, CA*

[12]*Nanyang Technological University, Singapore Centre on Environmental Life Sciences Engineering, School of Biological Sciences, and School of Civil and Environmental Engineering, Singapore, Singapore*

(inter-laboratory variability) of observed results were assessed. When standardized methodologies were used, intra- and inter-laboratory %CVs were generally low (median %CV 0.1 - 3.3% and 1.9 - 7.1%, respectively) and comparable to those observed in similar inter-laboratory validation studies performed on other methods of quantifying fecal indicator bacteria (FIB) in environmental samples. ANOVA of %CV values found three human-associated methods (Bsteri, BacHum, and HF183Taqman) to be similarly reproducible ($p > 0.05$) and significantly more reproducible ($p < 0.05$) than HumM2. This was attributed to the increased variability associated with low target concentrations detected by HumM2 (approximately 1 - 2 $\log_{10}$copies/filter lower) compared to other human-associated methods. Cow-associated methods (BacCow and CowM2) were similarly reproducible ($p > 0.05$). When using standardized protocols, variance component analysis indicated sample type (fecal source and concentration) to be the major contributor to total variability with that from replicate filters and inter-laboratory analysis to be within the same order of magnitude, but larger than inherent intra-laboratory variability. However, when reagents and protocols were not standardized, inter-laboratory %CV generally increased with a corresponding decline in reproducibility. Overall, these findings verify the repeatability and reproducibility of these MST methods and highlight the need for standardization of protocols and consumables prior to implementation of larger scale MST studies involving multiple laboratories.

Introduction

Many PCR-based methods for microbial source tracking (MST) have been developed in an effort to characterize the sources of fecal pollution in recreational waters (Santo Domingo *et al.* 2007). They have the potential to be implemented by water quality managers for source identification, source allocation, total maximum daily load (TMDL) determinations, and remediation of chronic contamination problems, as well as for quantitative microbial risk assessment (QMRA) applications (Soller *et al.* 2010). Certainly MST methods must be sensitive and specific to their target source, but for implementation in water quality management they also must be repeatable (demonstrate good intra-laboratory agreement) and reproducible (demonstrate good inter-laboratory agreement; Simpson *et al.* 2002, Stoeckel and Harwood 2007). Some qPCR-based MST methods have been validated within individual research laboratories (Shanks *et al.* 2010a,b). Inter-laboratory validation of these methods, however, has been minimal. The few comparative studies that have been performed on repeatability have generally evaluated library dependent methods, which have since been proven to be largely untenable (Griffith *et al.* 2003, Stoeckel *et al.* 2004).

Despite the absence of studies evaluating the reproducibility of qPCR-based MST methods, many previous studies have assessed the repeatability and reproducibility of other microbiological and molecular methods, offering metrics by which MST methods can be evaluated (Schulten *et al.* 2000; Langton *et al.* 2002; Stoeckel *et al.* 2004; Cao *et al.* 2011, 2013; Shanks *et al.* 2012). In food safety method validation studies, the repeatability and reproducibility of methods for the detection and enumeration of pathogens has been evaluated using the repeatability (r) and reproducibility (R) values for quantitative data (Schulten *et al.* 2000). Repeatability values represent the maximum expected difference (with 95% probability) between two independent test results, obtained with the same method on identical samples in the same laboratory (Schulten *et al.* 2000). Reproducibility values represent the same maximum expected difference in results across multiple laboratories. In addition to these metrics, intra- and inter-laboratory percent coefficients of variation (%CV) for related environmental water quality methods have been observed and offer benchmark values for comparison (Cao *et al.* 2011, Shanks *et al.* 2012).

Here, we report results from quantitative real-time PCR (qPCR) analyses performed by 10 laboratories using 9 qPCR-based MST methods. The laboratories were provided with replicate, blinded filter sets in order to evaluate repeatability and reproducibility. Repeatability was defined as the ability of a method to produce the same results for analyses of identical samples under the same conditions in the same laboratory (Schulten *et al.* 2000, Bustin *et al.* 2009). Reproducibility was defined as the ability to produce the same results for analyses of identical samples under the same conditions in different laboratories (Schulten *et al.* 2000, Bustin *et al.* 2009). Five of the participating laboratories used standardized protocols including centralized sources of reference DNA standards, as well as the same manufacturer preparations of nucleic acid isolation materials and PCR reagents so as to estimate repeatability and reproducibility exclusive of deviations

in protocols. The remaining laboratories used a variety of deviations from the standardized protocols and reagents for six of these nine methods. Taken together, these data enabled us to assess the importance of protocol standardization to overall method reproducibility.

The specific goals of the study were to investigate: 1) the repeatability of each MST method within laboratories employing standardized protocols and reagents, 2) the reproducibility of nine MST (qPCR) methods across several laboratories when protocols and reagents were standardized, and 3) the effects of protocol deviations on the reproducibility of qPCR MST methods. We also quantified the relative contribution of intra-laboratory and inter-laboratory variability to total variability, determined the influence of platform when standardized protocols and reagents were used, and identified the most repeatable and reproducible methods of those evaluated.

Methods

Sample Processing and Analysis

Blinded sets of 64 filters, comprised of 32 duplicate samples generated from 12 fecal sources, were prepared in replicate as described in Boehm and colleagues (Boehm *et al.* 2013). Fresh fecal sources included human feces, sewage, septage, cow, dog, deer, pigeon, seagull, goose, chicken, pig, and horse. Fecal samples were collected from multiple geographic locations across California and composited to better approximate the microbial community associated with each source. Geographic locations ranged throughout southern and central California (Boehm *et al.* 2013). Slurries containing composites of single sources (hereafter refered to as singletons) were prepared using sterile artificial freshwater (distilled water with 0.3 mM $MgCl_2$, 0.6 mM $CaCl_2$, and 1.4 mM $NaHCO_3$) and diluted to a concentration of approximately 1000 culturable *Enterococcus* per 50 ml "based on information gleaned from the literature and pilot studies on fecal enterococci concentrations" (Boehm *et al.* 2013). Slurries containing two fecal sources (hereafter refered to as doubletons) were prepared by combining singleton slurries in 90%:10% ratios by volume to approximate dominant and minor fecal pollution sources. All slurries were filtered through 0.45 µm polycarbonate filters (Millipore, Billerica, MA). Filtered volumes were adjusted depending on desired concentrations; 50 ml was filtered for full strength singleton and doubleton filters, while 5 ml was filtered to achieve 1:10 strength singleton filters (Boehm *et al.* 2013). In order to best evaluate the methods and source combinations relevant to California beaches, the 32 samples were comprised of 12 full strength singletons, 7 1:10 strength singletons, and 13 doubletons as described in detail in Boehm *et al*. (2013). Filters were immediately flash frozen in liquid nitrogen and stored at -80˚C. Replicate sets of frozen filters were shipped overnight on dry ice to participating laboratories for nucleic acid isolation and qPCR analysis.

The following 10 laboratories contributed data to this study: the Boehm Laboratory at Stanford University (Stanford, CA, USA), the Wuertz Laboratory at University of California, Davis (Davis, CA, USA), the Holden Laboratory at University of California, Santa Barbara (Santa Barbara, CA, USA), the Jay Laboratory at University of California, Los Angeles (Los Angeles, CA, USA), the Southern California Coastal Water Research Project (SCCWRP; Costa Mesa, CA, USA), two participants from the United States Environmental Protection Agency (USEPA) National Risk Management Research Laboratory (Cincinnati, OH, USA), two participants from the National Oceanographic and Atmospheric Administration (NOAA) Atlantic Oceanographic and Meteorological Laboratory (Miami, FL, USA; one participant stationed at the Southwest Fisheries Science Center, La Jolla, CA, USA), and the Ifremer Laboratoire de Microbiologie, EMP (Plouzane, France). Stanford, University of California Santa Barbara, University of California Los Angeles, SCCWRP and one laboratory from the USEPA were randomly assigned numbers 1 through 5. The remaining laboratories were randomly assigned numbers 6 through 10.

Nine qPCR methods were selected for repeatability and reproducibility analyses due to the fact that they were each performed by a total of four to six different laboratories (see Supplemental Information (SI) Table SI-1 available at ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_433_444SI.pdf). Laboratories 1 through 5 (hereafter referred to as core laboratories) used standardized protocols and reagents for nucleic acid isolation and qPCR analysis, while Laboratories 6 through 10 (hereafter referred to as non-core laboratories) utilized variable reagents and protocols (Tables SI-2 and SI-3). Samples were pre-screened by the coordinating laboratory, SCCWRP, for inhibition in each method by spiking

with a known concentration of reference DNA and assaying serial dilutions following the procedure described in Cao *et al*. (2012). Core laboratories ran each filter in triplicate and screened data using salmon testes DNA as a sample processing control (Haugland *et al.* 2005). Core laboratories also obtained all reagents from the same commercial vendor and all reference DNA standards from a centralized laboratory (Tables SI-4 and SI-5).

The MST methods evaluated in this study included the following: BacHum, HF183Taqman, HumM2, BsteriF1, DogBact, Gull2Taqman, Pig2Bac, CowM2 and BacCow (Table SI-1). These nine methods were performed by a minimum of three core laboratories so as to provide a large enough data set to reveal reliable relationships and results regarding the repeatability and reproducibility of these methods exclusive of protocol deviations. The *Enterococcus* spp. qPCR method, performed by three core laboratories, was also included for comparison with the MST methods (USEPA 2012).

qPCR Data Treatment

Raw qPCR quantification cycle (C_q) values from all core laboratories were treated with the same quality assurance and quality control protocols. For each method in each laboratory, standard curve data were pooled and regression analyses were used to identify outliers as those with a large standardized residual (absolute value >3. Outliers were iteratively removed until the adjusted standard curve data set no longer contained values with standardized residuals with absolute value >3. Pooled standard curves were truncated below the standard concentration at which 10 out of 12 replicates amplified.

Three core laboratories (Laboratories 1, 3, and 5) performed qPCR using a StepOnePlus™ platform (Life Technologies, Carlsbad, CA) while the remaining two (Laboratories 2 and 4) used a CFX-96 platform (Bio-Rad, Hercules, CA). To better allow comparison across all core laboratories for this study, platform-specific master standard curves were generated and used to determine the lower limit of quantification (LLOQ) for each platform method. These platform-specific master standard curves were generated by combining respective pooled standard curves and iteratively removing outliers with standardized residuals >|3|. The LLOQ for each platform method was then defined as the C_q value corresponding to 40 target copies per reaction. This concentration was chosen based on the lowest standard concentration able to be consistently quantified across core laboratories (12 of 12 replicates amplified). Laboratory sample data were subsequently qualified as in the range of quantification (ROQ) or below the lower limit of quantification (BLOQ) through application of the corresponding platform method LLOQ C_q. Laboratory sample data with C_q 40 (i.e., no amplification because 40 is the maximum number of cycles run for a qPCR assay) were qualified as not detected (ND). Finally, DNA target concentrations for each qPCR triplicate of each filter were estimated for each method using individual laboratory pooled standard curves. Since the DNA extraction protocol employed by core laboratories only resulted in partial recovery of lysate following the physical lysis step, DNA target concentrations were adjusted to more accurately estimate final copies/filter. Estimated DNA target concentrations were log transformed prior to statistical analysis. Only core laboratory replicates within the ROQ were used in subsequent analyses unless otherwise specified. Raw qPCR C_q values from remaining Laboratories 6 through 10 were treated by laboratory specific, non-standardized protocols. All assessments below were conducted using $\log_{10}$ transformed target concentration data, unless otherwise specified.

Repeatability within Laboratories using Standardized Protocols and Reagents

Repeatability was defined as the ability of a method to produce the same answer for replicate qPCR analyses in the same laboratory and was calculated for each method using $\log_{10}$ transformed target concentration data from the core laboratories. Repeatability (*r)* values representing the maximum expected difference between two qPCR replicates (with 95% probability) obtained in the same laboratory for each method were calculated using analysis of variance (Schulten *et al.* 2000). Additionally, within-laboratory percent coefficient of variation (%CV) was calculated among triplicate concentrations for each filter containing target source material for each MST method in each laboratory. Minimum, maximum and median of %CV were calculated.

Reproducibility among Laboratories using Standardized Protocols and Reagents

Reproducibility, defined as the ability of a method to produce the same answer for analyses

of replicate samples under the same conditions in different laboratories, was assessed for the core laboratories. Reproducibility (*R*) values representing the maximum expected difference between two results (with 95% probability) obtained with the same method in different laboratories were calculated using analysis of variance (Schulten *et al.* 2000), as described below. Additionally, mean sample concentrations (calculated from duplicate filter concentrations) within each laboratory were used to calculate the inter-laboratory %CV across all core laboratories for each MST method. Minimum, maximum, and median of %CV were calculated.

Reproducibility among Laboratories using Non-Standardized Protocols and Reagents

The use of two different thermal cycler platforms by the core laboratories described above allowed for the evaluation of variability contributed by platform. The role of platform in terms of its contribution to variability was evaluated by analysis of variance (ANOVA) via the GLM procedure in SAS (SAS Institute, Cary, NC).

The effects of other protocol and reagent deviations on the reproducibility of MST methods were determined by using data from the six qPCR methods for which data from both core and non-core laboratories were available (Table SI-1). Mean sample concentrations for each laboratory were used to calculate a new inter-laboratory %CV for each sample across all laboratories utilizing deviated protocols for each method. %CV minimum, maximum, and median were calculated, and compared to the inter-laboratory %CV distributions observed among the core laboratories.

Relative Contribution of Intra- and Inter-laboratory Variability to Total Variability among Laboratories using Standardized Protocols and Reagents

The relative contributions of intra-laboratory and inter-laboratory variability to total variability in estimated DNA target concentrations for all ten qPCR methods was determined by performing nested ANOVA with variance component analysis (SAS Institute, Cary, NC). Three factors and one error term were included in the ANOVA model: SampleType, Laboratory, Filter (nested under the Laboratory factor), and Random Error (i.e., qPCR repeatability). Intra- and inter-laboratory variability, defined above as *r* and *R*, respectively, were estimated as the variance contribution of Random Error and Laboratory, respectively, to the total variability. Laboratory and SampleType, which included fecal source as well as strength, were treated as random effects to infer the general inter-laboratory reproducibility of these methods. To determine if SampleType influenced inter-laboratory variability, an interaction term of the factors Laboratory and SampleType was included in the ANOVA model. Finally, the Filter factor assessed the variability contributed by both the sample filtration and DNA isolation steps within the laboratory.

Identification of the Most Reproducible Methods

To identify the most reproducible methods within each host group, inter-laboratory %CV values were ranked then analyzed via ANOVA with multiple comparison (Tukey's test, family-wise error = 0.05) (Cao *et al.* 2011). Inter-laboratory %CV values for this analysis were calculated utilizing triplicate target concentrations ($\log_{10}$ transformed) from both filter replicates. This procedure was carried out using data from the core laboratories to identify the most reproducible methods exclusive of protocol deviations.

Results

qPCR Data Treatment

All core laboratory C_q values were successfully processed according to the quality assurance and quality control protocols described above. Tabulated percentages of target (containing fecal source targeted by a method as defined in Boehm *et al*. 2013) and nontarget data, qualified as in the ROQ, BLOQ and ND were reported (Table 1). In general, ROQ DNA target concentrations were 3 to 9 $\log_{10}$ copies/filter for target filters (with the exception of CowM2 and HumM2 concentrations, which peaked at 7 $\log_{10}$ copies/filter) and between 3 and 5 $\log_{10}$ copies/filter for nontarget filters. BLOQ DNA target concentrations ranged between 2 and 3 $\log_{10}$ copies/filter for target filters, and between 1 and 3 $\log_{10}$ copies/filter for nontarget filters. DNA target source data were almost always in the ROQ, with the exception of approximately 36% of the HumM2 data (Table 1).

Table 1. Summary of Data Distribution Among Core Laboratories (1 - 5). Results were seperated into those for target versus non-target filters. Target was defined according to Boehm *et al*. 2013. Some non-target (false positive) amplification was observed, but fell mostly BLOQ and is discussed in depth in other publications (i.e., Boehm *et al*. in 2013).

Method	Target			NonTarget		
	%ROQ	%BLOQ	%ND	%ROQ	%BLOQ	%ND
BacCow	100.0	0.0	0.0	36.6	13.9	49.5
BacHum	98.2	1.8	0.0	26.5	35.9	37.6
BsteriF1	91.9	8.1	0.0	24.9	21.1	54.0
CowM2	100.0	0.0	0.0	0.4	5.4	94.2
DogBact	100.0	0.0	0.0	8.5	12.9	78.6
Gull2Taqman	100.0	0.0	0.0	5.1	10.9	84.0
HF183Taqman	88.9	11.1	0.0	7.2	37.5	55.3
HumM2	63.9	32.1	4.00	3.4	28.4	68.2
Pig2Bac	100.0	0.0	0.0	5.2	19.7	75.1

Repeatability within Laboratories using Standardized Protocols and Reagents

*Repeatability (*r*) Values*

Repeatability (*r)* values were generally low and ranged from 0.03 to 0.17 $\log_{10}$ copies/filter (Table 2). All human-associated methods had *r* values less than 0.10 $\log_{10}$ copies/filter. HF183 Taqman and HumM2 had the lowest *r* values of the human-associated methods with 0.05 $\log_{10}$ copies/filter, indicating better intra-laboratory agreement. CowM2 had the lowest *r* value and Gull2Taqman had the highest of all the methods. All *r* values for MST qPCR methods were observed to be of similar magnitude to the *r* value for *Enterococcus* spp. qPCR, indicating similar intra-laboratory agreement.

Table 2. Repeatability (*r)* and reproducibility (*R)* values for each method. Values represent the maximum expected difference (with 95% probability) between replicate qPCR results within the same laboratory (*r*) and replicate results in different laboratories (*R*).

Method	*r* ($\log_{10}$ copies/filter)	*R* ($\log_{10}$ copies/filter)
BacHum	0.09	0.19
HF183Taqman	0.05	0.09
BacCow	0.10	0.22
BsteriF1	0.06	0.27
CowM2	0.03	0.28
DogBact	0.12	0.21
Gull2Taqman	0.17	0.23
HumM2	0.05	0.37
Pig2Bac	0.07	0.09
Entero	0.11	0.66

qPCR Triplicate %CV Analysis

Median %CV values between qPCR triplicates ranged from 0.1 to 3.3%, indicating minimal variation in replicate qPCR measurements of the same filter and thus low inherent method variability (Figures 1 and 2). The %CV values were generally lower for methods targeting non-human sources compared to methods targeting human sources. Non-human methods, with the exception of Laboratories 1 and 3 performing the Gull2Taqman method, had maximum intra-laboratory %CV values well below 5.0%. Human methods, with the exception of BacHum, showed maximum intra-laboratory %CV values below 6.0%. For human-associated methods, the highest %CV values were frequently associated with doubleton samples containing sewage and/or septage as the minor contributor or 1:10 strength singletons. For non-human methods, the highest %CV values were associated with 1:10 strength singleton or doubleton samples where the DNA target source was the minor contributor.

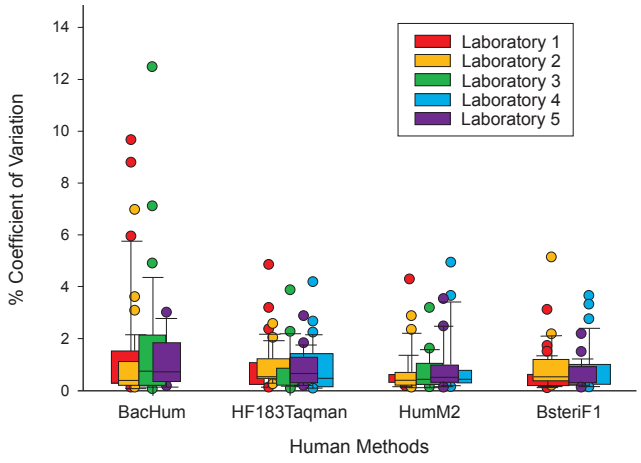

Figure 1. Intra-laboratory coefficients of variation (%CV) for human-associated methods. CV values were calculated between qPCR triplicate reactions for 38 target filters. Boxes represent the 25th to 75th percentile range, whiskers represent the 10th and 90th percentiles, and circles represent outliers.

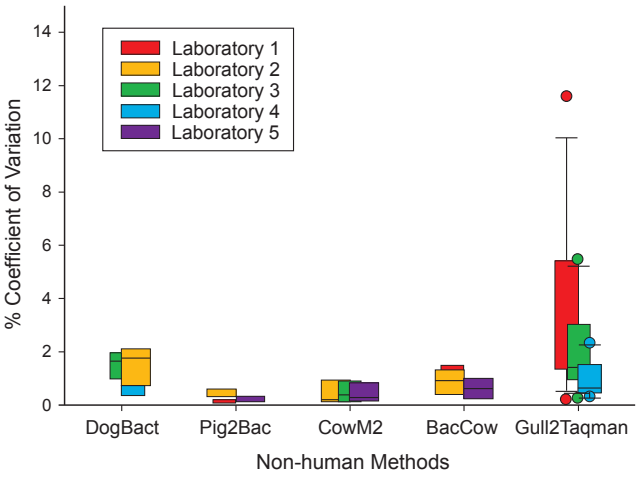

Figure 2. Intra-laboratory coefficients of variation (%CV) for nonhuman-associated methods. CV values were calculated between qPCR triplicate reactions for 8 (DogBact, Pig2Bac, CowM2, and BacCow) to 12 (Gull2Taqman) target filters. Boxes represent the 25th to 75th percentile range, whiskers represent the 10th and 90th percentiles, and circles represent outliers.

Reproducibility among Laboratories using Standardized Protocols and Reagents

*Reproducibility (*R*) Values*

Reproducibility (*R*) values ranged from 0.09 to 0.37 $\log_{10}$ copies/filter (Table 2). Among human-associated methods, HF183Taqman had the lowest *R* value while HumM2 had the highest (0.09 and 0.37 $\log_{10}$ copies/filter, respectively), indicating HF183Taqman showed the highest inter-laboratory agreement while HumM2 showed the lowest. Among non-human methods, Pig2Bac had the lowest *R* value, and of the two cow-associated methods, BacCow had a slightly lower *R* value than CowM2 (Table 2). All *R* values for MST qPCR methods were lower than the *Enterococcus* spp. qPCR *R* value, indicating similar or better inter-laboratory agreement among MST methods.

Inter-Laboratory %CV Analysis

Median inter-laboratory %CV for each method ranged from 1.9 to 7.1% (Figure 3). Inter-laboratory %CV values were lower for methods targeting non-human sources compared to methods targeting human sources. Non-human methods, with the exception of the Gull2Taqman method, had maximum inter-laboratory %CV well below 10.0%, while some human methods (Bsteri and HumM2) showed higher maximum inter-laboratory %CV values (up to 15.6%). As previously observed for qPCR triplicate %CV values, the highest %CV values for human-associated methods were almost always associated with doubleton samples containing sewage and/or septage as the minor contributor or 1:10 strength singletons. Again, for non-human methods, the highest %CV values were commonly associated with

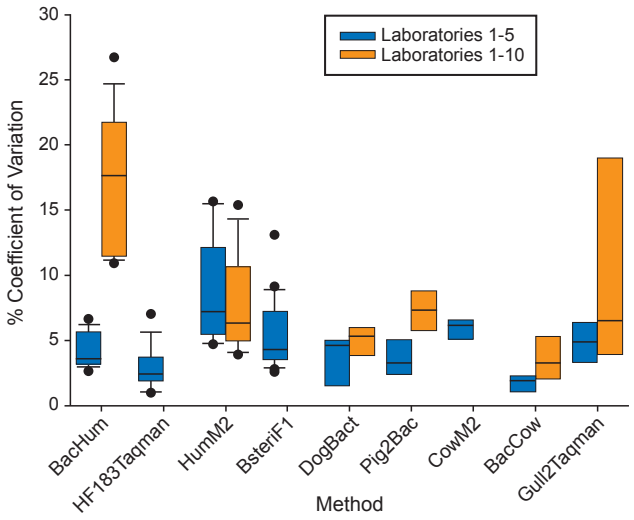

Figure 3. Inter-laboratory coefficients of variation (CV) among core laboratories (1 - 5) and among all laboratories (1 - 10). CV values were calculated between mean sample concentrations for each laboratory. Solid circles represent outliers. Methods lacking box plots for Laboratories 1 through 10 were not run in laboratories 6 through 10.

1:10 strength singleton or doubleton samples where the DNA target source was the minor contributor.

Reproducibility among Laboratories using Non-Standardized Protocols and Reagents

Platform variability was expected to have contributed to inter-laboratory variability based on preliminary visual inspection of initial regression analyses between results obtained on different platforms for the same method; a particular bias by ABI StepOnePlus appeared for most methods. Results of the GLM procedure in SAS showed that platform was not a significant contributor (p-value >0.05) to variability for any of the nine qPCR methods analyzed. However, other factors such as sample type (which includes both fecal source and filter strength) and filter-to-filter variability contributed significantly (p-value <0.01) to the total variability associated with these methods.

For the six methods for which data from both core and non-core laboratories were available, the median inter-laboratory %CV for each method ranged from 3.4 to 17.6% (Figure 3). Median %CV values and overall ranges were generally higher for all methods when protocol and reagents were varied, with the exception of the HumM2 method. The HumM2 method showed a slightly lower median %CV but a similar %CV distribution when protocol and reagents were varied (Figure 3). BacHum and Gull2Taqman showed the greatest increases in %CV when protocol and reagents were varied (Figure 3).

Table 3. Rank of variability contribution by different factors for each method. Sample type (fecal source and concentration) was always the largest contributor and filter-to-filter variability the second largest contributor to inter-laboratory variability.

Method	Rank of Variability Contribution
Pig2Bac	SampleType > Filter > Laboratory > Error*
HF183Taqman	SampleType > Filter > Laboratory > Error
CowM2	SampleType > Filter > Laboratory > Error
BacCow	SampleType > Filter > Laboratory > Error
Gull2Taqman	SampleType > Laboratory > Error > Filter
BacHum	SampleType > Laboratory > Filter > Error
DogBact	SampleType > Laboratory > Filter > Error
BsteriF1	SampleType > Laboratory > Filter > Error
HumM2	SampleType > Laboratory > Filter > Error

* Error represents random measurement error, i.e. intra-laboratory variability, (*r*) defined as qPCR triplicate variability

Relative Contribution of Intra- and Inter-Laboratory Variability to Total Variability among Laboratories using Standardized Protocols and Reagents

Nested ANOVA with variance component analysis showed the relative contributions of single

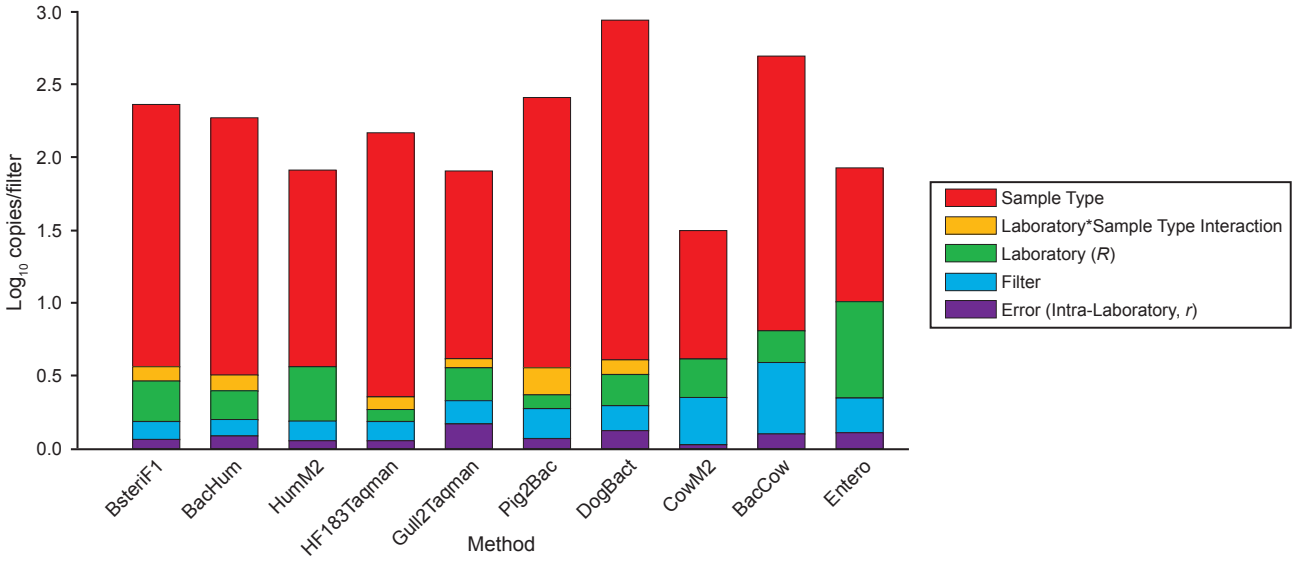

Figure 4. Relative contribution (in $\log_{10}$ copies/filter) of singular factors to total variability among core laboratories (1 - 5) for nine qPCR MST methods and *Enterococcus* spp. qPCR (USEPA 2012). Note that *r* = Error (intra-laboratory), *R* = Laboratory.

factors to the total variability in the whole dataset for each qPCR method (Figure 4). As expected, SampleType (fecal source and strength) contributed the most to total variability for each of these methods (Table 3; Figure 4), simply reflecting the fact that our challenge samples had various concentrations of target genetic markers for any given qPCR method. Inter-laboratory variability (i.e., the Laboratory factor) was just as likely as within-laboratory filter-to-filter variability (caused by filtration and DNA isolation) to be the second largest contributor to the total variance for MST qPCR methods (Table 3). Generally, the contribution of these two factors was greater than, but of similar order of magnitude to, the inherent qPCR method error (i.e., the *r* value, defined as intra-laboratory variability) for each method. Additionally, non-human methods generally had lower total variability than human-associated methods. The interaction term between sample type and laboratory indicated that for all human-associated methods, inter-laboratory variability was dependent on sample type, as would be expected considering the multiple types and dilutions of samples containing human target (Figure 4), with the exception of HumM2. This was also observed for the non-human methods BacCow and Gull2Taqman. Sample type variability did not affect the inter-laboratory variability for the Pig2Bac, CowM2 and DogBact methods (interaction terms were 0), indicating that for these methods the sample type did not affect inter-laboratories variability. Finally, MST qPCR methods showed a similar distribution of variability among single factors as *Enteroccoccus* spp. qPCR, as well as generally lower Laboratory (*R*) factor values.

Identification of the Most Reproducible Methods

Based on the %CV values produced among core laboratories, among the human-associated methods, formal statistical analysis showed that HumM2 was significantly less reproducible than the BacHum, BsteriF1 and HF183Taqman methods (p-value <0.05). The BacHum, BsteriF1, and HF183Taqman methods showed no significant differences from each other in terms of reproducibility. Among non-human sources, only the cow-associated methods offered an opportunity for comparison of methods because only one method was included for each of the other hosts in this study. It was found that BacCow and CowM2 were not significantly different (p-value >0.05) from each other in terms of reproducibility.

Discussion

The present study provides relevant information with regards to qPCR-based MST methods and their eventual implementation by beach managers and departments of public health. For laboratories employing standardized protocols, intra- and inter-laboratory %CV values were generally low, with median %CV for all MST methods evaluated ranging from 1.1 to 3.3% and 1.9 to 7.1%, respectively. These values were similar to those found during inter-laboratory validation of qPCR methods for enumeration of fecal indicator bacteria (FIB), as well as those found during inter-laboratory validation of methods for enumeration of FIB from sand (Cao *et al.* 2011, 2013; Shanks *et al.* 2012). The *r* and *R* values for MST qPCR methods were also similar to those observed for the *Enterococcus* spp. qPCR method. Variance component analysis showed filter-to-filter and inter-laboratory variance contribution for MST methods to be of similar magnitude to random error (0.03 - 0.56 $\log_{10}$copies/filter). It is also important to note that the filter-to-filter factor included variability associated with the filtration of sample replicates as well as the variability associated with nucleic acid isolation and purification. Considering the above, these results corroborate the integrity of both the reference challenge filter sets and the results of this study.

The observation that the highest %CV values were most often associated with 1:10 strength or minor contributor doubleton challenge samples is not surprising as these samples often contained lower concentrations of target DNA compared to full strength singletons, and higher variability as a function of higher C_q values would be expected (Cao *et al.* 2012). It is also important to note that the lower reproducibility observed with the HumM2 method is most likely due to a DNA target concentration effect rather than factors inherent to the protocol: HumM2 target concentrations were typically 1 to 2 $\log_{10}$ copies/filter lower than human target concentrations estimated by other methods. This may be attributed to the fact that HumM2 targets a single-copy functional gene, while the other human-associated methods target more conserved genes encoding 16S rRNA (Shanks *et al.* 2010a). Indeed, when the core laboratory LLOQ for HumM2 was readjusted to 100 copies per reaction (approximately 3.7 $\log_{10}$ copies/filter) instead of 40, the maximum inter-laboratory %CV value was appreciably lower (11.8% compared

to 15.3%), further supporting this theory. Interaction between the SampleType and Laboratory factors in the ANOVA model also indicated the concentration difference (BLOQ vs. ROQ) affected inter-laboratory reproducibility. This observation suggests that methods with higher analytical sensitivity, which would produce results at a relatively lower C_q are more reproducible. However, it remains undetermined how important the 1 to 2 $\log_{10}$ copies/filter difference in DNA target concentration between HumM2 and other human-associated assays is for MST applications. For example, it could be considered an advantage or liability for a more sensitive human-associated method to detect very low concentrations of human fecal pollution in waters depending on the particular water quality application. Overall, though, these observations further suggest that methods with higher analytical sensitivity, which would produce results at a relatively lower C_q, are more reproducible.

The finding that use of these two different amplification platforms contributed an insignificant amount of variability when all other aspects of protocols and reagents were standardized is also consistent with recent findings (Cao *et al.* 2013). However, it is unclear if this finding can be extended to platforms that have not otherwise been tested with these protocols and would potentially need to be re-evaluated if amplification platform was not standardized for method implementation.

It should be considered that this study was not exclusively designed with repeatability and reproducibility assessment in mind. Given the logistical implications of generating replicate reference challenge filter sets consisting of 64 filters for more than 25 participating laboratories, decisions had to be made regarding priority of pollution sources. The design of the challenge filter set was heavily weighted towards human sources: 38 of 64 filters (~60%) contained either solid human feces, sewage or septage, while non-human sources were present on between 2 and 12 filters of the 64 (Cao *et al.* 2013 (a,b); Boehm *et al.* 2013). Unequal sample size does influence the statistical comparison of %CV values between human and non-human methods (Cao *et al.* 2013 (a,b)). Additionally, the use of only solid feces with relatively high marker concentrations and no lower concentration effluent-type sources for non-human challenge samples may explain the generally lower %CV values (thus better reproducibility) and lower total variability observed for the non-human compared to the human-associated methods. This situation may account for the good reproducibility observed for the CowM2 method compared to the HumM2 method. Like the HumM2 method, the CowM2 method detected DNA target concentrations 1 to 2 $\log_{10}$ copies/filter lower than its comparative method (BacCow). Nonetheless, the CowM2 method was more reproducible compared to the HumM2 method, which may stem from the fact that the cow target samples were all generated from solid feces rather than effluents may account for its performance compared to the HumM2 method. A more balanced study design would be required to prove that nonhuman-associated methods were truly more reproducible than human-associated methods. These results, however, are an important first step towards better understanding whether some of these molecular methods lend themselves to field applications for the accurate identification of primary sources and the relative loadings with which they are associated.

The findings of this study indicate that deviations from a standardized protocol can have varying and substantial impacts on the reproducibility of qPCR-based MST methods. While this study was not designed to quantify the contribution of each protocol deviation to sample concentration variability (as in Shanks *et al*. 2011 and Cao *et al*. 2013), these findings affirm the need to establish standardized protocols in terms of laboratory and data analysis as well as centralized sources of reference materials to ensure successful implementation and technology transfer. For example, for the Pig2Bac method, Laboratory 8 employed a nucleic acid isolation approach different than the core laboratories, wherein no physical lysis step was used, and also utilized a different qPCR reagents than the core laboratories. The combination of extraction protocol and reagent deviations appeared to translate to appreciably higher inter-laboratory %CV values compared to when protocol and reagents were standardized (Figure 3). Conversely, for the HumM2 method, Laboratory 9 only deviated from the protocols and reagents employed by the core laboratories in its use of different qPCR reagents, and was able to produce very similar results, as indicated by the %CV distributions (Figure 3). This suggests that deviations at the nucleic acid isolation step may have an appreciable impact on reproducibility, which is consistent with the findings of other studies (Pan *et al.* 2010), however a more structured evaluation would be needed to confirm.

Regardless, the data indicate that lack of standardization is expected to increase variability in results.

Although efforts were made to provide varying reference challenge filter concentration levels, the samples analyzed in this study were comprised of relatively concentrated fecal, sewage, and septage source materials. For some samples, host-associated sequences were present at concentrations upwards of 10^9 gene copies/filter. Concentrations this high, though not impossible to find in some ambient waters, are not likely to occur in the natural environment and thus evaluation of repeatability and reproducibility at lower, more environmentally realistic concentrations is necessary. Also, considering the increased variability associated with the higher C_q values that would occur in more environmentally realistic concentrations, evaluation of reproducibility on challenge samples containing target concentrations at or near the LLOQ would be especially relevant. While this challenge filter set did contain a subset of lower strength samples that may have offered opportunity to further investigate reproducibility in this range, despite best effort, the actual singleton *Enterococcus* concentrations were highly variable, ranging from 148 to 4.7 x 10^6 MPN/50 ml, and did not provide sufficient data at low enough concentrations for separate analysis (Ervin *et al.* 2013). Furthermore, more complex matrices exist in the natural environment that may lead to amplification interference or result in false positives due to amplification of non-target sequences. Future multiple laboratory studies focusing on the reproducibility of MST methods should account for multiple types of sample matrices containing more environmentally relevant target concentrations. Finally, the relationship between target aging and reproducibility was not explored in this study, and is an important factor for field MST applications that must be examined in future work.

Literature Cited

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

Bustin, S.A., V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele and C.T. Wittwer. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* 55:611-622.

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., C.D. McGee, J.F. Griffith and S.B. Weisberg. 2011. Method repeatability for measuring *Enterococcus* in southern California beach sands. *Letters in Applied Microbiology* 53:656-659.

Cao, Y., M. Sivaganesan, J.L. Kinzelman, A.D. Blackwood, R. Noble, R. Haugland, J.F. Griffith and S.B. Weisberg. 2013. Effect of platform, reference material, and quantification model on enumeration of *Enterococcus* by quantitative PCR methods. *Water Research* 47:233-241.

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

Cao, Y., L.C.V.D. Werfhorst, E.A. Dubinsky, B.D. Badgley, M.J. Sadowsky, G.L. Andersen, J.F. Griffith and P.A. Holden. 2013(b). Evaluation of molecular community analysis methods for discerning fecal sources and human waste. *Water Research* 47:6862-6872.

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

Griffith, J.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* 1:141-151.

Haugland, R.A., S.C. Siefring, L.J. Wymer, K.P. Brenner and A.P. Dufour. 2005. Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase

chain reaction and membrane filter culture analysis. *Water Research* 39:559-568.

Langton, S.D., R. Chevennement, N. Nagelkerke and B. Lombard. 2002. Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. *International Journal of Food Microbiology* 79:175-181.

Pan, Y., L. Bodrossy, P. Frenzel, A.-G. Hestnes, S. Krause, C. Lüke, M. Meima-Franke, H. Siljanen, M.M. Svenning and P.L.E. Bodelier. 2010. Impacts of inter- and intralaboratory variations on the reproducibility of microbial community analyses. *Applied and Environmental Microbiology* 76:7451-7458.

Santo Domingo, J.W., D.G. Bambic, T.A. Edge and S. Wuertz. 2007. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Research* 41:3539-3552.

Schulten, S., P. in't Veld, N. Nagelkerke, S. Scotter, M. de Buyser, P. Rollier and C. Lahellec. 2000. Evaluation of the ISO 7932 standard for the enumeration of *Bacillus cereus* in foods. *International Journal of Food Microbiology* 57:53-61.

Shanks, O.C., M. Sivaganesan, L. Peed, C.A. Kelty, A.D. Blackwood, M.R. Greene, R.T. Noble, R.N. Bushon, E.A. Stelzer, J. Kinzelman, T. Anan'eva, C. Sinigalliano, D. Wanless, J. Griffith, Y. Cao, S. Weisberg, V.J. Harwood, C. Staley, K.H. Oshima, M. Varma and R.A. Haugland. 2012. Interlaboratory comparison of real-time PCR protocols for quantification of general fecal indicator bacteria. *Environmental Science & Technology* 46:945-953.

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

Shanks, O.C., K. White, C.A. Kelty, S. Hayes, M. Sivaganesan, M. Jenkins, M. Varma and R.A. Haugland. 2010b. Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. *Applied and Environmental Microbiology* 76:1359-1366.

Simpson, J.M., J.W. Santo Domingo and D.J. Reasoner. 2002. Microbial source tracking: State of the science. *Environmental Science & Technology* 36:5279-5288.

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. Mathes, K. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. O'Brien, T. Fenger, M. Samadpour, K. Strickler and B. 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). 2012. Method 1611: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay. EPA-821-R-12-008. USEPA Office of Water. Washington, DC.

Acknowledgements

The authors would like to acknowledge all particapants in the Source Identification Protocol Project (SIPP) study. Funding for this project has been provided in part through an agreement with the California State Water Resources Control Board. The contents of this article do not necessarily reflect the views and policies of the California State Water Resources Control Board, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Supplemental Information

Supplemental Information is available at ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2013AnnualReport/ar13_433_444SI.pdf.