Title: Quantitative PCR for genetic markers of human fecal pollution

Authors: Orin C. Shanks[1*], Catherine A. Kelty[1], Mano Sivaganesan[1], Manju Varma[2], and Richard A. Haugland[2]

[1] U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, 26 West Martin Luther King Drive, Cincinnati, OH 45268, USA

[2] U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 26 West Martin Luther King Drive, Cincinnati, OH. 45268, USA

Running Title: qPCR FOR HUMAN FECAL BACTERIA

*Corresponding Author: Tel.: +1 513 569 7314; Fax: +1 513 569 7328;
Email address: shanks.orin@epa.gov

Keywords: Quantitative PCR, risk assessment, microbial source tracking, water quality

ABSTRACT

Assessment of health risk and fecal bacteria loads associated with human fecal pollution requires reliable host-specific analytical methods and a rapid quantification approach. We report the development of quantitative PCR assays for quantification of two recently described human-specific genetic markers targeting *Bacteroidales*-like cell surface associated genes. Both assays exhibited a range of quantification from 10 to $1x10^6$ copies of target DNA. For each assay, internal amplification controls were developed to detect the presence or absence of amplification inhibitors. The assays predominantly detected human fecal specimens and exhibited specificity levels greater than 97% when tested against 265 fecal DNA extracts from 22 different animal species. The abundance of each human-specific genetic marker was measured in primary effluent wastewater samples collected from 20 geographically distinct locations and compared to quantities estimated by real-time PCR assays specific for ribosomal RNA gene sequences from total *Bacteroidales* and enterococci fecal microorganisms. Assay performances combined with the prevalence of DNA targets in sewage samples provide experimental evidence supporting the potential application of these quantitative methods for monitoring fecal pollution in ambient environmental waters.

INTRODUCTION

Waterborne diseases that originate from human fecal pollution remain a significant public health issue. As a result, a large number of methods have been developed to detect and quantify human fecal pollution (10, 12, 18, 21). The majority of

these methods are based on real-time quantitative PCR (qPCR) assays designed to estimate the concentration of 16S ribosomal RNA gene (rRNA) sequences from various subpopulations within the order *Bacteroidales*. This bacterial order constitutes a large proportion of the normal gut microbiota of most animals, including humans (3, 15, 27). Bacterial 16S rRNA genes are useful as markers because they have relatively low mutation rates (7) and are typically present in multiple operons, increasing template DNA levels available for detection (2, 11, 17, 29). While several studies have demonstrated the value of *Bacteroides* 16S rDNA-based qPCR assays, currently available assays cannot discriminate between several animal sources closely associated with humans, including cats, dogs, and/or swine (10, 12, 18, 21). Alternative qPCR assays targeting genes directly involved in host-specific interactions may be capable of increased discrimination of fecal pollution sources (23, 24) and are needed to complement existing qPCR-based approaches used to identify sources of human fecal pollution.

A recent metagenomic survey of a human fecal bacterial community using genome fragment enrichment has led to the identification of hundreds of candidate human fecal bacteria-specific DNA sequences (24). PCR assays targeting two gene sequences encoding for a hypothetical protein potentially involved in remodeling of bacterial surface polysaccharides and lipopolysaccharides (Assay 19) and a putative RNA polymerase extracytoplasmic function sigma factor (Assay 22) from *Bacteroidales*-like microorganisms exhibited a high level of specificity (100%) for human fecal material (24).. However, it remained to be determined whether these reported chromosomal DNA sequences are abundant and uniform enough within human populations to be detected once diluted in the environment. Based on these considerations, the next steps toward the

application of these gene sequences for water quality monitoring applications were to design qPCR assays for their detection and then to use these assays to evaluate the overall abundance and distribution of these sequences in human populations relative to those of ribosomal RNA gene sequences from different currently recognized fecal indicator bacteria groups.

Here we report the development of two qPCR assays for the quantification of the human-specific DNA sequences targeted by previously reported PCR assays 19 and 22 (24). Method performance characteristics including specificity, range of quantification (ROQ), limit of quantification, amplification efficiency, and analytical precision were defined for each assay. An internal amplification control (IAC) was designed to monitor for the presence of inhibitors commonly associated with environmental sampling that can confound DNA target copy number estimations. Finally, the abundance of each DNA target was measured by qPCR analysis in primary effluent wastewater samples representative of 20 geographically distinct human populations. In addition, the abundance of these human-specific DNA genes in wastewater was compared to those rRNA genes from *Bacteroidales* and enterococci, two general fecal indicator bacterial groups that have been widely used for water quality testing.

MATERIALS AND METHODS

Sample collection. Individual fecal samples (n = 265) and wastewater samples (n = 20) were collected for analysis as previously described (24). Primary effluent wastewater samples were collected on-site from 20 different wastewater treatment facilities across

95 the United States (Tab. 1). Facilities were selected based on population served and
96 geographic location. Briefly, 500 ml of primary effluent was collected from each facility
97 and immediately stored on ice. Samples were then packed in ice and shipped overnight
98 to Cincinnati, OH for laboratory testing. Twenty-five milliliters of primary effluent from
99 each facility was filtered through a 0.2 μm pore size Supor-200 filters (Whatman) and
100 each filter was placed in a sterile 1.5 ml microtube and stored at -80°C (< 6 months) until
101 time of DNA extraction and qPCR amplification.
102 Individual fecal samples were collected over a 12-month period at various
103 locations across the United States from 22 different animal species likely to impact
104 watersheds or beaches including *Homo sapiens* (human, n = 16), *Lama pacos* (alpaca; n =
105 2), *Anser* sp. (Canadian goose; n = 12), *Felis catus* (cat, n =10), *Gallus gallus* (chicken, n
106 = 10), *Bos taurus* (cow, n = 80), *Odocoileus virginianus* (white-tail deer, n = 15),
107 *Odocoileus hemionus* (mule deer, n = 5), *Cervus elaphus* (elk, n = 5), *Alces alces* (moose,
108 n = 1), *Antilocapra american* (pronghorn, n = 4), *Canis familiaris* (dog, n = 10), *Anas* sp.
109 (duck, n = 12), *Capra aegagrus* (goat, n = 7), *Equus caballus* (horse, n = 12), *Pelecanus*
110 sp. (pelican, n = 5), *Sus scrofa* (pig, n = 22), Laridae (gull, n = 12), *Ovis aries*, (sheep, n =
111 10), *Zalophus californianus* (sea lion, n = 5), Delphinidae (marine dolphin, n = 3), and
112 *Meleagris* sp. (turkey, n = 7). Each fecal sample was collected from a different
113 individual to maximize the opportunity to observe false positive amplifications.
114
115 **DNA extraction of fecal and primary effluent wastewater samples.** All DNA
116 extractions were performed with the FastDNA kit for soils (Q-Biogene, Carlsbad, CA) as
117 described (24) with the exception that a FastPrep®-24 instrument (MP, Solon, OH) at

118 setting of 6 m/s for 120 sec was used for cell lysis. DNA extraction yields were
119 determined with a NanoDrop ND-1000 UV spectrophotometer (NanoDrop Technologies;
120 Wilmington, DE). Filtration and extraction controls, with purified water substituted for
121 primary effluent, were performed each day samples were received or extracted to monitor
122 for potential contamination.
123
124 **Oligonucleotides and primers.** TaqMan® probe and primer assays targeting the rRNA
125 genes of *Bacteroidales* (GenBac3) and *Enterococcus* (Entero1) are reported elsewhere (8,
126 13). qPCR probe and primer sequences for the putative human-specific HumM2 and
127 HumM3 assays (Tab. 2) were designed with Primer Express software (Applied
128 Biosystems, Foster City, CA) based on the previously reported end-point PCR assays
129 HumM19 and HumM22, respectively (24). Primers and TaqMan® probes were designed
130 using the default parameters of the Primer Express software (Version 1.5; Applied
131 Biosystems). Fluorogenic probes were 5' labeled with 6FAM (6-carboxyfluorescein) or
132 VIC™ and 3' labeled with TAMRA (6-carboxytetramethylrhodamine). Optimal primer
133 and probe reaction concentrations were determined according to a standard Applied
134 Biosystems protocol (1). The HumM2 and HumM3 assay primer and probe sets (Tab. 2)
135 were tested for specificity with animal fecal and wastewater sample composites (5 ng
136 DNA template per PCR assay).
137
138 **DNA preparations from pure bacterial cultures.** American Type Culture Collection
139 (ATCC) bacterial strains were used to prepare DNA standards for the *Bacteroidales* and
140 *Enterococcus* qPCR assays. *E. faecalis* (ATCC #29212) was cultured as previously

141 described (8). *B. thetaiotaomicron* (ATCC # 29741) cells were grown in chopped meat
142 carbohydrate broth (Remel, Lenexa, KS) according to manufacturer's instructions. Both
143 cultures were harvested by centrifugation at 8,000 *x g* for 5 min, washed twice using
144 sterile phosphate buffered saline (Sigma, St. Louis, MO) and stored in aliquots at -40ºC.
145 Cell concentrations of each organism in the final washed suspensions were determined by
146 bright field microscopy at 40x magnification in disposable hemocytometer chambers
147 (Nexcelom Bioscience, #CP2-002, Lawrence, MA). DNA was isolated from the cell
148 suspensions using a bead beating extraction approach (8) and incubated for one hour at
149 37ºC with 0.017 µg/µl RNase A (Gentra Systems, USA). DNA purification was
150 performed using a silica column adsorption kit (DNA-EZ, GeneRite, Kendall Park, NJ.).
151 DNA concentrations of cell extracts were determined by spectrophotometric absorbance
152 readings at 260 nm (A_{260}) and purity of the DNA preparations was determined by
153 A_{260}/A_{280} ratios.
154
155 **Construction of IAC and plasmid DNA standards**. A plasmid DNA construct was
156 developed to function as an IAC DNA target that can be spiked into DNA extracts to
157 monitor for PCR inhibition and also as a plasmid DNA standard for calculation of
158 HumM2 and HumM3 qPCR assay calibration curves. The IAC construct was designed to
159 contain a single site for hybridization of a unique TaqMan® VIC™ labeled probe
160 sequence flanked by multiple primer binding sequences (Tab. 2, Fig. 1). To build the
161 human assay IAC construct, long oligonucleotides (> 100 bp, Tab. 1) containing multiple
162 primer sequences (24) were designed such that their 3' ends overlapped. The overlapping
163 fragments were then combined into a single DNA molecule using overlap extension PCR

(9). The IAC construct was inserted into a plasmid vector, purified, linearized, quantified, and diluted to generate samples ranging from approximately 10 to $1x10^6$ molecules of template DNA as described (22).

qPCR assays and quantification. The four qPCR assays used in this study were HumM2, HumM3, GenBac3, and Entero1 (Tab. 2). Amplification was performed in a 7900 HT Fast Real Time Sequence Detector (Applied Biosystems). Reaction conditions and thermal cycling parameters for GenBac3 and Entero1 are described elsewhere (20). For HumM2 and HumM3, reaction mixtures (25 µl) contained 1X TaqMan® Universal PCR Master Mix with AmpErase® uracil-N-glycosylase (UNG, Applied Biosystems), 0.2 mg/ml bovine serum albumin (Sigma), 1 µM of each primer, 80 nM FAM™ or VIC™ labeled TaqMan® probe (Applied Biosystems), and either 1 to 100 ng genomic DNA (fecal and wastewater samples) or 10 to $1x10^6$ target gene copies (human IAC plasmid DNA). Reaction mixtures for multiplex applications were the same as above with the additions of both 80 nM of VIC™ or TET labeled TaqMan® probes for IAC plasmid DNA and 80 nM of 6FAM™ labeled TaqMan® probe for native DNA targets. IAC spike concentrations were either 25 or 50 copies. All reactions were performed in triplicate in MicroAmp Optical 96-well reaction plates with MicroAmp Optical Caps (Applied Biosystems). Thermal conditions were 50ºC for 2 min to activate UNG followed by 10 min incubation at 95ºC to activate AmpliTaq Gold enzyme, the temperature profile then followed a forty cycle pattern with a short denaturation at 95ºC for 15 sec and a combined annealing and primer extension phase at 60ºC for 1 min. Data was initially analyzed with Sequence Detector Software (Version 2.2.2) at a threshold

187 determination of 0.08 for human-specific assays (HumM2 and HumM3) and 0.03 for
188 general fecal indicator bacteria assays (GenBac3 and Entero1). Threshold cycle (C_T)
189 values were exported to Microsoft Excel for further statistical analysis. A minimum of
190 three no-template amplifications with purified water substituted for template DNA were
191 performed for each 96-well qPCR experiment to monitor for potential contamination.
192
193 **Calculations and statistical analysis.** The specificity of HumM2 and HumM3 was
194 determined as: specificity = d/(b+d), where 'b' is false positives and 'd' is true negatives.
195 Master calibration curves, unknown DNA concentration estimates, and credible intervals
196 were determined using a Markov Chain Monte Carlo approach (25). Bayesian
197 calculations were performed using the publicly available software WinBUGS version
198 1.4.1 (http://www.mrc-bsu.cam.ac.uk/bugs) (14). See supplemental material for
199 WinBUGS program code and resulting data output used to develop master calibration
200 curves for HumM2 and HumM3 DNA standards. An analysis of covariance (ANCOVA)
201 model was used to compare the intercept and slope of individual standard curves used to
202 calculate the master calibration equations. One-way ANOVA tests comparing C_T values
203 from reactions with known amounts of standard target DNA were used to define the
204 range of quantification (ROQ) for each assay. The precision of C_T measurements
205 determined from DNA standards was expressed as a percent coefficient of variation (CV,
206 standard deviation expressed as a percentage of the mean). A one-way random effect
207 ANOVA model (with location as random factor) was used to test the hypotheses that the
208 variability between untreated wastewater sample locations was zero. A paired two

sample t-test was used to compare the overall mean difference in C_T values between HumM2, HumM3, Entero1, and GenBac3 qPCR mean C_T values for wastewater samples.

RESULTS

Master calibration curves and range of quantification. Overall fitted curves representing multiple independent runs of the DNA standards were compared using ANCOVA tests. Independent fitted curves for each qPCR assay demonstrated a significant difference in intercepts ($p < 0.05$), but no difference in slopes ($p > 0.05$). Calibration curve equations and performance characteristics of the four qPCR assays are shown in Table 3. Calibration curves for GenBac3 and Entero1 general fecal indicator bacteria assays were generated from eight independent runs using genomic DNA standards extracted from cultured cell suspensions whereas HumM2 and HumM3 fitted curves were generated from 12 independent fitted curves each using plasmid DNA standards. ROQs spanned the entire range of standard concentrations tested for all qPCR assays including 10 to $1x10^6$ copies for human-specific assays and 40 to $4x10^4$ copies for general fecal indicator bacteria assays. Precision of C_T measurements across defined ROQs for all assays was less than 3% CV and amplification efficiencies ranged from 1.87 to 1.99 (Tab. 3). No-template controls indicated the absence of contamination in 98.9% of qPCR experiments and all extraction blanks tested negative for presence of extraneous DNA molecules.

Evaluation of multiplex host-specific qPCR application. A composite synthetic internal control was developed for each host-specific assay to monitor fecal and wastewater DNA extracts for potential PCR inhibition. The IAC construct was designed with the intention to allow target DNA and an IAC to be coamplified with the same set of primers, under the same reaction conditions, in the same PCR tube. The target DNA and IAC product could then be detected and quantified simultaneously with different fluorescently labeled TaqMan® probes provided that (i) there is no significant difference ($p > 0.05$) between simplex and multiplex standard curve intercepts and slopes and (ii) a fixed amount of IAC could be quantified across a range of genomic DNA standard concentrations (22). An IAC spike of 50 copies was undetectable at human fecal DNA concentrations ranging from 1 to 100 ng for the HumM2 assay, while a significant difference between simplex and multiplex curve intercepts and slopes was observed for the HumM3 assay ($p < 0.05$) suggesting that neither of these assays is reliable as a multiplex reaction (data not shown). The failure of both assays to perform in a multiplex environment is most likely due to competition between genomic target DNA (6FAM labeled) and the IAC spike (VIC labeled). Thus, only the HumM3 IAC could be used to monitor for PCR inhibition and only in a simplex application.

Monitoring for PCR Inhibition in DNA Extracts. DNA isolation from wastewater and fecal samples may not remove all substances that can interfere with qPCR and the degree of interference may vary between samples. Therefore, internal controls designed to evaluate the suitability of isolated DNA for quantitative analysis were included for each DNA extract. All fecal DNA extracts were screened for inhibition of the HumM3 IAC

assay. The criterion for concluding no significant PCR inhibition of the HumM3 IAC assay by these samples was established as a $C_T = 34.6 \pm 1.65$, based on repeated experiments measuring the simplex mean C_T and standard deviation values for control reactions containing 50 copies of IAC in buffer (Fig. 2, panel A). Wastewater DNA extracts were also tested using the previously reported multiplex Entero1 application with a 25 copy IAC spike. The criterion for concluding no significant PCR inhibition in these assays was defined as a C_T of 34.0 ± 1.41 (Fig.2, panel B). IAC analyses indicated the absence of PCR inhibitors in all fecal and untreated wastewater DNA extracts based on both of these criteria.

Specificity of host-specific qPCR assays. Specificity of the HumM2 and HumM3 assays was tested with a reference collection of fecal samples from hundreds of non-target animals (Tab. 4). HumM2 and HumM3 assays exhibited specificity values of 99.2% and 97.2%, respectively. HumM2 elicited false positives with two chicken fecal samples (C_T 29.3 ± 0.16 and 29.1 ± 0.14), while HumM3 cross-reacted with a single elk sample (C_T 33.6 ± 0.35) and six sheep samples (C_T ranging from 24.4 ± 0.05 to 36.9 ± 0.73). Both assays successfully detected respective DNA targets in all human fecal and primary effluent wastewater DNA extracts (Tab. 4).

Quantification of fecal bacterial genes in untreated wastewater. Primary effluent wastewater samples were collected from 20 different geographic locations to characterize target DNA variability between localities and to compare the relative abundance of each target DNA to enteroccci and general *Bacteroidales* 16S rRNA genes. A one-way

277 random effect ANOVA model indicated that there is significant variability ($p < 0.05$) in
278 C_T values among all locations, for each assay. Variance (σ^2) between wastewater sample
279 locations ranged from 0.30 for HumM2, 1.06 for HumM3, 1.65 for Entero1 to 0.45 for
280 GenBac3. Target DNA relative abundance for each assay was compared by normalizing
281 data sets to 1 ng of template DNA and plotting $\log_{10}$ mean copy number estimates for
282 each wastewater sample by qPCR assay. A box-and-whisker diagram was used to
283 display differences between wastewater sample DNA target estimates for each qPCR
284 assay including the smallest observation, lower quartile (25^{th} percentile), median, upper
285 quartile (75^{th} percentile), largest observation, and outliers (Fig. 3).
286
287 **DISCUSSION**
288
289 **Human-specific qPCR.** We report on two qPCR assays that detect predominantly
290 human fecal DNA when tested against a panel of samples representing agriculturally
291 important animals such as cattle, poultry, and swine as well as many wildlife species.
292 These qPCR assays were designed to target the same gene sequences as two end-point
293 PCR assays (Assays 19 and 22) that were previously reported to be 100% human-specific
294 based on a fecal reference collection consisting of 160 individual samples representing 11
295 different animal species (24). The slight decrease in specificity of the real-time qPCR
296 assays compared to the end-point PCR assays may be due to the larger non-target fecal
297 sample reference library used to establish specificity values or factors associated with the
298 TaqMan® qPCR approach such as constraints in primer design, PCR reagent chemistry,
299 thermal cycling settings, and increased number of amplification thermal cycles.

Regardless of the reason, the HumM2 and HumM3 qPCR assays exhibit extremely high levels of specificity exceeding 97.5%.

Master calibration curves were used in this study due to the large numbers of fecal and wastewater samples processed and the need to maximize the number of samples in each experiment set-up and reduce expenses. Each master curve was compiled from up to 12 independent runs in order to reflect sources of intra- and inter-run variability. Master calibration curves were acceptable in this study because (i) there was no significant differences in the slopes of fitted curves between independent runs ($p > 0.05$), (ii) the analytical precision (%CV) over the ROQ between runs averaged less than 3%, and (iii) the Bayesian approach accounts for run-to-run variability with a 95% credible interval when generating fitted calibration curves (25).

Abundance of host-specific and fecal indicator genes. Little is known regarding the abundance and geographical distribution of human-specific genes in sewage. In this study, we tested 20 primary effluent wastewater samples collected from different geographic locations in the United States ranging from Hawaii to Florida. Wastewater samples were representative of approximately 4.1 million individuals responsible for generating an average of 5,180 million gallons of raw sewage per year and were ideal for estimating the abundance of host-specific gene targets in different human populations. Host-specific and general fecal indicator bacteria qPCR assays successfully detected respective genetic targets from 1 ng of DNA for 100% of the wastewater samples regardless of locality. The general *Bacteroidales* assay (GenBac3) detected the highest target gene concentration in all samples, which supports previous research reporting that

323 *Bacteroidales* often makes up a large portion of the human fecal bacterial community (6,
324 16, 28). The HumM2 and HumM3 gene targets were the next most abundant markers
325 and more prevalent than the enterococcal 23S rRNA genes (Fig. 3). Enterococci are
326 routinely detected in fecal polluted waters (26). The observation that host-specific gene
327 targets are more abundant than enterococcal 23S rRNA genes suggests that detectable
328 quantities of HumM2 and HumM3 gene targets may be present in ambient waters.
329 All qPCR assays exhibited less than 3.9% dispersion of C_T values from an overall
330 wastewater sample mean [(one-way random effect ANOVA qPCR standard
331 deviation/mean) x 100] regardless of gene target. In addition, a significant difference (p
332 < 0.05) was observed in concentrations of all qPCR gene targets between wastewater
333 geographic locations. Fluctuations in relative gene target concentrations between
334 wastewater samples could result from differences in local population diet, age, and/or
335 health, but could also reflect uncertainty associated with single sample events.
336 Regardless of the reason, low dispersion percentages (< 3.9%) suggest that the human-
337 specific gene targets can be detected with a similar level of confidence as 16S rRNA
338 general *Bacteroidales* and 23S rRNA enteroccocal gene targets.
339
340 **Implications for microbial source tracking (MST).** Recreational and drinking source
341 waters continue to be impacted by human fecal pollution and can impose a direct threat to
342 human health (4, 5, 19). In addition to human waste, many other agricultural and wildlife
343 animal sources can contribute to the total fecal load. Most MST methods attempt to
344 identify specific fecal sources to help local authorities prioritize polluted areas for
345 restoration. Recent advances in PCR-based methods now allow for the estimation of host-

346 specific DNA target concentrations. These quantitative approaches can extend the utility
347 of MST applications by supplying information regarding the concentration of host-
348 specific fecal pollution sources. To date, no qPCR-based method has been found to be
349 100% specific for human fecal pollution (10, 12, 18, 21). Animals that cohabitate with
350 humans such as cats and dogs, and animals that share similar digestive physiologies, such
351 as pigs, are the most problematic. Fecal pollution originating from pets can confound
352 MST studies where cat and dog waste is mixed with sewage and/or runoff after rain
353 events. A similar problem can arise in watersheds impacted by swine sources of fecal
354 pollution. HumM2 and HumM3 are the first qPCR assays available that can discriminate
355 between all three of these sources of fecal pollution. In addition, these assays can
356 quantify as few as 10 copies of target DNA per reaction with a high degree of precision.
357 DNA targets of these assays were widely distributed among 20 different human
358 populations and more abundant than fecal enterococci in almost all wastewater samples
359 tested.

360 To explore the potential of the HumM2 and HumM3 qPCR assays for
361 environmental monitoring, each assay underwent preliminary testing with DNA isolated
362 from river, stream, and storm water samples (n = 6). All six samples contained general
363 *Bacteroidales* target sequences (GenBac3 C_T values ranging from 34.3 ± 1.26 to 26.2 ±
364 0.10) suggesting the presence of fecal pollution. Two of these samples, both collected
365 from locations situated within 100 m downstream of a wastewater discharge pipe,
366 generated C_T values for both host-specific assays (C_T values ranging from 35.8 ± 0.46 to
367 34.7 ± 0.32). These preliminary results combined with the high levels of specificity and
368 broad distribution of their DNA targets in wastewater samples suggest that the HumM2

and HumM3 assays may have future utility in MST applications. However, to realize the full potential of these qPCR assays, several issues remain to be addressed. Future studies characterizing the survival of target DNA molecules through the wastewater treatment process and in the environment are needed to generate reliable estimates of the impact of these sources on ambient water samples. Research projects focusing on the relevance of each qPCR assay to current culture-based and qPCR-based fecal indicator methods (such as *E. coli* and enterococi) are also critical for successful MST applications. Finally, epidemiological studies are necessary to establish any links between the prevalence of host-specific DNA targets and relevant public health risks.

ACKNOWLEDGEMENTS

The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed, or partially funded and collaborated in, the research described herein. It has been subjected to the Agency's peer and administrative review and has been approved for external publication. Any opinions expressed in this paper are those of the author (s) and do not necessarily reflect the views of the Agency, therefore, no official endorsement should be inferred. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

1. **ABI.** 2002. TaqMan Universal PCR Master Mix Protocol. Applied Biosystems.
2. **Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj, and M. F. Polz.** 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. Journal of Bacteriology **186:**2629-2635.
3. **Andersson, A. F., M. Lindberg, H. Jackobsson, F. Backhed, P. Nyren, and L. Engstrand.** 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. Plos One **3:**e2836.
4. **Balarajan, R., V. Soni Raleigh, P. Yuen, D. Wheeler, D. Machin, and R. Caftwright.** 1991. Health risks associated with bathing in sea water. British Medical Journal **303:**1444-1445.
5. **Dufour, A. P.** 1984. Health effects criteria for fresh recreational waters. U.S. Environmental Protection Agency.

6. **Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman.** 2005. Diversity of human intestinal microbial flora. Science **308:**1635-1638.
7. **Grimont, F., and P. A. Grimont.** 1986. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. Annuals of the Institute of Pasteur in Microbiology **137B:**165-175.
8. **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.
9. **Higuchi, R., B. Krummel, and R. K. Saiki.** 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Research **16:**7351-7367.
10. **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.
11. **Kuwahra, T., A. Yamashita, H. Hirakawa, H. Nakayama, H. Toh, N. Okada, S. Kuhara, M. Hattori, T. Hayashi, and Y. Ohnishi.** 2004. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. Proceedings of the National Academy of Sciences of the United States of America **101:**14919-149-24.
12. **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.
13. **Ludwig, W., and K. H. Schleifer.** 2000. How quantitative is quantitative PCR with respect to cell counts? Systematic and Applied Microbiology **23:**556-562.
14. **Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter.** 2000. WinBUGS-A Bayesian modeling framework: Concepts, structure, and extensibility. Statistics and Computing **10:**325-337.
15. **Matsuki, T., K. Watanabe, J. Fujimoto, T. Takada, and R. Tanaka.** 2004. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. Applied and Environmental Microbiology **70:**7220-7228.
16. **Moore, W. E. C., and I. V. Holdeman.** 1974. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. Applied and Environmental Microbiology **27:**961-979.
17. **Nelson, K. E., R. D. Fleischmann, R. DeBoy, I. T. Paulsen, D. E. Fouts, J. A. Eisen, S. C. Daugherty, R. J. Dodson, A. S. Durkin, M. Gwinn, D. H. Haft, J. Kolonay, W. C. Nelson, T. Mason, L. Tallon, J. Gray, D. Granger, H. Tettelin, H. Dong, J. L. Galvin, M. J. Duncan, F. E. Dewhirst, and C. M. Fraser.** 2003. Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83. Journal of Bacteriology **185:**5591-5601.
18. **Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of

fecal pollution in freshwater. Applied Microbiology and Biotechnology **74:**890-901.

19. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002. Microbial source tracking: current methodology and future directions. Applied and Environmental Microbiology **68:**5796-5803.

20. **Seifring, S. C., M. Varma, E. Atikovic, L. J. Wymer, and R. A. Haugland.** 2008. Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. Journal of Water and Health **6:**225-237.

21. **Seurinck, S., T. Defoirdt, W. Verstraete, and S. Siciliano.** 2005. Detection and quantification of 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.

22. **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.

23. **Shanks, O. C., J. W. Santo Domingo, R. Lamendella, C. A. Kelty, and J. E. Graham.** 2006. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Applied and Environmental Microbiology **72:**4054-4060.

24. **Shanks, O. C., J. W. Santo Domingo, J. Lu, C. A. Kelty, and J. E. Graham.** 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. Applied and Environmental Microbiology **73:**2416-2422.

25. **Sivaganesan, M., S. Seifring, M. Varma, R. A. Haugland, and O. C. Shanks.** 2008. A Bayesian method for calculating real-time quantitative PCR calibration curves using absolute plasmid DNA standards. BMC Bioinformatics **9:**120.

26. **USEPA.** 2000. Improved enumeration methods for recreational water quality indicators: enterococci and *Escherichia coli* Report EPA-821-R-97-004. United States Environmental Protection Agency.

27. **Wang, R. F., M. L. Beggs, B. D. Erickson, and C. E. Cerniglia.** 2004. DNA microarray analysis of predominant human intestinal bacteria in fecal samples. Molecular and Cellular Probes **18:**223-234.

28. **Wood, J., K. P. Scott, G. Avgustin, C. J. Newbold, and H. J. Flint.** 1998. Estimation of the relative abundance of different *Bacteroides* and *Prevotella* ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. Applied and Environmental Microbiology **64:**3683-3689.

29. **Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. I. Gordon.** 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science **299:**2074-2076.

30. **Zhang, Y., D. Zhang, L. Wenquan, J. Chen, Y. Peng, and W. Cao.** 2003. A novel real-time quantitative PCR method using attached universal template probe. Nucleic Acids Research **31:**e123.

493 **Table 1:** Primary effluent wastewater sample information.
494

Facility	Location	Population Served	Inflow (mgd)
Sacramento RWTP	Sacramento, CA	1,200,000	168
Clarksburg WWTP	Clarksburg, WV	24,498	8.65
Lincoln Northeast WWTF	Lincoln, NE	55,000	5
Lower East Fork WWTP	Milford, OH	55,000	6.53
West Point WWTP	Seattle, WA	1,400,000	98.1
Crystal Lake WWTP No.2	Crystal Lake, IL	38,600	5.8
Little Falls WWTP	Little Falls, NY	49,000	5.14
Wildcat Hill WWTP	Flagstaff, AZ	60,000	3.3
Northwest Bergen County WWTP	Waldwick, NJ	102,448	10
Moorehead WWTP	Moorehead, KY	20,454	2.5
Buffalo WWTP	Buffalo, MO	6,000	0.72
Saginaw WWTP	Saginaw, MI	57,523	25
Bonner Springs WWTP	Bonner Springs, KS	7,500	0.53
Frankurt Sewer Department	Frankfurt, KY	48,000	6.69
Old Town PCF	Old Town, ME	9,500	1.2
Rutland WWTP	Rutland, VT	22,000	5.85
Maui County Kahului WWTF	Kahului, HI	41,720	4.3
City of St. Peter WWTP	St. Peter, MN	10,850	1.1
Las Vegas WWTP	Las Vegas, NV	815,207	68
Marshall St. Advanced WWTP	Clearwater, FL	65,000	5.3
	Totals	**4,088,300**	**431.7**

495
496 Inflow indicates the average rate of sewage influent at each treatment facility reported in million gallons
497 per day (mgd).

498 **Table 2:** Oligonucleotides, primers, and probes.
499

Assay	Primer and probe sequences (5' to 3')	Size (bp)	Reference
GenBac3	Forward: GGGGTTCTGAGAGGAAGGT; Reverse: AGTAGCGGAAGGATGACGG; Probe: FAM-CAATATTCCTCACTGCTGCCTCCCGTA-TAMRA	129	(20)
Entero1	Forward: AGAAATTCCAAACGAACTTG; Reverse: AATGATGGAGGTAGAGCAC; Probe: FAM-TGGTTCTCTCCGAAATAGCTTTAGGGCTA-TAMRA	92	(13)
HumM2	Hum2F: CGTCAGGTTTGTTTCGGTATTG; Hum2R: TCATCACGTAACTTATTTATATGCATTAGC; Probe: FAM-TATCGAAAATCTCACGGATTAACTCTTGTGTACGC-TAMRA	101	This Study
HumM3	Hum3F: GTAATTCGCGTTCTTCCTCACAT; Hum3R: GGAGGAAACAAGTATGAAGATAGAAGAATTAA; Probe: FAM-AGGTCTGTCCTTCGAAATAGCGGT-TAMRA	83	This Study
Human IAC	Frag1:GATCATGAGTTCACATGTCCGAGTAATTCGCGTTCTTCCTCACATACGTCAGGTTT GTTTCGGTATTG AGTTAGGAACAGGCGGCGACGAATG TTAATCTTCTATCTTC; Frag2:TCCGGTGATGTCTCGAGAGTGTCTCATCACGTAACTTATTTATATGCATTAGCGGT GAAGGTCTGGGAGGAAACAAGTATGAAGATAGAAGAATTAACATTCGTCGCCGC; Frag3:AGTTAGGAACAGGCGGCGACGAATGTTAATTCTTCTATCTTCATACTTGTTTCCTC CCAGACCTTCACCGCTAATGCATATAAATAAGTTACGTGATGAGACACTCTCGA; Frag4:CCGTCATCCTTCACGCTACTGATGTCTGCATGGTATATGTTGAGTGCAATGGGATT TTATCCGGTGAATCCGGTGATGTCTCGAGAGTGTCTCATCACGTAACTTATTTA; Probe: VIC-TAGGAACAGGCGGCGACGA-TAMRA[a]	258	(30); This Study

500
501 [a] The TaqMan probes was modified from the previously reported UT probe (30).
502

503 **Table 3:** Calibration curve equations and performance characteristics of qPCR assays.
504

Assay	Calibration Equation	Amplification Efficiency [a]	ROQ (copies) for target DNA	%CV across ROQ	Method [b]
Entero1	Y = 38.0 -3.42X	1.96	40-4x10^4	2.24	Multiplex
GenBac3	Y = 38.1 -3.34X	1.99	40-4x10^4	2.92	Simplex
HumM2	Y = 41.8 -3.67X	1.87	10-1x10^6	2.46	Simplex
HumM3	Y= 41.9 -3.66X	1.88	10-1x10^6	2.40	Simplex

505
506 [a] Amplification efficiency = $10^{(1/-slope)}$.
507 [b] Either a simplex approach or multiplex strategy where the target DNA was simultaneously detected with
508 an IAC.

509 **Table 4:** Specificity of HumM2 and HumM3 qPCR assays.
510

Animal Source	No.	HumM2		HumM3	
		Average C_T	*St. Dev.*	*Average* C_T	*St. Dev.*
Alpaca	2	.	.	.	.
Cow	80	.	.	.	.
Goat	7	.	.	.	.
Sheep	10	34.1	0.03	.	.
Horse	12	.	.	.	.
Pig	22	.	.	.	.
Antelope	4	.	.	.	.
Whitetail Deer	15	.	.	.	.
Mule Deer	5	.	.	.	.
Moose	1	.	.	.	.
Elk	5	35.7	0.24	.	.
Canadian Goose	12	.	.	.	.
Duck	12	.	.	.	.
Pelican	5	.	.	.	.
Gull	12	.	.	.	.
Turkey	7	.	.	.	.
Chicken	10	.	.	32.1	0.30
Marine Dolphin	3	.	.	.	.
California Sea Lion	5	.	.	.	.
Cat	10	.	.	.	.
Dog	10	.	.	.	.
Human	16	29.7	0.03	30.3	0.10
Wastewater	20	31.8	0.54	32.8	1.01
Total	**285**				

511
512 C_T values generated from 1 ng of total DNA.

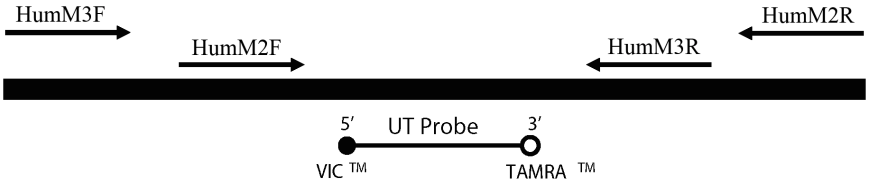
Fig. Captions:

Fig1: Diagram of human-specific plasmid DNA IAC composite construct. The IAC (258 bp) consists of a VIC-labeled universal probe binding site (30) flanked by primer sequences for HumM2 (101 bp) and HumM3 (83 bp) qPCR assays.

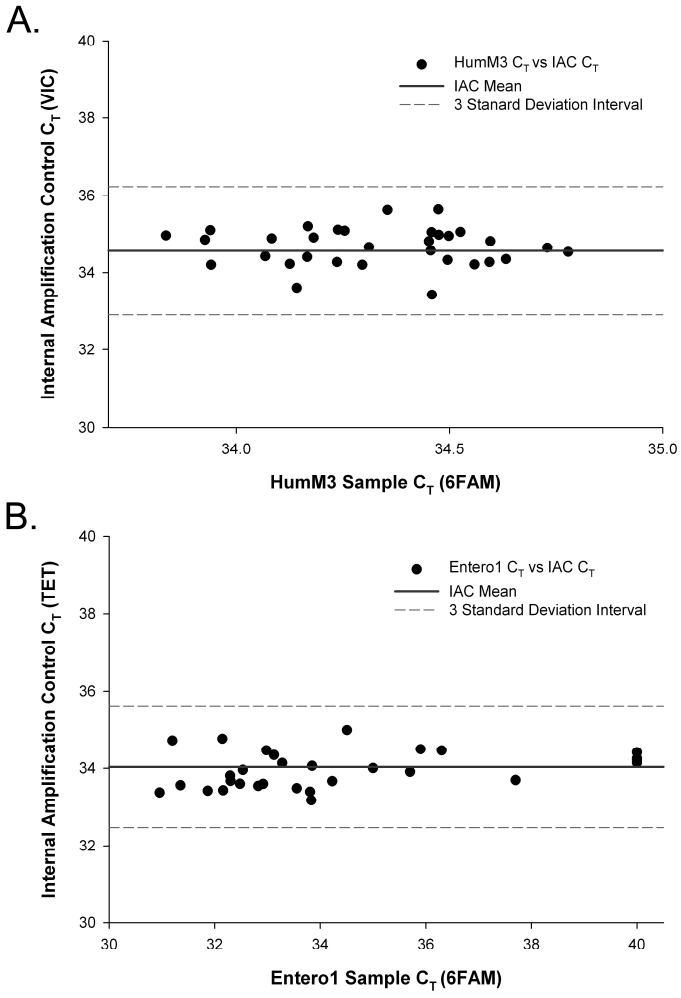
Fig2: Results of qPCR IAC inhibition tests for fecal and wastewater DNA extracts. Scatter plots show IAC (VIC® or TET probes) and genomic DNA (6FAM probe) C_T values from analyses of fecal DNA extracts using HumM3 (Panel A) and wastewater DNA extracts using Entero1 (Panel B). Confidence intervals (dashed lines) represent three standard deviations of the mean IAC C_T (solid lines; HumM3 C_T = 34.6 and Entero1 C_T = 34.0) established from repeated control experiments.

Fig3: Box-and-whisker diagram depicting the relative abundance of gene targets from HumM2, HumM3, Entero1, and GenBac3 qPCR assays from all primary effluent sewage sample locations. Estimated gene target concentrations are reported as $\log_{10}$ mean copy number per ng of total DNA. The boundary of the box closest to zero indicates the 25^{th} percentile, the line within the box represents the median, and the boundary of the box farthest from zero indicates the 75^{th} percentile. Whiskers above and below the box indicate the 10^{th} and 90^{th} percentiles. A "+" denotes outlier measurements.

Fig. 1:

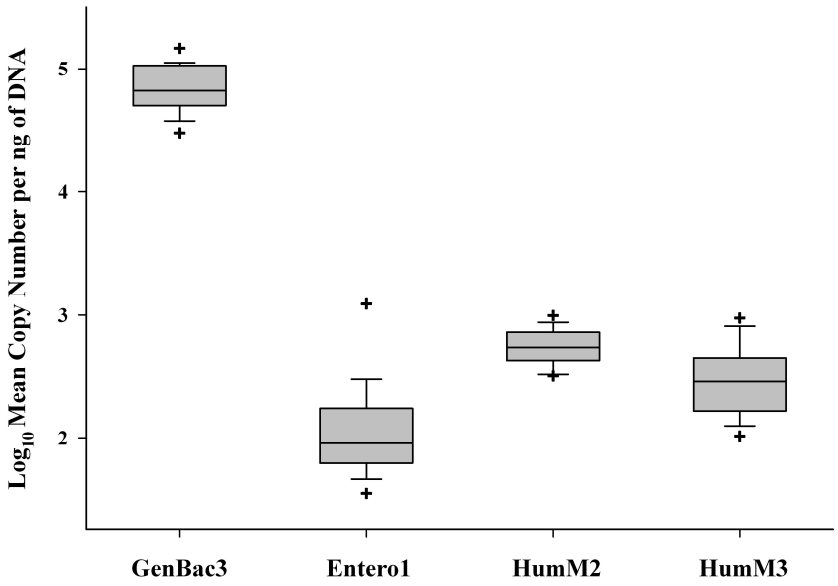


537 **Fig. 2:**
538



539
540

541 **Fig. 3:**
542



543
544