Quantitative PCR for Genetic Markers of Human Fecal Pollution


U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, and U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268. Received 6 February 2009/Accepted 2 July 2009

Assessment of health risk and fecal bacterial 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. Each assay exhibited a range of quantification from 10 to 1 × 10^6 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 in primary effluent wastewater samples collected from 20 geographically distinct locations was measured and compared to quantities estimated by real-time PCR assays specific for 16S rRNA 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.

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, 20). The majority of these methods are based on real-time quantitative PCR (qPCR) assays designed to estimate the concentrations of 16S rRNA gene 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 Bacteroidales 16S rRNA gene-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, 20). Alternative qPCR assays targeting genes directly involved in host-specific interactions may be capable of increased discrimination of fecal pollution sources (22, 23) 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 bacterium-specific DNA sequences (23). PCR assays targeting two gene sequences encoding 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 (23). 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. On the basis of 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 rRNA gene sequences from different currently recognized fecal indicator bacterial groups.

Here, we report the development of two qPCR assays for quantification of the human-specific DNA sequences targeted by previously reported PCR assays 19 and 22 (23). 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 in primary effluent wastewater samples representative of 20 geographically distinct human populations was measured by qPCR analysis. In addition, the abundances of these human-specific DNA genes in wastewater were compared to those of rRNA genes of 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 (23). Primary effluent wastewater samples were collected on-site from 20 different wastewater treat-
### TABLE 1. Primary effluent wastewater sample information

<table>
<thead>
<tr>
<th>Facility</th>
<th>Location</th>
<th>Population served</th>
<th>Inflow (mgd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacramento RWTP</td>
<td>Sacramento, CA</td>
<td>1,200,000</td>
<td>168</td>
</tr>
<tr>
<td>Clarksville WWTP</td>
<td>Clarksville, WV</td>
<td>24,498</td>
<td>8.65</td>
</tr>
<tr>
<td>Lincoln Northeast WWTF</td>
<td>Lincoln, NE</td>
<td>55,000</td>
<td>5</td>
</tr>
<tr>
<td>Lower East Fork WWTP</td>
<td>Milford, OH</td>
<td>55,000</td>
<td>6.5</td>
</tr>
<tr>
<td>West Point WWTP</td>
<td>Seattle, WA</td>
<td>1,400,000</td>
<td>98.1</td>
</tr>
<tr>
<td>Crystal Lake WWTP No.2</td>
<td>Crystal Lake, IL</td>
<td>38,600</td>
<td>5.8</td>
</tr>
<tr>
<td>Little Falls WWTP</td>
<td>Little Falls, NY</td>
<td>49,000</td>
<td>5.14</td>
</tr>
<tr>
<td>Wildcat Hill WWTP</td>
<td>Flagstaff, AZ</td>
<td>60,000</td>
<td>3.3</td>
</tr>
<tr>
<td>Northwest Bergen County WWTF</td>
<td>Waldwick, NJ</td>
<td>102,448</td>
<td>10</td>
</tr>
<tr>
<td>Moorhead WWTP</td>
<td>Moorhead, KY</td>
<td>20,454</td>
<td>2.5</td>
</tr>
<tr>
<td>Buffalo WWTP</td>
<td>Buffalo, MO</td>
<td>6,000</td>
<td>0.72</td>
</tr>
<tr>
<td>Saginaw WWTP</td>
<td>Saginaw, MI</td>
<td>57,525</td>
<td>25</td>
</tr>
<tr>
<td>Bonner Springs WWTP</td>
<td>Bonner Springs, KS</td>
<td>7,500</td>
<td>0.53</td>
</tr>
<tr>
<td>Frankfurt Sewer Department</td>
<td>Frankfurt, KY</td>
<td>48,000</td>
<td>6.99</td>
</tr>
<tr>
<td>Old Town PCF</td>
<td>Old Town, ME</td>
<td>9,500</td>
<td>1.2</td>
</tr>
<tr>
<td>Rutland WWTP</td>
<td>Rutland, VT</td>
<td>22,000</td>
<td>5.85</td>
</tr>
<tr>
<td>Maui County Kahului WWTF</td>
<td>Kahului, HI</td>
<td>41,720</td>
<td>4.3</td>
</tr>
<tr>
<td>City of St. Peter WWTP</td>
<td>St. Peter, MN</td>
<td>10,850</td>
<td>1.1</td>
</tr>
<tr>
<td>Las Vegas WWTP</td>
<td>Las Vegas, NV</td>
<td>815,207</td>
<td>68</td>
</tr>
<tr>
<td>Marshall St. Advanced WWTP</td>
<td>Clearwater, FL</td>
<td>65,000</td>
<td>5.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>4,088,300</strong></td>
<td><strong>431.7</strong></td>
</tr>
</tbody>
</table>

*a Inflow indicates the average rate of sewage influent at each treatment facility, reported in millions of gallons per day (mgd).*
human-specific HumM2 and HumM3 assays (Table 2) were designed with Primer Express software (Applied Biosystems, Foster City, CA) based on the previously reported end point PCR assays HumM19 and HumM22, respectively (23). Primers and TaqMan probes were designed using the default parameters of the Primer Express software (version 1.5; Applied Biosystems). Fluorogenic probes were 5’ labeled with 6-carboxyfluorescein (FAM) or VIC and 3’ labeled with 6-carboxytetramethylrhodamine (TAMRA). Optimal primer and probe reaction concentrations were determined according to a standard Applied Biosystems protocol (1). The HumM2 and HumM3 assay primer and probe sets (Table 2) were tested for specificity with animal fecal and wastewater sample composites (5 ng DNA template per PCR assay).

DNA preparations from pure bacterial cultures. American Type Culture Collection (ATCC) bacterial strains were used to prepare DNA standards for the Bacteroidales and Enterococcus qPCR assays. Enterococcus faecalis (ATCC 29212) was cultured as previously described (8). Bacteroides thetaiotaomicron (ATCC 29741) cells were grown in chopped-meat carbohydrate broth (Remel, Lenexa, KS) in accordance with the manufacturer’s instructions. Both cultures were harvested by centrifugation at 8,000 × g for 5 min, washed twice using sterile phosphate-buffered saline (Sigma, St. Louis, MO), and stored in aliquots at −40°C. The cell concentrations of each organism in the final washed suspensions were determined by bright-field microscopy at ×40 magnification in a disposable hemocytometer chambers (no. CP-002; Nexcelon Bioscience, Lawrence, MA). DNA was isolated from the cell suspensions by using a bead-beating extraction approach (8) and incubated for 1 hour at 57°C with 0.017 μg/ml RNase A (Genta Systems). DNA purification was performed using a silica column extraction approach (8) and incubated for 1 hour at 37°C with 0.017

FIG. 1. 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.

qPCR FOR HUMAN FECAL BACTERIA 5509

amplification efficiency = 10[(Ct - Ct,unknown)]

either a simplex approach or a multiplex strategy where the target DNA was simultaneously detected with an IAC.
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 $1 \times 10^6$ copies for human-specific assays and 40 to $4 \times 10^6$ copies for general fecal indicator bacterial assays. The 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 (Table 3). No-template controls indicated the absence of contamination in 98.9% of qPCR experiments, and all extraction blanks tested negative for the 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 of allowing 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 (21). 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 (FAM 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 that there was no significant PCR inhibition of the HumM3 IAC assay by these samples was established as a $C_T$ of $34.6 \pm 1.65$, based on repeated experiments measuring the simplex mean $C_T$ and standard deviation values for control reaction mixtures containing 50 copies of IAC in buffer (Fig. 2A). Wastewater DNA extracts were also tested using the previously reported multiplex Enterol1 application with a 25-copy IAC spike. The criterion for concluding that there was no significant PCR inhibition in these assays was defined as a $C_T$ of $34.0 \pm 1.41$ (Fig. 2B). IAC analyses indicated the absence of PCR inhibitors in all fecal and untreated wastewater DNA extracts on the basis of both of these criteria.

**Specificity of host-specific qPCR assays.** The specificities of the HumM2 and HumM3 assays were tested with a reference collection of fecal samples from hundreds of nontarget animals (Table 4). HumM2 and HumM3 assays exhibited specificity values of 99.2% and 97.2%, respectively. HumM2 elicited false positives with two chicken fecal samples ($C_T$ values of $29.3 \pm 0.16$ and $29.1 \pm 0.14$), while HumM3 cross-reacted with a single elk sample ($C_T$ of $33.6 \pm 0.35$) and six sheep samples ($C_T$ values ranging from $24.4 \pm 0.05$ to $36.9 \pm 0.73$). Both assays successfully detected respective DNA targets in all human fecal and primary effluent wastewater DNA extracts (Table 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 those of enterococci and general *Bacteroidales* 16S rRNA genes. A one-way random effect ANOVA model indicated that there is significant variability ($P < 0.05$) in $C_T$ values among all locations for each assay. Variance ($\sigma^2$) between wastewater sample locations ranged from 0.30 for HumM2, 1.06 for HumM3, and 1.65 for Enterol1 to 0.45 for GenBac3. Target DNA relative abundances for each assay were compared by normalizing data sets.
DISCUSSION

Human-specific qPCR. We report on two qPCR assays that detect predominantly human fecal DNA when tested against a panel of samples representing agriculturally important animals, such as cattle, poultry, and swine, as well as many wildlife species. These qPCR assays were designed to target the same gene sequences as two end point PCR assays (assays 19 and 22) that were previously reported to be 100% human specific based on a fecal reference collection consisting of 160 individual samples representing 11 different animal species (23). The slight decrease in the specificity of the real-time qPCR assays compared to the level for the end point PCR assays may be due to the larger nontarget fecal sample reference library used to establish specificity values or factors associated with the TaqMan qPCR approach, such as constraints in primer design, PCR reagent chemistry, thermal cycling settings, and an 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 setup and reduce expenses. Each master curve was compiled from up to 12 independent runs in order to reflect sources of intra- and interrun variability. Master calibration curves were acceptable in this study because (i) there were no significant differences in the slopes of fitted curves between independent runs ($P > 0.05$), (ii) the analytical precision (percent 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. The 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 abundances of host-specific gene targets in different human populations. Host-specific and general fecal indicator bacterial 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 concentrations in all samples, which supports previous research reporting that Bacteroidales often makes up a large portion of the human fecal bacterial community (6, 16, 28). The HumM2 and HumM3 gene targets were the next-most-abundant markers and more prevalent than the enterococcal 23S rRNA genes (Fig. 3). Enterococci are routinely detected in feces-polluted waters (26). The observation that host-specific gene targets are more abundant than enterococcal 23S rRNA genes suggests that detectable quantities of HumM2 and HumM3 gene targets may be present in ambient waters.

FIG. 3. Box-and-whisker diagram depicting the relative abundances of gene targets from HumM2, HumM3, Entero1, and GenBac3 qPCR assays for all primary effluent sewage sample locations. Estimated gene target concentrations are reported as log10 mean copy numbers per ng of DNA. The boundary of the box closest to zero indicates the 25th percentile, the line within the box represents the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 10th and 90th percentiles, respectively. “+” denotes outlier measurements.
All qPCR assays exhibited less than 3.9% dispersion of $C_T$ values from an overall wastewater sample mean [one-way random-effect ANOVA qPCR standard deviation/mean) × 100] regardless of gene target. In addition, a significant difference ($P < 0.05$) was observed in concentrations of all qPCR gene targets between wastewater geographic locations. Fluctuations in relative gene target concentrations between wastewater samples could result from differences in local population diet, age, and/or health but could also reflect uncertainty associated with single sample events. Regardless of the reason, low dispersion percentages (<3.9%) suggest that the human-specific gene targets can be detected with a level of confidence similar to those for 16S rRNA general Bacteroidales and 23S rRNA enterococcal gene targets.

Implications for MST. Recreational and drinking source waters continue to be impacted by human fecal pollution and can impose a direct threat to human health (4, 5, 19). In addition to human waste, many other agricultural and wildlife animal sources can contribute to the total fecal load. Most microbial source tracking (MST) methods attempt to identify specific fecal sources to help local authorities prioritize polluted areas for restoration. Recent advances in PCR-based methods now allow for the estimation of host-specific DNA target concentrations. These quantitative approaches can extend the utility of MST applications by supplying information regarding the concentration of host-specific fecal pollution sources. To date, no qPCR-based method has been found to be 100% specific for human fecal pollution (10, 12, 18, 20). Animals that cohabitate with humans, such as cats and dogs, and animals that share similar digestive physiologies, such as pigs, are the most problematic. Fecal pollution originating from pets can confound MST studies where cat and dog waste is mixed with sewage and/or runoff after rain events. A similar problem can arise in watersheds affected by swine sources of fecal pollution. HumM2 and HumM3 are the first qPCR assays available that can discriminate between all three of these sources of fecal pollution. In addition, these assays can quantify as few as 10 copies of target DNA per reaction with a high degree of precision. DNA targets of these assays were widely distributed among 20 different human populations and more abundant than fecal enterococci in almost all wastewater samples tested.

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

ACKNOWLEDGMENTS

The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed or partially funded collaborative research in the described herein. This research 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


