# Q-471

# Analysis of Molecular-Marker-Based Microbial Source Tracking for Fecal Load Estimation

Don Stoeckel, USGS, Columbus, OH 43229, stoeckel@usgs.gov; Jason Gregory, NOAA, Charleston, SC 29412;
Kerry Ritter, SCCWRP, Westminster, CA 92683; Jayson Wilbur, WPI, Worcester, MA 01609

## Introduction

The current state of microbial source tracking science is that fecal sources can be detected, but not quantified. Quantification of loads from specific sources will require one of the following technologies:

1. The ability to accurately classify fecal-indicator bacteria isolates to source (library dependent).
2. The ability to quantify a host-specific marker in the environmental matrix and relate that quantity to fecal load.

The inability of library-dependent methods to accurately classify fecal-indicator bacteria to source has been established (Griffith et al., 2003; Stoeckel et al., 2004; Moore et al., 2005). Thus, quantitative MST will require detection of markers (by qPCR, MPN, or other quantitative approaches) and relation of the marker density to fecal load.

### Objective:

To use measured error rates in a generalized model to estimate confidence intervals about MST molecular marker quantitation in water samples, and conversion to fecal concentration.

The section below includes descriptions of error sources in analysis of environmental samples, data regarding the extent of error, and development of processes that may reduce those sources of error.

## Approach

Steps at which error is introduced to the process of measuring target-sequence copy number in environmental waters were identified. Published and unpublished error rates at each step were compiled (see presentation of errors, below). Refinements to several steps were identified.

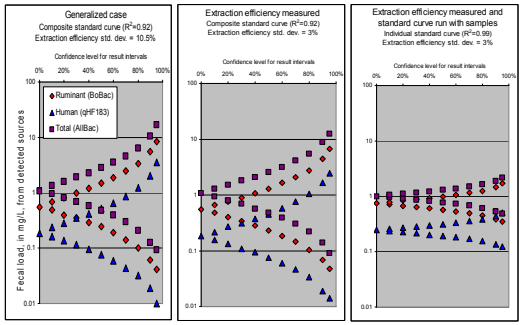
The expected qPCR measurement (C(t) value) was calculated for a hypothetical sample containing 0.25 mg/L human waste and 0.75 mg/L ruminant manure based on current knowledge of marker distribution in various hosts (Layton and others, 2006; Seurinck and others, 2005). The qPCR measurement of Ct was then used to estimate both the concentration of fecal material from the detected sources, and the confidence interval of that concentration.

The calculation was done using 1) generalized assumptions (extraction efficiency not measured, sample inhibition not corrected, standard curve not done with each run, replicate analyses not done), 2) incorporation of extraction efficiency measurement and 3) both extraction efficiency and run-specific standard curves.

Results were plotted as the distribution of confidence intervals at decadal values of alpha (type I error) for three cases. Artificial extended standard curves were generated with reduced $R^2$ values. Regression equations were within 0.2% of the more precise single-run standard curve.

## Results

Variation of the standard curve by as little as 0.2% (slope and y-intercept) substantially altered the resulting concentration estimates (from 0.75to 0.55 mg/L ruminant feces; from 0.25 to 0.18 mg/L human waste; first and second plot compared to third).
The generalized case assumes that the analyst uses accepted values for the standard curve and extraction efficiency. The true relative contribution by each source generally would not be detected (first plot).
Measurement of extraction efficiency in each run does little to reduce variability in results (second plot) when a low-precision standard curve is used.
Inclusion of a precise standard curve with each run greatly enhances the ability to detect relative contributions by each source (third plot).




## Interpretations and Conclusions

Error is measurable at each step of sample processing.

Some sources of error are correctable (sample inhibition, extraction efficiency). Others can be measured and minimized to reduce calculated uncertainty in the final result (Detection and Quantification variability). Still others are inherent and cannot be minimized (distribution of markers among source population), but must be accounted for.

Cumulative uncertainty can mask true differences in fecal contribution at the level of half-log difference when converting laboratory analysis results (C(t) values) to level of contamination in water (mg feces per L water).

Based on current estimates of uncertainty from the authors' original work and data from published literature, current procedures are inadequate to estimate the proportional fecal load to a stream from various host sources. When error is carefully measured and controlled, ranking of contributions by various sources may be feasible.

Incorporation of internal controls in processing steps allows
1) simultaneous evaluation of the steps involved in the source tracking process
2) assessment of intraassay and interassay variability
3) calibration to correct for losses due to inefficiency
4) evaluation of where error needs to be reduced.

## Description

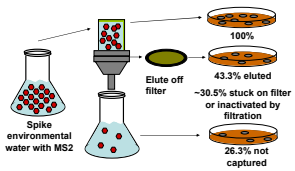
### Concentration

Concentration is an essential step in molecular analysis, and allows:

1)Enhancement of marker detection limit

2)Detection in volumes relevant to the study

Common methods of concentration include filtration and centrifugation. Regardless of the methodology, recovery of concentrated material is not completely efficient.

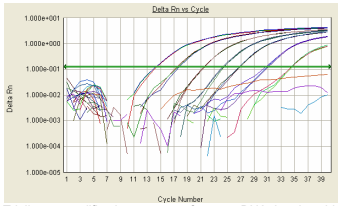



### Extraction and Purification

Before measurement by the PCR, DNA generally is extracted and purified using commercially available kits.
Variability in the Concentration and Extraction/Purification steps cannot easily be measured separately, but can be measured together from recovery of exogenous DNA or an internal standard.

### Detection

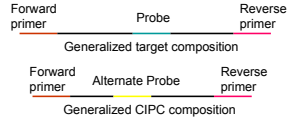
Analytical variability in marker detection can be measured as the difference in response in replicate analyses of the same sample.




Triplicate amplification curves of target DNA in plasmid. Fluorescence increases in direct proportion to the amount of cDNA template in reaction. Unknowns are plotted against the standard curve to allow quantification of markers or viral genomes.

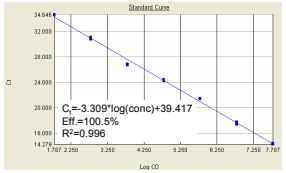
### Correction for sample inhibition

Sample volumes often are increased to enhance detection limits. PCR inhibitors are co-extracted with the target DNA. Competitive Internal Positive Control (CIPC) can be used to account for inhibition of detection (false negatives, skewed quantitation).
The CIPC incorporates the same primer binding sites as the target genome, allowing target-specific identification of inhibition, but differs in both size and internal sequence. The CIPC is distinguishable by gel electrophoresis (by size) and by qPCR (by melting curve or alternate probe).




### Quantification

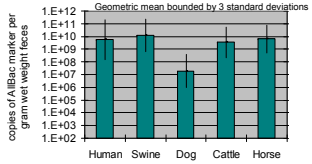
Relation of the cycle threshold to the concentration of marker present in the sample relies upon the precision of the qPCR standard curve.




Triplicate standard curve based on target DNA in plasmid

### Relation to amount of fecal material

In many studies, the end goal of the research is to relate marker detection to amount of fecal contamination from various sources. Fecal contamination can be indicated as density of fecal-indicator bacteria coming from that source or as mass of fecal contamination from that source. In either case, quantitation of fecal contamination from a particular source relies upon a relation between marker concentration and amount of feces, as well as the specificity of the marker.




Adapted from Layton et al., 2006

## Data

| ViroCap Filters | Pre filtration | Post filtration | Eluted |
|---|---|---|---|
| Average % Recovery | 100.0 | 26.3 | 43.3 |
| Standard Deviation % Recovery | 8.7 | 7.5 | 11.0 |

Results from triplicate spiked samples analyzed in duplicate.
0.45 µm HA nitrocellulose data similar, but not shown

Mumy and Findlay (2004) measured 28.3% efficiency ± 10.5% from soil.

Seurinck et al. (2005), on the other hand, measured 83% efficiency ± 3% from water.

Smith et al. (2006) and Seurinck et al. (2005) reported coefficients of variation of approximately 0.75% for repeated measurements of C(t). This level of variation is minor compared with the other sources of error in quantification of DNA in environmental matrices.

Gregory and others (2006) spiked aliquots of a target-free environmental water sample (100, 250, and 400 mL) with 1000 copies of enteroviral genome, filtered, and tested by qPCR. Without calibration, measured concentrations decreased with increasing sample volume.
By using the ΔCIPC as a calibrator and the following equation:
Correction factor = $(E + 1)^{\Delta CIPC\ Ct}$
the corrected enteroviral concentrations were more accurate (results are below).

Stoeckel (unpub. data) generated six standard curves from the same starting material over the course of two months

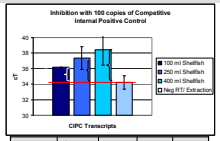
| Slope | Intercept | $R^2$ | Efficiency |
|---|---|---|---|
| -3.80 | 31.00 | 0.950 | 83% |
| -3.85 | 27.80 | 0.976 | 82% |
| -3.75 | 27.98 | 0.999 | 85% |
| -3.72 | 28.04 | 0.977 | 86% |
| -3.75 | 28.49 | 0.983 | 85% |
| -3.45 | 29.84 | 0.997 | 95% |
| Overall | | | |
| -3.82 | 28.67 | 0.881 | 83% |

## Development

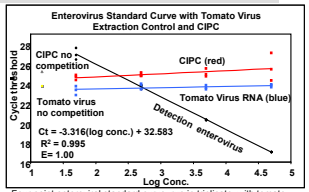
It may be possible to overcome low filtration efficiency by direct detection, without the filtration, extraction and purification. Layton et al. (2006) and Santo Domingo et al. (2003) both have used this approach. The standard deviation reported for C(t) values measured in this way was less than 2% of the C(t) values.

Total extraction efficiency can be estimated in each sample using an extraction control. This consists of co-extraction of a virus or bacteria with similar genetic structure, but not found in environmental samples, that can be quantified and compared to control extractions during the qPCR step (tomato spotted wilt virus, above).

No development directions identified on this topic




| | Control | 100 mL | 250 mL | 400 mL |
|---|---|---|---|---|
| Before correction | 914 | 265 | 114 | 68 |
| With CIPC correction | 914 | 944 | 839 | 983 |




Four point enteroviral standard curve, run in triplicate, with tomato spotted wilt virus (TSWV) extraction control and CIPC.

## References

Gregory, J.B., et al., 2006, Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples: Appl. Environ. Microbiol., v. 72, no. 6, p. 3960-3967.
Griffith, J.F., et al., 2003, Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. J. Water Health, v. 1, no. 4, p. 141-151.
Layton, A., et al., 2006, Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water: Appl. Environ. Microbiol., v. 72, no. 6, p. 4214-4224.
Moore, D.F., et al., 2005, Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed: J. Appl. Microbiol., v. 99, no. 3, p. 618-628.
Mumy, K.L., and Findlay, R.H., 2004, Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR: J. Microbiol. Methods, v. 57, no. 2, p. 259-268.
Santo Domingo, J. W., et al., 2003, Real-time PCR to method to detect *Enterococcus faecalis* in water: Biotechnol. Letters, v. 25, p. 261-265.
Seurinck, S., et al., 2005, Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater: Environ. Microbiol., v. 7, no. 2, p. 249-259.
Smith, C.J., et al., 2006, Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples: Environ. Microbiol., v. 8, no. 5, p. 804-815.
Stoeckel, D.M., et al., 2004, Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*: Environ. Sci. Technol., v. 38, no. 22, p. 6109-6117.

Use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.