Published in final edited form as:

Environ Sci Technol. 2015 February 03; 49(3): 1311–1318. doi:10.1021/es504579j.

Incorporating Expert Judgments in Utility Evaluation of *Bacteroidales* qPCR Assays for Microbial Source Tracking in a Drinking Water Source

Johan Åström[†], Thomas J. R. Pettersson[‡], Georg H. Reischer[§], Tommy Norberg^{||}, and Malte Hermansson^{*,‡}

[†]Tyréns AB, Lilla Badhusgatan 2, SE-411 21 Gothenburg, Sweden

[‡]Water Environment Technology, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

[§]Interuniversity Center Water & Health, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Straße 1a/166-5-2, A-1060 Vienna, Austria

^{||}Department of Mathematical Sciences, University of Gothenburg and Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

^{*}Department of Chemistry and Molecular Biology, Microbiology, University of Gothenburg, SE-405 30 Gothenburg, Sweden

Abstract

Several assays for the detection of host-specific genetic markers of the order *Bacteroidales* have been developed and used for microbial source tracking (MST) in environmental waters. It is recognized that the source-sensitivity and source-specificity are unknown and variable when introducing these assays in new geographic regions, which reduces their reliability and use. A Bayesian approach was developed to incorporate expert judgments with regional assay sensitivity and specificity assessments in a utility evaluation of a human and a ruminant-specific qPCR assay for MST in a drinking water source. Water samples from Lake Rådasjön were analyzed for *E. coli*, intestinal enterococci and somatic coliphages through cultivation and for human (BacH) and ruminant-specific (BacR) markers through qPCR assays. Expert judgments were collected regarding the probability of human and ruminant fecal contamination based on fecal indicator organism data and subjective information. Using Bayes formula, the conditional probability of a true human or ruminant fecal contamination given the presence of BacH or BacR was determined stochastically from expert judgments and regional qPCR assay performance, using Beta distributions to represent uncertainties. A web-based computational tool was developed for the procedure, which provides a measure of confidence to findings of host-specific markers and demonstrates the information value from these assays.

*Corresponding Author Phone +46(0)31 786 2575; fax +46(0)31 786 2599; malte.hermansson@cmb.gu.se.

Notes

The authors declare no competing financial interests.

Introduction

Microbial monitoring of fecal contamination in surface waters is used for managing drinking water treatment and to motivate remediation measures in the catchment. While the analysis of traditional fecal indicator organisms (FIO) such as *Escherichia coli* and intestinal enterococci is still the monitoring standard,¹ a wide range of microbial source tracking (MST) methods have been developed to provide information about the origin of fecal contamination.^{2,3} Several library-independent assays based on quantitative PCR (qPCR) have been established for the detection of genetic markers of the order *Bacteroidales* to identify fecal matter from different host groups, for example, humans and ruminants.^{4–7} A number of *Bacteroidales* qPCR assays targeting human and ruminant-associated genetic markers were recently evaluated on a transcontinental scale⁸ indicating that relatively stable target populations, especially for ruminant markers, occur around the globe.

For a credible answer on contribution from different sources, a limitation in these assays is that the source-sensitivity and source-specificity are often lower when introduced in new geographic regions, compared to the regions where the assays were originally developed.^{3,8} For instance, while the sensitivity and specificity of the human *Bacteroidales* qPCR BacH and the ruminant BacR assays were nearly 1.00 (i.e., 100%), when tested against fecal samples in Austrian catchments,^{6,7} an evaluation in 15 countries indicated an average sensitivity and specificity of the BacH of 0.77 and 0.53, respectively, and for the BacR assay 0.90 and 0.84 respectively. In fecal samples collected in Sweden the sensitivity and specificity were 0.86 and 0.62, respectively, for the BacH assay and 0.88 and 0.85, respectively, for the BacR assay, and similar results were observed in the evaluation of other human (BacHum) and ruminant specific assays (BacCow and BoBac).⁸ Development of highly specific and sensitive qPCR assays that are tailor-made for all regions in the world may be out of reach for practical as well as economic reasons. Instead, for most regions, a best-choice selection of available assays developed in other regions will be more realistic. However, in order for qPCR-based MST methods to gain a wider use as a credible tool for decision-making on remediation measures targeting fecal contamination, there is a need for procedures to assess the uncertainties with nonperfect specificities and sensitivities.

MST data will intuitively be judged in the light of previous knowledge about fecal sources in the local catchment area and traditional FIO data. While such subjective information is commonly used in an informal way,^{9–11} a procedure to systematically incorporate subjective information in the interpretation of MST data from qPCR assays is warranted and should incorporate the uncertainties both in the qPCR performance and in the subjective information. The assignment of conditional probabilities using Bayes formula has been suggested as a universally applicable measure for validating the utility of *Bacteroidales* qPCR assays for MST in new regions and environments.² In a Bayesian approach, prior knowledge regarding fecal sources in a watershed is accounted for as well as the source-sensitivity and specificity (i.e., the assay performance). Such prior knowledge has been defined by either the total marker frequency in the entire catchment⁴ or through the assessment of relative loads of fecal contamination from specific human and animal sources in relation to the total fecal loads.¹² So far, information that emanate from traditional expert judgments has not been used as prior knowledge to validate the utility of *Bacteroidales*

qPCR assays using Bayes formula. The fact that, for instance, the BacR assay originally developed in Austria,⁷ has already been used in other areas^{10,13–15} shows that development of such procedures are timely.

Here we developed and applied a Bayesian approach to incorporate expert judgments with assay performance data in a utility evaluation of a human and a ruminant-specific qPCR assay for MST in a drinking water source. As a proof of principle we used the approach in a Swedish region where the source-sensitivity and source-specificity of the BacH and BacR assays is reported to be lower than 1.00.⁸ The aim was to (i) model the prior knowledge collected from local expert judgments regarding fecal contamination at ten sites around Lake Rådasjön, Sweden together with the qPCR assay performance data using Beta distributions to specify uncertainties, (ii) determine the conditional probabilities of a true host-specific fecal contamination according to the Bayes formula and compare with results from a water sampling campaign, (iii) evaluate this Bayesian approach as a basis for assessing the local utility of the *Bacteroidales* qPCR assays, and (iv) provide a web-based computational tool based on this approach to be used when considering microbial source tracking in water sources around the world.

Materials and Methods

Study Area

The Lake Rådasjön on the Swedish west coast constitutes the main water source for the city of Mölndal, providing 60 000 people with drinking water and up to 600 000 people in the city of Gothenburg during shorter periods. The lake (Supporting Information (SI) Figure S1) has a surface area of close to 2.0 km², a catchment area of 14.9 km², and is part of the larger catchment area of the River Mölndalsån (268 km²). The Mölndal raw water intake (SI Figure S1; site 1) is located at 15 m depth. A few cattle and approximately 100 horses graze around the lake during summer. Emergency sewer overflows to the lake occur from the urban area south of the lake and upstream in River Mölndalsån, where on-site sewers are located.

Sampling Campaign

Ten sites surrounding Lake Rådasjön representing small streams and drainage points from point and diffuse fecal contamination were sampled (SI Figure S1, see SI Table S1 for details). Discrete water samples were collected biweekly from April to September 2008. After transport at 4 °C to the laboratory the samples were analyzed for FIO and a subset of samples were analyzed for human (BacH) and ruminant (BacR) *Bacteroidales* markers (see below).

DNA Extraction of Fecal Samples from Individuals in the Region

The regional performance in human and ruminant-targeted *Bacteroidales* qPCR assays was determined by analyzing fresh fecal matter from individuals in the region (see SI Table S3,⁸ J. Åström and G. H. Reischer, unpublished). DNA was extracted from approximately 0.25 g of each fecal sample, with the MoBio Power Soil DNA Isolation Kit. DNA concentration in undiluted DNA extracts was between 3 and 30 ng/μL, resulting in 1.5–15 ng of DNA in the

tested 1:4 dilution. DNA was shipped to Austria for qPCR analysis and further handled as described.⁸

Analyses of Fecal Indicator Organisms (FIO)

Water samples were analyzed for indicator bacteria within 6 h from sampling using standardized methods. Total coliforms and *E. coli* were analyzed by the Colilert Quantitray method (IDEXX Laboratories, Inc., ME), intestinal enterococci and sulfite-reducing clostridia by standard membrane filtration methods,^{16,17} excluding preheating in the latter and somatic coliphages by a plaque assay within 24 h from sampling using *E. coli* ATCC 13706 as host.¹⁸

DNA Extraction of Water Samples and qPCR Procedures

Water samples were filtered through Isopore 0.2 μm polycarbonate membrane filters (Millipore, Bedford, MA) within 24 h of sampling. As large a volume as possible during 2 h was filtered (up to 300 mL for each sample), and remaining unfiltered volume was decanted by a sterile pipet. The filters were immediately frozen and stored up to 1 month at $-20\text{ }^{\circ}\text{C}$ and at $-70\text{ }^{\circ}\text{C}$ for longer periods until DNA extraction. The extraction was performed as described,¹⁹ except that isopropanol was used for DNA precipitation and centrifugation at each step was performed at 13 000 rpm for 5 min. The washed and air-dried pelleted DNA was resuspended in 50 μL 10 mM Tris-EDTA buffer (pH 8.0). Controls for each round of extraction during the processing were negative. Extracted samples were stored at $-70\text{ }^{\circ}\text{C}$ until qPCR analyses.

qPCR analyses were performed on an iCycler iQ 5 (Bio-Rad, Hercules, CA) as earlier described for human BacH and ruminant BacR markers.^{6,7,14} PCR controls including filtration blanks, nontemplate controls (sterile deionized water and TRIS and EDTA solutions used in the DNA extractions) were negative. All qPCR runs had a PCR efficiency between 90 and 110% and the no-template controls were consistently negative. As before,^{6,7,14} several dilutions were made of each sample and the dilution that resulted in the highest qPCR result was selected.

Expert Judgment Data Collection and Modeling

A group of 11 experts was convened from different disciplines involving local drinking water managers, environmental inspectors and consultants. A questionnaire was provided to each person. Results on FIO from the 2008 sampling around Lake Rådasjön (% positive samples, median and maximum) and common information on fecal hosts in the area (SI Table S1) were presented separately for each sampling site for the group, blinding the results from the *Bacteroidales* qPCR analyses. Experts were requested to estimate, in terms of lowest and highest reasonable value, the probability of detecting human fecal matter and ruminant fecal matter at each sampling site in the water samples collected during summer 2008. Following a self-rating procedure²⁰ the experts rated their own knowledge concerning fecal sources upstream each site on a four-grade scale including the levels poor, moderate, high and excellent. These knowledge estimates were converted to values from 0.25 to 1.0, which were considered as weights in the further analysis. The questionnaire was answered

individually and without possibilities for discussions and the expert judgments are considered as independent.

The lowest and highest reasonable values in the expert judgments defined the 5- and 95-percentile in Beta distributions, from which the corresponding α and β -values were determined for each judgment and each sample site (SI Table S2). It is indeed reasonable to presume that an expert's uncertainty density is smooth, unimodal and, within the percentiles, reasonably symmetric. The Beta distribution has these properties and can represent a variety of different opinions by adjusting its two parameters α and β . A lowest reasonable estimate of 0 corresponded to $\alpha = 1$ and a highest reasonable estimate of 1 corresponded to $\beta = 1$. Weighted average values of α and β were calculated to produce consensus Beta distributions for each sampling site incorporating all expert judgments. These distributions serve as uncertainty distributions for the prior probabilities for human (P(H)) and ruminant (P(R)) fecal contamination. One of the respondents misunderstood the task, and data from this respondent (No. 6) was excluded.

Source-Sensitivity and Specificity Modeling

The source-sensitivity was defined as the percentage of target samples (i.e., human fecal samples for the human-targeted assay) giving a positive detection by qPCR, regardless of the marker concentration (all signals ≥ 1 copy per reaction; true positives). The source-specificity was defined as the percentage of nontarget samples not detected by qPCR (all signals < 1 copy per reaction; true negatives). Presently available data based on 28 single fecal samples (see above) from the region was used in the analyses with the entire transcontinental data set based on 280 fecal samples as a reference.⁸

Uncertainties in the source-sensitivity and specificity of qPCR assays for the BacH and BacR marker, related to the number of fecal samples, were represented in Beta distributions. The parameter α was defined by the number of successful fecal samples +0.5 and β by the number of unsuccessful samples +0.5 for each qPCR assay. Selecting the initial values of 0.5 for defining α and β gives the form of reference prior that has the least impact on the shape of the posterior.²¹

Conditional Probability Analysis Using Bayes Formula

For each sample site, the posterior probability $P(H|T)$ was determined by means of Bayes formula (eq 1) and Monte Carlo simulation was used to calculate its uncertainty distribution. Note that $P(H|T)$ is the probability of a true human source of fecal contamination (H) in a sample given a positive test result in the BacH assay (T).

$$P(H|T) = \frac{P(T|H) P(H)}{P(T|H) P(H) + P(T|H') P(H')} \quad (1)$$

where $P(T|H)$ is the sensitivity, that is, the probability of a positive BacH signal in a fecal sample from a human host. $P(H)$ is the prior probability, that is, the prior defined from the expert judgments, of detecting a human fecal contamination in a random water sample

during the sampling campaign, and $P(H')$ equals $1 - P(H)$. $P(T|H')$ is the probability of a positive BacH signal in a fecal sample from a nonhuman host, that is, $P(T|H')$ equals $1 - \text{specificity}$. Bayes formula was similarly used to determine the posterior $P(R|T)$ which is the probability of a ruminant source of fecal contamination, given a positive test result (T) with BacR and corresponding sensitivity and specificity for the BacR assay were incorporated in the calculations.

Furthermore, the posterior $P(H|T')$ was determined, which is the probability of a true human source of fecal contamination (H) in a sample given a negative test result (T') with the BacH assay (eq 2).

$$P(H|T') = \frac{P(T'|H) P(H)}{P(T'|H) P(H) + P(T'|H') P(H')} \quad (2)$$

where $P(T'|H)$ is the probability of a negative BacH signal in a fecal sample that is in fact human derived ($1 - \text{sensitivity}$), and $P(T'|H')$ is the probability of a negative BacH signal given a nonhuman fecal sample (specificity). Corresponding calculations were performed for the ruminant-targeted assay. Crystal Ball (Oracle) was used for the Monte Carlo simulations and for obtaining α or β values based on expert judgments. Results from all Monte Carlo simulations (10 000 iterations) were presented as 5-, 50- (median), and 95-percentiles.

MST Performance Tool

A tool was developed in Analytica (Lumina Decision Systems) and published as a web-based Analytica Cloud Player model (splash screen provided in Figure 1). The model is set up in a visual interface enabling the user to (1) calculate regional performance of human-specific and ruminant specific *Bacteroidales* markers, (2) define subjective information on fecal contamination at the sampling point, and (3) perform probability assessments regarding a human or ruminant-specific source of fecal contamination at the sampling point. The model is freely available and described together with URL in SI.

Results

Cumulative Probability of Beta Distributions

Modeling source-sensitivity and specificity of qPCR assays in Beta distributions demonstrated that increasing the numbers of fecal samples would lower the uncertainties as defined by the shape in the Beta distribution (Figure 2). The limited number of samples from the Swedish data set implies higher uncertainties compared to the transcontinental data set (see SI Table S3). The ruminant marker generally performed better than the human marker.

Detection of Fecal Indicator Organisms and Expert Judgments

The frequency and levels of FIO around Lake Rådasjön varied between the samples sites, with generally higher levels of *E. coli* than intestinal enterococci and somatic coliphages (Table 1). Based on this information and common information on fecal hosts in the area, expert judgments were collected for each sampling site (SI Table S2), including lowest and

highest reasonable estimates on the probability of human and ruminant fecal contamination. Most experts rated their knowledge concerning fecal sources upstream in the range from poor to high (0.25–0.75) and with the highest variation at the raw water intake (SI Figure S1, site 1). Expert No. 7 and 10 rated their knowledge as excellent for sites 1, 7, 17, and 18, whereas expert No. 3 rated the knowledge as poor for most sites. The highest discrepancy between the different experts was observed for site 1, while a consistency in the judgments were observed for, for example, site 3 (SI Figure S1).

Conditional Probabilities and Detection of Host-Specific Genetic Markers

The probability of a true human contamination given a positive result in the BacH assay was calculated according to Bayes formula, combining the prior with BacH and BacR assay performance. The consensus Beta distributions calculated for each site indicated the highest prior probability for human contamination at sites 3, 6, 7, and 18 (Table 2). The median posterior $P(H|T)$ at these sites were 0.80, 0.76, 0.89, and 0.78, with the highest range of uncertainty at site 18. In the MST analysis the human marker BacH was frequently detected at high levels at these sites (Table 3). The posterior for human fecal contamination given a negative BacH result, $P(H|T')$ was highest at site 7, indicating a probability at 0.5 of false negatives. For the ruminant fecal contamination, most prior probabilities $P(R)$ were lower compared to the human fecal contamination, except at site 17 (which was influenced by cattle; SI Table S1) where the prior was 0.78. For the ruminant contamination the median posterior values $P(R|T)$ were highest for sites 3, 4, 17, and 18. At these sites the ruminant marker was frequently detected at high levels except for site 4 (Table 3). The posterior for a true ruminant contamination, $P(R|T')$, given a negative result in the BacR assay was lower than 0.05 (median) for all sites, except for site 17, implying a low probability of detecting the BacR marker without having a ruminant fecal contamination (Table 3). The posterior probability based on expert judgments of likely source combined with BacH and BacR assay performance only changed the source considered dominant by the experts (prior probabilities) for one of 10 sites (site 10).

Discussion

Human and Ruminant-Specific qPCR Assay Performance and Expert Judgments

The utility of MST based on *Bacteroidales* qPCR assays is hampered in many regions in the world due to a lack of knowledge of the source-sensitivity and specificity of the method. This inhibits the use of these assays as a credible decision tool to address sources of fecal influence in drinking water sources.³ Given a low specificity and sensitivity, it can be questioned whether findings of host specific markers in water samples truly reflects host-specific fecal contamination. It can also be questioned to what extent the MST assays improve the knowledge on fecal sources obtained by traditional methods. Different modeling approaches have been used in MST, some of which include Bayesian procedures and algorithmic methods.²² In the Bayesian approach presented here, subjective information from expert judgments was incorporated in the prior.

In previous MST studies, statistical and geographical analyses have been performed to assess the relationship between fecal sources and host-specific markers in source

waters^{23,24} and it was recommended to combine different assays to strengthen the MST analyses.^{25,26} Several modeling techniques have been applied in MST, entailing different working assumptions and results.²² Logistic regression (LR) can be used to model relationships between predictor variables and response variables that are not normally distributed. For applying LR in our model area, information on land use, slope, soil, grazing animals, fertilizing, and on-site sewers are possible predictor variables, information that is lumped into the expert judgments in the approach presented here. Accounting for prior knowledge in the context of MST has previously been performed in various modeling techniques, including linear and quadratic discriminant analysis.²⁷ The Bayesian approach presented here is a structured way to incorporate knowledge on fecal sources and FIO data in the prior, which is updated with information on the assay performance. In comparison to alternative modeling techniques the approach is associated with low cost and ease of deployment to different geographic regions, which is a requested feature.²²

In this report, the source-sensitivity and specificity was based on a limited number of fecal samples from humans, ruminants, and other animals. A sensitivity analysis was undertaken on present simulations of posterior distributions for $P(H|T)$ and $P(R|T)$ (data not shown). This demonstrated higher rank correlation coefficients for the expert judgments compared to sensitivity and specificity parameters for all sites, except for $P(R|T)$ at site 7 and 10. The uncertainties associated with the limited number of fecal samples were thus of minor importance compared with the uncertainties in the expert judgments. To reduce the uncertainty in the posterior, it is therefore desirable to reduce the uncertainties in the prior, that is, the expert judgments. This uncertainty comes from the variability estimates for each sampling site and from the differences between the experts. Although it is advantageous to include experts from different disciplines as in this study, it is desirable to exclude representatives of some schools as being incompetent (not experts) to eliminate unnecessary uncertainties. Biases that may influence the results are motivational and availability bias.²⁸ Motivational bias can occur when an expert has a stake in the outcome of the study, for example, water utility personnel whose task it is to reduce pathogens in the raw water. Availability bias is the cognitive bias where events that are easily recalled (such as a sewage pipe break) are overestimated, while events that are difficult to recall or measure (e.g., diffuse contamination from ruminants) are underestimated.

The fecal samples from different hosts were selected to cover the main fecal sources in the region.⁸ Ideally, the fecal samples should represent the proportional contribution from the fecal hosts in the catchment. The equal number of human and ruminant samples was, for example, based on the assumption that these host groups produce equal loads of fecal matter. Catchment modeling can be used to evaluate the fecal load from both point- (sewage) and diffuse fecal sources (animals and birds).²⁹

Conditional Probabilities of Host-Specific Fecal Contamination in Comparison with Water Sampling Data

The frequency and concentrations of the BacH and BacR markers around Lake Rådasjön generally confirmed the posterior probabilities determined from Bayes formula. An exception was for site 4, where all samples were positive for the human marker while the

posterior was in median below 0.5. This site was affected by a nearby stream, contaminated from on-site sewers (SI Figure S1, site 3). At the raw water intake (site 1), the human, but never the ruminant marker, was detected. Given the low posterior, with $P(R|T)$ at 0.26, further BacR assay analyses would provide limited value of information. Sporadic detection of any of the markers in case of low posterior probability should be considered with caution, such as the BacH findings at site 15 and 17, and the findings of BacR at site 7. Human contamination at the raw water intake, with posterior $P(H|T)$ at 0.57, may originate from on-site sewers and from a sewer station south of the lake that overflows several times per year.

Comparing the conditional probabilities with results from the water sampling, human fecal contamination was of higher concern compared to ruminant fecal contamination. However, a high uncertainty in some of the expert judgments (Table 2) reduced the motivation to take remediation measures. Sites affected by multiple sources and transport routes generally implicated higher uncertainties in the expert judgments, for example, the raw water intake (site 1), while the experts were more confident in their assessment of sites with a few fecal sources from a short distance upstream (e.g., site 7). To reduce the uncertainties, additional assessments are therefore desirable before remediation measures. In another study, hydrodynamic modeling was used to assess the transport of the BacH and the BacR markers from different fecal sources around Lake Rådasjön to the raw water intake, demonstrating that the raw water intake was mainly affected by human fecal contamination.¹⁴

Evaluation of the Bayesian Approach for Assessing the Local Utility of the Bacteroidales qPCR Assays

While the BacH marker was associated with a quite low specificity when tested on regional fecal samples, the Bayesian approach demonstrated that this assay will still be adequate to track human fecal contamination, given that the experts have a high expectancy on human contamination. It follows from eq 1 that as long as the 50-percentile of the source-sensitivity and specificity are above 0.50, findings of a host-specific marker at a site will strengthen the evidence (i.e., the 50-percentile of the posterior will exceed the value 0.50) for a host specific fecal contamination compared to the prior knowledge. In contrast, as the sensitivity and specificity approaches 1.00, which was the case for the human and ruminant assays in Austrian catchments,^{6,7} the posterior distribution in relation to positive test results will move toward 1.00 and the prior knowledge will thus be of less importance.

The prior knowledge on fecal sources, which is accounted for as the prior in Bayes formula, may vary considerably between different drinking water sources and change over time as the knowledge increase. Microbial monitoring of FIO normally forms the knowledge base regarding fecal contamination, and regular operational monitoring in drinking water sources is strongly advocated.¹ However, since FIO and MST indicators are based on different principles they may vary in, for example, amounts shed from various animals and survival in the environment, which means that the relation between FIO and MST is not stable. It was recommended that no associations should be made between FIO and MST levels,³¹ if so, complementary information from experts, such as observations of animals upstream the sample point, would become even more relevant. Catchment modeling may further increase

this knowledge base and catchment pollution source profiling was used to assess the proportional contribution of fecal contamination from humans, wildlife and livestock in an Austrian karst spring catchment.¹² This profiling was used as the prior in Bayes formula to determine the confidence in findings of BacH and BacR qPCR assays.¹² A modeling approach may require a greater effort than asking experts to make subjective assessments of the relative contribution from different fecal sources. The expert judgments used as the prior in the present study, together with the self-rating procedure of knowledge, provides a lumped parameter that represents the present knowledge on fecal contamination at various sampling sites. A high frequency of FIO and other measurements may result in better expert judgments, and thus in better prior estimations (but see Discussion above on FIO usage). Similarly, more fecal samples used for the source-sensitivity and source-specificity, will increase the precision of the posterior

This study demonstrates a procedure whereby the MST data from *Bacteroidales* host-specific qPCR assays increase the knowledge on fecal contamination in surface waters, despite the fact that these assays exhibit shortcomings in terms of source-specificity or sensitivity. As the development of tailor-made host-specific qPCR assays is hampered in most regions in the world, due to economic and technical reasons, this procedure demonstrates the value of information that can be provided by these analyses. The societal value of this procedure lies in the fact that it demonstrates the additional information that can be provided by applying these MST assays.

In the MST performance tool the computations described here can be performed in any region of the world, using the web-based interface. Using Bayes formula, the conditional probabilities can be determined before decision is taken to perform the qPCR analysis, as the impact of source-sensitivity, source-specificity and the number of underlying fecal samples in relation to expert judgments can be tested site-specifically. Applying the tool before water samples are tested can save the cost of sample collections and avoiding qPCR analysis of samples from sites where the reliability of the results are low. Over time, molecular methods like these will be less costly and established worldwide. By the uncertainty assessments, this procedure may improve the utility of these MST methods as a tool to remediate fecal contamination in drinking water sources.

Supporting Information

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded jointly by Swedish Water & Wastewater Association (SWWA), the TECHNEAU project funded by the European Commission (contract 018320), the University of Gothenburg and Sven Tyréns foundation. Samplings and analyses were funded by the city of Gothenburg (bacterial indicators). We are very grateful for the support and input of the project management group: Inger Kjellberg, Olof Bergstedt, Henrik Rydberg, Reine Rohman, Dennis Yhr and Anders Bruce among others. We are grateful for assistance during the field sampling from Lovisa Olofsson (Mölnådal). We thank Kristian Kvint at the Department of Chemistry and Molecular Biology, University of Gothenburg, for assistance with the qPCR analyses.

References

- (1). WHO. Guidelines for Drinking-water Quality. 4th ed. World Health Organization; Geneva: 2011. http://whqlibdoc.who.int/publications/2011/9789241548151_eng.pdf?ua=1 [accessed December 10, 2014]
- (2). Santo Domingo JW, Bambic DG, Edge TA, Wuertz S. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Res.* 2007; 41(16): 3539–52. [PubMed: 17632210]
- (3). Hagedorn, C., Harwood, VJ., Blanch, AR. *Microbial Source Tracking: Methods, Applications, and Case Studies.* Springer; New York: 2011.
- (4). Kildare BJ, Leutenegger CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water Res.* 2007; 41(16):3701–15. [PubMed: 17644149]
- (5). Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 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.* 2006; 72(6):4214–24. [PubMed: 16751534]
- (6). Reischer GH, Kasper DC, Steinborn R, Farnleitner AH, Mach RL. A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Lett Appl Microbiol.* 2007; 44(4):351–6. [PubMed: 17397471]
- (7). Reischer GH, Kasper DC, Steinborn R, Mach RL, Farnleitner AH. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl Environ Microbiol.* 2006; 72(8):5610–14. [PubMed: 16885315]
- (8). Reischer GH, Ebdon JE, Bauer JM, Schuster N, Ahmed W, Åström J, Blanch AR, Bloschl G, Byamukama D, Coakley T, Ferguson C, et al. Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ Sci Technol.* 2013; 47(15):8548–56. [PubMed: 23755882]
- (9). USEPA. Microbial Source Tracking Guide Document. National Risk Management Research Laboratory, U. S. Environmental Protection Agency; Cincinnati, OH: 2005. <http://nepis.epa.gov/Adobe/PDF/2000D20V.pdf> [accessed December 2014]
- (10). Stapleton CM, Kay D, Wyer MD, Davies C, Watkins J, Kay C, McDonald AT, Porter J, Gawler A. Evaluating the operational utility of a Bacteroidales quantitative PCR-based MST approach in determining the source of faecal indicator organisms at a UK bathing water. *Water Res.* 2009; 43(19):4888–99. [PubMed: 19783026]
- (11). Seurinck S, Verdriev M, Verstraete W, Siciliano SD. Identification of human fecal pollution sources in a coastal area: A case study at Oostende (Belgium). *J Water Health.* 2006; 4(2):167–75. [PubMed: 16813010]
- (12). Reischer GH, Kollanur D, Vierheilig J, Wehrspaun C, Mach RL, Sommer R, Stadler H, Farnleitner AH. Hypothesis-driven approach for the identification of fecal pollution sources in water resources. *Environ Sci Technol.* 2011; 45(9):4038–45. [PubMed: 21466151]
- (13). Gourmelon M, Caprais MP, Le Mennec C, Mieszkina S, Ponthoreau C, Gendronneau M. Application of library-independent microbial source tracking methods for identifying the sources of faecal contamination in coastal areas. *Water Sci Technol.* 2010; 61(6):1401–9. [PubMed: 20351418]
- (14). Sokolova E, Åström J, Pettersson TJ, Bergstedt O, Hermansson M. Decay of Bacteroidales genetic markers in relation to traditional fecal indicators for water quality modeling of drinking water sources. *Environ Sci Technol.* 2012; 46(2):892–900. [PubMed: 22148545]
- (15). Åström J, Pettersson TJ, Reischer GH, Hermansson M. Short-term microbial release during rain events from on-site sewers and cattle in a surface water source. *J Water Health.* 2013; 11(3):430–442. [PubMed: 23981872]
- (16). ISO. ISO 7899-2, Water Quality—Detection and Enumeration of Intestinal Enterococci—Part 2: Membrane Filtration Method. International Organization for Standardization; Geneva: 2000.

- (17). ISO. ISO 6461-1, Water Quality—Detection and Enumeration of the Spores of Sulfite-Reducing Anaerobes (Clostridia)—Part 1: Method by Enrichment in a Liquid Medium. International Organization for Standardization; Geneva: 1986.
- (18). ISO. ISO 10705-2, Water Quality—Detection and Enumeration of Bacteriophages—Part 2: Enumeration of Somatic Coliphages. International Organization for Standardization; Geneva: 2000.
- (19). Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol.* 2000; 66(12):5488–91. [PubMed: 11097934]
- (20). Winkler RL. The consensus of subjective probability distributions. *Manage Sci.* 1968; 15(2):61–75.
- (21). Bernardo, JM. Bayesian statistics. Probability and Statistics. Viertl, R., editor. UNESCO; Oxford, UK: 2003. p. 1-45.
- (22). Belanche, LA., Blanch, AR. Statistical approaches for modeling in microbial source tracking. *Microbial Source Tracking: Methods, Applications, and Case Studies.* Hagedorn, C. Blanch, AR., Harwood, VJ., editors. Springer; New York: 2011. p. 207-227.
- (23). Shanks OC, Nietch C, Simonich M, Younger M, Reynolds D, Field KG. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay Oregon. *Appl Environ Microbiol.* 2006; 72(8):5537–46. [PubMed: 16885307]
- (24). Peed LA, Nietch CT, Keltz CA, Meckes M, Mooney T, Sivaganesan M, Shanks OC. Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. *Environ Sci Technol.* 2011; 45(13):5652. [PubMed: 21662992]
- (25). Blanch AR, Belanche-Munoz L, Bonjoch X, Ebdon J, Gantzer C, Lucena F, Ottoson J, Kourtis C, Iversen A, Kuhn I, Moce L, et al. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Appl Environ Microbiol.* 2006; 72(9):5915–26. [PubMed: 16957211]
- (26). Balleste E, Bonjoch X, Belanche LA, Blanch AR. Molecular indicators used in the development of predictive models for microbial source tracking. *Appl Environ Microbiol.* 2010; 76(6):1789–95. [PubMed: 20118380]
- (27). Wiggins BA. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. *Appl Environ Microbiol.* 1996; 62:3997–4002. [PubMed: 8899986]
- (28). Bedford, T., Cooke, R. Probabilistic Risk Analysis. Foundations and Methods. Cambridge University Press; Cambridge, U.K.: 2001.
- (29). Ferguson CM, Croke BF, Beatson PJ, Ashbolt NJ, Deere DA. Development of a process-based model to predict pathogen budgets for the Sydney drinking water catchment. *J Water Health.* 2007; 5(2):187–208. [PubMed: 17674569]
- (30). Sokolova E, Åström J, Pettersson TJ, Bergstedt O, Hermansson M. Estimation of pathogen concentrations in a drinking water source using hydrodynamic modelling and microbial source tracking. *J Water Health.* 2012; 10(3):358–70. [PubMed: 22960480]
- (31). Cao Y, Hagedorn C, Shanks OC, Wang D, Ervin J, Griffith JF, Layton BA, McGee CD, Riedel TE, Weisberg SB. Towards establishing a human fecal contamination index in microbial source tracking. *Int J Environ Eng Syst.* 2013; 4:46–58.

1. Regional performance of Bacteroidales marker

Use positive integers

Nodes 1

Human marker performance

Step 1 Add local values:

Human samples tested

Tested positive correctly

Nonhuman samples tested

Tested negative correctly

Step 2 Click to calculate:

Sensitivity mid

Sensitivity distribution mid

Specificity mid

Specificity distribution mid

Ruminant marker performance

Step 3 Add local values:

Ruminant samples tested

Tested positive correctly

Nonruminant samples tested

Tested negative correctly

Step 4 Click to calculate:

Sensitivity mid

Sensitivity distribution mid

Specificity mid

Specificity distribution mid

2. Define subjective information on fecal contamination at the sampling point

Use values in the range 0 to 1

Nodes 2

Number of expert judgments

Reset inputs

Expert No.	Estimated Probability of Human contamination		Estimated Probability of Ruminant contamination		Self-rated knowledge
	Lowest (5%)	Highest (95%)	Lowest (5%)	Highest (95%)	
1	<input type="text" value="0.2"/>	<input type="text" value="0.5"/>	<input type="text" value="0.02"/>	<input type="text" value="0.2"/>	<input type="text" value="High"/>
2	<input type="text" value="0.6"/>	<input type="text" value="0.8"/>	<input type="text" value="0.2"/>	<input type="text" value="0.8"/>	<input type="text" value="Moderate"/>
3	<input type="text" value="0.3"/>	<input type="text" value="0.6"/>	<input type="text" value="0.1"/>	<input type="text" value="0.25"/>	<input type="text" value="Poor"/>
4	<input type="text" value="0"/>	<input type="text" value="0.1"/>	<input type="text" value="0"/>	<input type="text" value="0.1"/>	<input type="text" value="High"/>
5	<input type="text" value="0.1"/>	<input type="text" value="0.4"/>	<input type="text" value="0.2"/>	<input type="text" value="0.6"/>	<input type="text" value="Poor"/>
6	<input type="text" value="0.1"/>	<input type="text" value="0.8"/>	<input type="text" value="0.05"/>	<input type="text" value="0.3"/>	<input type="text" value="High"/>
7	<input type="text" value="0.2"/>	<input type="text" value="0.7"/>	<input type="text" value="0.4"/>	<input type="text" value="0.8"/>	<input type="text" value="Excellent"/>
8	<input type="text" value="0"/>	<input type="text" value="1"/>	<input type="text" value="0"/>	<input type="text" value="0.01"/>	<input type="text" value="High"/>
9	<input type="text" value="0.15"/>	<input type="text" value="0.35"/>	<input type="text" value="0"/>	<input type="text" value="0.2"/>	<input type="text" value="Moderate"/>
10	<input type="text" value="0.15"/>	<input type="text" value="0.8"/>	<input type="text" value="0"/>	<input type="text" value="0.5"/>	<input type="text" value="Moderate"/>

3. Probability assessment

Nodes 3

Probability of a human source of fecal contamination at the sampling point

Human source according to the expert judgments (prior) $P(H)$ mid

Human source given a positive human marker detection (posterior) $P(H|T)$ mid

Human source given a negative human marker detection (posterior) $P(H|T')$ mid

Probability of a ruminant source of fecal contamination at the sampling point

Ruminant source according to the expert judgments (prior) $P(R)$ mid

Ruminant source given a positive ruminant marker detection (posterior) $P(R|T)$ mid

Ruminant source given a negative ruminant marker detection (posterior) $P(R|T')$ mid

Figure 1. Splash screen of the MST performance tool. Local values from tested fecal samples and from expert judgments are added in boxes 1 and 2, here exemplified for Site 1. Calculations are performed by clicking the Calc buttons in boxes 1 and 3.

Figure 2.

Cumulative probability of Beta distributions defining the source-sensitivity and specificity of the human (BacH, left panel) and the ruminant (BacR, right panel) qPCR assays, see text for details. The lines denote sensitivity for Swedish data set (solid line), specificity for Swedish data set (dotted line), sensitivity for transcontinental data set (dashed line) and specificity for transcontinental data set (dash-dotted line).

Table 1
Frequency and Levels of FIO at Sampling Sites Around Lake Rådasjön in Samples
Collected during the Summer Period April to September 2008

site no.	<i>E. coli</i>		intestinal enterococci		somatic coliphages	
	% positive samples ^a	median log CFU/100 mL	% positive samples ^a	median log CFU/100 mL	% positive samples ^a	median log CFU/100 mL
1	33	0.0 (0.5) ^b	17	0.5 (0.7) ^b	14	1.8 (1.8) ^b
3	88	2.3 (4.7)	93	2.9 (4.1)	86	2.4 (4.4)
4	86	1.2 (3.7)	86	1.3 (4.0)	83	1.3 (1.5)
6	92	0.5 (2.1)	69	1.0 (2.3)	67	1.2 (1.3)
7	100	4.0 (6.1)	95	2.9 (6.5)	100	2.3 (4.2)
10	82	0.5 (1.6)	73	0.7 (1.3)	33	1.2 (1.3)
15	73	2.5 (4.0)	91	1.9 (4.1)	83	2.3 (3.6)
17	82	2.5 (4.6)	100	1.9 (3.9)	83	2.6 (3.3)
18	92	2.0 (3.1)	92	1.9 (3.0)	67	2.1 (2.5)
19	92	1.1 (1.9)	75	1.3 (2.0)	33	1.4 (1.5)

^a% positive samples out of the total number of samples.

^bMaximum concentrations within parentheses.

Table 2
Probabilities of Human and Ruminant Fecal Contamination around Lake Rådasjön
Based on Expert Judgments (Prior) and Expert Judgments Combined with BacH and
BacR Assay Performance (Posterior)

site	prior		posterior			
	$P(H)^a$	$P(R)^a$	$P(H T)^b$	$P(H T')^b$	$P(R T)^c$	$P(R T')^c$
	0.39 ^d	0.06	0.57	0.13	0.26	0.01
1	(0.17–0.65) ^e	(0.01–0.20)	(0.28–0.82)	(0.02–0.42)	(0.04–0.65)	(0.00–0.06)
	0.65	0.24	0.80	0.32	0.64	0.05
3	(0.45–0.81)	(0.11–0.43)	(0.60–0.91)	(0.06–0.65)	(0.34–0.87)	(0.01–0.18)
	0.26	0.26	0.43	0.08	0.66	0.05
4	(0.14–0.40)	(0.11–0.48)	(0.23–0.63)	(0.01–0.24)	(0.34–0.89)	(0.01–0.20)
	0.60	0.15	0.79	0.27	0.49	0–03
6	(0.36–0.81)	(0.05–0.32)	(0.51–0.91)	(0.05–0.64)	(0.19–0.80)	(0.00–0.12)
	0.79	0.15	0.89	0.50	0.49	0.03
7	(0.61–0.92)	(0.09–0.24)	(0.74–0.96)	(0.11–0.81)	(0.27–0.76)	(0.00–0.09)
	0.39	0.20	0.57	0.14	0.58	0.04
10	(0.22–0.57)	(0.10–0.34)	(0.34–0.77)	(0.02–0.38)	(0.31–0.83)	(0.01–0.13)
	0.06	0.16	0.12	0.01	0.51	0.03
15	(0.01–0.17)	(0.06–0.33)	(0.02–0.33)	(0.00–0.08)	(0.21–0.81)	(0.00–0.12)
	0.09	0.78	0.18	0.02	0.95	0.37
17	(0.04–0.18)	(0.39–0.98)	(0.07–0.36)	(0.00–0.09)	(0.74–1.00)	(0.04–0.90)
	0.62	0.26	0.78	0.29	0.66	0.05
18	(0.25–0.91)	(0.09–0.52)	(0.40–0.96)	(0.04–0.76)	(0.30–0.90)	(0.01–0.22)
	0.38	0.12	0.57	0.13	0.44	0.02
19	(0.19–0.62)	(0.04–0.28)	(0.30–0.80)	(0.02–0.40)	(0.15–0.76)	(0.00–0.09)

^a Prior probabilities of a human $P(H)$ and ruminant $P(R)$ fecal contamination at each site determined from expert judgments.

^b $P(H|T)$ is the probability for a true human contamination, given a positive result in the BacH assay, and $P(H|T')$ is the probability given a negative result in the BacH assay.

^c $P(R|T)$ is the probability for a true ruminant contamination, given a positive result in the BacR assay, and $P(R|T')$ is the probability given a negative result in the BacR assay.

^d Median values.

^e Range indicated by 5- and 95-percentile values.

Table 3
Concentrations of Human and Ruminant Markers at Sites Around Rådasjön Measured
by the *Bacteroidales* BacH and BacR Assays

site	human marker (BacH)		ruminant marker (BacR)	
	no. of positive samples	ME/100 mL ^a	no. of positive samples	
1	2 (6) ^b	3.7 ^c (3.8) ^d	0 (6) ^b	
3 ^e	7 (7)	5.2 (6.1)	5 (7)	3.0 ^c (4.0) ^d
4	6 (6)	2.0 (4.3)	0 (6)	
6	4 (6)	3.0 (3.9)	0 (6)	
7 ^e	5 (5)	4.8 (6.5)	3 (5)	4.6 (6.0)
10	4 (6)	2.2 (4.1)	1 (6)	5.0 (5.0)
15	2 (6)	3.2 (3.5)	3 (6)	2.5 (5.9)
17 ^e	2 (6)	2.9 (3.2)	6 (6)	5.6 (7.4)
18	6 (6)	3.2 (4.4)	5 (6)	3.3 (4.3)
19	5 (6)	2.7 (3.0)	0 (6)	

^aMarker equivalents (ME)/100 mL.

^bTotal numbers of samples within parenthesis.

^cMedian values of positive samples.

^dMaximum value within parenthesis.

^eData from ref 14