Chapter 2
Performance Criteria

Valerie J. Harwood and Donald M. Stoeckel

Abstract The establishment of rigorous, consistent performance criteria for microbial source tracking (MST) methods is essential for their usefulness and widespread acceptance as research and regulatory tools. In this chapter, we focus on performance criteria for library-independent methods, although many aspects of the discussion are applicable to both library-independent and library-dependent methods. We separate these criteria into three levels for ease of discussion: (1) the intrinsic characteristics of the "marker" (target), (2) protocols for generating laboratory data, and (3) field applications. By ensuring that a consistent set of metrics for characteristics such as accuracy and precision be applied to field studies and published works, we can begin to circumscribe the set of MST tools that will be most useful for discriminating among fecal pollution sources in environmental waters.

Keywords qPCR • Performance • Efficiency • Accuracy • Precision • Error

2.1 Introduction

The nascent field of microbial source tracking has relied upon both library-dependent and library-independent approaches (see Chaps. 3 and 4, respectively) to detect fecal contamination from particular hosts. In particular, the library-dependent approach experienced a high level of application in first five or so years of the 21st century, which included the introduction of statistical methods such as discriminant analysis (Wiggins 1996), principle components analysis (Dombek et al. 2000), or nearest-neighbor analysis (Albert et al. 2003; Ritter et al. 2003;

V.J. Harwood (✉)
Department of Integrative Biology,
University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620, USA
e-mail: vharwood@usf.edu

C. Hagedorn et al. (eds.), *Microbial Source Tracking: Methods, Applications, and Case Studies*, DOI 10.1007/978-1-4419-9386-1_2,
© Springer Science+Business Media, LLC 2011

Robinson et al. 2007) to evaluate complex patterns generated by antibiotic resistance analysis (Hagedorn et al. 1999; Harwood et al. 2000; Wiggins 1996), rep-PCR (Carson et al. 2003; Dombek et al. 2000; McLellan et al. 2003), pulsed-field gel electrophoresis (Myoda et al. 2003), ribotyping(Parveen et al. 1999), and other methods. The validity of results from these library-dependent methods began to be questioned following proficiency testing with blind samples (Griffith et al. 2003; Harwood et al. 2003; Stoeckel et al. 2004). Other pressing concerns with library-dependent methods include the size and scope required for a "representative" library and concerns about broad geographic applicability and temporal stability (Stoeckel and Harwood 2007; US Environmental Protection Agency 2005; Wiggins et al. 2003).

As a result of these findings and concerns, library-independent methods, many of which showed better accuracy in limited proficiency testing compared with the library-dependent methods (Griffith et al. 2003; Harwood et al. 2003; Myoda et al. 2003), began to be more intensively developed and used in field studies. As was done with library-dependent methods, as these methods and markers emerge they should be routinely validated for provision of accurate results. The purpose of this chapter is to outline a strategy for method validation and proficiency testing that is applicable to library-independent MST methods, many of which utilize PCR and/or quantitative PCR (qPCR) to detect a host-associated target organism or gene. By establishment of rigorous performance criteria and application of proficiency tests, MST methods will be evaluated within a consistent framework, paving the way for more confident use in regulatory and legal contexts.

This chapter considers performance of MST methods separately at three levels – the genetic target or "marker," since interpretation of MST data for fecal source indication is dependent upon marker characteristics (sensitivity and specificity within the target population); the protocol for generating laboratory data, since without confidence in the data results cannot be interpreted; and field application, since interpretation of data collected from uncontrolled settings poses additional challenges beyond basic laboratory quality control. In this chapter, we use "performance" when referring to inherent characteristics of the method, e.g., sensitivity, specificity, evenness; and "proficiency" when referring to testing that is specifically designed to evaluate the quality and reliability of laboratory-generated data.

The use of common performance measures and validation strategies in the many studies that are expected over the next decade should facilitate rapid progress in this area, as we continue to work toward availability of reliable analyses, classification approaches, and interpretation strategies for tracking fecal contamination to its sources by use of MST tools. Although we focus here on methods that target specific genes via PCR, the general strategies and most of the considerations discussed here apply in some measure to all of the methodologies discussed in this book (see Chaps. 3 and 9 for criteria that are more appropriate for library- and chemical-based methods, respectively).

2.2 Evaluation of Target (MST Marker) Performance and Suitability

The various markers used for library-independent MST detect the presence of host-associated microbial populations. Sensitivity, or completeness of marker representation in the host population, along with specificity, or exclusivity of the host–microbe association, are critically important parameters (Table 2.1) (Stoeckel and Harwood 2007). Relatively poor sensitivity, which is associated with low-prevalence markers such as those that detect some pathogenic viruses (Noble et al. 2003; Stoeckel and Harwood 2007), frequently causes false-negative results. Incomplete specificity, which is associated with many existing genetic markers

Table 2.1 Characteristics of an ideal vs. a useful MST marker (Harwood 2007; US Environmental Protection Agency 2005)

Characteristic	Ideal marker	Useful marker
Specificity	Marker found only in target host species	Marker is differentially distributed among host species
Distribution in host population	Found in all members of all populations of target host species; contributes to sensitivity of method	Consistently found in host species whose feces could impact the target sites
Evenness	Quantity in the feces of individuals is similar	Quantity in aggregate sources, e.g., sewage influent; animal populations, is similar
Temporal stability in host	Frequency and concentration in host individuals and populations does not change over time	Despite variation in marker frequency and concentration in individuals, these characteristics are stable at the population level
Geographic range/ stability	The frequency and concentration in geographically separated host populations are similar	The marker can consistently be detected and quantified across the geographic area to be studied
Environmental persistence	Consistent decay rate in various matrices and habitats; no increase under any conditions; response to treatment processes and environmental insults is similar to that of pathogens	Predictable decay rate in various matrices and habitats; no increase under ambient conditions; response to treatment processes and environmental insults is characterized
Quantitative assessment	Can be accurately quantified	Accurately indicates presence/ absence of contamination source
Relevance to regulatory parameters	The marker is derived from an organism that is a regulatory tool	The marker is correlated with an organism that is a regulatory tool
Relevance to health risk	The marker is strongly correlated with risk of all types of waterborne disease, e.g., gastroenteritis, dermatitis, upper respiratory infections	The marker constitutes a health risk or is otherwise correlated with a subset of waterborne disease, e.g., viral gastroenteritis

(Harwood et al. 2009; Korajkic et al. 2009; Shanks et al. 2010), can cause false-positive results. The third major issue relevant to performance measurement for markers is evenness of marker distribution (in terms of prevalence and quantity), which applies both among populations and among individuals within a given host population. If the evenness of the marker is different from the evenness of fecal indicator bacteria or pathogens, then simple detection or even quantification of the marker may not be directly comparable to existing regulations or public health risk outcomes. These considerations are discussed in detail below.

2.2.1 Choosing the Tool(s) to Fit the Question

Potential applications of MST include (a) assessment of sources of fecal contamination in recreational or drinking source waters, (b) prioritization of impaired water bodies for total maximum daily load (TMDL) implementation or other interventions, (c) source apportionment for TMDL plans, and (d) forensic applications, i.e., assigning (or relieving) responsibility for pollution. The goals of a given study must be carefully considered when choosing or designing MST marker(s), and deciding whether conventional (presence/absence) PCR-based methods are sufficient or if quantitative PCR (qPCR) is required. For example, if one is most concerned about determining when and where contamination from human sources is present, a suite of human-specific markers may be chosen, and conventional PCR may be sufficient to achieve the study goals. If, however, one is attempting to apportion contributions from various fecal sources for TMDL applications, it would be necessary to use a suite of markers for the identified sources of fecal loading, and qPCR would be required.

Many authors have recommended toolbox or tiered approaches for MST study design, the first meaning that a group of MST methods is on hand and ready for deployment as the specific situation demands and the second meaning that lower cost methods that broadly measure contamination, such as conventional fecal indicator bacteria measurements, are used first, followed by more expensive, technically demanding methods such as PCR where they are needed to accomplish specific goals (Boehm et al. 2003; Lu et al. 2009; McQuaig et al. 2006; Noble et al. 2006; Vogel et al. 2007) (see also Chaps. 16 and 19). Another aspect of the toolbox approach is that multiple methods for detection of contamination from one source can be used to support one another (see below), alleviating the uncertainty that results from imperfections in all methods reported to date. On the contrary, the use of multiple tests increases the cost of a given study and can be unacceptably expensive for end users such as regulatory agencies. This situation can be a particular concern when multiple methods are used to identify one source.

One must also consider the performance characteristics of the methods and how they might affect interpretation of the results; for example, one could use a human-associated marker with high concentration in sewage but incomplete specificity to minimize the probability of false-negative results. Because use of such a marker could yield false-positive results, one might also use a highly human-specific marker that

has the drawback of lower concentration in sewage to back up the indication of human fecal pollution. Other performance characteristics of MST methods, such as sensitivity to inhibition from interfering compounds (discussed below) should also be considered in the context of the characteristics of the water bodies that are sampled.

2.2.2 Ideal Characteristics for MST Markers

Each of the MST markers described in the literature to date has both positive and negative qualities. The same caveat applies to available chemical markers (Chap. 9). Some pathogen markers, such as enteroviruses and adenoviruses, tend to have relatively high false-negative rates in sewage, and particularly in individual human fecal samples (Griffith et al. 2003; Noble et al. 2003), which can lead to low sensitivity. Other markers, such as the human-associated *Bacteroidales* 16S rRNA sequence delineated by PCR primers HF183 and 708R (Bernhard and Field 2000), display incomplete specificity (a low but detectable rate of false positives against feces from nonhuman animals) (Balleste et al. 2010; Harwood et al. 2009). The ideal MST marker is described previously (Table 2.1), and since a marker that meets all these criteria has not been identified for any host, the characteristics of a useful marker are also described as adapted from (US Environmental Protection Agency 2005). The following sections further discuss key characteristics and how they are experimentally assessed.

2.2.2.1 Specificity

The central hypothesis of MST is that some microorganisms have an exclusive or preferential association with the gastrointestinal tract of a particular host species or group, and that these host-associated microorganisms are shed in feces and can be detected in water bodies. The detected markers may be extremely host-specific, such as human polyomaviruses (Ahmed et al. 2009a; Harwood et al. 2009; McQuaig et al. 2006; McQuaig et al. 2009) or they may have limited host specificity, such as some of the human-associated markers targeting *Bacteroidales* 16S rRNA genes (Ahmed et al. 2009b; Harwood et al. 2009; Layton et al. 2006; Shanks et al. 2007).

The specificity of a marker is generally assessed by analyzing fecal and/or sewage samples from animals other than the targeted host (nontarget hosts) (Harwood 2007; Shanks et al. 2010; Stoeckel and Harwood 2007; US Environmental Protection Agency 2005). Although DNA sequences that are candidates for MST markers can undergo a preliminary, *in silico* specificity assessment (i.e., a computer-generated BLAST search against the NCBI database of sequences), such an analysis should not be substituted for testing against nontarget fecal material. A quantitative expression of specificity is 1 minus the proportion of nontarget fecal samples in which the marker is detected, which is also 1 minus the false-positive rate (as described and compiled in (Stoeckel and Harwood 2007)). Specificity is generally expressed as a percentage; therefore, the calculations above would be multiplied by 100. Specificity testing can be accomplished with individual fecal samples

or composite fecal samples (e.g., (Griffith et al. 2003)). Composite fecal samples can be further characterized as compilations of individual fecal samples made by the sampler, in which case one knows how many animals contributed to the sample, or large-scale composites such as sewage influent samples or slurries from dairy cattle operations. An obvious advantage of a composite sample is that, as long as tests negative for the marker, it can allow testing of more individuals with fewer negative-control analyses. Drawbacks to this approach are (a) if a false positive is obtained, one does not know how many individual scats contributed to that result, and (b) it is theoretically possible that the signal from one positive scat could be missed because it is diluted by the other scats in the composite. In the case of humans, it is highly recommended that sewage samples and, when applicable, onsite treatment and disposal systems (OSTDS, or septic systems) be tested because they are more likely to reach water bodies than waste from an individual, and because there is inherent selection for environmental survival or persistence within such systems (Gordon et al. 2002). A similar case can be made for some types of animal feces, e.g., slurries from cattle barns or egg layer poultry operations, or litter from broiler poultry production are very useful composite samples.

Box 2.2.2.1.1

The specificity of the marker commonly known as HF183 for human-source fecal contamination has been well documented both for conventional PCR (Bernhard and Field 2000; Harwood et al. 2009) and for the quantitative adaptations (Kildare et al. 2007; Seurinck et al. 2005; Shanks et al. 2009). Each of these reports is based on reference samples in North America or Europe. Specificity testing in New Zealand, however, indicated that the marker was commonly associated with a local species of opossum (Kirs et al. 2011). Furthermore, although the concentrations were not reported, extended sampling of nontarget sources not previously considered (e.g., fish) can identify additional sources of potential false-positive results (McLain et al. 2009).

Determination of the appropriate number of nontarget samples to include for specificity testing is not standardized, but should be based on the geographic area of the study, the intended use of the marker, and the distribution of host species in the study area that are reasonably expected to impact water quality. The USEPA MST Guide Document (2005) recommends that at least ten animals per host type are sampled for specificity. While it is not practical to sample the feces of more than a small subset of all individuals in a given area, a good faith effort should be made to capture the diversity among relevant host populations. For example, sampling the feces of five cattle from one farm in a study intended to characterize fecal sources in a watershed that is potentially impacted by ten cattle farms is clearly an inadequate effort. A more inclusive strategy in such a case would be to

sample from three or more farms and to make five or more composite samples from 5 to 10 animals each for each farm. If one is testing the specificity of a marker with the goal of using it across a broad geographic region, the number and scope of samples tested should be broadened accordingly. For example, Shanks et al. (2009) tested the specificity of several qPCR methods for human markers using 265 fecal samples from 22 nontarget species collected across USA. The specificity of a qPCR assay for human polyomaviruses was tested with 127 fecal samples from 14 nontarget species (McQuaig et al. 2009).

The specificity parameters reported in the literature may not accurately reflect marker characteristics in a particular study due to factors such as geographic variability in marker distribution and/or the completeness of previous sampling effort(s). For these reasons, validation of the method protocol against reference fecal samples as part of the experimental start-up procedure is strongly recommended (see Sect. 2.3 for details on evaluation of data). Even relatively well-characterized MST markers such as the human-associated *Bacteroidales* 16S rRNA marker HF183 are subject to new findings when tested in a new geographic area or against previously untested host species (Box 2.2.2.1.1).

It is very important to characterize the error rate associated with specificity as thoroughly as is practically possible. There is no universally accepted criterion for the minimum specificity required of a useful MST marker. Of course, 100% specificity is ideal, but is rarely achieved. Even when observed in one study, this figure is rarely maintained over subsequent studies. It is generally agreed that methods with less than 80% specificity are not useful in most circumstances (US Environmental Protection Agency 2005), and the majority of recently published or frequently used methods have 90% or greater measured specificity, at least in the geographic area(s) for which they are characterized (Ahmed et al. 2009a; McQuaig et al. 2009; Shanks et al. 2009; Weidhaas et al. 2010).

When amplification of a particular marker from nontarget sources is noted (generally termed "false positive" in the literature) it may occur because the target sequence is present in the nontarget fecal or sewage sample, e.g., (Harwood et al. 2009). However, any number of other reasons may cause apparent false-positive results, including an uncalibrated thermocycler (annealing temperature too low), the existence of very similar, but demonstrably different, sequences in the sample, or contamination of the sample. These mistakes should be guarded against with adequate controls and method performance evaluations (see Sects. 2.3–2.6), and amplicons should be sequenced to determine whether they (a) are identical to the target, (b) are similar to the target, or (c) represent an unrelated PCR artifact. The latter concern is not as great for probe-based qPCR methods, as probe as well as primer must match the target sequence.

2.2.2.2 Distribution and Sensitivity

The distribution of a marker in the feces of individual members of a host species is a major contributor to method sensitivity. As discussed in Stoeckel and Harwood (2007), the sensitivity of a marker can be defined as the proportion of positive-control

fecal or sewage samples (target samples) that yield positive results. Like specificity values, sensitivity is generally expressed as a percentage, so the figure obtained from the calculation above would be multiplied by 100. A more sensitive marker will be more frequently detectable in polluted water samples than a less sensitive marker, unless it is very widely distributed in the target host population although it is at low concentration.

Like specificity testing, sensitivity can be assessed in individual fecal samples or in composites. The small-scale composites described in specificity testing are not generally used for sensitivity assessment; however, large-scale composites can be particularly valuable when they represent the form of fecal material that is most likely to contaminate a water body. Sewage influent, septage in pump-out trucks or holding tanks, cattle or pig waste lagoons, and litter from poultry barns are some examples of useful, large-scale composite samples for assessment of method sensitivity. Individual fecal samples are useful for determining the evenness of marker distribution within a population (see below). Ideally, one would analyze both individual and composite fecal samples for better characterization of marker distribution, with the caveat that this practice makes specificity testing more costly. Furthermore, human fecal samples from healthy individuals (not clinical samples) can be very difficult to obtain, and at least in USA, permission to obtain such samples also can be a logistical challenge. Sewage samples, on the contrary, generally are very easy to obtain.

The number of samples needed to adequately assess sensitivity is another evaluator of method performance that has been approached in an *ad hoc* fashion. Many recent studies have included 20 or more sewage/septage samples when testing sensitivity of human markers (Harwood et al. 2009; McQuaig et al. 2009; Shanks et al. 2009). A study of the use of bovine polyomaviruses for detection of cattle waste tested 26 individual urine samples and ten individual fecal samples (Hundesa et al. 2010). Certainly, one must be cognizant of the geographic area represented by a given study and attempt to collect samples that adequately represent that area. For example, a study that examined the usefulness of MST markers for use across the US Gulf Coast states tested human sewage and septage from the Florida peninsula ($n=24$), the Florida panhandle several hundred miles away ($n=18$), and Mississippi ($n=11$) (Harwood et al. 2009). Data were also obtained from Louisiana and Texas. One hundred percent sensitivity was observed for the three human-associated MST markers (human polyomaviruses, HF183 *Bacteroidales*, and *M. smithii*), providing a strong indication that these markers are prevalent across the Gulf Coast of USA. In practice, the initial sensitivity testing for most new MST methods is more limited, but broadens as others use the methods and as more comprehensive studies are developed. It is highly recommended, however, that markers with limited or unknown specificity be fully vetted before publication of results or recommendation for wider usage.

The evenness of marker distribution among individuals within host populations can influence its usefulness in various locales or geographic ranges (among population distribution) and also becomes important when relatively small numbers of animals may impact a water body. Evenness within a host population is less important, however, in cases where homogenized waste is the source. For example, evenness is

less important when leakage from a dairy lagoon is concerned compared with direct deposition of fecal material to a stream by individuals in a small herd of cattle, as the dairy lagoon waste is a composite from many individuals. A recent study (Shanks et al. 2010) of seven PCR methods that target bovine feces has found that marker prevalence and quantity varied widely among herds, even between herds housed at the same facility. These findings support the recommendation of Stoeckel and Harwood (2007) that preliminary testing of marker suitability in a given area is an extremely useful step for determining whether a given MST marker should be further vetted for a particular application and/or in a particular geographic region.

2.3 Evaluation of Data Quality

The previous section discusses considerations for selection of an appropriate analysis for MST based on method performance on laboratory samples. It is important to recognize that collection of reliable data about a given MST target in environmental waters is a daunting task that is quite a bit more complex than working with fecal and sewage samples. Regardless of whether the data are quantitative or qualitative, the researcher must start by evaluating the effectiveness of his or her analytical detection methods in the type(s) of environmental waters included in the study. Though it may be tempting to directly extrapolate bench-level results to environmental scenarios, potential errors introduced in the intermediate steps also must be considered. This section begins with a discussion of data quality assurance for laboratory results, i.e., the data delivered by the analytical protocol. In the next section, assurance of protocol consistency across extended time frames is evaluated. Various complexities introduced by the processes of sample concentration, purification, and storage are then discussed along with approaches to measure and correct for potential error added during these steps. Many of the problems and solutions presented in this section are couched in terms of marker detection by use of quantitative PCR; however, analogous situations and solutions should be apparent for other protocols.

2.3.1 Quality of Data Delivered by the Analytical Protocol and Other Preliminary Considerations

Before environmental samples are analyzed, it is critical to ensure that data of acceptable quality (that will provide meaningful results) can be generated by the analytical protocol. For example, qualitative presence/absence data are meaningless unless the laboratory has confidence in the consistency of detection on positive-control samples. Absence – more appropriately referred to as failure to detect – is much more meaningful when bounded by the limit of detection. Further, the integrity of the values provided by an analytical instrument must be supported by basic laboratory quality-control practices. The purpose of this section is to briefly present and describe the

analytical quality-control measures that are necessary for provision of quality MST data. The "Quality Assurance/Quality Control Guidance for Laboratories Performing PCR Analyses on Environmental Samples" (US Environmental Protection Agency 2004) is a resource that might be consulted for related information.

2.3.1.1 Use and Composition of Negative and Positive Controls

Negative controls should be included in all protocols to guard against reagent contamination and (or) false-positive measurements by the instrument. The so-called "no-template control" for PCR is generally done by substituting reagent-grade water for test sample in the PCR reaction. In analyses other than the PCR, this type of negative control frequently is called the reagent blank. Positive results in the no-template control indicate contamination of reagents or equipment and the analysis should be repeated after identifying and correcting the source of contamination. A reagent blank is necessary with each batch of sample analyses to guard against false-positive results (US Environmental Protection Agency 2004). Recommendations for frequency of no-template controls range from inclusion in each standard curve run (Sigma qPCR Tech manual at http://www.sigmaaldrich.com/life-science/molecular-biology/pcr/quantitative-pcr/qpcr-technical-guide.html) to a rate representing 1 per 10 environmental samples analyzed (US Environmental Protection Agency 2004).

Inclusion of appropriate positive-control samples is also necessary during the initial evaluation of the PCR protocol, and, at minimum, each day the samples are run in the laboratory. Initial positive-control tests ideally should be done with well-defined material obtained from a colleague or culture collection (such as a pure culture of a target organism, if applicable), a plasmid containing the target, or a known-positive DNA extract or amplicon. In the event that such a control is not available, one could substitute sewage or feces from the target source; however, the resultant amplicon must be sequenced to determine that the correct product has been produced. The verified product from fecal material can then be cloned into a plasmid vector and subsequently used as positive control material. For quantitative methods, preliminary positive controls would include a dilution series on the positive-control material to evaluate amplification efficiency and the analytical limit of detection (described in more detail in Sect. 2.3.1.3).

After the protocol is demonstrated to consistently generate true-positive reactions, further tests must be done to characterize the method performance. It is essential that reaction positive controls be included with each set of test samples for analyses to guard against false-negative results. For quantitative methods, the standard curve (described in Sect. 2.3.1.2) may serve as the reaction positive control. For qualitative methods, a reaction positive control is necessary for each batch (as described in, for example, USEPA 2004). The concentration of analyte in the reaction positive control should be high enough to be consistently detected. On the contrary, the concentration in the reaction positive control must not be so high as to lack relevance to the environmental samples (and, as a practical note, excessive amounts of amplicon are more likely to produce laboratory contamination, which can be very difficult to eradicate). In our experience, use of a synthetic sample, such as a target-carrying plasmid,

at 3–10× the protocol detection limit as the reaction positive control is useful for this purpose. Once the protocol is established, similar positive-control measures incorporated to establish ongoing method performance (Sect. 2.4) may be suitable or used in place of these initial positive-control samples.

2.3.1.2 Composition and Performance Characteristics of the Standard Curve

The primary purpose of the qPCR standard curve is to allow quantification of the target concentration in the purified DNA extract. To fulfill this purpose, the performance of the method (as indicated by the standard curve) should be evaluated. In cases where the model response is well understood – as, for example, in the qPCR – the slope of the standard curve can be used as a diagnostic test of protocol performance. In any case, the standard curve can be used to confirm that the dynamic range of the analytical protocol (the range of concentrations over which the target can be accurately quantified) is suitable for sample analysis. It is important to insure that plasmid control DNA is well purified and free of chromosomal DNA; otherwise, the concentration of target DNA will be overestimated and the standard curve will be erroneous (skewed high). The standard curve should include not fewer than three concentration levels (in addition to the blank) to ensure linearity of response. The concentration levels should be evenly distributed across the relevant range of concentrations that one wishes to detect, such as decimal (tenfold) dilutions for the qPCR.

Lack of linearity at the high or low end of the standard curve indicates that the curve extends beyond the dynamic range of either the measuring instrument or the protocol chemistry. When lack of linearity is noted, either by visual observation of the standard curve or by a coefficient of determination (R^2 value) less than 0.985 (Sigma qPCR tech manual), the detection method must be optimized or the standard curve must be truncated to the linear portion. Truncation of the standard curve may necessitate dilution of the sample extract to bring the sample concentration within the upper limit of detection. Observations higher and lower than the standard curve cannot be treated as reliable quantitative data.

The slope of the standard curve can be used as a performance criterion. In the PCR, for example, doubling of the target DNA is expected during each cycle. When the threshold cycle (Ct) is plotted against $\log_{10}$ (concentration) of target DNA, this leads to an idealized standard curve with a slope of −3.32. This slope generally is converted to amplification efficiency ($E = 10^{(-1/\text{slope})} - 1$) when used as a performance criterion. Amplification efficiency reflects the relationship between the change in target concentration and the change in fluorescence measured; efficiencies between 0.8 and 1.1 often are considered to be acceptable (Sigma qPCR tech manual).

Replicate measurements for assessing the precision of the assay under ideal conditions (in buffer and water) are provided by the standard curve. Shanks et al. (2010) assessed the precision of standard curve values for several qPCR assays targeting human waste by calculating the mean percent coefficient of variation (CV) of the various data points included in the standard curve. Percent CV is the standard deviation divided by the mean and multiplied by 100 for expression as a

Box 2.3.1.2 The Importance of Accurate Dilutions

The slope of the standard curve should indicate the overall efficiency of the reaction. What if the standard curve reflects dilution error instead of nonideal reaction efficiency? The following table illustrates the error that would result from accepting and using a standard curve with a nonideal slope that indicates amplification efficiency of 0.8 (80%) or 1.1 (110%), if the true efficiency were 1.0 (100%). The results are presented in terms of log (concentration), the native output from the qPCR, as well as untransformed concentration to demonstrate the effect of slope on percent error of the measurement. Measurement error can be quite high, and because the standard curve is log normal, the measurement error increases as true concentration increases.

True concentration	Measured C_T	Calculated concentration (per µL)				Calculated measurement error (%)	
		Log(conc)		Concentration			
		$E=0.8$	$E=1.1$	$E=0.8$	$E=1.1$	$E=0.8$	$E=1.1$
10	36.667	0.85	1.08	7	12	40	16
100	33.333	1.71	2.15	51	141	95	29
1,000	30.000	2.56	3.23	367	1,682	173	41
10,000	26.667	3.42	4.30	2,623	20,002	281	50

percentage. Variability in the % CV for the various assays was observed, ranging from 1.03 to 3.00%.

Another important consideration related to standard curves is an understanding of what is being measured. A typical standard curve measures only the response of the instrument when exposed to a given concentration of analyte. For this purpose, qPCR standards can be prepared from extracted, transformed, or synthesized DNA for various targets. If more specific information on performance is desired then more complex approaches must be developed. For example, amplification efficiency within the specific sample matrix (i.e., DNA extracted from an environmental water sample) might be measured by spiking the sample with a known amount of target DNA (Sect. 2.6.1); this approach is helpful for determining whether the DNA extract contains substances that are inhibitory to the PCR. As another example, considering that the measurement typically is back-calculated to represent concentration in the original environmental matrix, some researchers choose to use standardized cell suspensions, rather than extracted DNA, for creation of the dilution series. Each diluted cell suspension must then be extracted independently prior to analysis. In this way, some of the uncertainties in sample processing are incorporated into the standard curve (see Sect. 2.4.2 for full discussion of the uncertainties in sample processing and ways to address those uncertainties).

For the purposes of this section, the standard curve is meant solely to represent the instrument response to a given concentration of analyte. Creation of a standardized solution of analyte can be challenging. For the example of qPCR, a suspension

might be made of genomic DNA from the target organism, plasmid DNA with a target amplicon inserted, or synthesized oligomeric DNA representing the target. In any of these cases, determination of copy number concentration can be less than straightforward. If the total DNA is measured to standardize a genomic DNA extract, it is critical to know the molecular weight of the genome and confirm that the DNA extract is free from extrachromosomal DNA. Similarly, if the total DNA is measured for use to standardize a plasmid DNA extract, the molecular weight of the transformed plasmid must be known and the absence of unintended plasmids and of chromosomal DNA should be confirmed. If synthesized DNA is used as the standardized stock material, the lyophilized product must be rehydrated at a high enough concentration to allow quantification. Extreme care must be taken with this solution, as well as less-concentrated extracts of genomic or plasmid DNA, to avoid contamination of the working space with positive-control material. Care must also be taken with the standardized stock material to ensure that the concentration remains consistent throughout the experimental time frame, i.e., it should be aliquoted into volumes for one-time use and stored at −80°C if possible.

2.3.1.3 Estimation of the Protocol Limit of Detection (LOD)

Once a standardized solution of positive control material is created, the method protocol limit of detection can be measured by dilution to extinction. This protocol limit of detection can be used as a quantitative upper bound when reporting and analyzing no-detect data (e.g., <10 gene copies), as described by Helsel (Helsel 1990).

2.3.1.4 Estimation of the Protocol Limit of Quantification (LOQ)

The protocol limit of quantification cannot be lower than the lowest standard of the standard curve because, lacking evidence of linearity, one cannot say that the standard regression is applicable to C_T values above that representing the lowest standard. In many cases, the more dilute standard(s) of the standard curve generate variable responses. To evaluate whether a dilute standard generates less-precise estimates of original concentration relative to other standards, it is useful to test whether the standard error of the mean of replicate measurements is significantly higher than that for other samples. If so, the limit of quantification might more appropriately be set to the most dilute sample for which the standard error of the mean C_T among replicate measurements is consistent with the other standards. When presenting data, results that fall below the LOQ but above the LOD represent a quandary since the marker was detected; however, it was not quantifiable. It is good practice to censor these data and replace them with a notation to indicate detected, but not quantified (along with the LOQ), or to qualify the concentration calculated as “estimated.”

2.3.2 Evaluation of Recovery Methods

The previous sections dealt with performance of the analytical protocol for detection of selected MST markers. In addition to these considerations, the processing of environmental samples (sampling, concentration, extraction, and purification) also must be considered. Decisions such as the use of centrifugation, filtration, or selective capture to concentrate a sample, and the protocol for extraction and purification of DNA-based or other markers, should be made based on both careful research and collection of performance data. The selected sample processing protocol should consider completeness of recovery, reproducibility of recovery, and purity of the final extract according to the data needs of the study. A test suspension of known-source fecal material, such as human wastewater for protocols designed for general or human-associated fecal contamination, can be useful in these tests.

2.3.2.1 Completeness of Recovery

Aliquots from a single sample should be processed by each candidate method in duplicate or triplicate. The analytical protocol developed using positive-control material (Sect. 2.4.1.2) should be used to detect the selected marker. Some processing methods do not result in full volumetric recovery of sample at each step as, for example, when the DNA extraction protocol requires leaving behind some portion of the total volume to prevent carryover of contaminants. To avoid bias in the final interpretation of recovery, the final results should be calculated in terms of copies or mass per volume of original sample. For example, the researcher may choose to test three protocols with the parameters against 100-mL aliquots of a dilute sewage sample, as shown in Table 2.2. If the researcher is content to evaluate the three protocols solely against the calculated target concentration per μL extract, he or she might choose to use Protocol 1, with nearly 5 times the apparent recovery (concentration in extract) relative to protocol 2 and 1.5 times the apparent recovery relative to protocol 3. If volumetric losses in the concentration, extraction, and purification steps are considered, however, it becomes apparent that protocol 3 outperformed the other two protocols.

If determination of absolute rather than relative recovery through the protocol is necessary, a standardized suspension of cells, i.e., (Stoeckel et al. 2009) or DNA i.e., (Siefring et al. 2008) must be used. In most cases, the target used to measure absolute recovery efficiency will not be the same as the MST marker. Rather, the target will be a surrogate measure and the degree to which it accurately reflects the recovery of MST marker must be tested.

2.3.2.2 Precision of Recovery

In cases where consistency is as or more important than completeness of recovery, the criterion described in the previous section changes from selection for highest

Table 2.2 The influence of recovery of extracted volume from environmental samples on the estimate of analyte concentration

Parameter	Protocol 1	Protocol 2	Protocol 3	Units
Initial volume	100	100	100	mL
Volume lysate	1.5	1	1.5	mL
Volume to extraction	1.3	1	1.3	mL
Volume to purification	1	1	1	mL
Volumetric percent recovery	67%	100%	67%	
Final extract	200	200	500	µL
C_T	27.254	29.547	26.548	Cycles
Standard curve	$C_T = -3.452 \times \log(\text{conc}) + 38.542$			
Log_{10} Conc	3.27	2.61	3.47	per rxn
Volume reaction	5	5	5	µL
Dilution factor	10	10	10	
Conc (in extract)	4.66E+02	1.01E+02	2.98E+02	per µL
Conc (in water)	1.40E+03	2.02E+02	2.24E+03	per 100 mL

recovery to selection for greatest precision (lowest variability among replicate measurements). Precision is frequently expressed as the relative standard deviation, which is the standard deviation of at least three replicated divided by the mean, then multiplied by 100 for expression as a percentage (US Environmental Protection Agency 2004).

2.3.2.3 Purity of DNA Extract

There is no clear-cut, simple test for the suitability of DNA extract for PCR amplification, i.e., freedom from inhibitory substances. If the researcher desires to evaluate extraction purity as a protocol selection criterion, then the candidate methods should be challenged by a series of samples that represents the range of contaminants that might be encountered in the experimental samples. Tests for matrix inhibition (described in Sect. 2.5.2) should be done to identify the methods that result in extracts that must be diluted or further purified to overcome inhibition.

2.4 Consistency of Results: Ongoing Proficiency Testing

Once the sample processing and analytical protocols are selected and characterized for use in the individual laboratory (as described in Sect. 2.2), they will be applied to environmental samples over the course of the study. Before discussing assurance of data quality for environmental samples (Sect. 2.5), this section examines ongoing quality-control tests that can be used to ensure that results across the entire study are comparable. Changes in equipment, staffing, or reagents may result in unexpected changes in method performance.

2.4.1 Use of Process Blanks to Establish a Demonstrated Limit of Detection

In every laboratory, no matter how careful the analyst is, there exists the possibility of false-positive results caused by cross-contamination by test samples or positive-control samples. Extra care must be taken to guard against contamination by the amplicons generated in the standard curve and in areas where concentrated positive-control materials are manipulated. Negative-control samples, generally called process blanks to distinguish them from analytical negative-control samples (frequently termed no-template controls), should be liberally included at all steps of the analysis. Field blanks test for contamination in collection and transport. These are samples in which uncontaminated laboratory water is collected, transported, and processed alongside the test sample. Extraction blanks test for contamination during concentration, extraction, and purification in the laboratory. For these samples, uncontaminated water is processed alongside test samples. Analytical blanks (as described in Sect. 2.3.1.1) test for contamination during the analytical protocol. In addition to detecting contamination problems at various stages of sample manipulation, results from these samples can be used to protect against false-positive results caused by laboratory contamination.

For qPCR-based analyses, the results of all blank analyses can be compiled over a reference time frame. Most of these analytical results likely will be nondetects; however, there likely will be several incidences of low-level detection. Depending on project needs, a level of acceptable error might be established. If, for instance, the researcher wishes to ensure that contamination is identified and rejected at least 95% of the time it causes detection, the fifth percentile C_T value among all blank analyses might be used in place of the analytical limit of detection. If the data pattern indicates that contamination was sporadic and concentrated in a particular time frame, and there are sufficient blank analyses in the time frame, then a rolling LOD or time-specific LOD might be applied to specific time frames. This strategy can be effective to discount low-level detections that have a high probability of being caused by contamination.

2.4.2 Amplification Efficiency

Amplification efficiency is an expression of the doubling rate of DNA during the exponential phase of the PCR. A perfectly efficient reaction precisely doubles the DNA at each cycle, resulting in 100% efficiency (Karlen et al. 2007). Amplification efficiency can be quite variable from one specific qPCR assay to the next, i.e., Shanks et al. (Shanks et al. 2010) observed values ranging from 89 to 99.5% for a variety of assays targeting human waste.

Variation in efficiency, and thus in the slope of the standard curve, can magnify into highly variable results (see box, Sect. 2.3.1.2). To guard against imprecision in

slope measurement that is caused by inaccurate dilution, a control chart can be created (i.e., http://www.itl.nist.gov/div898/handbook/mpc/section3/mpc37.htm). In essence, the expected value (slope of the standard curve) and typical variation around the expected value are calculated. Runs in which the measured slope does not fall within the expected range can be subject to closer scrutiny, and may indeed be subject to reanalysis. In this way, false estimation of concentration can be avoided.

It is valuable to consider the possible causes of variation in amplification efficiency when determining good laboratory practices or troubleshooting a problem. The thermal cycler might be out of calibration, a situation that can be addressed by recalibration according to manufacturer's specifications. More difficult to address are problems with the standardized positive-control material and dilutions thereof that are used for the standard curve. Poorly calibrated pipettors and (or) poor technique can contribute to bias in the dilution series (dilution by a fixed factor other than 10; see Box 2.3.1.2). Stored standards may degrade, lose water to sublimation and (or) evaporation, or be sorbed out of solution.

To avoid these problems, proper calibration of thermal cycler and pipettors must be maintained. Technical staff should not only be properly trained, but they should be encouraged to demonstrate proficiency by pipetting control volumes. Concentrated positive control stocks should be stored frozen at −80°C in sealed cryogenic vials. Freeze–thaw cycles for stored stocks should be rigorously minimized – the practice of aliquoting storage material into multiple tubes with volumes suitable for one use is a good strategy for minimizing freeze–thaw cycles. DNA extracts should be stored in low-binding tubes and, particularly for dilute (low concentration) extracts, and incorporation of carrier DNA might prove to be a useful strategy.

2.4.3 Reproducibility (Precision) Over Time

Another way to evaluate data quality over time is to include repeated measures in each analytical batch. The repeated measure sample should be spiked with sewage or other target fecal source and could be DNA extracted from an environmental water sample, in other words, it should present the same challenges for analysis as typical test samples. The material chosen for the repeated measure should be aliqoted into multiple cryogenic vials and stored frozen at −80°C until use to prevent freeze–thaw degradation.

2.5 Performance in Environmental Samples

As described in the previous sections, reliable detection of MST marker presence and (or) concentration in a DNA extract using the PCR can be complex. The considerations outlined in Sects. 2.3 and 2.4 enhance our ability to accomplish this task in laboratory-generated positive-control samples. For test samples consisting of

environmental waters or other matrices, additional effects should be considered. The efficiency of marker recovery through sample processing steps must be measured or assumed based on previous measurements. The potential for poor amplification efficiency in the individual sample, sometimes caused by inhibitors in the extract matrix, must be evaluated. Finally, variability must be measured to allow comparison of values or detection rates across samples.

2.5.1 Recovery Efficiency

In the rare instances that recovery efficiency through the processes of concentration, extraction, and purification has been measured, wide differences have been observed within sample sets. Several recent studies have used spike-and-recovery controls to estimate recovery efficiency. In one study (Stoeckel et al. 2009), recovery efficiency ranges were 1–5% in one sample set, 3–12% in another, and 0.2–30% in a third. These wide ranges in recovery efficiency are not unique. One study (Koike et al. 2007) reported a recovery range of 20–200% and another (Lebuhn et al. 2004) measured recoveries in the range of 10–70%. It is clear from these reports that variable recovery efficiencies can greatly affect the measured concentration of analyte, and lack of recovery efficiency data can impair cross-sample and cross-study comparisons by (1) masking true differences because of high variability and (2) showing false differences because of different recovery efficiencies in different studies. Use of measured recovery efficiency to adjust observed concentrations to account for analyte lost in processing can reduce variability in replicate measures (Stoeckel et al. 2009) and may be essential for accurate comparison of results between studies.

These considerations are not limited to detection using the qPCR. Divergent recovery efficiencies can complicate the interpretation of presence–absence analysis for endpoint PCR. For example, several samples that contain marker at approximately 5 times the limit of detection might be encountered. If the range of recovery efficiencies for a particular protocol is 10–60%, then the marker will be detected in some of the samples, but not all of the samples. Particularly in descriptive experiments with little replication, this scenario can lead to erroneous conclusions unless the limit of detection can be adjusted according to the recovery efficiency. With adjustment, it would be apparent to the researcher that the negative results were more common in samples with lower recovery efficiency, indicating that those samples were not necessarily less contaminated compared with the samples that tested positive.

2.5.2 Matrix Inhibition

An assumption of the quantitative PCR is that the efficiency of each reaction is the same as the efficiency measured in the standard curve. The matrix of the individual reaction tube may influence the efficiency of the PCR (Cankar et al. 2006), In the

case of conventional PCR, differences in amplification efficiencies between tubes can alter the detection limit, making interpretation of nondetect results difficult. These considerations generally are discussed in terms of "matrix effect" (US Environmental Protection Agency 2004).

Four approaches are sometimes used to address the matrix effect on amplification efficiency. The first depends on generation of measureable product in the qPCR. The slope of the linear portion of the amplification curve is compared with the slope of the standard curve. If the amplification efficiency is consistent, the slopes will not be different (Ramakers et al. 2003).

The second approach is applicable only to the qPCR using TaqMan or related chemistry with a probe. A competitive internal positive control (CIPC) sequence, also known as internal amplification control (IAC) (Gregory et al. 2006; Sen et al. 2007), is developed with the same flanking primers as the marker but a different (artificial) internal sequence for probe hybridization. This positive control sequence is created and inserted into a carrier plasmid. A second probe, labeled with a different fluor and, therefore, detectable on a different channel from the marker detection probe, is added to the qPCR mix to create a multiplex qPCR using one primer pair and two probes. After carefully optimizing the multiplex reaction, the CIPC is added to every reaction mixture at low concentration (approximately 100 copies per reaction). Failure to detect the CIPC at the expected concentration indicates low amplification efficiency and inhibition in a reaction. In cases where the failure is due to matrix inhibition, dilution may result in a more accurate estimate of original marker concentration.

The third approach, also applicable to qPCR, is to use a dilution series to detect inhibition by amplification. Using the same premise as the standard curve, in which the slope of a series of dilutions measures amplification efficiency, a dilution series on a test sample also can be prepared. The difference between threshold cycles measured for decimal dilutions of a given sample should be approximately 3.3. Large differences in amplification efficiency are apparent when the difference is substantially less than 3.3. A benefit of this approach is that the appropriate dilution is known upon completion of the qPCR analysis. The assay does not have to be repeated, and the test simply indicates which dilution was appropriate. On the negative side, the dilution method is time-, labor-, and reagent-intensive; it cannot be used for samples in which the ambient concentration is near the detection limit, and it cannot detect slight deviations from the assumed amplification efficiency.

A traditional matrix spike is the fourth possible approach (USEPA 2004) and can serve a similar purpose for both conventional (end point) and quantitative PCR. The PCR master mix for a run is created and divided into two aliquots. One aliquot is used to generate the standard curves (if desired) and no-template control and the other is amended with a finite amount of the positive-control standard. A spiked no-template control must be run, and all test reactions also use the spiked master mix. Matrix inhibition is apparent in cases where the C_T of the test reaction is higher than the C_T of the spiked no-template control, or where a conventional PCR fails to amplify. This approach leaves one open to the possibility of partial inhibition (amplification present at low efficiency). It is theoretically possible to quantify the spiked PCR reaction and obtain an estimate of the spike concentration by

subtracting the concentration in a parallel, unspiked test reaction. However, given the variability in concentration measurement by qPCR, it is virtually impossible to detect small differences, in the order of 100 copies, between two measurements that are substantially higher.

2.5.3 Processing Precision

Error measurement is critically important, particularly if the results might be challenged in a court of law (Chap. 13). One strategy by which processing and analytical error can be evaluated is incorporation of split replicate samples in the experimental design. A selected number of samples (in the order of 1 in 20, depending on project needs) is split into two samples, with unique identifying labels, and processed as separate test samples. The average standard error of the means can be used as a useful measure of precision in the overall process of sample processing and analysis.

2.6 Validation of Method Accuracy

Topics related to selection and (or) validation of an appropriate MST marker, initial evaluation of protocol performance against quality-control samples, ongoing tests of protocol performance, and performance measures to enhance the validity of data collected on test samples were discussed in detail in prior sections. None of those considerations is valuable if the data do not lead the researcher to correct information about environmental contaminants and processes. The purpose of this section is to describe broader quality-control measures to test the validity of interpretations.

2.6.1 Confirmatory Reference Materials

It is useful to check natural reference materials (i.e., feces or sewage) for sensitivity and specificity on a regular basis, particularly if a method is to be used over several years' period. It is essential to gather new reference materials for performance testing when a study is initiated in a new geographic area, as the distribution of the marker in hosts may vary. Note that quantitative data, e.g., target copies derived from qPCR, should be normalized to a common parameter such as grams of fecal material.

2.6.2 Aqueous Suspensions

The practice of suspending reference material in water and then testing the results provides a useful "reality check" for MST method performance (Field et al. 2003; Griffith et al. 2003; Harwood et al. 2003; 2009). Sensitivity, specificity, and a relative,

semiquantitative assessment of the method limit of detection can be gained in this manner. For the latter, sewage (i.e., municipal sewage or water from animal-waste holding ponds) can be subjected to serial dilutions and added to buffer or an environmental water sample. These mixtures are then concentrated and DNA is extracted according to the laboratory's standard protocols. In this manner, the extent to which a pollution source can be diluted and still detected (Harwood et al. 2009) or quantified can be assessed. One can also add mixed sewage sources to aqueous samples and the detection and/or quantification of multiple targets can be assessed. An important consideration in interpreting results from multiple-fecal source samples is that the various targets may well be present at different concentrations in the reference material; therefore, if qPCR is used, the relative proportion of multiple targets quantified in the sample will not necessarily reflect the proportions of reference material added to the sample.

2.7 Conclusions

As the science of MST continues to grow and become more sophisticated, performance criteria will doubtlessly become more codified and consistent. All of the methods and applications discussed in this book depend upon the rigorous application of performance criteria and validation strategies to ensure the accuracy and usefulness of results. In a period of about two decades, we have made major strides toward legitimizing the use of MST for applications ranging from total maximum daily load assessment, to protection of the health of recreational water users, to pollution source identification in the legal area. We trust that this chapter will be obsolete within a decade, as further improvements to method performance criteria will doubtlessly be discovered and implemented by MST practitioners.

References

Ahmed, W., Goonetilleke, A., Powell, D., Chauhan, K., and Gardner, T. (2009a). Comparison of molecular markers to detect fresh sewage in environmental waters. *Water Res* **43**(19), 4908–17.

Ahmed, W., Goonetilleke, A., Powell, D., and Gardner, T. (2009b). Evaluation of multiple sewage-associated *Bacteroides* PCR markers for sewage pollution tracking. *Water Res* **43**(19), 4872–7.

Albert, J. M., Munakata-Marr, J., Tenorio, L., and Siegrist, R. L. (2003). Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli*. *Environ Sci Technol* **37**(20), 4554–60.

Balleste, E., Bonjoch, X., Belanche, L. A., and Blanch, A. R. (2010). Molecular indicators used in the development of predictive models for microbial source tracking. *Appl Environ Microbiol* **76**(6), 1789–95.

Bernhard, A. E., and Field, K. G. (2000). A PCR assay To discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Appl Environ Microbiol* **66**(10), 4571–4.

Boehm, A. B., Fuhrman, J. A., Mrse, R. D., and Grant, S. B. (2003). Tiered approach for identification of a human fecal pollution source at a recreational beach: case study at Avalon Bay, Catalina Island, California. *Environ Sci Technol* **37**(4), 673–80.

Cankar, K., Stebih, D., Dreo, T., Zel, J., and Gruden, K. (2006). Critical points of DNA quantification by real-time PCR--effects of DNA extraction method and sample matrix on quantification of genetically modified organisms. *BMC Biotechnol* **6**, 37.

Carson, C. A., Shear, B. L., Ellersieck, M. R., and Schnell, J. D. (2003). Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl Environ Microbiol* **69**(3), 1836–9.

Dombek, P. E., Johnson, L. K., Zimmerley, S. T., and Sadowsky, M. J. (2000). Use of repetitive DNA sequences and the PCR To differentiate *Escherichia coli* isolates from human and animal sources. *Appl Environ Microbiol* **66**(6), 2572–7.

Field, K. G., Chern, E. C., Dick, L. K., Fuhrmann, J., Griffith, J., Holden, P., LaMontagne, M. G., Le, J., Olson, B., and Simonich, M. T. (2003). A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. *Journal of Water and Health* **01**(4), 181–94.

Gordon, D. M., Bauer, S., and Johnson, J. R. (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology* **148**(Pt 5), 1513–22.

Gregory, J. B., Litaker, R. W., and Noble, R. T. (2006). Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. *Appl Environ Microbiol* **72**(6), 3960–7.

Griffith, J. F., Weisberg, S. B., and McGee, C. D. (2003). Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *Journal of Water and Health* **01**, 141–151.

Hagedorn, C., Robinson, S. L., Filtz, J. R., Grubbs, S. M., Angier, T. A., and Reneau, R. B., Jr. (1999). Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Appl Environ Microbiol* **65**(12), 5522–31.

Harwood, V. J. (2007). Assumptions and limitations of microbial source tracking methods. *In* "Microbial Source Tracking" (J. Santo-Domingo, Sadowsky, M., Ed.). ASM Press, Washington, D.C.

Harwood, V. J., Brownell, M., Wang, S., Lepo, J., Ellender, R. D., Ajidahun, A., Hellein, K. N., Kennedy, E., Ye, X., and Flood, C. (2009). Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. *Water Res* **43**(19), 4812–9.

Harwood, V. J., Whitlock, J., and Withington, V. (2000). Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. *Appl Environ Microbiol* **66**(9), 3698–704.

Harwood, V. J., Wiggins, B., Hagedorn, C., Ellender, R. D., Gooch, J., Kern, J., Samadpour, M., Chapman, A. C. H., Robinson, B. J., and Thompson, B. C. (2003). Phenotypic library-based microbial source tracking methods: Efficacy in the California collaborative study. *J. Water Health* **01**, 153–66.

Helsel, D. R. (1990). Less than obvious: Statistical treatment of data below detection limit. *Environmental Science and Technology* (24), 1767–1744.

Hundesa, A., Bofill-Mas, S., Maluquer de Motes, C., Rodriguez-Manzano, J., Bach, A., Casas, M., and Girones, R. (2010). Development of a quantitative PCR assay for the quantitation of bovine polyomavirus as a microbial source-tracking tool. *J Virol Methods* **163**(2), 385–9.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C., and Mermod, N. (2007). Statistical significance of quantitative PCR. *BMC Bioinformatics* **8**, 131.

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. *Water Res* **41**(16), 3701–15.

Kirs, M., Harwood, V. J., Fidler, A. E., Gillespie, P. A., Fyfe, W. R., Blackwood, A. D., and Cornelisen, C. C. (2011). Source tracking faecal contamination in an urbanized and a rural waterway in the Nelson-Tasman region, New Zealand. *New Zealand J Freshwater Res* **45**:53–8.

Koike, S., Krapac, I. G., Oliver, H. D., Yannarell, A. C., Chee-Sanford, J. C., Aminov, R. I., and Mackie, R. I. (2007). Monitoring and source tracking of tetracycline resistance genes in lagoons and groundwater adjacent to swine production facilities over a 3-year period. *Appl Environ Microbiol* **73**(15), 4813–23.

Korajkic, A., Badgley, B. D., Brownell, M. J., and Harwood, V. J. (2009). Application of microbial source tracking methods in a Gulf of Mexico field setting. *J Appl Microbiol* **107**(5), 1518–27.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., and Sayler, G. (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* **72**(6)**,** 4214–24.

Lebuhn, M., Effenberger, M., Garces, G., Gronauer, A., and Wilderer, P. A. (2004). Evaluating real-time PCR for the quantification of distinct pathogens and indicator organisms in environmental samples. *Water Sci Technol* **50**(1), 263–70.

Lu, J., Santo Domingo, J. W., Hill, S., and Edge, T. A. (2009). Microbial diversity and host-specific sequences of Canada goose feces. *Appl Environ Microbiol* **75**(18), 5919–26.

McLain, J. E., Ryu, H., Kabiri-Badr, L., Rock, C. M., and Abbaszadegan, M. (2009). Lack of specificity for PCR assays targeting human Bacteroides 16S rRNA gene: cross-amplification with fish feces. *FEMS Microbiol Lett* **299**:38–43.

McLellan, S. L., Daniels, A. D., and Salmore, A. K. (2003). Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. *Appl Environ Microbiol* **69**(5), 2587–94.

McQuaig, S. M., Scott, T. M., Harwood, V. J., Farrah, S. R., and Lukasik, J. O. (2006). Detection of human-derived fecal pollution in environmental waters by use of a PCR-based human polyomavirus assay. *Appl Environ Microbiol* **72**(12), 7567–74.

McQuaig, S. M., Scott, T. M., Lukasik, J. O., Paul, J. H., and Harwood, V. J. (2009). Quantification of human polyomaviruses JC Virus and BK Virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples. *Appl Environ Microbiol* **75**(11), 3379–88.

Myoda, S. P., Carson, C. A., Fuhrmann, J. J., Hahm, B.-K., Hartel, P. G., Yampara-Iquise, H., Johnson, L., Kuntz, R. L., Nakatsu, C. H., Sadowsky, M. J., and Samadpour, M. (2003). Comparison of genotypic-based microbial source tracking methods requiring a host origin database. *Journal of Water and Health* **01**(4), 167–180.

Noble, R. T., Allen, S. M., Blackwood, A. D., Chu, W., Jiang, S. C., Lovelace, G. L., Sobsey, M. D., Stewart, J. R., and Wait, D. A. (2003). Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study. *J Water Health* **1**(4), 195–207.

Noble, R. T., Griffith, J. F., Blackwood, A. D., Fuhrman, J. A., Gregory, J. B., Hernandez, X., Liang, X., Bera, A. A., and Schiff, K. (2006). Multitiered approach using quantitative PCR to track sources of fecal pollution affecting Santa Monica Bay, California. *Appl Environ Microbiol* **72**(2), 1604–12.

Parveen, S., Portier, K. M., Robinson, K., Edmiston, L., and Tamplin, M. L. (1999). Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. *Appl Environ Microbiol* **65**(7), 3142–7.

Ramakers, C., Ruijter, J. M., Deprez, R. H., and Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339**(1), 62–6.

Ritter, K. J., Carruthers, E., Carson, C. A., Ellender, R. D., Harwood, V. J., Kingsley, K., Nakatsu, C., Sadowsky, M., Shear, B., West, B., Whitlock, J. E., Wiggins, B. A., and Wilbur, J. D. (2003). Assessment of statistical methods used in library-based approaches to microbial source tracking. *J Water Health* **1**(4), 209–23.

Robinson, B. J., Ritter, K. J., and Ellender, R. D. (2007). A statistical appraisal of disproportional versus proportional microbial source tracking libraries. *J Water Health* **5**(4), 503–9.

Sen, K., Schable, N. A., and Lye, D. J. (2007). Development of an internal control for evaluation and standardization of a quantitative PCR assay for detection of Helicobacter pylori in drinking water. *Appl Environ Microbiol* **73**(22), 7380–7.

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (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* **7**(2), 249–59.

Shanks, O. C., Domingo, J. W., Lu, J., Kelty, C. A., and Graham, J. E. (2007). Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Appl Environ Microbiol* **73**(8), 2416–22.

Shanks, O. C., Kelty, C. A., Sivaganesan, M., Varma, M., and Haugland, R. A. (2009). Quantitative PCR for genetic markers of human fecal pollution. *Appl Environ Microbiol* **75**(17), 5507–13.

Shanks, O. C., White, K., Kelty, C. A., Sivaganesan, M., Blannon, J., Meckes, M., Varma, M., and Haugland, R. A. (2010). Performance of PCR-based assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples. *Environ Sci Technol* **44**(16), 6281–8.

Siefring, S., Varma, M., Atikovic, E., Wymer, L., and Haugland, R. A. (2008). Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. *J Water Health* **6**(2), 225–37.

Stoeckel, D. M., and Harwood, V. J. (2007). Performance, design, and analysis in microbial source tracking studies. *Appl Environ Microbiol* **73**(8), 2405–15.

Stoeckel, D. M., Mathes, M. V., Hyer, K. E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T. L., Fenger, T. W., Samadpour, M., Strickler, K. M., and Wiggins, B. A. (2004). Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. *Environ Sci Technol* **38**(22), 6109–17.

Stoeckel, D. M., Stelzer, E. A., and Dick, L. K. (2009). Evaluation of two spike-and-recovery controls for assessment of extraction efficiency in microbial source tracking studies. *Water Res* **43**(19), 4820–7.

U.S. Environmental Protection Agency (2004). Quality assurance/quality control guidance for laboratories performing PCR analyses on environmental samples. U.S. Environmental Protection Agency. EPA 815-B-04–001.

U.S. Environmental Protection Agency (2005). Microbial source tracking guide document. U.S. Environmental Protection Agency. EPA/600/R-05/064. June 2005.

Vogel, J. R., Stoeckel, D. M., Lamendella, R., Zelt, R. B., Santo Domingo, J. W., Walker, S. R., and Oerther, D. B. (2007). Identifying fecal sources in a selected catchment reach using multiple source-tracking tools. *J Environ Qual* **36**(3), 718–29.

Weidhaas, J. L., Macbeth, T. W., Olsen, R. L., Sadowsky, M. J., Norat, D., and Harwood, V. J. (2010). Identification of a Brevibacterium marker gene specific to poultry litter and development of a quantitative PCR assay. *J Appl Microbiol* **109**:334 – 47.

Wiggins, B. A. (1996). 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* **62**(11), 3997–4002.

Wiggins, B. A., Cash, P. W., Creamer, W. S., Dart, S. E., Garcia, P. P., Gerecke, T. M., Han, J., Henry, B. L., Hoover, K. B., Johnson, E. L., Jones, K. C., McCarthy, J. G., McDonough, J. A., Mercer, S. A., Noto, M. J., Park, H., Phillips, M. S., Purner, S. M., Smith, B. M., Stevens, E. N., and Varner, A. K. (2003). Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. *Appl Environ Microbiol* **69**(6), 3399 – 405.

http://www.springer.com/978-1-4419-9385-4

Microbial Source Tracking: Methods, Applications, and Case Studies

Hagedorn, C.; Blanch, A.R.; Harwood, V.J. (Eds.)
2011, XIV, 642 p., Hardcover
ISBN: 978-1-4419-9385-4