

Development and Implementation of Rapid Microbiological Methods for Measuring Recreational Water Quality

John F. Griffith

Southern California Coastal Water Research Project (SCCWRP), Costa Mesa, California

Introduction

California's beach water quality monitoring programs are the most comprehensive in the nation. State Health Services regulations implemented in response to California Assembly Bill 411 require measurement of three indicator bacteria (total coliforms, fecal coliforms and enterococci) on at least a weekly basis at high use beaches. Regulations further require that the public be warned of possible health risk if any of these indicator bacteria exceed threshold values that were established through epidemiological studies.

While these beach monitoring programs are extensive, there remains opportunity for improvement. Current laboratory measurement methods used to enumerate indicator bacteria require up to 96 hours. Contaminated beaches remain open during this processing period, but may already have returned to acceptable levels by the time laboratory results are available and warning signs are posted (Leecaster and Weisberg 2001, Boehm et al. 2002). Thus, swimmers may be exposed to contamination during the sample processing period and warning signs informing the public of possible exposure to waterborne pathogens are often posted after waters are already clean. This time lapse also inhibits tracking of contamination sources, because the signal can dissipate before upstream tracking is initiated. Consequently, lacking a more rapid method, investigators are unable to follow the trail of contamination back to its origin.

Continued advances and improvements in molecular and immunological techniques provide new opportunities for measuring bacteria more rapidly (Noble and Weisberg 2005). While current methods rely on bacterial growth and metabolic activity, these new methods allow direct measurement of cellular attributes such as genetic material or surface immunological properties. By eliminating the necessity for a lengthy incubation step, some of these methods provide results in less than four hours, a short enough time for managers to take action to protect public health (i.e. post or close a beach) on the same day that water samples are collected.

While promising, these rapid methods require extensive independent testing before they can be adopted for use in beach water quality monitoring. Analyzing environmental water samples presents challenges not frequently encountered in other fields, such as complex sample matrices and the presence of other potentially confounding native bacterial species. The State of California has requested that the Southern California Coastal Water Research Project (SCCWRP) conduct such evaluation tests. SCCWRP has conducted three such evaluation studies to date. This paper provides an overview of each these studies.

2004 Method Evaluation Study

This study evaluated four new rapid methodologies (Griffith et al. 2004). The first was immunomagnetic separation coupled with ATP Bioluminescence (IMS/ATP) performed by the University of Michigan. The IMS/ATP method uses magnetic beads coated with antibodies specific to enterococci to bind and capture the target bacteria in a water sample. Once bound, the magnetic bead/antibody/bacteria complexes are pulled out of solution using a powerful magnet. Now separated from the rest of the bacterial population, the captured cells are enzymatically-lysed, releasing their ATP into solution. This ATP is then quantified via a bioluminescent assay. The amount of ATP is converted to the number of bacteria captured following a calibration curve established using stock cultures of known concentration.

The second was a flow cytometry (FC) method, which was performed by Advanced Analytical Technologies and employed the Advanced Analytical Technologies (Ames, IA) RBD 3000 instrument. Like IMS/ATP, this method derives its specificity from antibodies specific to enterococci. Once captured by the first antibody, the target cells are labeled using a secondary antibody containing a fluorogenic tag that, when excited by a laser in the instrument, allows the cells to be enumerated in the flow cytometer.

The third method was quantitative polymerase chain reaction (QPCR) performed by U.S. EPA, which detects and enumerates unique genetic targets found in enterococci (Santo Domingo et al. 2003, Haugland et al. 2005). The bacteria are first captured on a filter. The filter containing the bacteria is then subjected to bead beating, which mechanically lyses the cells and releases their DNA into solution. This DNA then is used in the quantification step, where specific DNA targets are simultaneously amplified and measured using the Taqman® system of fluorescent probes (Applied Biosystems, Foster City, CA) and the advanced optics of the Q-PCR instrument (Cepheid, Sunnyvale, CA).

The last method was dual-wavelength fluorimetry (DWF), performed by the University of Connecticut, which relies on the same sugar-fluorophore substrate as is used in the commercially available IDEXX™ assays, but advances the detection process. Through use of a novel fluorometer, the method simultaneously measures the rate at which bacteria take up the chromogenic substrate as well as the rate at which the fluorescent byproduct of substrate metabolism appears. This ratiometric measurement allows detection and enumeration of target bacteria in a matter of only four hours.

The evaluation approach was to assess equivalency with existing methods through simultaneous processing of water samples using both new and existing methods of enumerating fecal indicator bacteria. Samples processed included both natural samples and laboratory-created samples, to ensure that a range of conditions was evaluated. Laboratory-created samples were included because they offer the ability to control the number of indicator organisms and potentially interfering contaminants present, but they do not completely mimic natural conditions. Environmental water samples were included because they contain complex combinations of interferences that cannot be duplicated in

artificial samples, though they offer less control over specific variables that need to be evaluated.

Testing occurred in two phases. The first phase involved application of the new methods by the experts that developed them. The second phase involved application of the methods by senior members of several local microbiology laboratories, who would become likely users of these methods should they be approved. The goal of the second phase was to assess whether these new technologies are readily transferable to local practitioners.

During both phases, all samples also were processed by five local laboratories using methods they employ in their routine processing. For enterococci, this included both defined Enterolert® defined substrate (DS) (IDEXX, Westbrook, ME) and membrane filtration (MF) methods.

Results and Discussion

Of the methods evaluated, QPCR was the most comparable to the reference labs. However, it tended to overestimate levels of enterococci, producing a higher mean value than traditional methods for two-thirds of the samples. The method also exhibited relatively poor precision between replicate samples, particularly when levels of enterococci were low. The greatest variability was observed for blanks and seawater samples containing interferences. Variability was lowest for samples containing a simple seawater matrix spiked with sewage.

While Q-PCR was the most accurate of the methods, it generally overestimated concentrations of enterococci relative to the culture-based methods. This is consistent with previous Q-PCR applications in environmental water samples, where culture-based methods have rightly or wrongly been used as a kind of “gold standard”, and may reflect the fact that measurements of non growth-related attributes of target bacteria do not differentiate between cultivable and non-cultivable cells (Duprey et al. 1997, Frahm and Obst 2003, Brinkman et al. 2003). This is also consistent with the greater success of this method with sewage-inoculated samples than with urban runoff inoculated samples, since sewage is a fresher fecal source and should have a higher percentage of cultivable cells.

DWF had the best precision among the methods and in some cases, better than some of the laboratories using standard methods. This was particularly impressive, given that the precision of the reference laboratories in this study likely was better than that for most microbiology laboratories because each participated in at least three prior intercalibration studies designed to facilitate comparability among southern California laboratories (Noble et al. 2003a, Noble et al. 2003b, Griffith et al. 2006). Variability within and across laboratories prior to these exercises was typically higher than that observed for any of the new methods evaluated in this study.

While DWF was highly repeatable, it severely overestimated levels of enterococci in samples containing urban runoff, either as the matrix or as an inoculum. This was true

even when the urban runoff had been passed through a 0.2 μm filter, suggesting that non-biological processes might have been responsible for cleavage of the chromogenic substrate in these samples. This was perplexing, as DFW employs the same fluorogenic substrate, 4-methylumbelliferyl- β -D-glucoside, as used in the Enterolert® method and none of the reference laboratories using this method exhibited a similar trend for the same samples.

The FC method registered high values (ca. 1000 cells/100 ml) for almost all samples, including most blanks. This was particularly problematic, as it did not provide for discrimination between contaminated and non-contaminated sites. Subsequent discussions with the manufacturer led to the conclusion that the sensitivity of the method was not great enough to accurately enumerate the low levels of enterococci required for beach monitoring purposes without concentrating the samples at least ten-fold.

The IMS/ATP method had the opposite problem, measuring values near zero for most samples. It was unclear why this method performed so poorly. The method had produced results comparable to existing methods in previous freshwater testing (Lee and Deininger 2004) and in informal participation in previous comparative testing with California marine samples (Griffith et al. 2006).

Despite the poor performance of two of the methods, QPCR and DWF performed well enough to be optimistic about their eventual chances for adoption, especially for use as screening tools to detect extreme events. It was recognized that further work would be necessary to assess and correct the causes for overestimation. However, it was also recognized that overestimation by rapid methods is preferable to underestimation. While there are costs associated with false positives, these costs may be outweighed by extra protection against an extreme circumstance. For example, a sewage spill occurring at a high-use beach could go undetected, potentially exposing bathers to human pathogens for at least a day before results from current EPA-approved assays became available. These rapid measurement methods would provide the ability to detect a sewage spill the same day.

While none of the new rapid methods produced results equivalent to those of the reference laboratories, all of the methods were transferred easily to local personnel with minimal training investment. All of the local participants indicated a comfort level in implementing the methods. Moreover, the precision of their results was comparable to that when the method developers performed the methods. While the local practitioners generally were laboratory managers and had more experience than the typical technician, many had no specific previous training in many of the areas employed here, such as micropipetting or PCR. While there is need for performance improvement in each of the new methods, technology transfer does not seem to be an important impediment to method adoption.

2005 Method Evaluation Study

In response to a favorable showing in the first test, several participants from the 2004 study expressed interest in participating in further testing. Additionally, several other groups developing rapid detection technologies approached SCCWRP to discuss inclusion of their methods in future tests. Toward that end, SCCWRP and the Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET) developed a cooperative relationship and initiated a second evaluation test in June 2005.

The study involved assessing equivalency with traditional water quality monitoring methods through simultaneous processing of water samples using both new and EPA approved methods. Samples processed included both natural samples and laboratory-created samples, to ensure that a range of conditions was evaluated. Laboratory-created samples were included because they offer the ability to control the number of indicator organisms and potentially interfering contaminants present, but they do not completely mimic natural conditions. Environmental water samples were included because they contain complex combinations of interferences that cannot be duplicated in artificial samples, though they offer less control over specific variables that need to be evaluated.

Four types of methods, implemented by six investigators, were evaluated. Several investigators implemented multiple permutations of their method resulting in a total of 11 methods being tested. The first method type, quantitative polymerase chain reaction (QPCR), was implemented by three research groups using nine permutations of a similar approach. The basic steps included capturing bacteria on a filter and then using either bead beating, or bead beating coupled with chemical treatment to lyse the cells and release the target deoxyribonucleic acid (DNA). The differences among the methods generally involved the enzymes, primers and probes used or the methods used to release and capture the bacterial DNA from the target cells.

The University of North Carolina (UNC) performed three QPCR method permutations and applied them to both *E. coli* and enterococci. The first two methods differed only in the way the samples were prepared. In the UNC Extracted method, filters were prepared using a commercial DNA extraction kit that included bead beating of the filter and DNA capture and purification. In the UNC Bead Beaten method, the UNC team used bead beaten samples provided by the USEPA National Exposure Research Laboratory (NERL) team that were not subjected to the DNA concentration and purification step. The UNC Alternate Cycling Time method was identical to the UNC Bead Beaten method, except that it used a longer annealing time.

The USEPA NERL performed three QPCR method permutations and applied them to enterococci: The ABI-Taqman method (Haugland et al. 2005), the Omni-TaqMan method, and the Omni-TaqMan method conducted at an alternate temperature. The main difference between the ABI-TaqMan method and the Omni-TaqMan method was in the DNA polymerase used. The ABI-TaqMan method uses AmpliTaq™ DNA polymerase and reagents along with TaqMan™ probes (Applied Biosystems, Foster City, CA), while the Omni-TaqMan method uses OmniMix™, a freeze-dried all-in-one polymerase and

reagent system containing TaKaRa DNA polymerase (Cepheid, Sunnyvale, CA) with the same TaqMan™ probes.

USEPA Region 1 (R1) performed three QPCR method permutations applied to both enterococci and *E. coli*. The R1 permutations differed only in the manner in which the result was quantified. The first permutation was Absolute Quantitation, in which the gene copy number in each sample was interpolated to cell counts via a standard curve created using DNA standards. The second permutation was Adjusted Absolute Quantitation, which differed from Absolute Quantitation by assuming an altered number of gene copies per cell. The third permutation was Relative Quantitation, which used a standard curve generated from actual cell suspensions rather than a DNA standard to interpolate cell concentrations in samples.

The second method type was transcription mediated Amplification (TMA), which was performed by Gen-Probe and applied to enterococci. TMA is similar in concept to QPCR in that it amplifies a genetic target in the bacteria and uses a fluorescent probe for detection (Persimoni *et al.* 2002). TMA differs from QPCR in that it targets bacterial RNA, rather than DNA, and uses two enzymes (RNA polymerase and reverse transcriptase) to amplify the target sequence. In this application, the assay targeted Group II and Group III *Enterococcus* spp. As with the QPCR method, bacteria were first captured on a filter. Following this step the filter was washed to remove interfering substances and chemical and enzymatic treatment used to lyse the bacteria. A specific capture DNA probe was then used to bind enterococcal rRNA which was subsequently captured by complementary sequences bound to magnetic beads. The bead/probe/RNA complex was then pulled out of solution using a powerful magnet, leaving cellular debris and other contaminants behind. The captured RNA template was then combined with amplification reagents and inserted into a real-time amplification engine.

The third class of method was dual-wavelength fluorimetry (DWF), which was conducted by Rosewood Industries and applied to both enterococci and *E. coli*. DWF was performed in the same manner as described above for the 2004 evaluation.

The fourth method type was an immunological dipstick manufactured by Silver Lake Research. In this method, sample water is combined with liquid growth media and incubated for 4 – 6 hours at constant temperature on an orbital shaker. Once incubation is complete, the dipstick, which contains antibodies specific to *E. coli* is immersed in the growth media. This method produces a binary answer. If the original concentration of *E. coli* was greater than 400 per 100 mL of water, then a black bar becomes visible on the dipstick, indicating a positive result.

Results from the new methods were compared to those from the traditional methods employed by the reference laboratories in several ways. First, we assessed the number of individual samples from each new method that differed by half a log unit from the reference laboratory median. Half a log unit was selected because previous laboratory intercalibration studies have demonstrated that this is the typical range of variability observed for traditional methodologies, both within and among laboratories (Noble *et al.*

2004; Griffith et al. 2006). For this analysis, blank samples were counted as outside of range when values exceeded 50 cells/100 ml.

Second, results were evaluated for false positives and false negatives relative to the State of California standard of 104 cells/100 ml for enterococci and 400 cells/100 ml for *E. coli*, as the State requires posting warning signs for any sample that occurs above this level. The decision of whether a sign should have been posted that day, against which the new methods were being evaluated, was based on the median concentration for that sample as measured by the reference laboratories.

The third analysis assessed precision of the measurements, which we could accomplish because each sample was processed three times as blind replicates. Precision was quantified as coefficient of variation (CV) and was compared between the new methods and the reference laboratories.

The fourth analysis assessed the variability of the methods through linear regression across all samples. Each method was compared with the grand median of the reference laboratories and with each of the other methods.

The final analysis was an integrated evaluation designed to discern how often the results from the new tests would result in a public health officer making the same or a different decision regarding issuing a public health warning. Here triplicate results from each sample processed by new methods were compared to those of the reference laboratories and categorized as “Equivalent”, “Not Materially Different”, or “Materially Different” than traditional water quality monitoring methods.

To be considered equivalent to current methods, a sample had to exhibit the following characteristics:

- 2 of 3 replicates and the median were correct with respect to the AB411 standard
- 2 of 3 replicates were within $\frac{1}{2}$ log unit of reference lab median
- Replicate exhibited a smaller variance than worst reference lab

To be deemed materially different than current methods the criteria were:

- 2 of 3 replicates were incorrect with respect to standard
- Median value differed by $>1/2$ log unit from reference lab median
- Coefficient of variation is twice that of the worst reference lab

Samples deemed “Not materially different than current methods” failed none of the materially different criteria, but did not meet equivalency criteria.

Each of these evaluations was also performed for the traditional measurement methods by using DS as a reference for MF, and vice versa. This provided context for the certainty of EPA approved methods within which to interpret viability of the new methods. This analysis was only possible for enterococci, though, as the reference labs used only DS to measure *E. coli*.

The time elapsed between samples being provided and results submitted was also quantified. A target of four hours or less was established, as this turnaround time allows beach managers to collect samples in the morning and post warning signs by noon if bacterial concentrations are elevated.

Results and Discussion

The most encouraging result of the 2005 study was the dramatic improvement exhibited by several of the genetic methods over the methods tested in 2004 (Griffith et al. 2004). Of these, TMA and the QPCR methods developed by USEPA NERL and UNC in conjunction with Cepheid performed best in terms of agreement with routine EPA approved water quality monitoring methods. The QPCR methods were faster on average than was TMA. This was especially true of those QPCR methods that employed only bead beating prior to the amplification step.

The bead beating QPCR methods were among the most accurate with respect to the State's standard for enterococci and scored well in terms of equivalency to traditional methods in regard to making beach management decisions. However, as in the 2004 test, bead beating QPCR methods tended to overestimate enterococci compared to traditional growth-based methods and exhibited greater variability between replicate samples than did growth-based or other QPCR methods.

The UNC Extracted Method for enterococci was among the most accurate of all methods when concentrations of enterococci were low, but had a marked tendency toward underestimation when bacterial concentrations were $>10^3/100$ mL. Still, this method was still able to produce repeatable results at concentrations near the State standard, which is the range of most concern to beach managers.

While the UNC Extracted method performed reasonably well for enterococci, it outperformed all other methods in this study for measurement of *E. coli*. Results were deemed equivalent or not materially different than traditional methods for making beach management decisions for almost 90% of the samples. Despite this high performance, the UNC Extracted QPCR method for *E. coli* did encounter some difficulty with underestimation for one set of ambient samples.

The USEPA R1 methods were the most repeatable of the QPCR techniques, but they severely overestimated levels of both *E. coli* and enterococci, possibly due to systematic error in the calibration or calculation steps. Despite this, these methods were the most repeatable of any evaluated in the study and exhibited the highest correlation with traditional methods when compared across several different concentrations of the same inoculant category. High precision and correlation to traditional methods, yet with a significant overestimation bias, suggests a problem with calibration.

TMA was the most accurate of the genetic methods with respect to the State standard for enterococci. This method also performed well in the integrated evaluation, but tended to underestimate concentrations of enterococci, which is the opposite problem from most of

the QPCR applications. There are several possible explanations for this, the most likely of which is that TMA targets cellular ribosomal ribonucleic acid (rRNA), rather than ribosomal deoxyribonucleic acid (rDNA). rRNA levels are known to vary dramatically depending on the physiological state of the cell (Brock et al. 1994) and empirical evidence suggesting that the physiological state of the target bacteria played a role in underestimation of enterococci concentrations was observed in the results from laboratory created samples. Further, TMA underestimated enterococci concentrations only slightly for samples inoculated with laboratory cultures or primary sewage influent, where the fecal material was only hours old, but underestimated by almost an order of magnitude for the urban runoff inoculum, in which the age of fecal material and physiological state of the bacteria was less certain.

A second factor that might have led to underestimation is that the TMA targets Group 2 and Group 3 enterococci, which include *E. faecalis* and *E. faecium* and are the dominant species of enterococci found in the human gut (Geldreich et al. 1978). This high level of specificity is designed to minimize cross-reactivity with non-enterococcal organisms. In contrast, traditional growth-based methods target a larger set of enterococcal species. However, speciation was conducted on isolates measured by traditional methods from samples used in the evaluation testing and species targeted by TMA consistently accounted for the majority of the enterococci in the test samples. While this difference is real, in most cases it would only be a small contributor to the underestimation of TMA relative to traditional methods.

A third factor that may have affected the TMA results is contamination by laboratory dust. During the first day of testing, a large number of observers were present and there was extensive use of an outside door near where this method was being implemented. The TMA method is thought to be sensitive to airborne particles, which can inhibit the efficiency of nucleic acid amplification. While we cannot quantify this potential inhibition, the possible presence of airborne contaminants in the laboratory should be considered when contemplating use of this method in future applications. This is especially important since many environmental microbiology laboratories may not meet the cleanliness standards found in the clinical laboratories where TMA is more routinely used.

DWF performed poorly compared to most other methods and to its own performance in testing conducted the previous year (Griffith et al. 2004). The performance of this method was hampered by both software and equipment problems. The software failed on the first day of testing. Repairs were made the following morning, but this required that the first day samples be held overnight before they could be run. As a result, we could only make a comparison to DS results from one laboratory that kindly volunteered to perform an additional analysis specifically for this purpose. The second day results were much improved, but on the third day, a hardware failure again compromised the results. A forensic investigation of the analytical instrument identified a voltage problem that produced spikes in the signal coming out of the detector. Further, testing will be necessary before an accurate evaluation of this method is possible.

The Immunological Dipstick method performed best on ambient samples, recording no false positives or false negatives. However, this method exhibited considerable difficulty with laboratory created samples, particularly those inoculated with cultured cells. On the surface, these samples should be simpler as they have less potential interference from the matrix or from other bacterial species. This method has great appeal because it could potentially be used to make water quality decisions in the field. However, to reach its potential, more research will be necessary to assess why there were false negatives for so many of the created samples.

In the overall assessment, methods performed well enough for researchers to be optimistic regarding possible implementation in the near future. In particular, results from the UNC QPCR, the NERL QPCR and the GenProbe TMA methods would have led to the same beach management decisions for more than 75% of the enterococci measurements. While this implies an error rate of approximately 25%, it must be considered in context of the error rate for traditional methods. Traditional methods are highly variable, with a typical confidence interval that is 50% or more of the measured value (Griffith et al. 2006, Noble et al. 2004). In this study, we used both DS and MF as traditional methods for the enterococci samples and found that results from the two traditional methods would have resulted in a different management decision for 11% of the samples. Thus, the error rate for the new methods is only approximately twice that of traditional methods and is likely to improve as the developers learn where biases exist from evaluations such as this one.

For the UNC Extracted *E. coli* method, the results were even more promising, with almost 90% agreement in terms of beach management decisions, comparable to the rate of agreement between the two traditional methods.

The new methods measure different bacterial properties than traditional methods and are not expected to produce completely equivalent results. A preferred approach to method validation is to conduct epidemiological studies that establish a relationship between new method measurements and health risk, but this is a lengthy and expensive process. It potentially requires modification of the State's bacterial standards based on the results of such studies. Assessing equivalency is a more expeditious approach for incorporating rapid methods into a warning system that is presently hampered by extended processing time.

Thus, readiness of the new methods for routine use is a subjective determination that involves balancing equivalency between new and traditional methods with the desire to incorporate new rapid methods into a beach water quality warning system in need of improvement. To assist with this decision process, we consulted with Beach Water Quality Work Group (BWQWG) of the State Water Resources Control Board, which includes members of the regulatory, public health, scientific and environmental communities, and has historically been instrumental in providing the State of California recommendations regarding approaches to beach water quality monitoring. The BWQWG identified six applications for rapid indicators, including tracking spatial progress and dilution of inland sewage spills as they move toward the beach; re-opening

beaches subsequent to a sewage spill; routine beach monitoring; tracking fecal contamination back to its source; National Pollutant Discharge Elimination System (NPDES) regulatory compliance; and tracking trends in beach condition. They established desired levels of equivalency for each application.

The BWQWG determined that the QPCR and TMA methods might presently be ready for use in several, but not all, of the applications. However, they identified the desirability for additional testing focused on ambient samples to ensure that the methods were evaluated under a full range of potential confounding factors. They also expressed preference that further testing be conducted by personnel from local laboratories, rather than the method developers, to ensure that the methods are readily transferable to personnel that would be responsible for their routine use once approved.

2006 Beta Testing Study

Following the 2005 evaluation study, California's Beach Water Quality Workgroup (BWQWG) was asked to evaluate the results to assess whether the new methods were ready for adoption within California. The BWQWG determined that while the testing was an appropriate first step for evaluating new rapid indicator measurement methods, it was insufficient alone for making recommendations as to whether the new methods were ready for adoption. They identified two shortcomings in the testing. First, the test placed emphasis on laboratory-created samples, which were used to ensure that testing with a limited number of samples covered a wide range of bacterial concentrations. The BWQWG felt that the tests needed to be expanded to include a greater number of ambient samples because they may contain confounding factors less likely to be encountered in laboratory-created samples. Second, they were concerned that the method developers processed the test samples using molecular methods that are not currently commonplace in water quality testing laboratories, and identified the need to establish that local water quality personnel could produce comparable results when running the assays themselves.

The BWQWG suggested that there should be an additional type of testing, referred to as beta testing, in which local practitioners perform the new methods on typical ambient beach water samples in parallel with existing methods. They felt that this should be done by at least two local laboratories, processing at least 100 samples. Developers of two of the methods that fared best in the 2005 test agreed to collaborate with SCCWRP and two local water quality monitoring laboratories to conduct beta testing of these methods.

The study involved simultaneous processing of samples using two new rapid methods and two existing methods. The rapid methods tested were Quantitative Polymerase Chain Reaction (QPCR) as developed by Dr. Rachel Noble of the University of North Carolina and Transcription-Mediated Amplification (TMA) as developed by Gen-Probe Incorporated (San Diego, CA). The existing methods included EPA Method 1600 (mEI agar) and the IDEXX defined substrate method (Westbrook, ME). QPCR was performed for both enterococci and *E. coli* by the Orange County Sanitation District. TMA was performed for enterococci by the Orange County Public Health Agency.

There were 163 samples processed, 138 of which were ambient samples collected from 41 locations. The remaining samples consisted of seawater spiked with primary sewage influent or secondary effluent. All samples were processed in duplicate using both sets of methods. Testing was conducted from February through July 2006.

Ambient water samples were collected from five categories of locations: open-ocean beaches distant from storm drains (that serve as outlets for land-based runoff); open-ocean beaches near storm drains; enclosed embayment beaches; locations within flowing storm drains; wet weather samples from open ocean beaches. Sewage spiked samples were created by inoculating clean ocean water with varying concentrations of either primary sewage influent or secondary sewage effluent and stirring continuously for a minimum of 15 minutes.

All samples were split in thirds for processing. The first third was analyzed for enterococci and *E. coli* using existing methods by the laboratory that collected the ambient sample or created the spiked sample. The second third was used by the collecting agency for processing one of the new methods, QPCR in the case of OCSD or TMA in the case of OCPHL. The last third was transported immediately on ice to the non-collecting agency for analysis by the other of the new methods.

QPCR was performed using the method developed by Dr. Rachel Noble as described previously for the 2005 method evaluation study. QPCR was performed on the initial 33 samples using the extracted method. However, the multi-step purification and concentration portion of the method proved logistically difficult for personnel at the test laboratories to perform. As a result, a decision was made to switch sample preparation to the less labor-intensive bead-beaten approach. The balance of samples was analyzed for both enterococci and *E. coli* by the bead-beaten QPCR method, except for eleven samples lost to a laboratory accident. In addition, the laboratory personnel began storing samples for batch analysis at $-80\text{ }^{\circ}\text{C}$ following the initial filtration. This increased efficiency, as the thermocycler was able to accommodate a week's worth of samples in a single run.

For absolute quantification of the ENT and EC targets using the QPCR assay, water quality personnel initially utilized a standard curve approach. Because of workflow issues, however, this approach was changed to using a "calibrator" for each QPCR run, i.e. known quantities of either *E. faecalis* or *E. coli* were added to blank filters, bead beaten, and analyzed in the same manner as the unknown water samples, providing a basis for quantification. Results from the calibrators were then used 1) as positive controls for the *Enterococcus* target or the *E. coli* target in the QPCR assays, and 2) as the basis for target sequence quantification using the dCt method of Haugland et al. (2005). The calibrators provided a means for quantification of run-to-run variability, sequence recovery and PCR efficiency.

Two types of internal controls were also incorporated into the QPCR assay. The first of the internal controls was a Sample Processing Control (SPC) in the form of *Lactococcus lactis* cells, which allowed assessment of target recovery from the bead beating process. The second type of internal control was the Super Smart Internal Control (SSIC) developed by Cepheid (Sunnyvale, CA) as part of the SmartBead® system of reagents and was included in all enterococci and *E. coli* QPCR reactions and is designed to detect inhibition of the PCR reaction that could lead to underestimation of cell concentrations in the sample.

The initial calculation of QPCR results employed the ΔC_T (comparative cycle threshold) calculation method. This calculation method, derived by Applied Biosystems (Anonymous 1997) for calculating the ratios of target sequences in two DNA samples (e.g. a calibration and water filtrate sample) normalizes for differences in total DNA recovery from samples using QPCR analysis C_T values for the SPC sequence. A second set of results adjusted for amplification efficiency was calculated by applying an amplification factor (AF) to the original results produced by the ΔC_T method. The AF is a measure of the average efficiency

at which target DNA is amplified during QPCR cycling and was calculated by using the average of many standard curves conducted by the users at OCSD.

TMA was performed using the method developed by Gen-Probe as described previously for the 2005 method evaluation study. Following the amplification process, a standard curve was used to calculate the number of enterococci RNA copies in each sample. An equation was used to calculate the number of enterococci per 100 ml based upon the expected number of rRNA copies per cell. No internal controls were utilized in the TMA assays during this study, although such controls have been developed for this assay subsequently.

TMA was performed on all 163 samples. Similar to QPCR, most samples were frozen following the initial purification step and processed in batches at the end of each week to maximize laboratory efficiency.

In addition to the testing described above, a small set of laboratory-created samples containing secondary sewage effluent was treated with chlorine that was subsequently neutralized, and assayed over time. The purpose of these samples was to test the ability of the genetic methods to accurately enumerate enterococci in chlorine disinfected wastewater. s

As before, the primary means of data analysis was to compare the results from the new rapid methods to those produced by EPA-approved methods. This was done in four ways. First, we assessed the number of individual samples from each new method that differed by half a log unit from the reference method median. Second, results were evaluated for false positives and false negatives relative to the State of California standard of 104 cells/100 ml for enterococci and 400 cells/100 ml for *E. coli*. Third precision of the measurements was quantified as the coefficient of variation (CV) and was compared between the new and existing methods.

The fourth analysis was an integrated evaluation designed to discern how often the results from the new tests would result in a public health officer making the same or a different decision regarding issuing a public health warning. Here duplicate results from each sample processed by new methods were compared to those of the reference labs and categorized as “Equivalent”, “Not Materially Different”, or “Materially Different” than existing methods.

To be considered equivalent to current methods, a sample had to exhibit the following characteristics:

- Both replicates and the median were correct with respect to the state standard
- Both replicates were within ½ log unit of existing methods median
- Replicates exhibited a variance smaller than 2x the variance of existing methods

To be deemed materially different than current methods the criteria were:

- Average of replicates incorrect with respect to standard
- Average of replicates differed by $>1/2$ log unit from existing methods
- Coefficient of variation is $> 4x$ that of existing methods

Samples that failed none of the materially different criteria, but did not meet all of the equivalency criteria, were classified as “Not materially different than current methods”.

Results and Discussion

Technology transfer of rapid measurement methods to working water quality laboratories was largely successful. Except for the QPCR extracted method, which was abandoned because OCSD staff found it difficult and time-consuming, personnel from both beta testing laboratories were able to competently execute the rapid methods. While both laboratories identified numerous opportunities for automating several steps in the process, they found the methods workable and were able to achieve results that were comparable to those produced by expert operators in the previous evaluation studies (Griffith and Weisberg 2006).

Both methods produced results for enterococci that would have led to a public health officer making an equivalent decision as if the measurement had been made with an EPA-approved method about 70% of the time. This is only slightly less than the 75% agreement rate between Enterolert and EPA Method 1600 for the same samples. However, the differences between the two current methods resulted from random measurement precision error, whereas the rapid methods were biased low relative to current methods. This led to a false negative rate almost three times higher than when the two current methods were compared to each other.

A false negative, in which a sample that actually exceeds standards is measured as below standards, is problematic in a public warning system. Beach managers place a high priority on ensuring that bathers aren't swimming in contaminated water. When a false negative occurs, there is no other mechanism for capturing the fact that a problem exists. In contrast, a false positive could lead to an inappropriate posting, but this would be quickly remedied by additional sampling with alternative methods or alternative indicators that would be triggered by the positive measurement.

The most likely explanation for underestimation is inhibition of nucleic acid amplification reactions associated with both methods. Inhibition typically occurs when there are compounds in the source water, often high molecular weight compounds like humic acids and other complex carbohydrates, that combine with metal ions to sequester nucleic acids from polymerases and prevent amplification (De Boer et al. 1995; Kreader 1996). However, the TMA method possesses a target capture step that has been shown to be effective at removing inhibitors prior to the amplification reaction. Despite this, we observed greater underestimation for storm drain samples. This is consistent with inhibition, as PCR inhibiting compounds are likely to occur in storm drains. These drains discharge a complex mixture of inputs from an urbanized landscape, including humic and

fulvic acids, as well as other organic compounds from the breakdown of plant material, and a wide variety of metal ions from automobiles.

The addition of the extra cleanup steps in the QPCR extracted method was intended to reduce inhibition, but the method was neither repeatable nor simple for the laboratory personnel to master. The extracted method had a coefficient of variation between replicates that was three times higher than that for the bead-beaten or for existing methods. Inclusion of the extra steps also had little effect on underestimation, though this may have had more to do with implementation than with the underlying efficacy of the method. The implementation difficulty resulted from the many pipetting steps of the extracted method that introduced opportunities for imprecision, but part of the difficulty with implementation likely stemmed from the manner in which the laboratory attempted to integrate the method into their regular workflow. Their approach was to break the protocol into segments that could be performed by a succession of technicians. This likely introduced error due to varying levels of familiarity with the purification method and opportunities for miscommunication between technicians performing successive steps of the protocol. Perhaps the precision, convenience and suitability of this method for reducing inhibition will improve over time as method developers establish a means for automating steps in the process.

Developers of the QPCR method incorporated two internal controls into their assay that were intended to provide a means to measure and compensate for inhibition or recovery issues. One of these controls, the SSIC, was used in calculating the “adjusted” enterococci values. While the adjustment for controls reduced the false negative error rate, it led to an overall increase in level of error. For instance, the percent of samples in which a materially different decision would have been made compared to existing methods bead-beaten method increased from 27% to 77% after adjustment. It is unclear why the adjustment failed to improve comparability with existing methods. One possibility is that the target primers and probe that comprise the SSIC have a different amplification efficiency than those used to detect enterococci and *E. coli*. For example, the SSIC may be more sensitive to inhibitory compounds than are the enterococci and *E. coli* assays because it is designed to be particularly sensitive toward inhibition. If this is the case, then extrapolating inhibition rates from the SSIC might lead to the type of overestimation that was observed. Regardless of the cause, developers of both methods have indicated that they are placing priority on developing improved internal controls for implementation in future evaluation studies.

It is also possible that freezing of samples, rather than inhibition, led to some of the underestimation. Samples were frozen in both beta testing laboratories because the number of samples collected daily for this effort was too small to cost-effectively quantify them on a daily basis. Freezing would not be part of normal practice if these methods were adopted for routine use. However, it is unlikely that freezing after extraction was a major contributor to the underestimation. Previous studies have shown that genetic material remains intact when frozen at -80°C as were the samples used in this study (Smith 2005). Moreover, we did not see an improvement in classification accuracy for those samples that were processed without freezing.

Previous evaluations of genetically based methods have suggested that overestimation relative to existing methods is a greater concern than underestimation (Noble and Weisberg 2005). Overestimation would result because nucleic acid methods are based on measuring the presence of specific genetic fragments without assessing the presence of live, viable cells. This concern was certainly manifested in the chlorination portion of the study, in which no viable cells were measured by existing methods, but bacterial concentrations, as measured by the genetic methods, were relatively unchanged following chlorination. This indicates that the new methods should not be used for measurement near a chlorinated effluent, however, we did not see evidence that overestimation due to viability was a concern at other locations. Many of our sample sites were located in or near storm drain effluents, which are likely to contain several day old fecal sources originating upstream. Perhaps the lack of overestimation relative to existing methods reflects moderation by inhibition, but an alternative explanation is that genetic material breaks down at rates comparable enough to inactivation of live cells that the ratio of live:dead cells is not large enough to be of concern. Certainly this is an area that requires further investigation.

Though the rapid methods beta tested in this study are not yet ready to adopt for beach water quality monitoring, they are approaching the accuracy of some existing EPA approved methods. Further, our study showed that the technology needed to perform these methods can be successfully transferred to working water quality laboratories and performed by technicians with little or no prior experience with molecular methods. Despite these encouraging findings, there is still work to be done. It is apparent that the internal controls developed for the QPCR method to guard against false negative results are not yet perfected and that given the tendency of TMA toward underestimation, an internal control is needed for this method as well. Finally, labor requirements are an impediment to implementing these methods in their current iterations. The challenge will be for method developers to introduce a level of automation to their methods that minimizes errors and reduces labor costs while still keeping instrumentation affordable.

Literature Cited

Anonymous. 1997. User Bulletin #2. ABI Prism 7700 Sequence Detection System. Foster City, CA, Applied Biosystems.

Boehm, A.B., S.B. Grant, J.H. Kim, S.L. Mowbray, C.D. McGee, C.D. Clark, D.M. Foley and D.E. Wellman. 2002. Decadal and shorter period variability and surf zone water quality at Huntington Beach, California. *Environmental Science and Technology* 36: 3885-3892.

Brinkman, N.E., R.A. Haugland, L.J. Wymer, B. Murulrdhara, R.L. Whitman and S.J. Vesper. 2003. Evaluation of a rapid, quantitative real-time PCR method for enumeration of pathogenic *Candida* cells in water. *Applied and Environmental Microbiology* 69: 1775-1782.

De Boer, S.H., Ward, L.J., Li, X., and S. Chittaranjan. 1995. Attenuation of PCR inhibition in the presence of plant compounds by addition of BLOTTO. *Nucleic Acid Research* 23:2567-2568

Dupray, E., M.P. Caprais, A. Derrien and P. Fach. 1997. Salmonella DNA persistence in natural seawaters using PCR analysis. *Journal of Applied Microbiology* 82: 507-10.

Frahm, E. and U. Obst. 2003. Application of the fluorogenic probe technique (TaqMan PCR) to the detection of *Enterococcus* spp. and *Escherichia coli* in water samples. *Journal of Microbiology, Methods* 52:123-131.

Griffith, J.F., Weisberg, S.B. and C.D. McGee. 2004. Evaluation of new, rapid methods for measuring microbiological water quality. Southern California Coastal Water Research Project Annual Report. 354-362.

Griffith, J.F., L.A. Aumand, I.M. Lee, C.D. McGee, L.L. Othman, K.J. Ritter, K.O. Walker, and S.B. Weisberg. 2006. Comparison and verification of bacterial water quality indicator measurement methods using ambient coastal water samples. *Environmental Monitoring and Assessment* 116:335-44.

Griffith, J.F. and S.B. Weisberg. 2006. Evaluation of Rapid Microbiological Methods for Measuring Recreational Water Quality. Southern California Coastal Water Research Project, Westminster, CA.

Haugland, R.A., S.C. Siefring, L.J. Wymer, K.P. Brenner and A.P. Dufour. 2005. Comparison of *Enterococcus* density measurements by quantitative polymerase chain reaction and membrane filter culture analysis at two freshwater recreational beaches. *Water Research* 39:559-568.

Kreader, C.A. 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl. Envir. Microbiol.* 62: 1102-1106.

Lee, J.Y. and R. A. Deininger. 2004. Detection of *E. coli* in beach water within 1 hour using immunomagnetic separation and ATP bioluminescence. *Luminescence* 19: 31-36.

Leecaster, M.K. and S.B. Weisberg. 2001. Effect of sampling frequency on shoreline microbiology assessments. *Marine Pollution Bulletin* 42: 1150-1154.

Noble, R.T., S.B. Weisberg, M.K. Leecaster, C.D. McGee, K. Ritter, K.O. Walker and P.M. Vainik. 2003a. Comparison of beach bacterial water quality indicator measurement methods. *Environmental Monitoring and Assessment* 81: 301-312.

Noble, R.T., D.F. Moore, M. Leecaster, C.D. McGee and S.B. Weisberg. 2003b. Comparison of total coliform, fecal coliform, and enterococcus bacterial indicator response for ocean recreational water quality testing. *Water Research* 37: 1637-1643.

Noble, R.T., M. Leecaster, C.D. McGee, S.B. Weisberg and K. Ritter. 2004. Comparison of bacterial indicator analysis methods in stormwater-affected coastal waters. *Water Research* 38: 1183-1188.

Noble, R.T. and S.B. Weisberg. 2005. A review of technologies for rapid detection of bacteria. *Journal of Water and Health* 3:381-391

Piersimoni, C., C. Scarparo, P. Piccoli, A. Rigon, G. Ruggiero, D. Nista and S. Bornigia. 2002. Performance assessment of two commercial amplification assays for direct detection of *Mycobacteria tuberculosis* complex from respiratory and extrapulmonary specimens. *Journal of Clinical Microbiology* 40:4138-4142.

Santo Domingo, J.W., S.C. Siefring and R.A. Haugland. 2003. Real-time PCR method to detect *Enterococcus faecalis* in water. *Biotechnology Letters* 25: 261-5.

Smith, S. and P.A. Morin. 2005. Optimal storage conditions for highly dilute DNA samples: A role for Trehalose as a preserving agent. *Journal of Forensic Science*. 50: ISSN: 0022-1198