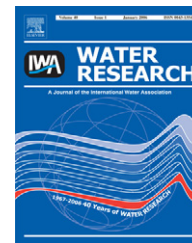


Available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/watres

16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: A Bayesian approach

Beverly J. Kildare^{a,1}, Christian M. Leutenegger^{b,2}, Belinda S. McSwain^{a,3},
Dustin G. Bambic^{c,4}, Veronica B. Rajal^{a,5}, Stefan Wuertz^{a,*}

^aDepartment of Civil and Environmental Engineering, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

^bLucy Whittier Molecular & Diagnostic Core Facility, TaqMan(R) Service, Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

^cLarry Walker Associates, 707 Fourth Street, Davis, CA 95616, USA

ARTICLE INFO

Article history:

Received 20 February 2007

Received in revised form

8 June 2007

Accepted 13 June 2007

Available online 21 June 2007

Keywords:

Quantitative PCR

Microbial source tracking

Bacteroidales

ABSTRACT

We report the design and validation of new TaqMan[®] assays for microbial source tracking based on the amplification of fecal 16S rRNA marker sequences from uncultured cells of the order *Bacteroidales*. The assays were developed for the detection and enumeration of non-point source input of fecal pollution to watersheds. The quantitative “universal” *Bacteroidales* assay BacUni-UCD detected all tested stool samples from human volunteers (18 out of 18), cat (7 out of 7), dog (8 out of 8), seagull (10/10), cow (8/8), horse (8/8), and wastewater effluent (14/14). The human assay BacHum-UCD discriminated fully between human and cow stool samples but did not detect all stool samples from human volunteers (12/18). In addition, there was 12.5% detection of dog stool (1/8), but no cross-reactivity with cat, horse, or seagull fecal samples. In contrast, all wastewater samples were positive for the BacHum-UCD marker, supporting its designation as 100% sensitive for mixed-human source identification. The cow-specific assay BacCow-UCD fully discriminated between cow and human stool samples. There was 38% detection of horse stool (3/8), but no cross-specificity with any of the other animal stool samples tested. The dog assay BacCan-UCD discriminated fully between dog and cow stool or seagull guano samples and detected 62.5% stool samples from dogs (5/8). There was some cross-reactivity with 22.2% detection of human stool (4/18), 14.3% detection of cat stool (1/7), and 28.6% detection of wastewater samples (4/14). After validation using stool samples, single-blind tests were used to further demonstrate the efficacy of the developed markers; all assays were sensitive, reproducible, and accurate in the quantification of mixed fecal sources present in aqueous samples. Finally, the new assays were compared with previously published sequences, which

*Corresponding author. Tel.: +1 530 754 6407; fax: +1 530 752 7872.

E-mail address: swuertz@ucdavis.edu (S. Wuertz).

¹ Current address: Carollo Engineers, 2500 Venture Oaks Way, Suite 320, Sacramento, CA 95833, USA. Tel.: +1 916 565 4888; fax: +1 916 565 4880.

² Current address: IDEXX Laboratories, Molecular Diagnostics, 2825 KOVR Drive, West Sacramento, CA 95605, USA.

³ Current address: Department of Civil, Environmental, and Architectural Engineering, The University of Kansas, 2150 Learned Hall, 1530 W. 15th Street, Lawrence, KS 66045-7609, USA. Tel.: +1 785 864 3731; fax: +1 785 864 5379.

⁴ Current address: AMEC Earth & Environmental 3800 Ezell Road, Suite 100, Nashville, TN 37211, USA. Tel.: +1 615 333 0630x417; fax: +1 615 781 0655.

⁵ Current address: INIQUI (CONICET), Facultad de Ingeniería, Universidad Nacional de Salta, Buenos Aires 177, Salta 4400, Argentina. Tel./fax: +54 387 425 1006.

0043-1354/\$ - see front matter © 2007 Published by Elsevier Ltd.

doi:10.1016/j.watres.2007.06.037

showed the new methodologies to be more specific and sensitive. Using Bayes' Theorem, we calculated the conditional probability that the four assays would correctly identify general and host-specific fecal pollution in a specific watershed in California for which 73 water samples had been analyzed. Such an approach allows for a direct comparison of the efficacy of different MST methods, including those based on library-dependent methodologies. For the universal marker BacUni-UCD, the probability that fecal pollution is present when the marker is detected was 1.00; the probability that host-specific pollution is present was 0.98, 0.84, and 0.89 for the human assay HF160F, the cow assay BacCow-UCD, and the dog assay BacCan-UCD, respectively. The application of these markers should provide meaningful information to assist with efforts to identify and control sources of fecal pollution to impaired watersheds.

© 2007 Published by Elsevier Ltd.

1. Introduction

The water quality of many waterways and coastal waters is deteriorating due to point and non-point sources of fecal contamination caused by human and animal sources (Handler et al., 2006). Microbial source tracking (MST) is an increasingly used approach to determine host-specific contributions of fecal contamination to waterbodies, thus helping resolve these unknown sources. This information can, in turn, guide decisions regarding the appropriate corrective measures for affected waters. One emerging MST method is the detection of host-specific 16S rRNA markers that target the order *Bacteroidales*, which are found exclusively in feces, animal rumen, and other cavities of humans and animals (Paster et al., 1994), often in greater abundance than traditionally used coliform bacteria.

Kreader developed PCR-based assays to amplify genes from three cultivated strains of *Bacteroides* to monitor human fecal pollution in water (Kreader, 1995), and evaluated their persistence in the environment (Kreader, 1998). Others further advanced this approach by identifying host-specific *Bacteroidales* 16S rDNA markers for humans and cows based on DNA sequences representing uncultivated fecal diversity. Initially, researchers relied on length heterogeneity-PCR (LH-PCR) and terminal restriction length polymorphism (T-RFLP) analysis to screen fecal bacterial DNA extracts from environmental waters (Bernhard and Field, 2000a). With information obtained from this study, they identified cluster-specific primer sets that are useful in discriminating human and ruminant feces (Bernhard and Field, 2000a). The system of qualitative assays for presence/absence detection was further tested and confirmed to reliably and specifically detect host-specific markers from feces and polluted water samples (Field et al., 2003). These PCR targets were classified as belonging to the phylum *Bacteroidetes*, of which the cultivated fecal members are in the order *Bacteroidales* and the genera *Bacteroides* and *Prevotella* (Dick et al., 2005a). Early advances by Field and colleagues were followed by (i) a quantitative PCR (qPCR) assay for the detection of general fecal pollution based on a 16S rRNA marker specific for many fecal *Bacteroidales* sequences (Dick and Field, 2004), (ii) further conventional PCR assays for the detection of pig- and horse-specific fecal pollution (Dick et al., 2005a) as well as dog- and elk-specific fecal pollution (Dick et al., 2005b), and (iii) a real-time PCR

assay using SYBR Green 1 for the detection of the previously identified human-specific marker sequence (Seurinck et al., 2005). Recently, additional qPCR assays have been developed for "total" (Layton et al., 2006), bovine-associated (Layton et al., 2006; Okabe et al., 2007; Reischer et al., 2006; Stricker et al., in press), pig-associated (Okabe et al., 2007), and human-associated *Bacteroidales* 16S rRNA genes (Layton et al., 2006; Okabe et al., 2007; Reischer et al., 2007; Stricker et al., in press).

To date, the development of molecular target detection assays for *Bacteroidales* has provided a fast, reliable, and relatively inexpensive means by which to diagnose the source(s) of fecal contamination to natural waters. A comparison study showed *Bacteroidales* to be the most accurate MST method for discriminating human vs. non-human impacts using tests of mixed fecal sources in aqueous samples (Griffith et al., 2003). There has been some discussion concerning the benefits in terms of sensitivity and specificity of TaqMan over SYBR Green assays, and the relative sensitivities of the two technologies have been debated (Morrison et al., 1998; Wittwer et al., 1997). With regard to specificity, TaqMan assays are considered the better choice for differentiation of host-specific DNA sequences when environmental samples are considered. This is because non-specific products and mRNAs with high sequence identity may be detected with the SYBR Green chemistry. Melting curve analysis can assist in proper identification of these false-positive sequences, but accurate quantification may be compromised, and a comparison of numerical data generated by SYBR Green with quantitative results obtained by TaqMan analysis has not been performed for *Bacteroidales*. By combining multiple species-specific TaqMan PCR assays, specificity can be expanded within a genus if new sequence information becomes available.

The ultimate goal of MST is to determine the relative amounts of host-specific fecal contributions to a water sample. In the case of *Bacteroidales* methodologies, the goal may be to directly quantify the relative contributions of specific hosts (e.g. humans) based on a relationship between host-specific markers and the total concentration of *Bacteroidales* DNA sequences detected. These data would provide watershed managers enhanced information with which to formulate more accurate plans to reduce the loading of fecal pollution to receiving waters. Therefore, there is a need for

Bacteroidales methodologies to not only become more quantitative with respect to the amplified target sequences, but also for such data to be analyzed statistically to calculate conditional probabilities of correctly identifying sources of fecal pollution in a watershed, given that a particular host-specific assay tests positive in environmental samples. The specific objectives of the present study were the design of a new, more inclusive, quantitative universal *Bacteroidales* assay, as well as three new quantitative host-specific *Bacteroidales* assays, allowing the relative determination of the amount of human-, dog-, and cow-specific fecal contaminations over space and time. Further, we developed a conditional probability approach to estimate the likelihood of correctly identifying fecal sources of contamination in a water sample given the detection of the new markers.

2. Materials and methods

2.1. Assay design

Real-time TaqMan PCR systems were designed against *Bacteroidales* DNA sequences found in the Genbank Database (Pittsburgh Supercomputing Centers, Pittsburgh, PA). Sequence alignment of several of the submitted *Bacteroidales* isolates revealed a conserved region in which the universal *Bacteroidales* assay (BacUni-UCD) was designed. Alignment of separate DNA sequences for host-specific *Bacteroidales* isolates indicated conserved regions for mixed human-specific (BacHum-UCD), dog-specific (BacCan-UCD), and cow-specific (BacCow-UCD) assays. The cow assay used a previously published conventional PCR forward primer specific to ruminant *Bacteroidales* (Bernhard and Field, 2000b) along with a newly developed reverse primer and probe. Assays were designed using Primer Express software (Applied Biosystems, Foster City, CA). Each TaqMan probe was labelled at the 5' end with the reporter dye 6-FAM (6-carboxy-fluorescein) and at the 3' end with the quencher dye TAMRA (6-carboxytetramethyl-rhodamine). The probe designed for the universal assay has a minor groove binder attached at the 3' end along with the quencher dye.

All oligonucleotide sequences were chosen to allow concurrent amplification of all assays under standard conditions: 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C.

2.2. Sequence comparisons and phylogenetic tree construction

Sequences were aligned and a bootstrap consensus tree was created using the neighbor-joining method (Molecular Evolutionary Genetics Analysis (MEGA) software, version 3.1). The root was determined using *Cytophaga fermentans* 16S rDNA sequence (M58766). No correction was used in the distance calculations.

2.3. Clone construction of marker DNA sequences

Separate PCR products amplified with marker DNA sequences specific to each assay were purified using Qiagen columns

(QIAquick PCR Purification Kit, Qiagen, Valencia, CA) according to the manufacturer's recommendations. Each set of PCR products was cloned into a PCR-TOPO plasmid using the TA cloning strategy (Invitrogen, Carlsbad, CA). The plasmid was transformed into *Escherichia coli* DH5 α and recombinant bacteria were selected on ampicillin-containing LB agar. Real-time TaqMan PCR was used to screen white colonies, and two positive clones were selected for overnight liquid culture propagation. Plasmids were extracted from *E. coli* using Qiagen columns, and inserts were re-sequenced to confirm analytical specificity. Plasmids were quantified by spectrophotometry to generate standard curves in triplicate by real-time TaqMan PCR for each assay.

2.4. Fecal sample collection

Fecal samples were collected with sterile utensils and placed in sterile 50-ml tubes. Cow, horse, dog, and cat samples were collected from several different locations throughout California. Three cow, one horse, and one dog sample originated in a remote mountain region of Ecuador. Human fecal samples were donated by a variety of individuals of varying age, sex, and ethnic backgrounds. Two human fecal samples were also donated by visitors to the United States. The seagull samples tested represent a well-mixed sample of ten different seagull sources. Additionally, one of the ten cow fecal samples, one of the eight horse fecal samples, and one of the eight dog fecal samples each represent a mixture of ten individual fecal samples (of the corresponding host) which were collected separately, but homogenized completely prior to extraction and analysis. Details of samples are shown in Table S1 in Supplementary Data. All other samples were from individual hosts. All samples were transported to the laboratory on ice, where a subsample was mixed with lysis buffer (ASL, QIAmp Stool Kit, Qiagen, Valencia, CA) and immediately stored at –20 °C until extracted.

2.5. Primary wastewater influent sample collection from wastewater treatment plants

Samples of primary influent were collected from various municipal wastewater treatment plants in sterile 250-ml bottles, and transported on ice to the laboratory. Samples were then centrifuged at 4000g for 10 min at 4 °C to pellet the biomass in the sample. The pellet was removed from the bottle with a sterile utensil, and bacterial DNA was extracted immediately.

2.6. Water sample collection

Seventy-three grab samples of water from six sites in Calleguas Creek Watershed (CCW) in Ventura County, California were collected monthly over a period of 12 months in five clean, rinsed, 20-l polypropylene carboys for MST. The water in the five carboys from each site was pooled and processed by hollow fiber ultrafiltration as described previously (Rajal et al., 2007). Final retentates were stored at –20 °C. Furthermore, lab duplicates, field duplicates, and field blanks were submitted on occasion as a quality control measure. Additional water samples were collected at each

site and analyzed for total suspended solids according to Standard Method 2540D and traditional microbiological analyses (total and fecal coliform by SM9221, *E. coli* by SM9223, and *Enterococcus* by SM9320).

2.7. Nucleic acid extraction

DNA from stool and concentrated environmental water samples and from the resultant pellet of the centrifugation of the primary wastewater influent samples was extracted using the QIAamp DNA Stool kit (Qiagen) according to the manufacturer's directions. Final eluted volumes were approximately 200 μ l. Nucleic acid extraction for the blindly submitted water samples for field validation of the four new assays reported herein was performed with a modified extraction process. Specifically, 10 ml of well-mixed liquid sample was added to a 200-ml conical plastic centrifuge bottle containing 40 ml of lysis buffer (Boom et al., 1990), and the solution was pulse vortexed 15 times. After a 10-min incubation period at room temperature, 40 ml of absolute ethanol was added, and again pulse vortexed 15 times. The entire supernatant was added to a QIAamp Maxi Spin column (Qiagen) using a vacuum manifold (Qiagen) under a suction of 800 mbar. The extraction column was washed once with 5 ml buffer AW1 (Qiagen), followed by an additional washing step with 5 ml buffer AW2 (Qiagen). The extraction column was then placed into a sterile 50-ml collection tube, centrifuged at 4000g for 15 min, and then incubated at 70 °C for 10 min to remove traces of AW1. Nucleic acid was eluted twice with 600 μ l of DEPC-treated water at 4000g for 5 min.

2.8. Assay limits of detection and quantification standards

Development of standard curves and determination of the amplification efficiencies for each new *Bacteroidales* assay, as well as for those assays previously developed by others and used in this study, were performed with plasmids containing the specific marker sequence. Assay limits of quantification (A_{LOQ}) were determined during the production of the standard curves for each assay, as the lowest concentration of marker sequence within the linear range of quantification during TaqMan amplification.

Assay limits of detection (A_{LOD}) were determined according to the approach suggested by the US EPA for determining method detection limits (40 CFR 136, Appendix B). A solution of plasmid containing marker sequence that was the approximate concentration of the predetermined A_{LOQ} was prepared in PCR-grade water and quantified in at least seven replicate wells by TaqMan PCR. At the same time, a second solution with a slightly different concentration of plasmid containing marker sequence was prepared and quantified in at least seven replicate wells. The standard deviation of the results obtained from the two similar concentrations was calculated, and an F-test for two-sample variance was performed to determine whether the difference between the variances was statistically insignificant. Concentrations tested that produced statistically insignificant variances were used to calculate the assay A_{LOD} . Sample variances from these tests were pooled and the A_{LOD} was calculated according to

the formula $A_{LOD} = \text{Student's } t \text{ value times the pooled standard deviation}$. It should be noted that the estimated A_{LOD} is defined only for pure water and not for an environmental water matrix.

2.9. Validation and cross-reactivity testing

An approach for normalizing the testing of each new assay was utilized that involved initial screening of each fecal and sewage extract with the BacUni-UCD universal *Bacteroidales* assay. Upon determination of the exact individual dilutions needed for each separate extract that corresponded to a *Bacteroidales* concentration of approximately 500 copies/ μ l (as measured by the universal *Bacteroidales* assay BacUni-UCD), the same dilution was tested with the remaining three host-specific assays for sensitivity and cross-reactivity testing. This approach was used in an effort to establish the true specificity and sensitivity for each assay given a constant number of *Bacteroidales* 16S rRNA gene copies. Reactions were performed in duplicate and appropriate controls for each assay were included in every TaqMan run performed during the validation process.

For all TaqMan assays, 10 μ l of nucleic acid extract was assayed in a final reaction volume of 25 μ l. The samples were placed in 384-well plates and amplified in an automated fluorometer (ABI PRISM 7900 Sequence Detection System, Applied Biosystems). Standard amplification conditions were used: 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Fluorescent signals were collected during the annealing temperature and cycle threshold (Ct) values were calculated using a baseline range from 3 to 15 and adjusted manually if Ct values below 15 were obtained. Each 25- μ l reaction contained a final concentration of 1 \times TaqMan PCR master mix (Eurogentec) with 400 nM of the forward and each reverse primer, and 80 nM of probe. Ct values were evaluated with a threshold that was specific to each assay; 0.07 for universal *Bacteroidales* marker (BacUni-UCD) detection, 0.15 for mixed human-specific *Bacteroidales* marker (BacHum-UCD) detection, and 0.11 for both cow-specific *Bacteroidales* marker (BacCow-UCD) detection and for dog-specific *Bacteroidales* marker (BacCan-UCD) detection.

2.10. Field validation of assays using single-blind test

A field validation study was performed to further verify the developed MST assays, and demonstrate the accuracy and reliability of utilized host-specific molecular markers. Specifically, assays were applied to blindly submitted aqueous test samples with known amounts of fecal material to test for the following parameters: reproducibility of results, relative ability to accurately quantify target marker sequences, and host specificity in mixed samples. Ten aqueous test samples were blindly submitted to UC Davis by staff from Larry Walker Associates, Inc. (Lafayette, CA). Aqueous test solutions containing several host types were as follows:

Mixed-human solution: One liter of screened primary influent was collected in a clean glass bottle and stored in a dark refrigerator until preparation of aqueous test samples.

Cow solution: Fecal samples were collected from a cattle herd grazing near Simi Valley, CA. The herd consisted of 11

cows and one calf. A total of 10 samples were collected from individual cow “pies”. A composite sample was then prepared by placing roughly equal amounts of individual fecal samples into a clean Ziploc bag and thoroughly mixing. Then 1.043 g of the mixed material was diluted into 1 l of deionized water in a clean HDPE bottle, and stored for less than 1 h prior to creating aqueous test samples.

Dog solution: Fecal samples were collected from a kennel near Camarillo, CA. A total of six samples were collected, each from a different canine host. All samples had been deposited within the previous 12 h. A composite sample was then prepared by placing roughly equal amounts of individual fecal samples into a clean Ziploc bag and thoroughly mixing. Then 1.130 g of the mixed material was diluted into 1 l of deionized water in a clean HDPE bottle, and stored for less than 1 h prior to creating aqueous test samples.

Horse solution: Fecal samples were collected from a horse stable near Moorpark, CA. A total of 10 samples were collected, each from a different horse host. Samples had been deposited within the previous 4 h. A composite sample was then prepared by placing roughly equal amounts of individual fecal samples into a clean Ziploc bag and thoroughly mixing. Then 1.249 g of the mixed material was diluted into 1 l of deionized water in a clean HDPE bottle, and stored for less than 1 h prior to creating aqueous test samples.

Seagull solution: Fecal samples were collected from a flock of seagulls perched in Arroyo Simi near Moorpark, CA. A total of seven fresh samples were collected, combined with roughly equal amounts from each host in a clean Ziploc bag, and thoroughly mixed. Then 0.901 g of the mixed material was diluted into 1 l of deionized water in a clean HDPE bottle, and stored for less than 1 h prior to creating aqueous test samples.

All *host-specific solutions:* Mixed aqueous test samples were prepared and submitted blindly as filled 1-l HDPE bottles labeled with numbers from 1 to 10 (see Table 5 for details). The aqueous test samples were designed to reflect one and two orders of magnitude differences in host-specific fecal concentrations. After samples were given to UC Davis (without knowledge of their contents), each water sample was stored on ice during transport to the laboratory. Nucleic acid extraction and TaqMan analysis were performed as described in Section 2.6. Enumeration of marker sequences detected in each sample was performed for the lowest dilution that was found to be out of the range of inhibition as described previously (Rajal et al., 2007).

2.11. Method comparison protocol

Upon appropriate validation of each new assay, an experiment was performed to compare the new universal *Bacteroidales* (BacUni-UCD) assay with two previously developed total *Bacteroidales* assays (Dick and Field, 2004; Layton et al., 2006) used to enumerate fecal contamination. Likewise, a comparison test was performed between the new human-specific (BacHum-UCD) assay, the previously developed human-specific (HF183) SYBR Green assay (Seurinck et al., 2005), and the TaqMan-based human-specific assay (HuBac) (Layton et al., 2006). The new cow-specific (BacCow-UCD) assay was compared with the BoBac assay (Layton et al., 2006). Each of the assays was tested against all fecal and wastewater influent

extracts at a concentration of 0.5 pg total nucleic acid per reaction as measured by fluorometry, in an effort to normalize the approach in a non-biased format. Reactions were performed in duplicate and appropriate controls for each assay were included for every TaqMan run performed during the comparison process.

Each sample was analyzed with the TaqMan assays at the same time in an effort to produce comparable results. This was not possible with the SYBR Green assay due to variable cycling conditions. As a result, samples analyzed by TaqMan were stored at 4 °C for no more than 3 h prior to analysis by the SYBR Green assay. It was assumed that the original marker concentration was unaffected by this waiting period.

TaqMan protocols for all four new assays, as well as the previously published total *Bacteroidales* assays, were followed as shown above. Ct values were evaluated with a threshold of 0.16 for the total *Bacteroidales* assay. Human *Bacteroidales* marker sequences (HF183) based on the SYBR Green assays were detected according to the procedure outlined by Seurinck et al. (2005). Briefly, each 25- μ l reaction contained 1 \times qPCR Mastermix Plus for SYBR Green 1 (Eurogentec) with an optimized concentration of 0.1 μ M for both forward and reverse primers. The cycle times were also adjusted to 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles at 95 °C for 15 s, 53 °C for 45 s, and 60 °C for 1 min. Ct values were evaluated with a threshold set to 0.18. Dissociation temperatures for each tested well were analyzed. Those wells that showed amplification and had a dissociation temperature in the range of 76–79 °C were considered positive for the human-specific HF183 *Bacteroides* 16S rRNA genetic marker.

2.12. Calculation of host-specific marker concentrations in fecal and water samples

Amplification efficiency and linearity of amplification was tested using six 10-fold serial dilutions of known amounts of the plasmid containing the cloned target sequence. A PCR reaction that amplifies the target sequence with 100% efficiency (E) would, in theory, double the amount of PCR products with each cycle. The amount of PCR products (C_n) from C_0 input target molecules after n cycles could be calculated according to

$$C_n = C_0(1 + E)^n \quad (1)$$

Amplification efficiencies were calculated according to the formula

$$s = -\frac{1}{\log(1 + E)}, \quad (2)$$

where s is the slope of the standard curve; therefore,

$$E = 10^{1/-s} - 1. \quad (3)$$

2.13. Conditional probability analysis using Bayes Theorem

Bayes' Theorem was used to calculate the probability that any detection of the universal, human, cow, or dog genetic marker was the result of a true positive. To apply Bayes' Theorem, assays were assumed to be independent discrete random

variables. For the determination of sensitivity, such an assumption is appropriate: either the marker was present (true) or absent (false). To demonstrate our approach, we calculated the background probability of detecting the universal or one of the three host-specific *Bacteroidales* markers in CCW in southern California, based on the analysis of 73 environmental water samples. As an example, Eq. (4) illustrates the use of Bayes' Theorem to calculate the conditional probability of the human-specific assay to detect the marker sequence BacHum-UCD, originating from humans in a water sample (true positive), and not fecal *Bacteroidales* sequences originating from another known source (false positive). In other words, Eq. (4) estimates $P(H|T)$, the probability of the event that there is a human source of contamination (H) in an analyzed water sample given the event that the test signals positive (T) with the human-specific BacHum-UCD assay:

$$P(H|T) = \frac{P(T|H)P(H)}{P(T|H)P(H) + P(T|H')P(H')}, \quad (4)$$

where $P(T|H)$ is the probability of a positive signal with the human-specific assay in a fecal sample that is human derived; $P(T|H')$ is the probability of positive signal with the human-specific assay in a fecal sample that is not human derived; and $P(H)$ is the background probability of detecting the BacHum-UCD marker in a specific watershed. An a priori assumption is that the background probability can be established by environmental sampling of host-specific *Bacteroidales* sequences in a watershed. For CCW, this value was set to 0.89 since the marker was detected in 65 of 73 samples. The equivalent case in a medical disease or disorder would be the background probability of that specific condition in the human population of, say, the United States.

$P(H')$ is the background probability that the BacHum-UCD marker is absent in CCW. This value is $1 - P(H)$ or 0.11. Additional background information on the application of Bayes' Theorem to MST based on quantification of gene copy numbers is available online as Supplementary data.

2.14. Statistical analysis

Linear regression analysis, t-tests, and analysis of variance (ANOVA) tests were performed using SigmaPlot 8.0 and SigmaStat 3.5 software (SPSS Inc.)

3. Results

3.1. Oligonucleotide sequences

The primers and internal probes for the newly designed TaqMan PCR assays are listed in Table 1. A universal *Bacteroidales* assay (BacUni-UCD) was designed to incorporate all *Bacteroidales* sequences from known fecal hosts that were available from public databases (NCBI and EMBL Nucleotide Sequence Database) in 2005. BacUni-520f overlaps qBac560F (Okabe et al., 2007) by 13 bases and was designed independently prior to publication of qBac560F. BacUni-690r1 is a perfect match for Bac708R, a previously published universal *Bacteroidales* primer (Bernhard and Field, 2000a), but BacUni-

690r1 has seven additional bases. BacUni-690r1 was also independently designed but is identical to qBac725R (Okabe et al., 2007). BacUni-520f has 100% sequence overlap and sequence identity with 44 environmental marine metagenome sequences that have been placed in GenBank. Neither of the two reverse primers, BacUni-690r1 and BacUni-690r2, has 100% overlap and sequence identity with any environmental sequence. It is not known how many of the metagenome sequences represent free-living non-fecal bacteria, but the fact that none of the reverse primers amplifies any of the environmental sequences suggests that the universal assay can be applied in coastal waters for the detection of fecal *Bacteroidales*.

Quantitative TaqMan assays were developed to target human-specific (BacHum-UCD), cow-specific (BacCow-UCD), and dog-specific (BacCan-UCD) *Bacteroidales* DNA sequences. BacHum-160F has a 16-base perfect overlap with Bac183F, a previously published human-specific primer (Bernhard and Field, 2000b). BacHum-241r matches a previously published, human-specific QPCR reverse primer (Seurinck et al., 2005) but has three additional bases at the 5' end. The cow assay developed for this study, called BacCow-UCD, was performed with a forward primer previously published as CF128F (Bernhard and Field, 2000b) and a newly designed reverse primer and internal probe.

3.2. Phylogenetic tree

Bacteroidales 16S rRNA sequences from GeneBank were used for sequence alignments and phylogenetic analysis (Fig. 1). Most of the sequences had greater than 90% sequence identity. Target sequences of bovine origin clustered into one well-separated group and human sequences clustered into two groups separated by gull sequences. The dog sequences also formed a distinct group deeply branching from sequences of cultivated *Bacteroides* species isolated from humans (X83952 and M58762) (Fig. 1). Only three gull sequences are included in the phylogenetic tree for reasons of clarity of presentation; additional sequences did not overlap with target groups of sequences. Given the relative rarity of gull 16S rRNA sequences in GenBank, this group could be underrepresented, which would increase the biological sequence variability. In addition, some of the outlying sequences could represent pass-through isolates obtained from other species and therefore appear misplaced in the phylogenetic tree. Similar observations were made with pig and elk sequences. For these species, significantly more sequences need to be generated in order to resolve their placement within a phylogenetic tree. The relative lack of specific sequence information is a bias for the design of highly specific TaqMan PCR systems and further sequences will have to be incorporated to develop additional assays.

3.3. Limits of detection and quantification for TaqMan PCR systems

The standard curve equations for each new assay (Fig. 2) and amplification efficiencies in pure water, corresponding assay limits of detection (A_{LOD}), and assay limits of quantification (A_{LOQ}) are listed in Table 2. Amplification efficiencies were

Table 1 – Real-time PCR assays used to detect *Bacteroidales* 16S rRNA genes including the primers and probe used for each assay and the expected size of the amplified product

Assay ^a	Oligonucleotide sequence (5'–3')	Size of product (bp)	Annealing T (°C)	References
Universal <i>Bacteroidales</i>				
(BacUni-UCD)		170	60	
BacUni-520f	CGTTATCCGGATTTATTGGGTTTA			This study Okabe et al. (2007); This study
BacUni-690r1 ^b	CAATCGGAGTTCTTCGTGATATCTA			
BacUni-690r2	AATCGGAGTTCCTCGTGATATCTA			This study
BacUni-656p	6-FAM-TGGTGTAGCGGTGAAA-TAMRA-MGB			This study
Human <i>Bacteroidales</i>				
(BacHum-UCD)		81	60	
BacHum-160f	TGAGTTCACATGTCCGCATGA			This study This study
BacHum-241r	CGTTACCCCGCCTACTATCTAATG			
BacHum-193p	6-FAM-TCCGGTAGACGATGGGGATGCGTT-TAMRA			This study
Cow <i>Bacteroidales</i>				
(BacCow-UCD)		177	60	
BacCow-CF128 ^f	CCAACYTTCCCGWTACTC			Bernhard and Field, (2000a, b) This study
BacCow-305r	GGACCGTGTCTCAGTTCAGTG			
BacCow-257p	6-FAM-TAGGGGTTCTGAGAGGAAGTCCCCC-TAMRA			This study
Dog <i>Bacteroidales</i>				
(BacCan-UCD)		145	60	
BacCan-545f1	GGAGCGCAGACGGGTTTT			This study This study
BacUni-690r1 ^b	CAATCGGAGTTCTTCGTGATATCTA			
BacUni-690r2	AATCGGAGTTCCTCGTGATATCTA			This study
BacUni-656p	6-FAM-TGGTGTAGCGGTGAAA-TAMRA-MGB			This study

^a Numbers within the primer/probe name indicate the nucleotide position within the *Bacteroidales* 16S rRNA gene.

^b An identical primer was independently designed by Okabe et al. (2007) and named qBac725R.

^c Previously published primer (Bernhard and Field, 2000b).

excellent and ranged from 91% to 100%. Additionally, a plasmid standard was constructed for one of the previously developed total *Bacteroidales* TaqMan assays used in this study (Dick and Field, 2004). A plasmid standard for the SYBR Green human-specific assay was supplied by the original authors (Seurinck et al., 2005). Both standards performed according to the original publications. Assay limits of detection were comparable among all tested assays.

3.4. Assay validation

Specificity and sensitivity were initially estimated *in silico* by Genescan Blast searches at the NCBI (Altschul et al., 1990) for each new TaqMan PCR assay. Each assay exclusively targeted *Bacteroidales* 16S rDNA sequences and mostly species-specific *Bacteroidales* 16S rDNA sequences originating from fecal samples of the respective targeted hosts. Specificity was further tested with bacterial DNA from cultures of other fecal and non-fecal bacteria including *E. coli* (ATCC 15997), *Pseudomonas aeruginosa* (ATCC 15692), *Delftia acidovorans* (ATCC 15668), and *Acinetobacter baylyi* ADP1 (Vanechoutte et al., 2006), as well as *Bacteroides fragilis* (ATCC 25285) and *Bacter-*

oides galacturonicus (ATCC 43244). Each assay was tested using 0.5 pg total DNA per reaction, as measured by fluorometry. No DNA from cultures tested outside of the *Bacteroides* genus was amplified with any of the four new *Bacteroidales* assays. DNA from *Bacteroides galacturonicus* was amplified by only the universal *Bacteroidales* assay, while *B. fragilis* DNA was amplified by both the universal (BacUni-UCD) and the human-specific (BacHum-UCD) assay.

Host-specific assays were further tested by amplification of DNA extracted from target and non-target host fecal samples, and primary influent from several wastewater treatment plants (Table 3). This experimental approach established sensitivity and specificity of the new assays. All samples tested produced positive and quantifiable results for the universal BacUni-UCD assay. The results for the BacHum-UCD assay (which was designed to target human-specific fecal contamination) require closer analysis. The assay amplified 67% of the 18 human fecal samples tested, and 100% of the 14 raw sewage samples tested. This latter finding is important in an environmental monitoring context because human fecal contamination to rivers and streams is more likely the cause of mixed-human fecal input (as with sewage

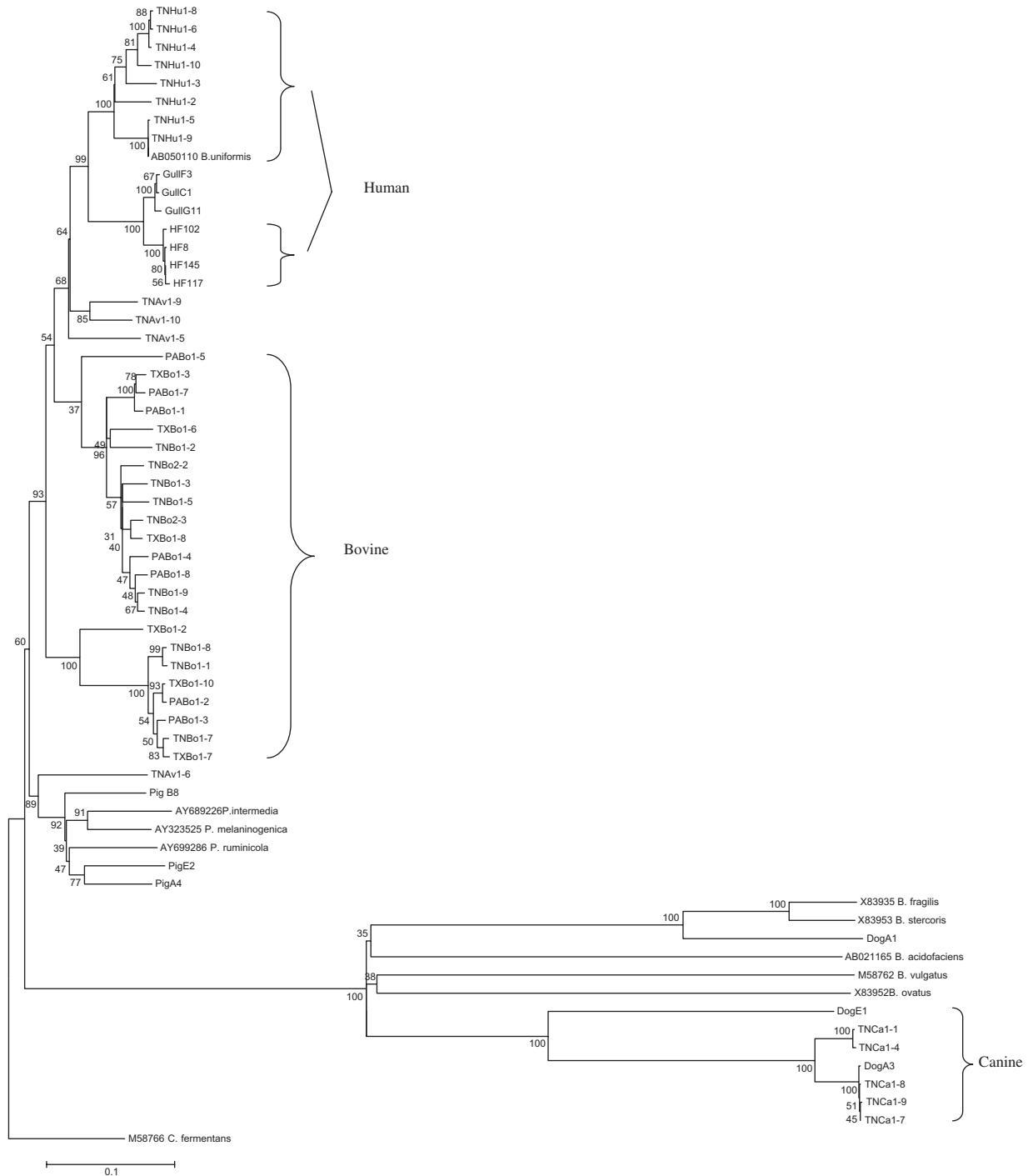


Fig. 1 – Phylogenetic relationships among 16S rRNA sequences from human (TNHu and HF), cow (TNBo, PABo, and TXBo), avian (gull (Gull) and chicken (TNAv)), pig (Pig), and dog (TNCa and Dog) fecal sources. Sequence designations were adopted from GenBank sequences submitted in previous publications (Layton et al., 2006; Bernhard and Field, 2000a, b). Sequences were aligned and a bootstrap consensus tree was created using the neighbor-joining method (Molecular Evolutionary Genetics Analysis (MEGA) software, version 3.1). The root was determined using *Cytophaga fermentans* 16S rRNA sequence (M58766). No correction was used in the distance calculations. References for cultured and uncultured *Bacteroides* sequences included in the tree were *Bacteroides fragilis*, *B. ovatus*, *B. stercoris*, *B. uniformis*, *B. acidofaciens*, and *B. vulgatus*. Brackets indicate targets of specific real-time PCR assays.

overflows) than of individual human input. Therefore, to be consistent with its validation, the BacHum-UCD marker is categorized as a “mixed-human” marker. It should be noted

that the assay also amplified a small number (13%, or 1 of 8) of canine samples tested. As discussed below, other researchers have reported difficulty differentiating among human and

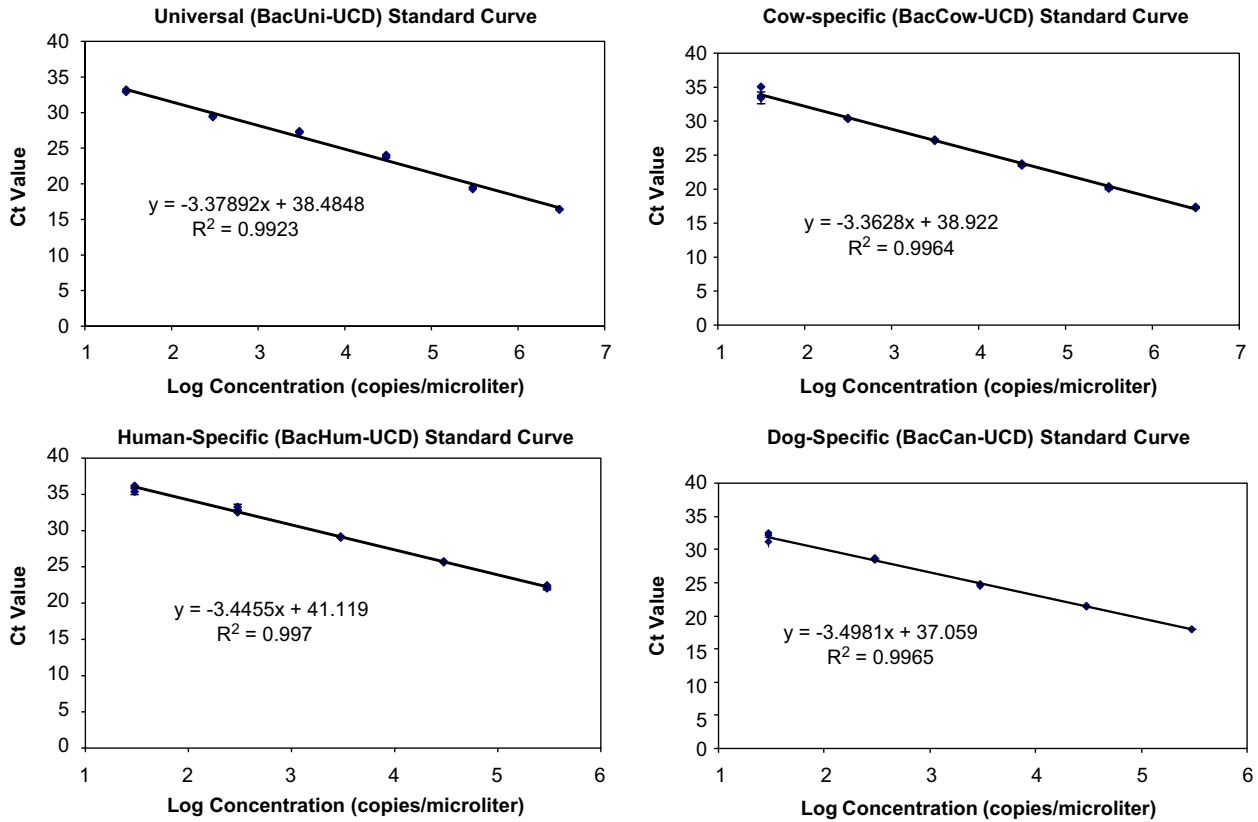


Fig. 2 – Standard curves of measured Ct values ±SD versus log plasmid marker concentration for four newly developed *Bacteroidales* assays. SD error bars were small and are contained in the symbols.

Table 2 – New *Bacteroidales* assay limits of detection and quantification and comparison with two published assays

<i>Bacteroidales</i> assay	Standard curve	Amplification efficiency (%)	A _{LOD} (gc ^a /reaction)	A _{LOQ} (gc/reaction)
Universal (BacUni-UCD)	Y = -3.38X+38.46	98	2	30
Human-specific (BacHum-UCD)	Y = -3.45X+41.12	95	3.5	30
Cow-specific (BacCow-UCD)	Y = -3.36X+38.91	98	3.7	31
Dog-specific (BacCan-UCD)	Y = -3.49X+37.06	93	1	30
Total <i>Bacteroidales</i> ^c	Y = -3.56X+39.82	91	ND ^b	3
Human-specific (HF183) ^d	Y = -3.29X+33.96	100	0.8	2.7

^a Number of gene copies.
^b Not determined.
^c Previously published assay by Dick and Field (2004).
^d Previously published assay by Seurinck et al. (2005) based on SYBR Green chemistry.

canine hosts. Additional host species related to humans including pigs were not tested.

The cow-specific *Bacteroidales* assay BacCow-UCD amplified 100% of the cow fecal extracts tested, and 38% (3/8) of the horse fecal extracts tested, but none of the other extracted fecal DNA that originated from dog, cat, seagull, raw wastewater, and human feces. Due to reactivity with horses, the BacCow-UCD assay could be referred to as a “cow/horse” marker. For many studies, the differentiation among horses and cows may not be important. Thus, the efficacy

of this assay is very promising, although further testing on fecal samples from wildlife animals should be performed to determine if the assay could be used to differentiate fecal pollution originating from livestock from that originating from wildlife animals (e.g., deer). These results confirm a differentiation between human and non-human input, and even between some livestock and domestic animal inputs.

Of the newly developed assays, the dog-specific BacCan-UCD assay was less sensitive, amplifying 63% of the canine

Table 3 – Comparison of sensitivity and specificity of newly developed *Bacteroidales* assays with that of previously published assays^a

Source	Percentage of positive stool or wastewater samples with listed assay (no of samples positive/no of samples tested) targeting								
	All tested animal hosts of fecal <i>Bacteroidales</i> (%)			Human-specific fecal <i>Bacteroidales</i> (%)			Cow- or bovine-specific fecal <i>Bacteroidales</i> (%)		Dog-specific <i>Bacteroidales</i> (%)
	BacUni-UCD ^b	Total <i>Bacteroidales</i> ^c	AllBac ^d	BacHum-UCD ^b	HuBac ^d	HF183 ^e	BacCow-UCD ^b	BoBac ^d	BacCan-UCD ^b
<i>Fecal material</i>									
Human	100 (18/18)	94.4 (17/18)	100 (18/18)	66.7 (12/18)	88.9 (16/18)	61.1 (11/18)	0.00 (0/18)	11.1 (2/18)	22.2 (4/18)
Cow	100 (8/8)	87.5 (7/8)	100 (8/8)	0.00 (0/8)	37.5 (3/8)	0.00 (0/8)	100 (8/8)	100 (8/8)	0.00 (0/8)
Horse	100 (8/8)	100 (8/8)	100 (8/8)	0.00 (0/8)	12.5 (1/8)	0.00 (0/8)	37.5 (3/8)	0.00 (0/8)	0.00 (0/8)
Dog	100 (8/8)	87.5 (7/8)	100 (8/8)	13.0 (1/8)	87.5 (7/8)	25.0 (2/8)	0.00 (0/8)	0.00 (0/8)	62.5 (5/8)
Cat	100 (7/7)	100 (7/7)	100 (7/7)	0.00 (0/7)	71.4 (5/7)	14.3 (1/7)	0.00 (0/7)	0.00 (0/7)	14.3 (1/7)
Seagull	100 (10/10)	100 (10/10)	100 (10/10)	0.00 (0/10)	0.00 (0/10)	0.00 (0/10)	0.00 (0/10)	0.00 (0/10)	0.00 (0/10)
WWTP ^f influent	100 (14/14)	100 (14/14)	100 (14/14)	100 (14/14)	100 (14/14)	100 (14/14)	0.00 (0/14)	0.00 (0/14)	28.6 (4/14)

^a Using 0.5 pg of stool DNA extract.

^b Developed in this study.

^c Published by Dick and Field (2004).

^d Published by Layton et al. (2006).

^e Published by Seurinck et al. (2005).

^f WWTP, wastewater treatment plant.

fecal samples, and it was also less specific because 22%, 33%, and 14% of the human, raw sewage, and cat fecal DNA samples were amplified, respectively. However, note that none of the samples from cow or seagull were amplified, allowing for differentiation among livestock and wildlife sources, though additional wildlife samples need to be tested to further support this conclusion.

3.5. Using Bayes' Theorem to estimate the probability of detecting human fecal contamination

It is unlikely that any designed PCR assay can be absolutely host specific and differentiate between all possible hosts of fecal *Bacteroidales*. For this reason it is necessary to estimate the probabilities of detecting feces originating from host-specific sources. We used Bayesian statistics to compute the conditional probabilities associated with the four new assays. For example, the probability of a human source of contamination (H) in an analyzed water sample given a positive test result (T) with the human-specific BacHum-UCD assay is $P(H|T)$. It can be estimated using Eq. (4) (see Section 2.13) if the background probability, $P(H)$, of detecting the BacHum-UCD marker in a certain watershed is known. For CCW, this value was set to 0.89 since the marker was detected in 65 of 73 environmental water samples.

$P(H)$ is the background probability that the BacHum-UCD marker is absent in CCW. This value is $1 - P(H)$, or 0.11. $P(T|H)$ is the probability of a positive signal with the "mixed" human-specific assay in a fecal sample that is human derived. This value was obtained from the laboratory validation study as 1.00 due to the 100% detection of wastewater

samples screened with this assay. We refer to the human assay as "mixed" only when we use the wastewater validation data to determine $P(T|H)$. The value could also have been set to 0.67 because this was the proportion of individual human stool samples that tested positive with the human assay (Table 3). In that case the assay is simply referred to as human assay (see Supplementary data). $P(T|H')$ is the probability of a positive signal with the (mixed) human-specific assay in a fecal sample that is not human derived. This value was obtained from the laboratory validation study (Table 3) as 0.13 due to the 13% detection of dog-derived fecal sources by this assay.

In the case of the mixed-human marker, Eq. (4) becomes

$$P(H|T) = \frac{(1.00)(0.89)}{(1.00)(0.89) + (0.13)(0.11)} = 0.98.$$

In other words, based on water and fecal samples analyzed with the methods employed during this study, there is a 98% probability that a detection of the BacHum-UCD marker in a water sample from CCW in California is due to mixed human contamination and not fecal *Bacteroidales* sequences originating from dogs.

The diagnostic sensitivity is the ratio of the number of samples that correctly tested positive to all those samples that actually experienced fecal contamination of mixed human origin (see Supplementary data for a more detailed explanation).

$$\text{Sensitivity} = \frac{\text{TPC}}{(\text{TPC} + \text{TNI})} = \frac{0.89}{(0.89 + 0)} = 1.00.$$

Diagnostic specificity is the ratio of the number of samples that correctly tested negative to the total number of samples that actually did not experience fecal contamination of

Table 4 – Conditional probability, sensitivity, specificity, positive and negative predictive value, and prevailing rate for the four 16S rRNA-based Taqman assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales in a Californian watershed^a

Assay	Conditional probability ^b	Sensitivity	Specificity	Predictive value		Prevailing rate ^c
				Positive	Negative	
Universal	1.00	1.00	1.00	1.00	1.00	1.00
Mixed human	0.98	1.00	0.87	0.98	1.00	0.89
Cow	0.84	1.00	0.62	0.84	1.00	0.66
Dog	0.89	0.63	0.57	0.89	0.23	0.84

^a Seventy-three water samples from Calleguas Creek Watershed (CCW) were analyzed with each developed assay.

^b Probability analysis based on Bayes' Theorem. The conditional probability is identical to the positive predictive value.

^c Proportion of the total number of water samples taken from CCW that were positive for the genetic marker. It is numerically identical to the background probability of a specific marker.

mixed-human origin.

$$\text{Specificity} = \frac{\text{TNC}}{(\text{TNC} + \text{TPI})} = \frac{0.0957}{(0.0957 + 0.0143)} = 0.87.$$

The positive predictive value of the test is the ratio of the number of samples that correctly tested positive to the total number of samples that tested positive. It is numerically identical to the conditional probability, $P(H|T)$.

$$\begin{aligned} \text{Positive predictive value} &= \frac{\text{TPC}}{(\text{TPC} + \text{TPI})} \\ &= \frac{0.89}{(0.89 + 0.0143)} = 0.98. \end{aligned}$$

The negative predictive value of the test is the ratio of the number of samples that correctly tested negative to the total number of samples that tested negative.

$$\begin{aligned} \text{Negative predictive value} &= \frac{\text{TNC}}{(\text{TNC} + \text{TNI})} \\ &= \frac{0.0957}{(0.0957 + 0)} = 1.00. \end{aligned}$$

The prevailing rate is the proportion of the total number of samples that actually experienced mixed human fecal contamination.

$$\begin{aligned} \text{Prevailing rate} &= \frac{(\text{TPC} + \text{TNI})}{(\text{TPC} + \text{TNI} + \text{TPI} + \text{TNC})} \\ &= \frac{(0.89 + 0)}{(0.89 + 0 + 0.0143 + 0.0957)} = 0.89. \end{aligned}$$

Note that if $P(T|H)$ is set to 0.67 instead of 1.00 (to correspond to testing of individual, as opposed to mixed, human fecal sample testing), $P(H|T)$ decreases only slightly from 0.9841 to 0.9766. On the other hand, the negative predictive value decreases from 1.00 to 0.25, a significant change. The prevailing rate or background probability remains at 0.89 (see Supplementary data). In a wastewater effluent sample, there are fecal sequences originating from anywhere between a few 100 and several 100,000 person equivalents. When the task of the analysis is to trace treated or untreated effluent originating as a point source it seems appropriate to use $P(T|H) = 1$. When the source of effluent is a septic tank or renegade trailer it would be more prudent to

use $P(T|H) = 0.67$. It should be noted that the latter probability is also expected to change as more human stool samples are incorporated into the database.

The same approach was applied to determine the probabilities associated with the other three newly developed assays (Table 4). The universal BacUni-UCD assay produced no false positives or false negatives during the validation process (Table 3). Therefore, the estimated probability of a positive signal in a water sample being due to fecal input when a positive signal is produced by the universal assay BacUni-UCD is 100%. For cows and dogs, there is an 84% and 89% probability that a detection of the BacCow-UCD and BacCan-UCD marker, respectively, in a water sample is due to actual contamination from the respective hosts (Table 4). Note that the conditional probability of the BacCow-UCD marker would be unity if cows and horses were grouped together as "livestock".

3.6. Comparison of new markers with published Bacteroidales host-specific markers

The new Bacteroidales assays reported herein were directly compared with other quantitative Bacteroidales assays using 73 fecal samples (Table 3). We compared the new universal BacUni-UCD assay with two previously published total Bacteroidales assays (Dick and Field, 2004; Layton et al., 2006). The new human-specific assay (BacHum-UCD) was compared with the published human-specific (HF183F) assay (Seurinck et al., 2005) and the HuBac566F assay (Layton et al., 2006); the new cow-specific assay (BacCow-UCD) was compared with the BoBac367F assay (Layton et al., 2006).

The new universal Bacteroidales assay (BacUni-UCD) amplified 100% of all samples tested, while the total Bacteroidales (Dick and Field, 2004) assay failed to amplify one of the 18 human samples, one of the eight dog fecal samples, and one of the eight cow fecal samples. The Allbac assay for total Bacteroidales (Layton et al., 2006) performed similar to the BacUni-UCD assay with 100% amplification of fecal samples.

As described in the previous section, the new human specific (BacHum-UCD) assay amplified 67% of the human

fecal samples and 13% of the dog fecal samples tested. The previously published human-specific HuBac566F assay amplified 90% of the human fecal samples, but also amplified 38% of cow, 13% of horse, 88% of dog, and 71% of cat fecal samples, respectively. The previously published human-specific HF183F assay amplified 61% of the human fecal samples, but also amplified 25% and 14% of the dog and cat fecal samples, respectively. It is unlikely that we did not use this assay optimally in our study because (i) amplification efficiencies were as reported in the literature and (ii) standard curves were obtained using the identical plasmid, kindly supplied by the original authors, *Seurinck et al. (2005)*. We designed the TaqMan probe for the BacHum-UCD assay to distinguish between sequences that had been amplified by the specific primers (i.e., we ran the probe against those amplified sequences and not all the sequences in the database). This enabled us to increase the ability to screen out non-target host-specific sequences such as those originating from cats.

Based on the limited number of fecal stool samples tested, the new human-specific assay (BacHum-UCD) had the highest sensitivity and specificity of the available human-specific assays. Only 10 of 18 human fecal samples were amplified by all three human-specific assays, while 2 of the 18 human fecal samples did not amplify with any human-specific assay used. One of the human fecal samples tested amplified with the HF183F assay, but not the BacHum-UCD assay, and 2 separate human fecal samples amplified with the BacHum-UCD assay but not the HF183F assay.

The concentration of a particular DNA marker sequence in a specific host was variable between hosts when detected. These results suggest that arrays of markers may be necessary to reliably detect inputs from certain individuals, but as discussed in Section 3.4, it is likely that sources are mixed (as opposed to individual humans) and thus one marker may be sufficient.

During the analysis of the comparison test results, it became apparent that the new universal (BacUni-UCD) and mixed human-specific (BacHum-UCD) *Bacteroidales* assays resulted in higher copy number detections in most fecal and wastewater influent samples tested when compared with the previously published total and human-specific (HF183F) *Bacteroidales* assays (results not shown). To determine if this difference was significant, a pairwise t-test was performed for each comparison dataset following log transformation (data were log-normally distributed) of the detected copy numbers for all samples tested by each assay (73 samples total). In comparing the sequence copy numbers detected in all of the samples tested by the universal (BacUni-UCD) with the total *Bacteroidales* assay, it was found that there was a significantly higher number ($P \leq 0.001$) detected by the universal (BacUni-UCD) assay. Similarly, based on fecal and wastewater samples, the newly developed human BacHum-UCD assay detected significantly higher copy numbers than the human HF183 assay ($P \leq 0.001$). The same was not true when BacUni-UCD and BacHum-UCD assays were compared with the HuBac566F and Allbac assays (*Layton et al., 2006*), respectively; the latter assays had lower Ct values (i.e., higher total gene copies detected) but they were not as specific as the newly developed assays as explained above.

The efficacy of the bovine markers was also evaluated. BoBac367F amplified 100% of cow stool samples, 25% of human stool samples (2/8), and no other tested fecal samples, compared with 100% of cow stool samples, 0% of human stool samples, and 38% of horse stool samples amplified by the new cow-specific BacCow-UCD assay. This means that the previously published BoBac367F assay was somewhat more specific than the BacCow-UCD assay, but it did not fully discriminate between cow and human fecal pollution (as did the BacCow-UCD marker). The only cross-reactivity for the newly developed cow assay was with horse samples, and the discrimination of horse and cow inputs may not be important for some watershed studies (e.g., those in highly urbanized areas). The number of copy numbers of the relevant specific marker determined in fecal samples was comparable for the BoBac367F assay and the BacCow-UCD assay (results not shown). Finally, note that there were no other published dog assays available, so it was not possible to compare the newly developed dog (BacCan-UCD) assay with other markers.

3.7. Single-blind field validation

As reported in Section 3.6, the initial screening of the assays suggested some cross-sensitivity for the human, cow, and especially for the dog-specific BacCan-UCD assay. In an effort to further test the utility of our new assays in environmental applications, a validation study was conducted on blindly submitted, mixed-aqueous solutions of differing host-specific fecal inputs. The solutions were tested as described in Section 2.10, and measured copy number concentrations for each marker sequence were determined (Table 3). Aqueous samples were created to reflect one-to-two order of magnitude differences in fecal material. Measured *Bacteroidales* concentrations were compared, by a second party, with those of the actual relative fecal input compositions of each sample tested. As a major outcome of this test, measured concentrations were shown to accurately reflect the relative abundance of host-specific material in each sample, while correctly identifying the source of fecal material in each blind sample, and there were no cases of false positives or false negatives, as shown in Table 5.

Samples #1 and #2 had human, cow, and universal marker concentrations that were all within the same order of magnitude. The universal BacUni-UCD marker concentrations were less than 10% different between the samples. Sample #3 contained approximately two orders of magnitude more human-derived fecal solution, and two orders of magnitude less cow-derived fecal solution, when compared with samples #1 and #2, and this relationship was confirmed with the *Bacteroidales* genetic markers. Samples #4 and #5 contained only dog-derived fecal solution and were duplicates of each other. BacUni-UCD and BacCan-UCD concentrations were within 8% and 18% of each other, respectively, confirming that relationship. The human and cow markers were negative. Sample #6 was a 100-fold dilution of samples #4 and #5. Approximately 21 times fewer gene copies were detected with BacCan-UCD compared with samples #4 and #5. Samples #7 contained 10-fold less horse feces than sample #8 and approximately 17 times fewer gene copies were found. Sample #9 consisted only of seagull guano and there were no

Table 5 – Results of blind-field validation of microbial source tracking (*Bacteroidales*) markers

Blind sample number	Percent of host solution (v/v)	Measured <i>Bacteroidales</i> marker concentrations (gene copies/ml)			
		Universal (BacUni-UCD)	Human (BacHum-UCD)	Cow (BacCow-UCD)	Dog (BacCan-UCD)
1	0.310 human, 30.00 cow	2.03×10^6	4.58×10^4	7.91×10^5	ND
2	0.295 human, 30.00 cow	1.88×10^6	1.18×10^5	5.40×10^5	ND
3	30.00 human, 0.320 cow	3.61×10^7	7.42×10^6	3.95×10^3	ND
4	30.00 dog	3.40×10^7	ND ^a	ND	2.17×10^6
5	30.00 dog	3.66×10^7	ND	ND	1.79×10^6
6	0.295 dog	7.89×10^5	ND	ND	9.10×10^4
7	3.000 horse	9.58×10^4	ND	4.76×10^2	ND
8	30.00 horse	6.60×10^5	ND	8.13×10^3	ND
9	3.070 seagull	4.04×10^2	ND	ND	ND
10	Blank (no fecal material)	ND	ND	ND	ND

^a Not detected.

false positives for the three host-specific markers; only the universal BacUni-UCD was positive for bird fecal matter. Sample #10 was a field blank and a non-detect for all markers.

4. Discussion

We identified a new “universal” marker sequence for the quantitative detection of fecal *Bacteroidales*, which is present at higher copy numbers, and may be a more comprehensive indicator of fecal contamination than the previously published total *Bacteroidales* assay (Dick and Field, 2004) and the AllBac *Bacteroidales* assay (Layton et al., 2006). Moreover, three new species-specific assays were developed for the quantitative detection of mixed human-, cow-, and dog-specific *Bacteroidales*. The mixed human-specific *Bacteroidales* marker sequence was also shown to be present at higher copy numbers than the previously published human-specific (HF183) marker sequence as detected with SYBR Green chemistry (Seurinck et al., 2005). Validation of each new assay was performed with fecal DNA extracts from several host species as well as primary influent wastewater samples from a number of different treatment plants. In a previous report, the detection of the HF183 marker sequence was 11 out of 13 samples tested, or about 85% when testing 2–4 ng (10^{-9} g) of fecal DNA (Bernhard and Field, 2000a). Seurinck et al. (2005) reported the detection of this same marker sequence, by conventional assay, as 67% (4 of 6) of human fecal samples tested, and 83% (5 of 6) by SYBR Green assay. By comparison, in the present study the HF183 marker was detected by Taq nuclease assay at 0.5 pg (10^{-12} g) of fecal DNA in 61% (11 of 18) of human fecal samples tested. Similarly, the new BacHum-UCD marker sequence was detected by Taq nuclease assay at the same concentration in 67% (12 of 18) of the human fecal samples tested. Also, although the HF183 marker had not been previously detected in any of 5 dog fecal samples tested (Seurinck et al., 2005), it was detected in 2 of the 8 dog fecal samples tested in the present study. One of these dog fecal extracts also amplified with the new BacHum-UCD assay. Others have reported difficulties in differentiating

human *Bacteroidales* populations from those of their domestic pet counterparts (Dick et al., 2005a). Further, 1 of 7 cat fecal samples produced positive results with the HF183 assay, while none of the same samples amplified with the BacHum-UCD assay. The results of the single-blind field validation study provided an even higher level of confidence in the application of the new assays than was suggested by the results of the initial assay validation, indicating that the new quantitative *Bacteroidales* assays are quite accurate, reproducible, and host-specific. A comparatively low number of gene copies were detected in seagull guano with the universal assay. This observation suggests that it will be more difficult to design a specific real-time PCR assay that can quantify low levels of seagull-specific fecal pollution in waterways and coastal areas.

Although our study proved the new assays to be relatively host-specific, the results of the initial assay validation tests underscore the difficulty in differentiating human and domestic animal fecal waste by molecular analysis. Our conditional probability approach differs from other published analyses and the US EPA guide document on MST (EPA, 2005) because it explicitly assigns as background probability the distribution of a genetic marker in a watershed. As more data are collected from a watershed the background probability should stabilize. It would be preferable if the background probability for MST assays could be determined by different means—for example, by using a chemical source tracking method—but we believe that ours is a workable first step toward applying the concepts of conditional probability to MST. Alternatively, one could estimate average background probabilities of human fecal pollution in specific geographical areas based on general public health information, comparable to obtaining statistical information regarding the prevalence of a medical condition in the population of the United States. However, there is a lack of such information for (animal) host-specific fecal pollution.

At present, the ability to measure gene copy numbers in fresh feces does not translate into the ability to perform quantitative MST. That is, there are data gaps that must be filled before we will know if truly quantitative MST is feasible

(Santo Domingo et al., in press). They include host specificity and relative abundance as well as temporal and geographical stability of specific library-independent genetic markers. It should also be noted that all existing and new host-specific *Bacteroidales* assays only target a certain percentage of the total *Bacteroidales* DNA that is present in a particular fecal extract sample. This is because multiple host-specific marker sequences comprise the “total” amount in feces. As such, a “baseline ratio” of total DNA in a fecal sample that is targeted by each assay, relative to the total DNA targeted by the universal *Bacteroidales* assay, should be evaluated. It is unknown if qPCR assays can be used to report the actual percentage of host-specific *Bacteroidales* in a water sample, similar to the reporting style often used by library-dependent methods. Quantitative data need to be translated into probabilities of host-specific pollution and corresponding fecal loads. In addition, furthering the sequence information in public databases will allow the design of assays, which target more specific sequences from a given species. With more sequences available, an estimation of so-called “pass-through” strains of *Bacteroidales* from a target species in a non-target species would shed light on sequence outliers in the phylogenetic analysis. These outliers may significantly affect the design of TaqMan assays by adding a specificity bias leading to cross-reactive assays.

5. Conclusions

Four new molecular *Bacteroidales* assays targeting regions of the 16S rRNA gene were developed and validated against different non-target host species and in a single-blind trial. Microbial source tracking (MST) studies should benefit as follows:

- Order of magnitude differences in fecal material concentrations were accurately reflected by reported values, which suggests the markers can be used to quantify relative fecal loading over space and time.
- Duplicates were well within accepted limits (and in fact differences suggest that MST assays may be more precise than traditional indicator tests, which are known to vary by a factor of 2–3 in laboratory duplicates).
- Single-blind field validation revealed no false positives or negatives, further confirming the specificity and sensitivity of the developed host-specific markers.

A conditional probability approach using Bayes' Theorem allowed the calculation of probabilities of correctly identifying human or animal sources in a watershed. These reported Bayesian probabilities are not fixed but are subject to variations based on background probabilities in other watersheds and further testing of animal stool samples.

The probability analysis is applicable to any MST method, including library-based techniques. To be truly useful for the management of watersheds, it is necessary to analyze probability distributions (i.e., test the efficacy) of applied quantitative *Bacteroidales* assays.

Acknowledgments

We thank N. Boon and W. Verstraete for providing a reference plasmid for the quantification of fecal *Bacteroidales* using their real-time HF183 PCR assay. K. Nelson and S. Stafford contributed dog fecal samples. D.E. Thompson is acknowledged for technical assistance in assay validation, and G. Reide for field assistance during the single-blind field validation study. This work was supported by the Environmental Division of the California Department of Transportation Contract no. 43A0168, TO 03, and TO 23, and by the Calleguas Creek Watershed Management Plan. V.B.R. was awarded a postdoctoral fellowship by the Fogarty International Center (NIH-FIC training Grant TW-05718) at the University of California, Davis.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.watres.2007.06.037.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410.
- Bernhard, A.E., Field, K.G., 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* 66 (4), 1587–1594.
- Bernhard, A.E., Field, K.G., 2000b. 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), 4571–4574.
- Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-vandillen, P.M.E., Vandernoordaa, J., 1990. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* 28 (3), 495–503.
- Dick, L.K., Field, K.G., 2004. Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR assay for 16S rRNA genes. *Appl. Environ. Microbiol.* 70 (9), 5695–5697.
- Dick, L.K., Bernhard, A.E., Brodeur, T.J., Domingo, J.W.S., Simpson, J.M., Walters, S.P., Field, K.G., 2005a. Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71 (6), 3184–3191.
- Dick, L.K., Simonich, M.T., Field, K.G., 2005b. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71 (6), 3179–3183.
- EPA, U., 2005. Microbial Source Tracking Guide Document. Office of Research and Development National Risk Management Research Laboratory, EPA/600/R-05/064, 123pp.
- Field, K.G., Bernhard, A.E., Brodeur, T.J., 2003. Molecular approaches to microbiological monitoring: fecal source detection. *Environ. Monitor. Assess.* 81 (1–3), 313–326.
- Griffith, J.F., Weisberg, S.B., McGee, C.D., 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *J. Water Health* 1, 141–151.
- Handler, N.B., Paytan, A., Higgens, C.P., Luthy, R.G., Boehm, A.B., 2006. Human development is linked to multiple water-body

- impairments along the California Coast. *Estuaries Coasts* (formerly *Estuaries*) 29 (5), 860–870.
- Kreader, C.A., 1995. Design and evaluation of *Bacteroides* DNA probes for the specific detection of human fecal pollution. *Appl. Environ. Microbiol.* 61 (4), 1171–1179.
- Kreader, C.A., 1998. Persistence of PCR-detectable *Bacteroides distasonis* from human feces in river water. *Appl. Environ. Microbiol.* 64 (10), 4103–4105.
- Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., Saylor, 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), 4214–4224.
- Morrison, T.B., Weis, J.J., Wittwer, C.T., 1998. Quantification of low-copy transcripts by continuous SYBR (R) green I monitoring during amplification. *Biotechniques* 24 (6), 954.
- Okabe, S., Okayama, N., Savichtcheva, O., Ito, T., 2007. Quantification of host-specific *Bacteroides*–*Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl. Microbiol. Biotechnol.* 74 (4), 890–901.
- Paster, B.J., Dewhirst, F.E., Olsen, I., Fraser, G.J., 1994. Phylogeny of *Bacteroides*, *Prevotella*, and *Porphyromonas* spp. and related bacteria. *J. Bacteriol.* 176 (3), 725–732.
- Rajal, V.B., McSwain, B.S., Thompson, D.E., Leutenegger, C.M., Kildare, B.J., Wuertz, S., 2007. Validation of hollow fiber ultrafiltration and real time PCR using bacteriophage PP7 as surrogate for the quantification of viruses from water samples. *Water Res.* 41, 1411–1422.
- Reischer, G.H., Kasper, D.C., Steinborn, R., Mach, R.L., Farnleitner, A.H., 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72 (8), 5610–5614.
- Reischer, G.H., Kasper, D.C., Steinborn, R., Farnleitner, A.H., Mach, R.L., 2007. A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Lett. Appl. Microbiol.* 44 (4), 351–356.
- Santo Domingo, J.W., Bambic, D.G., Edge, T.A., Wuertz, S., in press. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Res.*, in press, doi:10.1016/j.watres.2007.06.001.
- Seurinck, S., Defoirdt, T., Verstraete, W., 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), 249–259.
- Stricker, A.R., Wilhartitz, I., Farnleitner, A.H., Mach, R.L., in press. Development of a Scorpion probe-based real-time PCR for the sensitive quantification of *Bacteroides* sp. ribosomal DNA from human and cattle origin and evaluation in spring water matrices. *Microbiol. Res.* (corrected proof).
- Vaneechoutte, M., Young, D.M., Ornston, L.N., De Baere, T., Nemec, A., Van Der Reijden, T., Carr, E., Tjernberg, I., Dijkshoorn, L., 2006. Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. *Appl. Environ. Microbiol.* 72 (1), 932–936.
- Wittwer, C.T., Herrmann, M.G., Moss, A.A., Rasmussen, R.P., 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* 22 (1), 130.