Identification of culturable stream water bacteria from urban, agricultural, and forested watersheds using 16S rRNA gene sequencing

Kenneth T. Belt, Christina Hohn, Aiah Gbakima and James A. Higgins

ABSTRACT

Bacteria present in water samples taken on a weekly basis, from June 2004 through June 2005, from three streams, were cultured on Coliscan® Easygel® agar plates. Colonies representative of a variety of colors and morphologies were subjected to amplification and sequencing of a 1000–1100 nt portion of the 16S rRNA gene. A total of 528 colonies were sequenced; these categorized into 26 genera and 78 species. Of 175 dark blue/purple colonies presumed to be *E. coli*, sequence analysis indicated that 45 (25%) were actually other genera. For the urban stream Gwynns Falls Gwynns Run, *E. coli* was the most common genus/species encountered, followed by *Klebsiella* and *Aeromonas*. For Pond Branch, a stream located in a forested watershed, it was *Serratia*, followed by *Yersinia* and *Aeromonas*. For McDonogh (MCDN), a stream associated with *Zea mays* (corn) row crop agriculture, *E. coli* was the most frequently isolated genus/species, followed by *Aeromonas* and *Enterobacter*. ERIC-PCR genotyping of isolates from the most prevalent genera/species, indicated a high degree of diversity within-stream for *E. coli* and *K. pneumoniae*. Conversely, genotyping of *Y. enterocolitica* isolates indicated that some were shared between different streams.

Key words | 16S rRNA, bacterial identification, freshwater bacteria, stream flora

Kenneth T. Belt
USDA Forest Service,
UMBC CUERE/BES LTER,
Baltimore, MD 21227,
USA

Christina Hohn
USDA-ARS,
Beltsville, Maryland 20705,
USA

Aiah Gbakima
Morgan State University,
Baltimore, Maryland 21251,
USA

James A. Higgins (corresponding author)
USDA – ARS,
Bldg 173, 10300 Baltimore Blvd,
Beltsville MD 20705,
USA
Tel.: 301-504-6443
Fax: 301-504-6608
E-mail: *jhiggins@anri.barc.usda.gov*

INTRODUCTION

Methods utilizing the detection and enumeration of select bacterial species, historically referred to as 'coliform' bacteria, continue to be important tools in monitoring the quality of fresh and marine waters in the US and Europe. Water bodies with culturable bacterial numbers in excess of given thresholds may be considered unfit for consumption and recreational usage (National Research Council 2004). Conversely, reduction in coliform counts may be regarded as evidence that a previously polluted body of water has benefited from remediation and anti-pollution efforts. Most water quality testing laboratories use one of any number of commercial or standardized test methods to culture coliforms from samples, usually 100 ml in volume (Rompré *et al.* 2002).

While the assemblage of species making up the coliforms has undergone revision over the years, at present,

the consensus among most microbiologists is that it is predominantly comprised of members of the family Enterobacteriaceae; in their review Leclerc *et al.* (2001) list 32 genera, among them *Buttiauxella*, *Enterobacter*, *Hafnia*, *Morganella*, *Pantoea*, *Rahnella*, and *Yersinia*. It is recognized that not all of these taxa are necessarily exclusively fecal in origin, and many of them can be considered normal members of the aquatic flora (Taylor 2003). It is also true that many fecal bacterial species are not represented in the coliforms category. These considerations have resulted in the substitution of other assays, for example, those targeting enterococci, for the traditional coliform count. However, despite its drawbacks the coliform count remains widely used by water quality laboratories due to its ease of use, economy, and acceptance by regulatory institutions, such

the United States Environmental Protection Agency (USEPA) (National Research Council 2004).

Over the past five years we have been using a commercial, USEPA-approved coliform count assay, the Coliscan® Easygel® protocol from Micrology Laboratories (www.micrologylabs.com), to enumerate *E. coli* in stream samples. These samples have been collected on a weekly basis from nine core stream sites under the auspices of the Baltimore Ecosystem Study (BES), a National Science Foundation (NSF) sponsored Long Term Ecology Research (LTER) project. The Coliscan protocol uses a proprietary formulation of chromogenic agar to allow the end user to visually differentiate between bacterial genera and species growing on the same agar plate. Purple/dark blue colonies represent *E. coli*, while pink, white, blue-green, and yellow colonies represent other species.

We were interested in using a molecular method, 16S rRNA gene sequencing, to identify to the genus/species level individual colonies growing on Coliscan plates. In so doing, we hoped to gain an increased understanding of the diversity of culturable bacteria in water samples representative of agricultural, forested, and urban watersheds. We were also interested in learning if the putative *E. coli* colonies identified on the plate were confirmable using 16S rRNA sequence data, as well as if known pathogenic genera and species of enteric bacteria were present in these streams. Another goal was to assemble panels of frequently-occurring genera and/or species, and subject them to in-depth genotyping assays, thereby obtaining some appreciation of the genetic diversity among these isolates. Here, we provide the results of our investigations into these topics.

METHODS

Sampling

Three Baltimore, MD streams (Piedmont province; 76°30′, 39°15′) were sampled weekly from June 2004 to June 2005 (map in Higgins *et al.* 2005). McDonogh (MCDN (Agr), 7.8 ha.) is a small western Baltimore County stream with corn (*Zea mays*) no-till, row-crop agriculture and forested land use (no residents) with high levels of nutrients (i.e. N and P). The corn fields are not irrigated, and receive applications of horse manure as a fertilizer at 6–8 week intervals during the period from harvest to planting (usually mid-September to early May). Pond Branch (POBR (For), 32.3 ha.) is a small low-nutrient stream in an uninhabited forested catchment in Oregon Ridge Park (Baltimore County). The third stream sampled, Gwynns Run (GFGR (Urb), 557 ha.), is a somewhat larger inner city urban catchment, impacted by sewage leaks, that flows into a major urban park in Baltimore City. Included among the GFGR (Urb) samples were a smaller subset taken from Gwynns Falls (GFCP) just downstream of the confluence with Gwynns Run; inclusion of GFCP - derived bacteria was sometimes necessitated when the densities of colonies in the GFGR (Urb) sample were too numerous to permit selection of individual colonies from the surface of the agar plate, or when cold weather samples of GFGR (Urb) water failed to yield purple/dark blue colonies indicative of *E. coli*.

Thirty five ml samples of stream water were transported to the laboratory in coolers (on ice) to the United States Forest Service/Center for Urban Environmental Research and Education laboratory at the University of Maryland Baltimore County in Catonsville, MD within 5 hours of collection. There, 1 ml or 5 ml aliquots were mixed with 13 ml of Coliscan Easygel® media (dilutions with sterile water for GFGR of 1:5, 1:10 or 1:100 were done when concentrations of bacteria were anticipated to be high, as in the warmer months), poured into empty 90 mm diameter Coliscan petri dishes, allowed to congeal, and incubated overnight at 37°C. Putative *E. coli* (PEC) colonies were enumerated within 18–20 hours and plates were stored at 4°C for DNA extraction.

DNA extraction, 16S rRNA PCR, and sequencing

The sampling protocol called for selection of ∼16 colonies from each of the three stream sites per month, for a total of ∼48 colonies per month, over a one-year period. For each month, an effort was made to obtain half of the 48 colonies from the first two weeks of the month, and the remaining colonies from the second half of the month. Whenever possible, 40%–50% of a given stream site's monthly colony selection was comprised of dark blue or purple colonies (i.e. PEC). The balance of the remaining samples were selected at random from the pink, red, yellow, green, and white colonies present on the Coliscan plates. We observed that in

the warmer months of the year (i.e. May through September) blue colonies were plentiful on plates for all three stream sites; however, during the Winter months (i.e. December through March) blue colonies were less plentiful, and sometimes the bulk of sampled colonies for the MCDN (Agr) and POBR (For) sites necessarily consisted of non-blue colonies.

For June and July 2004, samples were obtained from both Coliscan and CHROMagar (Hardy Diagnostics, Santa Maria, CA) plates; afterwards Coliscan plates were used exclusively.

Sterile glass Pasteur pipettes were used to core individual bacterial colonies from the Coliscan Easygel plates; the cores were deposited into 100 μl volumes of Instagene matrix (Bio-Rad, Hercules, CA), vortexed, and subjected to consecutive heating steps of 56°C for 15 minutes and 95°C for 8 minutes, after which samples were briefly centrifuged to pellet the Instagene matrix. Five μl of Instagene-treated DNA was used as template for PCR with the conserved 16S rRNA primer set of Kazor *et al.* (2003), which amplified ~1400 bp of the gene from the majority of our specimens. Thermal cycling conditions and reagents were as previously described (Higgins *et al.* 2005). PCR products were electrophoresed on agarose gels, stained with ethidium bromide, and visually inspected to confirm the presence of amplicons with appropriate size and quantity. Five μl aliquots of PCR product were treated with 2 μl Exo-SAP™ reagent (USB Corp., Cleveland, OH) according to the manufacturer's recommendations, and 2 μl portions of the Exo-SAP treated PCR product was used as a template for dye-terminator cycle sequencing with the Big Dye™ 3.0 kit (PE Biosystems, Foster City, CA). Sequencing reactions were electrophoresed on an Applied Biosystems ABI 3100 model automated fluorescent sequencer, and resultant data processed using DNAStar™ software (Madison, Wisconsin). Forward and reverse reads were submitted to the Ribosomal Database Project II website (rdp.cme.msu.edu) (Cole *et al.* 2005) for database query and selection of nearest matches.

Enterobacterial repetitive intergenic consensus (ERIC) PCR

Three μl of bacterial DNA was subjected to PCR using the primer pair ERIC1R and ERIC2 (Versalovic *et al.* 1991) at 50 pmol each, with 2.0 mM MgCl$_2$, 2 *U* DNA polymerase

(Q-Biogene, Irvine, CA), and 10 mM each dNTP in a 50 μl reaction volume. Two μl aliquots of PCR product were quantitated using a NanoDrop ND-1000 model spectrophotometer (NanoDrop Technologies, Wilmington, DE), and ~3.7–4.0 μg amounts of each product electrophoresed on 2% agarose gels. Images were saved as tiff files and analyzed using the Bionumerics 2.5 software package (Austin, TX). Following normalization of bands, clustering analysis, dendrogram construction and multidimensional scaling (MDS) plots were generated using UPGMA with Pearson correlation similarity coefficient, using the Bionumerics software.

Since MDS plots are somewhat rarely encountered in most published phylogenetic studies, a brief description may be in order: while traditional phylogenetic diagrams depict taxa in a two-dimensional space, with the taxa labeled at the end of "tree branches", in MDS plots taxa are represented as spheres in a three-dimensional space. Spheres in close proximity to one another may be considered to have a high degree of similarity, while conversely, spheres located at a visibly large distance from one another may be regarded as having a high degree of dissimilarity. Taxa with a very high degree of similarity may be represented by partially fused spheres, while 100% similarity is rendered by a single sphere with bi-colored hemispheres.

Monitoring of amplified sequences for adventitious contaminants

One of the drawbacks associated with the routine usage of universal 16S rRNA primers is that it can enhance opportunities for contamination of template DNA, and PCR and cycle-sequencing reactions, with adventitiously amplified contaminants. These contamination events can occur despite the conscientious use of good laboratory practices; for example, Tanner *et al.* (1998) examined contaminant 16S rRNA sequences present in PCR and cloning reactions conducted in their laboratory, and identified a diverse assemblage of genera and species. Among the organisms they identified are many that occur in water and/or soil: *Duganella*, *Pseudomonas*, *E. coli*, *Stenotrophomonas*, *Leptothrix*, and *Shigella*. Many of the contaminant sequences had previously been reported as constituents of published clone libraries from various habitats, including acid mine drainage, infected guinea

pig lung tissue, Siberian tundra soil, and a contaminated aquifer (Tanner *et al.* 1998). The presence of these putative contaminants in such a variety of sample types raises the possibility that they are ubiquitous in laboratory environments, and thus, may be present in our samples and reagents as well. This is of concern in light of the fact that over the yearlong course of our project we observed faint bands in ~15% of our "no template" PCR controls, indicating that contaminant DNA was present in our amplification reagents. To address the issue of whether any of our stream-derived bacterial 16S rRNA sequences were in fact 'universal' contaminants similar to those observed by Tanner *et al.*, we randomly selected sequences for 19 *E. coli* isolates, 8 *Acinetobacter* isolates, and one *Duganella* isolate, obtained from our stream water samples. We then compared these 16S rRNA sequences to those 16S rRNA sequences identified as contaminants by Tanner *et al.*: AF047649 (similar to *Stenotrophomonas maltophilia*), AF050528 (*Duganella zoogloeoides*), L20812 (*Acinetobacter* spp.), U34035 (*D. zoogloeoides*), X91528 (*S. maltophilia*), and X91531 (*E. coli*); their lengths ranged from ~350 nt to ~1450 nt.

RESULTS

16S rRNA sequence- based identification of bacterial colonies

Over 99% of colony DNA preparations yielded DNA capable of amplification with the universal 16S rRNA primers, and 95% of those amplicons yielded forward and reverse run sequence data. In general, most runs generated 500–600 nt of quality read (i.e. clearly separated electropherogram peaks with minimal background and minimal unassigned nucleotide calls) with a total 16S rRNA gene read length of 1000–1200 nt, equivalent to ~66%–80% of the entire gene sequence.

Query of the RDP II database with our sequences resulted in similarity scores of >99% for all submitted samples; our submitted sequence was assigned its genus and species identity based on the highest-scoring taxa retrieved by the RDP II database, which was usually ≥99.7%.

A total of 26 genera and 78 species were observed among 528 colonies analyzed from June 2004 through June

2005. Figure 1 shows the per cent distribution of the most frequently isolated genera for each of the three stream sites. For the polluted stream site GFGR (Urb), *E. coli* was the most common genus encountered on the agar plates at 39%, followed by *Klebsiella* (15%) and *Aeromonas* (13%). For POBR (For), which is located in a forested watershed with (presumably) little or no input from anthropogenic sources, it was *Serratia* (16%), followed by *Yersinia* (14%) and *Aeromonas* (12%). MCDN (Agr) was similar to GFGR in that *E. coli* was the most frequently isolated genus (20%), followed by *Aeromonas* (12%) and *Enterobacter* (11%).

Of a total of 175 blue colonies screened via 16S rRNA sequencing, 45 (25%) were non -*E. coli*. These other taxa presenting as blue colonies included *Citrobacter*, *Buttiauxella*, *Shigella*, *Enterobacter*, *Klebsiella*, *Brevundimonas*, *Serratia*, *Pseudomonas*, and *Aeromonas*. The likelihood of a blue colony originating from a species other than *E. coli* was higher in the warmer months: for all three stream sites combined, the percentage of false-positive PEC from the period from December through February was 12.8% ($n = 39$), while for the period May through September, the percentage of PEC that were false-positive was 48.3% ($n = 60$). The implication of these "false positive" counts, upon coliform/*E. coli* counting data used to monitor water quality, is uncertain. For example, weekly concentrations of *E. coli* in the three streams, as determined by counting blue colonies, are presented in Figure 2; because of sewage input (presumably) concentrations in the polluted urban stream GFGR (Urb) are markedly higher than for MCDN (Agr) and POBR (For). Even if we posit that one-fourth of the colony counts used to determine concentration for a given sample date for GFGR (Urb) are invalid, 'legitimate' *E. coli* concentrations remain in violation of recreational water quality standards for the state of Maryland.

ERIC-PCR analysis of bacterial species present in multiple streams

While our ERIC-PCR analyses are presented as traditional dendrograms (Figures 3–6), we have also provided to the reader multidimensional scaling plots of these dendrograms (Figure 7). Labeling individual genotypes (i.e. spheres) in each of the plots in a legible manner would be helpful in their interpretation, but would not be feasible without
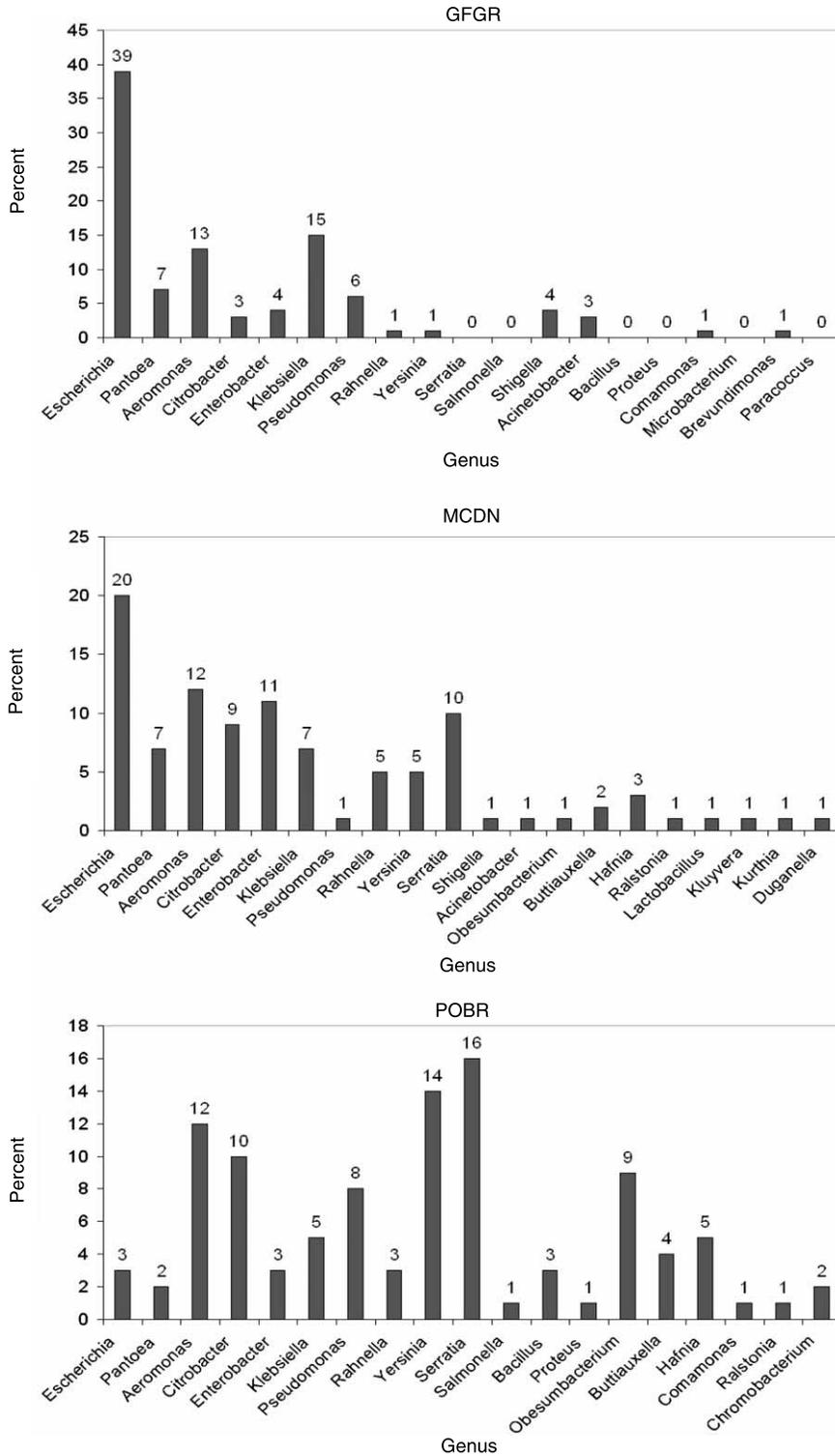
Figure 1 │ Bar charts showing per cent distribution of most frequently isolated genera derived from colonies identified by 16S rRNA gene sequencing from three stream sites: Gwynns Falls Gwynns Run (GFGR; urban), McDonogh (MCDN; agricultural), and Pond Branch (POBR; forested). Approximately 176 colonies, cultured on Coliscan plates, were screened for each stream site over the period June 2004 - June 2005. Percentages are provided atop each bar; note that values < 1.0% are indicated as 0%.
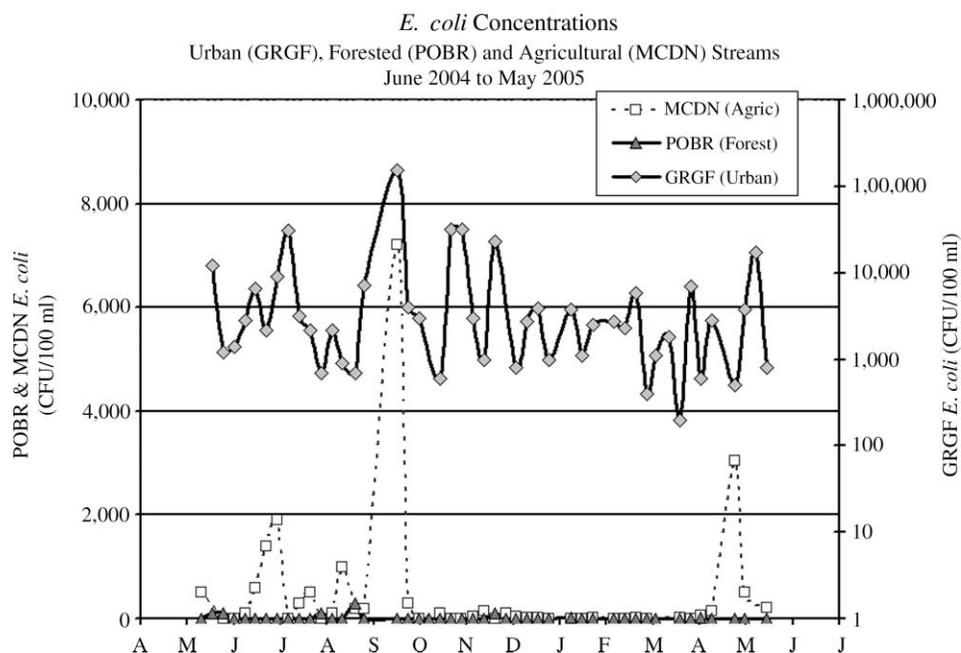
Figure 2 | Weekly concentrations of *E. coli* (as determined by counts of dark blue and purple colonies on Coliscan plates) in samples obtained from three streams: McDonogh (MCDN; agricultural), Pond Branch (POBR; forested) and Gwynns Falls Gwynns Run (GFGR; urban), for the period June 2004 – June 2005. Time (in months) is plotted on the x-axis and colony forming units (CFU) per 100 ml water on the y-axes.

expanding each plot to an entire page in order to accommodate such labeling. However, we hope that even in the absence of labels, the plots will provide a visually accessible summation of the clustering analysis results for each of our bacterial assemblages.

Yersinia enterocolitica was prevalent in the colder months of the year in water samples from POBR; isolates were also recovered from MCDN and GFGR. In order to gain some understanding of the genetic diversity among our isolates, we subjected 17 of them to ERIC-PCR; the results are presented in Figures 3 and 7A. The GFGR isolate from March 1, 2005, and the MCDN isolate from October 12, 2004, failed to cluster with other isolates and presumably represent genotypes unique to these streams. POBR isolates from October 2004 – January 2005 showed the greatest degree of similarity based on ERIC PCR banding patterns, while several MCDN isolates clustered with POBR isolates, suggesting that some genotypes of *Y. enterocolitica* may be present at these two different stream types. The ATCC reference strain of *Y. enterocolitica* (No. 23715, a clinical isolate), did not display a significant degree of clustering with the stream isolates.

Pantoea agglomerans was present for several months of the year in both MCDN and GFGR; a dendrogram and MDS plot, generated via ERIC-PCR profiles of 19 isolates from these two sites, are presented in Figures 4 and 7B. Appreciable levels of clustering were observed only within-stream for this species, for example, the August 17, 2004 'B' replicate and the August 31, 2004 isolates from GFGR. Some isolates, such as the June 22 2004 MCDN and August 17 2004 GFGR samples, appear to represent highly diverse lineages of *P. agglomerans*.

ERIC-PCR analysis of bacterial species present in one stream

We were interested in genotyping multiple isolates of one species from a single stream site. For MCDN (Agr), *E. coli* was present in all months save February 2005, and was selected for analysis. A total of 18 isolates, spanning June 2004 to June 2005, were assayed and results of cluster analysis of ERIC-PCR banding patterns are presented in Figures 5 and 7C. Isolates from September 21 2004 and October 12 2004 showed a high degree of similarity, as did
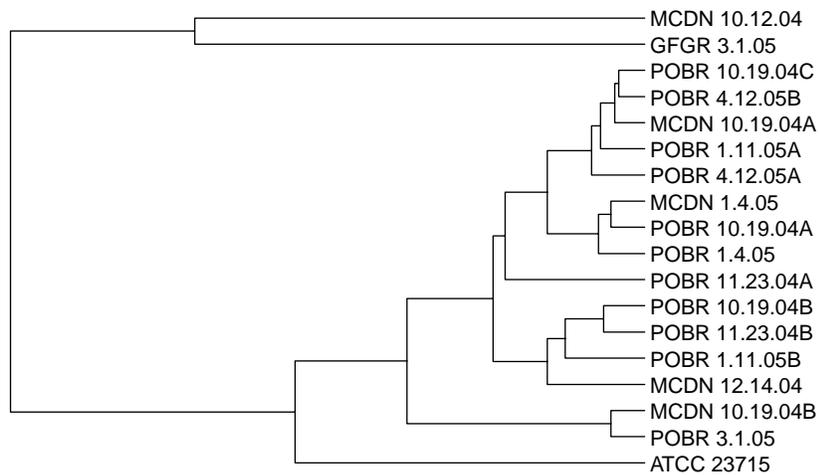
Figure 3 │ Dendrogram generated from ERIC-PCR profiles of isolates of *Yersinia enterocolitica* cultured from McDonogh (MCDN; agricultural), Pond Branch (POBR; forested), and Gwynns Falls Gwynns Run (GFGR; urban) stream water samples taken between June 2004 and June 2005. Sample ATCC 23715 is an isolate obtained from the American Type Tissue Collection. Date format is: Month.date.year; multiple isolates generated from one water sample are indicated with an 'A', 'B', etc, designation after the year.

isolates from April 26 2005 and May 11 2005, and two from June 14 2005. However, the majority of ERIC-PCR banding patterns of the *E. coli* isolates from MCDN (Agr) failed to segregate into discrete end-branch clusters, suggesting that a variety of genotypes were present in the water samples.

We observed that *Klebsiella pneumoniae* was prevalent in GFGR (Urb) samples from August, 2004 through April 2005; accordingly, we subjected 18 isolates to ERIC-PCR-based genotyping. A dendrogram/MDS plot constructed

from the resultant analysis is provided in Figures 6 and 7D. Isolates segregated into two large groupings; interestingly, both within-month (September 8 2004 and April 12 2005) and consecutive-month (August – September and November – December 2004) isolates showed a high degree of similarity. However, the remaining isolates displayed great variability in their ERIC-PCR banding profiles and some, such as the October 19 2004 replicate 'B' and June 22 2004 replicate 'A' isolates, probably represented distinct lineages.
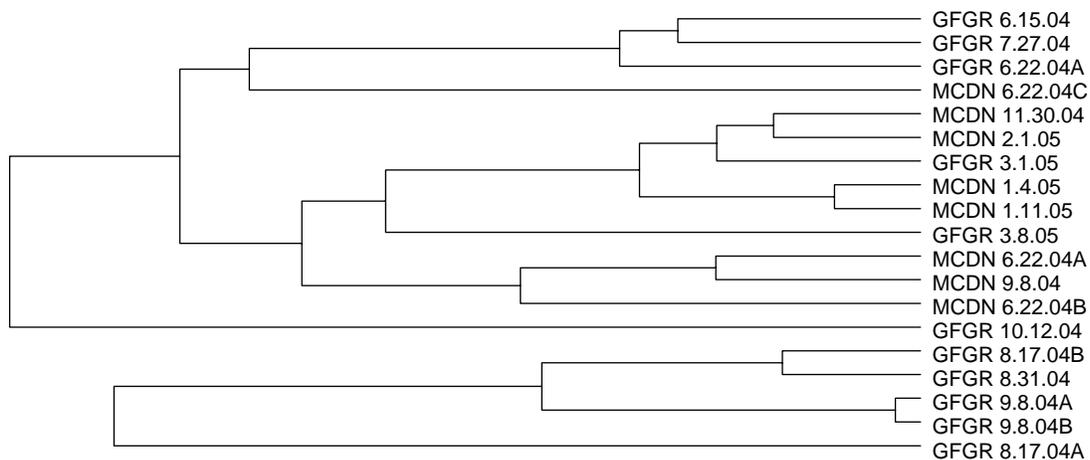
Figure 4 │ Dendrogram generated from ERIC-PCR profiles of isolates of *Pantoea agglomerans* cultured from McDonogh (MCDN; agricultural) and Gwynns Falls Gwynns Run (GFGR; urban) stream water samples taken between June 2004 and June 2005. Date format is: Month.date.year; multiple isolates generated from one water sample are indicated with an 'A', 'B', etc, designation after the year.
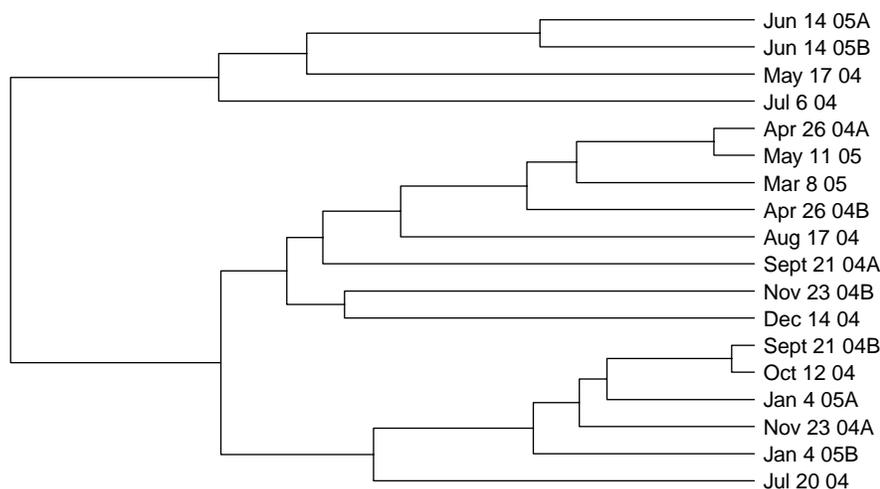
Figure 5 | Dendrogram generated from ERIC-PCR profiles of isolates of *Escherichia coli* cultured from McDonogh (MCDN; agricultural) stream water samples taken between June 2004 and June 2005. Multiple isolates generated from one water sample are indicated with an 'A', 'B', etc, designation after the year.

COMPARISON OF STREAM BACTERIAL 16S rRNA SEQUENCES WITH THOSE OF LABORATORY CONTAMINANTS

None of our randomly selected 19 *E. coli* sequences had a similarity score in excess of 91.4% with that of the contaminant identified as *E. coli* by Tanner *et al.* (1998). For our 8 *Acinetobacter* sequences, the similarity scores ranged from 96.4% to a maximum of 97.8% with the *Acinetobacter* identified by Tanner *et al.* (1998). Our

Duganella sequence from MCDN (Agr) (isolated Sept. 21, 2004) had 74.8% similarity with entry AF050528 and 96.4% similarity with entry U34035, 16S rRNA sequences identified as *Duganella* by Tanner *et al.* (1998).

One of our negative control 16S rRNA PCR reactions had a band present in sufficient quantity for sequencing; when queried at the RDP II website, the resultant 582 nt sequence scored 97.3% with Genbank accession X91518, "uncultured bacterium, clone group K3".
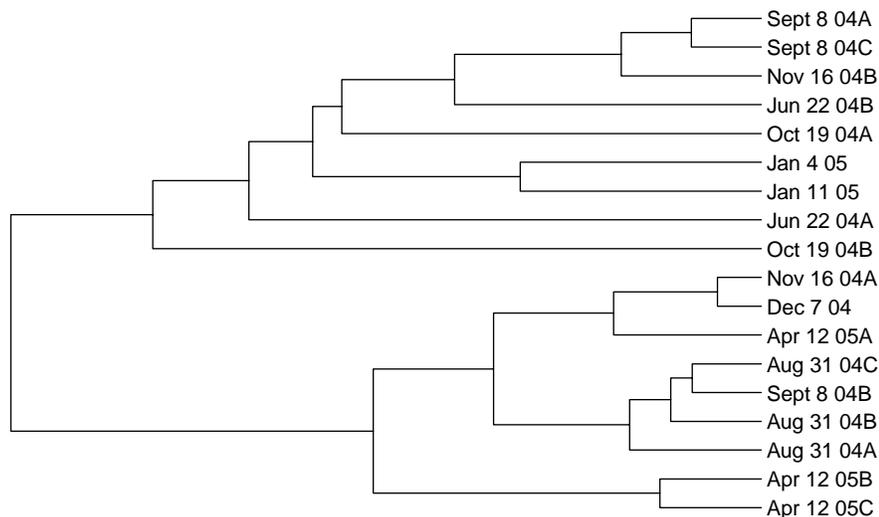

Figure 6 | Dendrogram generated from ERIC-PCR profiles of isolates of *Klebsiella pneumoniae* cultured from Gwynns Falls Gwynns Run (GFGR; urban) stream water samples taken between June 2004 and June 2005. Multiple isolates generated from one water sample are indicated with an 'A', 'B', etc, designation after the year.
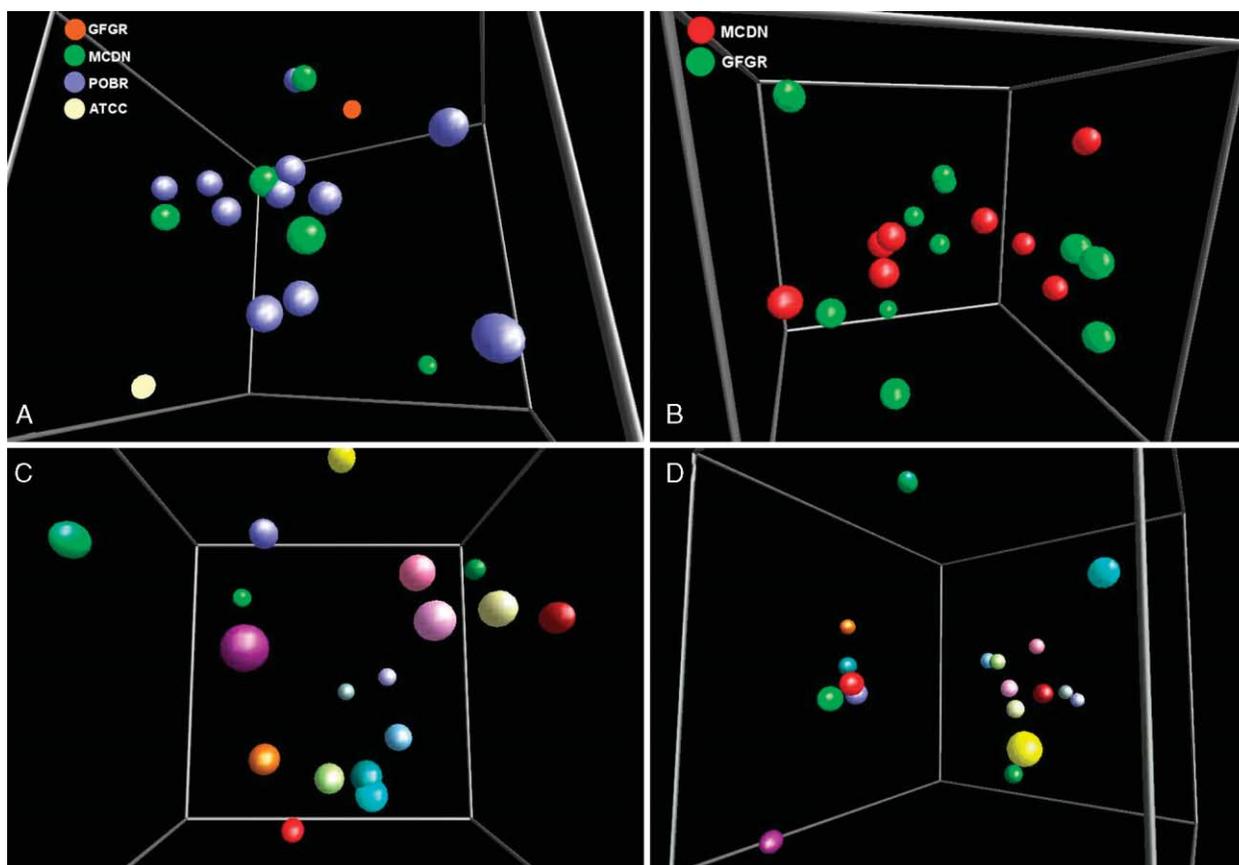
Figure 7 | Multidimensional scaling plots (MDS) for dendrograms constructed from ERIC-PCR generated banding profiles for *Y. enterocolitica* (Panel A), *Pantoea agglomerans* (Panel B), *E. coli* in Mc Donogh stream (Panel C), and *K. pneumoniae* in Gwynns Falls Gwynns Run stream (Panel D). Each isolate's ERIC-PCR banding pattern, or genotype, is represented by an individual sphere. Spheres in close proximity to one another, or partially fused, represent isolates/genotypes with high degrees of similarity, while spheres at some distance from one another represent more dissimilar isolates. Note that in panels A and B sphere color schemes are used to represent streams from which the isolates were obtained, while in panels c and D each isolate is represented by a unique sphere color.

DISCUSSION

The Coliscan method is one of a number of commercially available protocols that can be used to detect and enumerate coliform bacteria in source and finished water samples. These protocols rely on enzyme-substrate reactions, derived from observations on bacterial physiology, as their mechanism of operation. While there are drawbacks to this approach, the ease of use and inexpensive nature of these assays makes them popular for use in studies of bacterial content in both source and finished waters (Rompré *et al.* 2002). We are of course mindful that molecular-based approaches to surveying fresh water bacterial composition are preferable when identifying species that are difficult (if not impossible) to culture on

conventional agar formulations (O'Sullivan *et al.* 2002). However, widespread use of PCR and DNA sequencing-based methods for detecting and enumerating coliforms remains, at present, too expensive for routine use by most institutions and organizations responsible for monitoring water quality. Since we do have resources for sequence-based identification approaches to water monitoring, we were interested in using molecular methods to identify and characterize bacteria cultured from stream water samples on Coliscan plates.

Most of the 26 culturable genera/78 species we detected in our stream water samples are members of the 'coliform' group and confirm that the Coliscan agar is capable of supporting growth of this varied assemblage of microorganisms. To the best of our knowledge, our study

represents the first use of 16S rRNA sequencing data to identify large numbers of organisms cultured on this agar formulation; previous identifications have been derived from biochemical profiling. While our sampling regimen was not of sufficient size to allow us to make statistically valid conclusions about the temporal/seasonal distribution of the various genera and species observed, we do feel that the data support some general observations; for example, certain genera, such as *Yersinia* and *Hafnia*, are more likely to be cultured in water samples taken in the colder months, while other genera (*E. coli*, *Pantoea*, and *Aeromonas*) can be cultured year-round (*refer to Supporting Document "Sequencing results"*) Whether this phenomenon is due simply to the reduced concentration of bacteria in cold-weather samples, which allows "hardier" genera an improved opportunity to grow under laboratory conditions, is unclear and may require more intensive sampling over more defined times of the year.

The genus most frequently cultured from all three streams over the greatest number of months was *Aeromonas*, which (historically) is not considered a member of the coliform group (Leclerc *et al.* 2001). This bacterium is widespread in aquatic environments and its presence not unsurprising (Pettibone 1998). *E. coli* was the second most-prevalent species encountered, and again, its presence (particularly in the polluted urban stream GFGR) was not unexpected. We observed potentially pathogenic species of bacteria from our forested watershed stream, POBR; these included 11 isolates of *Yersinia enterocolitica*, and one of *Salmonella typhimurium*. The origin of these isolates is unclear, but anthropogenic sources appear to be unlikely. The detection of *S. typhimurium*, *Shigella flexneri*, *S. boydii*, and *S. sonnei* in GFGR (Urb) may be explained by sewage influx, since we have observed other enteric pathogens (e.g. *Giardia* spp.) in this stream. The presence of *Y. enterocolitica* and *S. flexneri* in MCDN (Agr) cannot be explained by sewage influx, as this stream is located between a forested area and corn fields; however, the periodic deposition of horse manure onto the adjoining corn fields potentially may serve as a source of *Yersinia* spp. To date, only humans and primates have been identified as hosts for *Shigella* spp., therefore, it is unlikely that the horse manure served as a source for this organism. Consequently,

the origin of the *S. flexneri* detected in MCDN (agr) remains unclear.

The public health importance of detecting potentially pathogenic species of *Salmonella*, *Yersinia*, and *Shigella* in our sampled streams is difficult to assess, in lieu of the lack of any reportage - in the mass media or scientific literature- of outbreaks of waterborne disease in the metropolitan Baltimore area during our sampling period. We are aware that recreational activities, such as fishing and wading, do take place in GFGR despite its obviously polluted nature (and despite prominently placed signs alerting the public of this fact). While POBR (For) ultimately flows into Loch Raven reservoir, which supplies drinking water to Baltimore city, chlorination should remove any health concerns associated with the possible deposition of pathogenic bacterial species into the reservoir. However, based on our results, we would caution individuals engaged in recreational activities near POBR (For) to refrain from drinking the stream water without performing a disinfection step and it may be prudent to place signs indicating this fact alongside the stream.

Our sequence queries to the Ribosomal Database Project website elicited similarity scores in excess of 99%; we are confident that the genus designations assigned to our stream water bacterial isolates are accurate. Whether the species assignments derived from our sequence submissions all are 100% accurate is less clear. At present there does not appear to be a universal standard among diagnostic microbiology laboratories, for the minimum length of 16S rRNA gene sequence necessary to arrive at a species designation with, say, $\geq$99% confidence (Kolbert & Persing 1999; Clarridge III 2004). We are aware that for some genera (for example *Bacillus*; Blackwood *et al.* 2004), the entire sequence of the 16S rRNA gene may not provide enough nucleotide substitutions to allow for discrimination between species. On the other hand, Clarridge III (2004) demonstrates that sequencing either of a 500 or 1500 bp region of the 16S rRNA gene allows for discrimination of species and strains of *Brevibacterium*; minor changes in phylogenetic trees constructed from these sequences can be attributed to the differences in sequence read length. And, for *Haemophilus influenzae*, Sacchi *et al.* (2005) found that there was sufficient heterogeneity in 16S rRNA sequences to not only allow for species discrimination, but also for differentiation

to the serotype level, in a manner comparable to multilocus sequence typing (MLST) of portions of the *adk*, *atpG*, *frdB*, *fucK*, *mdh*, *pgi*, and *recA* genes. In addition, a commercial protocol for bacterial identification (the Applied Biosystems MicroSeq®) relies on a 500 bp region of the 16S rRNA gene for species assignation. In light of these considerations, we feel our species designations for our stream water isolates are reasonably accurate.

The presence of nucleic acids derived from contaminant microorganisms can present a significant obstacle to obtaining accurate identifications of bacteria, particularly when a conserved gene target, such as the 16S rRNA, is employed. Tanner *et al.* (1998) had identified a number of species and genera that appear to be ubiquitous in laboratory environments and could possibly confound interpretation of sequence-based identification of culturable bacteria. In an effort to determine if some of our sequences were derived from these contaminants, we compared them to a database of previously identified bacterial contaminant 16S sequences. None of our sequences displayed >98% identity with the contaminants reported by Tanner *et al.* (1998). We also selected a band present in one of our negative control reactions at sufficient intensity to provide enough template, to sequencing; its highest % similarity was with an uncultured bacterium not represented in our stream water bacterial assemblage.

Accordingly, we feel that the likelihood of our database containing false-positive entries is low.

Any approach that relies on a qualitative variable, particularly one with a degree of subjectivity such as colony color, to select samples for analysis is at some risk of skewing data collection. We took pains to select, as putative *E. coli* colonies, those with a dark blue/purple color and fast-growing (i.e. colony appearance after 12 hr incubation) nature, as recommended by the Coliscan instructions. Forty-five (25%) of 175 such colonies generated 16S rRNA sequence of non - *E. coli* organisms. It is difficult, in the absence of comparative studies with other chromogenic agar formulations, to determine if this is an acceptable level of false-positive identification. We note that studies conducted using the Colilert® system, which relies on an enzymatic reaction in a fluid-based format to identify *E. coli* and coliforms, have reported false positive reactions when marine and freshwater samples were assayed (Pisciotta *et al.* 2002; Chao *et al.* 2004). In our hands also, the Colilert method is capable of generating false-positives from stream water-derived bacteria. To be fair, vendors of the Coliscan and Colilert assays do not guarantee that their products are 100% sensitive and specific. Based on our observations, we might suggest that investigators using the Coliscan method to enumerate *E. coli* in samples use some degree of caution in interpreting their results, particularly in the warmer months of the year when coliform concentrations in water may be high.

In conclusion, we surveyed culturable bacteria from three streams in the metropolitan Baltimore, Maryland area over a one-year period, using 16S rRNA sequence to identify the organisms to the genus/species level. The observed bacteria were mostly comprised of members of the coliform group, and included as well potential pathogens such as *Y. enterocolitica*, *S. typhimurium*, and *Shigella* spp. Our results demonstrate that a sizeable proportion (25%) of the Coliscan colonies presumed to be *E. coli* were in fact other species and genera of bacteria. Users of the Coliscan reagent may therefore want to contemplate the use of ancillary methods (further biochemical or physiological tests, or molecular biology-based protocols) to confirm the identity of putative *E. coli* colonies.

CONCLUSIONS

- Pathogenic species of bacteria (*Salmonella*, *Yersinia*, and *Shigella*) were recovered from polluted urban, rural agricultural and forested watersheds.
- For all three sampled streams, *Aeromonas* was among the three most-prevalent genera recovered.
- Twenty-five per cent of putative *E. coli* colonies sampled from Coliscan plates were, upon 16S rRNA sequence-based identification, other genera of bacteria.
- Genetic diversity was high among in-stream isolates of *E. coli* and *Klebsiella pneumoniae*, while between-stream isolates of *Yersinia enterocolitica* had similar genotyping profiles.

ACKNOWLEDGEMENTS

REFERENCES

Blackwood, K. S., Turenne, C. Y., Harmsen, D. & Kabani, A. M. 2004 Reassessment of sequence-based targets for identification of *Bacillus* species. *J. Clin. Microbiol.* **42**, 1626–1630.

Chao, K.-K., Chao, C.-C. & Chao, W.-L. 2004 Evaluation of Colilert-18 for detection of coliforms and *Escherichia coli* in subtropical freshwater. *Appl. Environ. Microbiol.* **70**, 1242–1244.

Clarridge, J. E., III 2004 Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious disease. *Clin. Microbiol. Rev.* **17**, 840–862.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., Garrity, G. M. & Tiedje, J. M. 2005 The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **1**, 33.

Higgins, J. A., Belt, K. T., Karns, J. S., Russell-Anelli, J. & Shelton, D. R. 2005 *tir* and *stx* positive *Escherichia coli* in stream waters in a metropolitan area. *Appl. Environ. Microbiol.* **71**, 2511–2519.

Kazor, C. E., Mitchell, P. M., Lee, A. M., Stokes, L. N., Loesche, W. J., Dewhirst, F. E. & Paster, B. J. 2003 Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients. *J. Clin. Microbiol.* **41**, 558–563.

Kolbert, C. P. & Persing, D. H. 1999 Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. *Curr. Opin. Microbiol.* **2**, 299–305.

Leclerc, H., Mossel, D. A. A., Edberg, S. C. & Struijk, C. B. 2001 Advances in the bacteriology of the coliforms group: their suitability as markers of microbial water safety. *Ann. Rev. Microbiol.* **55**, 201–234.

National Research Council 2004 *Indicators for Waterborne Pathogens*. The National Academies Press, Washington, DC.

O' Sullivan, L. A., Weightman, A. J. & Fry, J. C. 2002 New degenerate *Cytophaga-Flexibacter-Bacteroides*-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in River Taff epilithon. *Appl. Environ. Microbiol.* **68**, 201–210.

Pettibone, G. W. 1998 Population dynamics of *Aeromonas* spp. in an urban river watershed. *J. Appl. Microbiol.* **85**, 723–730.

Pisciotta, J. M., Rath, D. F., Stanek, P. A., Flannery, D. M. & Harwood, V. J. 2002 Marine bacteria cause false-positive results in the Colilert-18 rapid identification test for *Escherichia coli* in Florida waters. *Appl. Environ. Microbiol.* **68**, 539–544.

Rompré, A., Servais, P., Baudart, J., de-Roubin, M. -R. & Laurent, P. 2002 Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. *J. Microbiol. Meth.* **49**, 31–54.

Sacchi, C. T., Alber, D., Dull, P., Mothershed, E. A., Whitney, A. M., Barnett, G. A., Popovic, T. & Mayer, L. W. 2005 High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. *J. Clin. Microbiol.* **43**, 3734–3742.

Tanner, M. A., Goebel, B. M., Dojka, M. A. & Pace, N. R. 1998 Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* **64**, 3110–3113.

Taylor, H. 2003 Surface waters. In *The Handbook of Water and Wastewater Microbiology* (ed. D. Mara & N. Horan), Academic Press, New York. pp. 612–626.

Versalovic, J., Koeuth, T. & Lupski, J. R. 1991 Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acid Res.* **19**, 6823–6831.

Available online March 2007