Molecular techniques and data integration: investigating distribution system coliform events

S. J. Payne, M. C. Besner, J. Lavoie, C. Krentz, L. Truelstrup Hansen, M. Friedman, M. W. LeChevallier, M. Prévost and G. A. Gagnon

ABSTRACT

A retrospective analysis of coliform events in a drinking water distribution system was conducted. A total of 13 coliform isolates with specific geographic/sampling records were analysed using microbial source tracking (MST) to establish genetic (based on DNA) and phenotypic (based on substrate utilization) relatedness between isolates, and a data integration approach (geo-spatial and event analysis) to identify potential causes for their presence. Phenotypic analysis, using the API® 20E system, identified two major genera: *Enterobacter* and *Citrobacter*, showing a high degree of phenotypic similarity within each species. However genotypic analysis (using pulsed field gel electrophoresis) of the *Enterobacter* and *Citrobacter* isolates demonstrated a wide genetic diversity within each species. The data integration approach revealed the probable causes for coliform events were the installation of new mains (8 samples) and rehabilitation of existing main (1 sample). The data integration and MST approach (genotyping) suggested that these coliform events were a result of sporadic intrusion from maintenance events and not chronic biofilm re-growth. The combination of MST and the data integration approach offers utilities a potentially powerful tool for the examination and determination of coliform occurrences in a distribution system.

Key words | coliform, data integration, distribution system, drinking water, microbial source tracking, pulsed field gel electrophoresis

S. J. Payne (corresponding author)
G. A. Gagnon
Department of Civil and Resource Engineering,
Dalhousie University, 1360 Barrington Street,
Halifax NS B3J 1Z1,
Canada
Tel.: 1 902 494 3006
Fax: 1 902 494 3108
E-mail: *sjopayne@dal.ca*

M. C. Besner
M. Prévost
Génies Civil, Géologique et des Mines,
École Polytechnique de Montréal,
C.P. 6079, Succ Centre-Ville, Montréal QC,
Canada

J. Lavoie
Ville de Laval, Service de l'Environnement,
2550 boul. Industriel, Laval QC H7V 3Z4,
Canada

C. Krentz
Department of Pathology & Laboratory Medicine,
G227-2211 Wesbrook Mall, Vancouver BC V6T 2B5,
Canada

L. Truelstrup Hansen
Deparment of Process Engineering and Applied Science, Dalhousie University,
1360 Barrington Street D 321,
Halifax NS B3J 1Z1,
Canada

M. Friedman
Confluence Engineering Group,
Confluence Engineering Group LLC,
517 NE 92nd Street, Seattle WA 98115,
USA

M. W. LeChevallier
American Water,
1025 Laurel Oak Rd., P.O. Box 1770,
Voorhees NJ 08043,
USA

INTRODUCTION

Total coliform bacteria are used as indicators of microbial contamination in water distribution systems. Although there has been some debate about the validity of using total coliforms as indicators of fecal contamination (Edberg *et al.* 2000; Australian Government 2003), these bacteria can be used for operational monitoring, likely indicating a breach in treatment or in the distribution system. In the United States, total coliforms measurement

doi: 10.2166/aqua.2010.026

and monitoring are a principal component of the Total Coliform Rule regulation (USEPA 1989). The Total Coliform Rule (TCR) is currently under review and it is suspected that the approach to manage coliform bacteria will be evolving. In Canada, the Drinking Water Guidelines as set out by Health Canada recommends use of the maximum allowable concentration of *Escherichia coli* and total coliforms for drinking water monitoring (Health Canada 2006). Occurrence of coliform bacteria in water samples collected from a distribution system may therefore represent an operational and/or a regulatory issue for a water utility. The knowledge of the causes or sources of these coliform bacteria can provide clues in developing remediation and improved operational strategies.

Microbial source tracking (MST) is generally used to associate the occurrence of a microbial agent with its original reservoir (host species such as human or animal hosts). The tools used in MST provide characterization data concerning a target microorganism (Moussa & Massengale 2008). For instance, the genetic and/or phenotypic traits of the target microorganism can be used to determine if it is similar to or different than other isolates in the same environment.

Four types of MST approaches are used: genotypic library based methods, phenotypic library based methods, non-library based culture independent methods, and direct measurement methods. Genotypic library based methods include: ribotyping, pulsed field gel electrophoresis (PFGE), denaturing gradient gel electrophoresis (DGGE) and repetitive polymerase chain reaction (REP-PCR) (Simpson *et al.* 2002). Phenotypic library based methods include: antibiotic resistance analysis (ARA), carbon source utilization (Harwood *et al.* 2000) and manual commercially available enzyme-based identification kits (O'Hara 2005). Non-library based culture independent methods include techniques such as PCR followed by terminal restriction length polymorphism (T-RELP) (Field *et al.* 2003). Direct measurement methods include real-time/quantitative polymerase chain reaction (QPCR) of specific target organisms (Noble *et al.* 2003). Using MST methods in a multi-tier approach can provide a clearer description (identification and characterization) of the target organism (Noble *et al.* 2006) and ultimately determine strategies to mitigate its occurrence. Noble *et al.* (2006) used a combination of culture based methods, genetic methods (quantitative PCR), and genetic sequencing to track sources of fecal contamination in a watershed. The culture based methods revealed only the population of organisms culturable under the selected conditions and time. The quantitative PCR method was a faster detection and enumeration method (3 hours), and counted both viable (culturable) and non-viable bacteria.

MST has been applied in epidemiological investigations of foodborne microbial contamination (Giammanco *et al.* 2002; Mullane *et al.* 2007), as well as in watershed monitoring (Noble *et al.* 2006; Kelsey *et al.* 2008; Moussa & Massengale 2008). Although MST is not often applied to the analysis of drinking water it has been used to investigate waterborne outbreaks (Moyer *et al.* 1992). Moyer *et al.* (1992) used ribotyping, a genotypic method, to investigate a gastroenteritis outbreak caused by aeromonads. Ribotyping profiles enabled the investigators to genetically associate *Aeromonas* spp. found throughout the distribution system, which confirmed their suspicions that the water treatment filters had been backwashed with unchlorinated water. However, the genetic profiles of the environmental aeromonads isolated from the water treatment plant did not have the same ribotyping profiles as the clinical isolates taken from outbreak patients. The authors noted that the sampling at the water plant occurred 2–3 months after the clinical *Aeromonas* spp. outbreak had been discovered, so the water plant could not be conclusively eliminated as the potential source of *Aeromonas* spp. contamination.

As an alternative to MST, identification of the likely sources (or causes) of positive coliform samples in a distribution system is possible through the use of tools that consider available utility data. Such tools may include hydraulic modeling (McMath & Casey 2000), geographic information systems (GIS), and data integration (Fang *et al.* 1997; Fricker *et al.* 2006; Besner *et al.* 2007). In comparison to MST, these types of data integration tools do not allow for identification of the source of a contamination as being of human or fecal origin or delineate its degree of genetic or phenotypic relatedness to other isolates. However, they can be useful in identifying the operational characteristics of a system that can lead to the occurrence of coliform positive samples.

This study was completed as part of Awwa Research Foundation project 3116: Strategy to Manage and Respond to Total Coliforms and *E. coli* in the Distribution System. There were two primary objectives for this part of the study. The first objective was: to evaluate the usefulness of two MST strategies for phenotypic characterization and determining the genetic relatedness among coliform isolates obtained from a drinking water distribution system. The PFGE and API® (API 20E, BioMérieux, Marcy L'Etoile, France) methods were selected for use in this retrospective study at a workshop held by Awwa Research Foundation project 3116. The workshop highlighted the potential use of several methods including ribotyping, PFGE and several phenotyping methods. PFGE is a DNA "fingerprinting" tool that can identify genetic variability in clinical and environmental isolates of the same species of bacteria. PFGE is considered the gold standard in microbial identification as it is highly reproducible and has been proven to be more discriminatory than other MST methods (Meays *et al.* 2004). However, in some cases PFGE methods may not be discriminatory enough to distinguish between subtypes of the same serogroup (Kool *et al.* 2000). Phenotypic methods can include methods that identify a microorganism based on expressed enzymes and substrate utilization under standard conditions. API® 20E is a phenotypic identification system that uses 20 key substrates to generate a species-level identification of a coliform. The usefulness of substrate utilization profiles is somewhat limited as they assess bacteria capable of metabolism under experimental conditions, and as such do not reflect the carbon utilization profile the target organism would use *in situ* (Kirk *et al.* 2004).

The second objective was to determine if the results from the applied MST strategies and the utility data integration approach could be combined to improve the assessment of the causes of coliform events identified in a distribution system.

MATERIALS AND METHODS

Site description

The City of Laval (Québec, Canada) has a population of approximately 380,000, and is served by three drinking water treatment plants using surface water as their raw water source. Treatment processes include coagulation, flocculation, filtration, ozonation and an additional step of activated carbon filtration is used at one of the plant. Free chlorine is the residual disinfectant used in the distribution system. The drinking water distribution system is a total of 1,590 km long, and is constituted of pipes ranging in diameter from 150 to 1,200 mm (Ville de Laval 2007). Twelve fixed sampling points are used to monitor water quality in the distribution system and sampling is typically conducted from four to seven times per week depending upon the period of the year. In addition to routine monitoring, coliform analyses are conducted for sanitary release samples (samples collected to assess microbial water quality prior to the connection of new or rehabilitated mains to the existing system) and for some samples collected after consumers' complaints, if necessary.

Sixteen coliform isolates from Laval were characterized using MST analysis. These 16 isolates were selected from a preserved set of 53 isolates collected by Laval between 2001 and 2004 (including isolates collected as part of routine sampling, prior to returning new or rehabilitated mains into service, or following customer complaints). The 16 isolates were selected as they had similar API® profiles, but were not identified at the time of sampling. Although all the isolates selected for MST analysis had a temporal localization, only 13 out of 16 had a spatial localization. Two isolates only had a street name (major boulevard) (QC 3 and QC 7) and one (QC 15) did not have any information allowing geo-referencing of the event. Because both the temporal and spatial components are necessary to establish relationships between operational events in a distribution system and changes in water quality, the data integration approach was used to investigate the possible causes of the remaining 13 coliform isolates from the Laval distribution system collected from 2002 to 2004. The selected isolates are listed in Table 1 (coliform events are listed according to the type of sampling location). Seven positive coliform samples were collected as part of routine water quality sampling at dedicated sampling locations, five positive samples were collected at hydrants and finally there were four samples from an operation and maintenance activity for which the type of sampling point was not specified (probably collected from hydrants but this cannot be confirmed).

Table 1 | Results of the application of the data integration approach to identify causes of coliform occurrences. Source: Friedman *et al.* 2009. © 2009 Water Research Foundation. Reprinted with permission.

Date	Location	ID#	Type of sampling point	Cause identified with data integration approach	API® identification
2002-05-08	D	QC 9	Routine*	Unknown	*C. freundii*
2002-05-30	D	QC 8	Routine	Unknown	*C. freundii*
2003-08-07	A	QC 14	Routine	New main installation	*E. cloacae*
2003-08-08	A	QC 11	Routine	New main installation	*E. cloacae*
2003-08-11	A	QC 12	Routine	New main installation	*E. cloacae*
2003-08-12	A	QC 13	Routine	New main installation	*E. cloacae*
2003-08-19	A	QC 1	Routine	New main installation	*C. braakii*
2003-07-21	B	QC 2	Hydrant	Unknown	*C. braakii*
2003-08-26	H	QC 4	Hydrant	New main installation (*non-compliant sanitary release sample*)	*E. aminigenus 2*
2003-09-08	E	QC 5	Hydrant	New main installation *(non-compliant sanitary release sample)*	*E. cloacae*
2004-07-06	C	QC 6	Hydrant	New main installation (*non-compliant sanitary release sample*)	*E. cloacae*
2004-08-27	G	QC 16	Hydrant	Main rehabilitation (*non-compliant sanitary release sample*)	*E. cloacae*
2004-06-28	F	QC 10	Unspecified	Unknown	*E. cloacae*
2003-10-08	Street name only (major boulevard)	QC 3	Unspecified	Not assessed	
2004-06-07	Street name only (major boulevard)	QC 7	Unspecified	Not assessed	
2004-10-21	Unspecified	QC 15	Unspecified	Not assessed	

*Routine sampling points represent dedicated sampling locations used for monitoring of water quality in the distribution system, as per the monitoring program established by the water utility.

Detection and isolation of coliform bacteria

Distribution system water samples were sampled in 250 mL bottle and quenched with sodium thiosulphate; the samples were then analysed in a 100 mL Colilert® (IDEXX Laboratories Inc., Westbrook, Maine, USA) presence-absence test for total coliforms. Coliform positive samples were streaked onto MacConkey agar for confirmation. The isolation of the colonies was performed according to *Standard Methods* as described in APHA & AWWA (1998).

Isolated colonies from each site were selected for further substrate utilization analysis by API® 20E and genotypic analysis using PFGE. These isolates were kept frozen on glass beads at −20°C in the presence of 50% glycerol by the water utility since their time of collection. The quality assurance and quality control protocols mandated by the Quebec Ministry of the Environment for laboratory accreditation were used to ensure the integrity of the isolates.

Identification and biochemical characterization

Enteric gram negative isolates were identified using the API® 20E substrate utilization test kit. Isolates were plated on TSA and grown overnight at 37°C. Isolates were analyzed according to the manufacturer's instructions. The reactions were scored as positive or negative and used to generate a seven digit numeric profile, which was searched against the manufacturer's web database (API web™, BioMérieux, Marcy L'Etoile, France) to determine the most probable identity of the isolate. Two of the isolates were analyzed in duplicate (12.5% of samples), and one was analyzed in triplicate (18.75% of samples).

Pulsed field gel electrophoresis

PFGE was used to generate a DNA fingerprint for bacterial isolates. Bacteria were cultured on TSA at 37°C overnight. Cells were scraped off the plate using a sterile pipette tip and suspended in 1.5 mL of cell suspension buffer (100 mM Tris:100 mM EDTA, pH 8.0). Cell density was adjusted until the sample absorbance at 610 nm was 1.4 ± 0.1. 120 μL of the cell suspension was transferred to an eppendorf tube and mixed with 6 μL of 20 mg/mL Proteinase K solution and 120 μL of agarose solution (1% SeaKem Gold:1% SDS in TE Buffer, pH 8.0). The mixture was immediately dispensed into the well of a disposable plug mold and allowed to solidify. Agarose plugs were placed in tube containing 5 mL of cell lysis buffer (50 mM Tris:50 mM EDTA, pH 8.0; 1% Sarcosyl) and 25 μL of 20 mg/mL Proteinase K solution. Samples were incubated at 54°C for two hours. Agarose plugs were washed twice in 10 mL of pre-warmed deionized water for 10 minutes at 50°C, and three times in 10 mL of pre-warmed TE buffer for 10 minutes at 50°C. A 2 mm slice of the agarose plug was digested with *XbaI* (Roche Applied Sciences, Roche Diagnostics Corporation, Basel, Switzerland) for 2 hours and 37°C. Plug slices were washed once with 200 μL of 0.5X TBE and loaded into the wells of a 1% SeaKem Gold gel. The wells were sealed with a 1% SeaKem Gold solution and the gel was run in 0.5X TBE buffer for 19 hours at 14°C using a PFGE electrophoretic box (Chef Mapper®, Bio-Rad Laboratories Inc., Hercules, California, USA). The PFGE program used was an auto algorithm with low molecular weight of 30 kb and high molecular weight of 600 kb, and initial and final switch times of 2.16 sec and 54.17 sec respectively, at a gradient of 6 V/cm, and an included angle of 120°. *Salmonella Braenderup* H9812 (ATCC *Salmonella Braenderup* H9812 BAA-664, ATCC, Manassas, Virginia, USA) was used as a size standard and *Enterobacter cloacae* (ATCC *Enterobacter cloacae* 23355, ATCC, Manassas, Virginia, USA) was used as a control to monitor reproducibility of gel runs. The gel was stained using ethidium bromide and photographed under UV light. Two of the isolates were analyzed in duplicate.

Cluster analysis of PFGE results and band sizing

The *Salmonella Braenderup* H9812 standard was used to construct a standard curve that was used to size the unknown samples based on the distance traveled through the gel. The bands of samples with the same PFGE profiles on different gels were averaged in order to account for variation between gels.

Construction of band size groups

For each site all possible band size positions were identified. Band size groupings were constructed such that each band size fit into one discrete grouping, and a group could not include two separate bands from the same sample isolate. The lower and upper limits were less than or equal to one standard deviation from each sample isolate. In most cases the spacing between groups was at least 5 kb.

Cladograms

Each band size group was used to generate the template for similarity scoring. The presence of a band size was given a positive score of one, and the absence of a band size was given a negative score of zero. This data was entered into Systat 11 (Systat Software Inc.) and a multi-variate cluster hierarchical analysis was performed. The resulting cladogram showed percent difference in genotypic profiles for all the isolates (results not shown).

Data integration approach to identify causes of coliform events

A data integration approach including the simultaneous consideration of water quality data, network structural data, system operation and maintenance (O&M) data, information from a hydraulic model and data visualization with a geographical information system (GIS) was used to investigate the possible causes of the coliform events. The approach was specifically developed to study historical coliform data from numerous water utilities (Besner *et al.* 2007). The steps necessary for the application of the approach are briefly described below. A detailed description of the approach, the tools, and the step by step methodology used to investigate water quality problems in distribution systems have previously been described in the reports by Besner *et al.* (2007) and Martel *et al.* (2005). Available data from the water utility were collected for the period

2001–2004. These included: water quality data from the three water treatment plants and from the twelve distribution system routine sampling locations, spatio-temporal data for repairs of pipe leaks and breaks, hydrant operation, valve operation, installation of new pipes, pipe rehabilitation, hydrant flushing, customer complaints and events at the treatment plants. The hydraulic model of the distribution system (Epanet file format (Rossman 2000)) was obtained from the water utility and included all mains (no skeletonization was performed). The size of this model is about 30,000 nodes and pipes. For each coliform isolate, an hydraulic simulation was run to determine specific hydraulic paths (i.e. all possible water paths supplying the location where the positive sample was collected), eventually allowing to identify operational events taking place upstream of the sampling location when superimposing the different types of information available. A GIS street layer was also available. Data were formatted (validation of format and content) and spatially-referenced data were geocoded (assigned a geographical coordinates (x and y) for GIS visualization. The concept of data visualization and database query was used to facilitate the identification of the spatio-temporal relationship between the coliform data and the other types of data available. The interactive data analyzer tool (Besner *et al.* 2007) was used to perform multi-table (simultaneous queries in different databases) and multi-criterion (according to hydraulic, spatial, temporal and specific parameter values) database queries for each positive coliform occurrence. Results of the queries were directly viewed with the GIS software, allowing the identification of all the events that could have affected the water quality at the studied location.

RESULTS

Microbial source tracking

The identification of sixteen coliform bacterial isolates taken from eight different points in the Laval distribution system was performed using the phenotypic API® 20E and genotypic PFGE analyses. A total of eleven different substrate utilization profiles and fifteen different PFGE

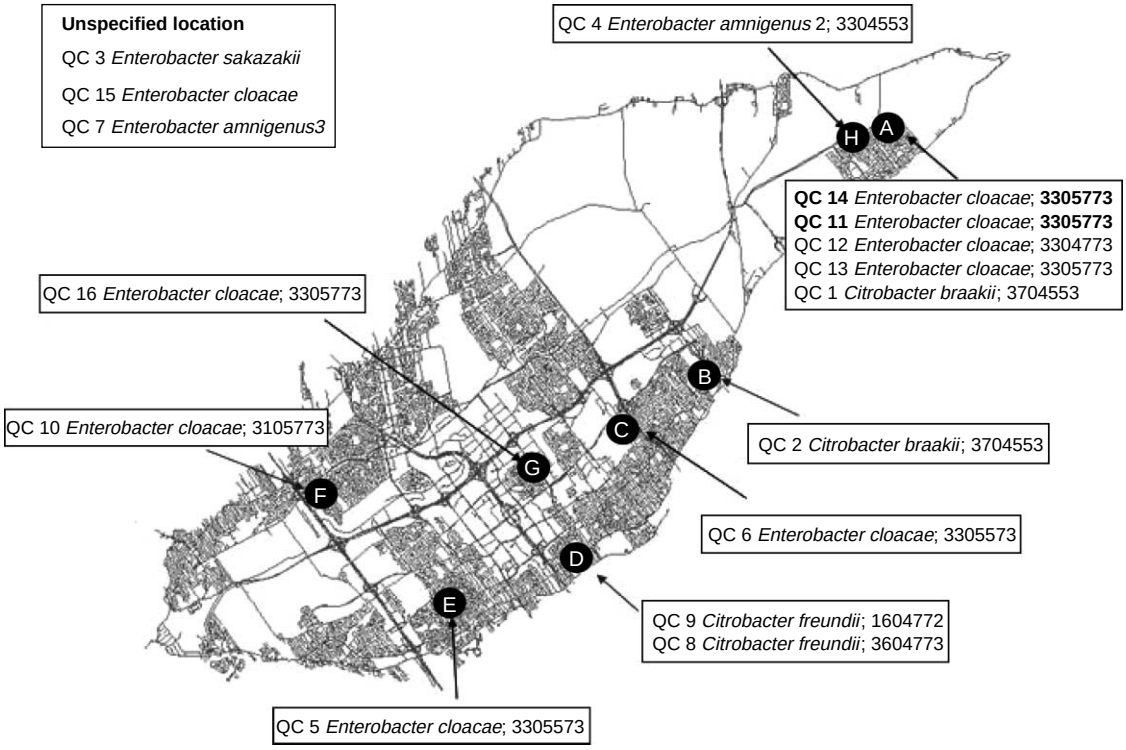

Figure 1 | Identification of the isolates from each sampling site. The API® species identification and API®profile are shown in the box. Bolded text indicates that the isolates had the same profile as one another as determined by PFGE. Source: Friedman *et al.* 2009. © 2009 Water Research Foundation. Reprinted with permission.

profiles were observed. The sampling points and the corresponding isolates are shown in Figure 1. The substrate utilization profiles are shown in Figures 2 and 3. Substrate utilization profiles showed strong phenotypic similarity among isolates; however, the genotyping by PFGE demonstrated that the isolates were genetically diverse. Analysis revealed the presence of two main genera of coliforms in the distribution system: *Enterobacter* and *Citrobacter*. Although the virulence properties of these bacteria in water have not been fully elucidated, it is important to note that *E. cloacae* and *C. freundii* are considered to be opportunistic pathogens and may cause nosocomial infections in immunocompromised individuals (Janda & Abbott 2006).

Enterobacter

Twelve isolates were identified as members of *Enterobacter*. Analysis of the *Enterobacter* isolates showed eight different substrate utilization profiles and 11 different PFGE profiles (Figure 2). The substrate utilization profile "3305573" was identified for two isolates (QC 5, and QC 6); however, these two isolates were shown to have unique PFGE profiles. The substrate utilization profile "3305773" was found for four of the isolates (QC 11, QC 13, QC 14, and QC 16); however, of these isolates only two had identical PFGE profiles (QC 11 and QC 14). In fact, the latter isolates, which were taken from the same sampling point one day apart, were the only strains with identical PFGE profiles in this study.

Citrobacter

Four isolates were identified as members of the genus *Citrobacter*. Analysis of the *Citrobacter* isolates identified three substrate utilization profiles and four distinct PFGE profiles (Figure 3). Two isolates were shown to have the substrate utilization profile 3704553; however, these two isolates had different PFGE profiles.

Discrimination of substrate utilization and PFGE profiles

Overall the substrate utilization profiles showed more similarity between the isolates than the PFGE profiles

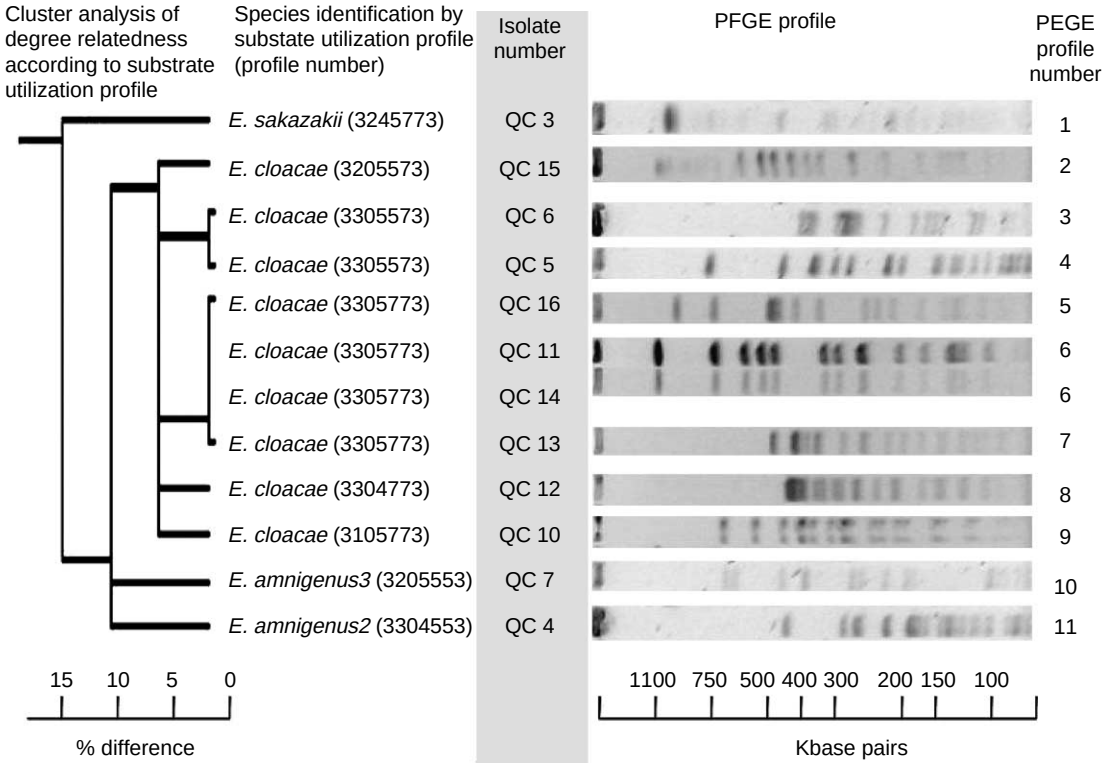

Figure 2 | Coliform isolates identified as *Enterobacter* spp. Cluster analysis shows percent difference in substrate utilization profiles. The PFGE profile number and image are also shown for the coliform isolates.

(Figure 4). Of the *Enterobacter* isolates, 50% were shown to have non-unique substrate utilization profiles (same substrate utilization profile as at least one other isolate). Only two (16.7%) of *Enterobacter* isolates shared the same PFGE profile. Likewise *Citrobacter* isolates also showed less discrimination using substrate utilization profiles as two out of four were shown to have non-unique substrate utilization profiles (same substrate utilization profile as at least one other isolate). None of the *Citrobacter* isolates shared identical PFGE profiles.

Data integration

The causes of 13 positive coliform samples were investigated using the data integration approach. Overall, the likely cause of 9 of positive coliform samples (69% of the samples) could be identified as summarized in Table 1. For the remaining 4 samples, no specific event could be identified as being the cause of the coliform occurrence from the available data. The main cause of the selected total coliform events was the installation of new mains (for 8 samples) and the rehabilitation of existing mains (for 1 sample). Four samples were found to correspond to non-compliant sanitary release samples collected prior to the connection of the new/rehabilitated mains to the existing system. Figure 5 illustrates the results obtained from the database study and queries for the 5 coliform samples collected within a 12-day period at site A. A new main installed directly upstream of the routine sampling site was sampled on July 30 and opened on August 1. Total coliforms were detected in the water starting on August 7. Because of the temporal, spatial and hydraulic proximity between the opening of the new main and the coliform events, it seems reasonable to assume that the new main is the cause of the microbial occurrences at this sampling site. Figure 6 illustrates the other type of cause identified, i.e. non-compliant sanitary release samples collected prior to connecting a new/rehabilitated main to the existing system. For these cases, the date of collection of the coliform positive sample was always found to correspond to one of the sampling dates of the "isolated" water main. Therefore in these cases, the contaminated drinking water was not distributed to the consumers.

DISCUSSION

Results from PFGE corroborate the data integration approach with the type of cause identified

The installation/rehabilitation of new mains was found to be the probable cause of the coliform events investigated in this study. Because the isolates were selected from a list that included samples collected as part of routine sampling, prior to returning new or rehabilitated mains into service, or following customer complaints, it was expected that the

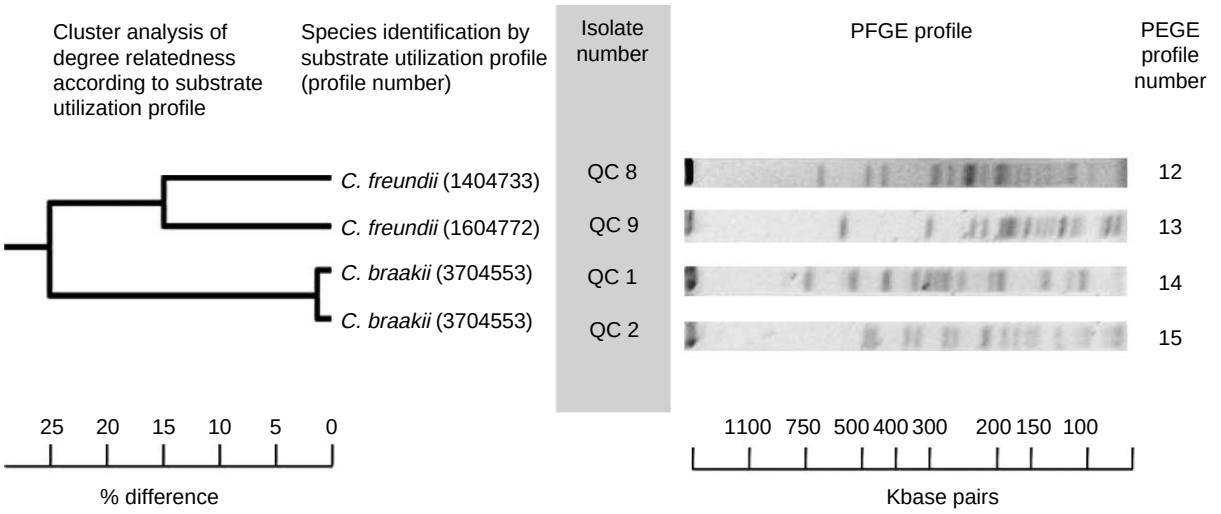

Figure 3 | Coliform isolates identified as *Citrobacter* spp. Cluster analysis shows percent difference in substrate utilization profiles. The PFGE profile number and image are also shown for the coliform isolates.

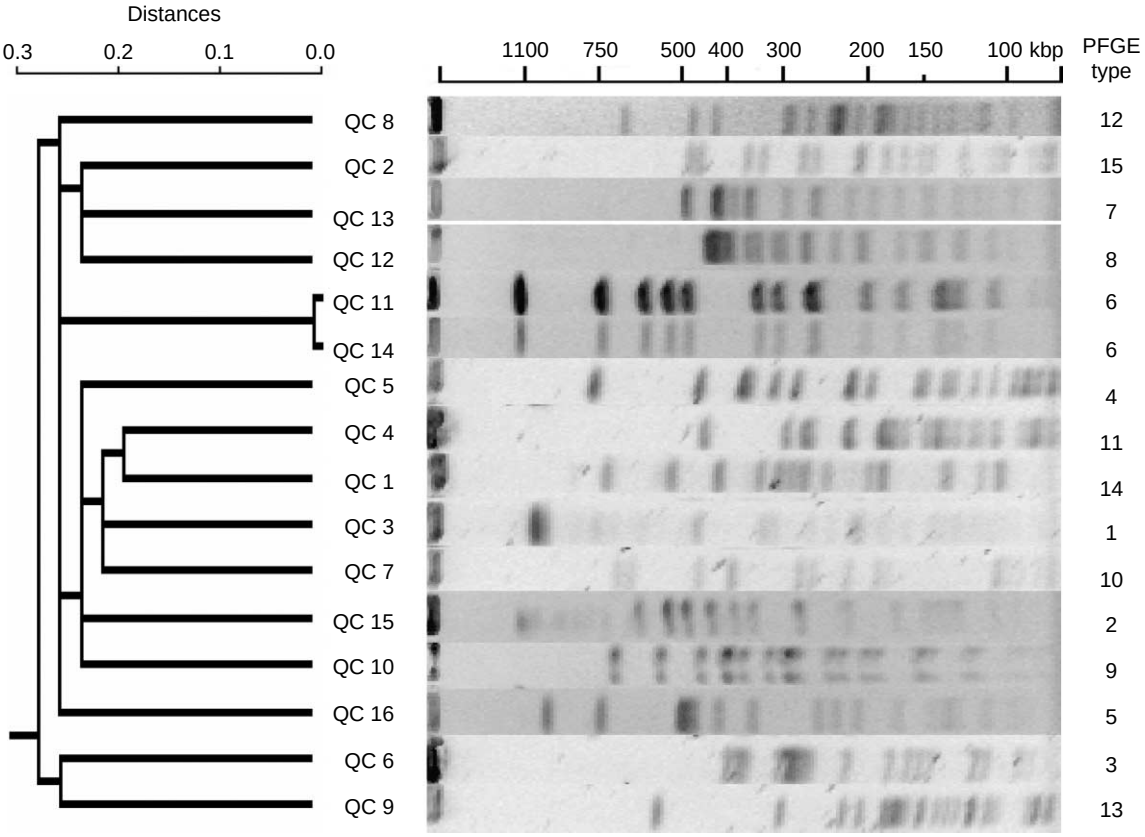

Figure 4 | Cluster analysis of the PFGE profiles obtained for *Enterobacter* and *Citrobacter* spp. isolated from the water distribution system. The cladogram shows the genotypic relationships between isolates determined by PFGE analysis. Source: Friedman *et al.* 2009. © 2009 Water Research Foundation. Reprinted with permission.

likely cause would probably be related to events taking place in the distribution system. PFGE analysis revealed that the coliform isolates identified over the study period were genetically unique (with only one exception). The microbial source tracking approach demonstrated that there was only one identified occurrence of isolates with the same PFGE and phenotypic profile (QC 11 and 14) during the study period. The genetic diversity of the coliforms found in the Laval distribution system is believed to be likely a reflection of the dynamic microbial ecology reported in distribution systems. Similar studies have identified the occurrence of *Aeromonas* spp. (Chauret *et al.* 2001), *Legionella* and *Mycobacterium* (Pryor *et al.* 2004) and *E. coli* (Hrudey *et al.* 2003) in distribution systems. The cause of the microbial occurrence can be connected to chronic biofilms and/or intrusion events, such as distribution systems operations (e.g. water main rehabilitation). This study contributes strong evidence that distribution system operations contributed to the microbial diversity presented in these data.

PFGE provides more information than substrate utilization profiles

Our results showed PFGE to be discriminatory between two isolates identified as the same species using substrate utilization tests, because its sensitivity to genetic differences between bacterial genomes. Similar results were obtained by Edberg *et al.* (1994) who used PFGE to determine if *Enterobacter cloacae* isolates from a distribution system (i) were all identical or different, (ii) were identical from isolates from the source water and (iii) were the same or different to isolates collected from hospitals at the same period. The authors found that while there was only one clone of *E. cloacae* in the distribution system, there was no homogeneity between source and distribution isolates and

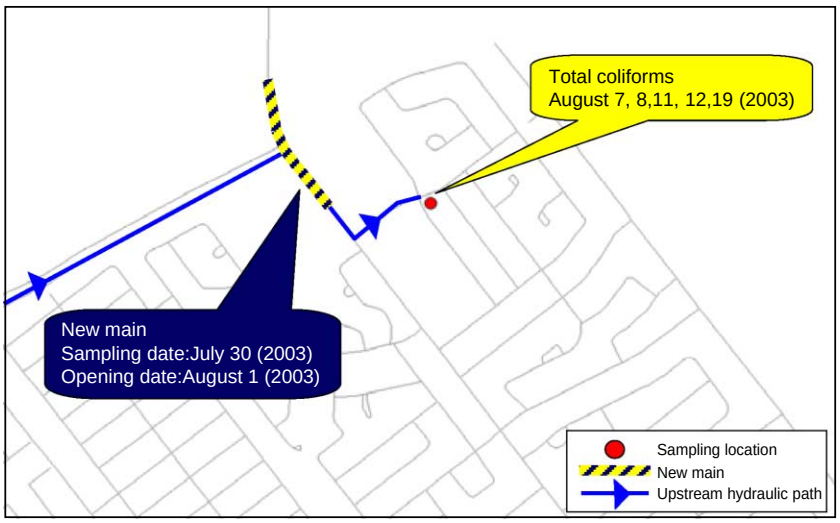

Figure 5 | Installation of a new main identified as the likely cause of the presence of 5 (3 unrelated [QC 1, 12, 13] and 2 related [QC 11, 14]) positive coliform samples over a 12 day period at a routine sampling site. Source: Friedman *et al.* 2009. © 2009 Water Research Foundation. Reprinted with permission.

between clinical and distribution isolates. The PFGE analysis allowed the authors to determine that a regrowth bloom was taking place in the distribution system (Edberg *et al.* 1994).

PFGE has also been used in epidemiological applications to determine the source or relatedness of disease clusters. Giammanco *et al.* (2002) used PFGE to characterize strains of *E. coli* O157:H7 that were isolated from cattle, food and haemolytic uremic syndrome patients to determine relatedness. The PFGE analysis was able to identify 54 distinct profiles from the 57 strains. Ultimately it was shown that epidemiologically related strains had an identical or similar characterization profiles, and the epidemiologically un-related isolates showed distinctly different patterns (Giammanco *et al.* 2002). Likewise, Ho *et al.* (2007) used PFGE to obtain three distinct profiles of community associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA), and found a geographic association with each of the strains as well as the strain transmission in family groups.

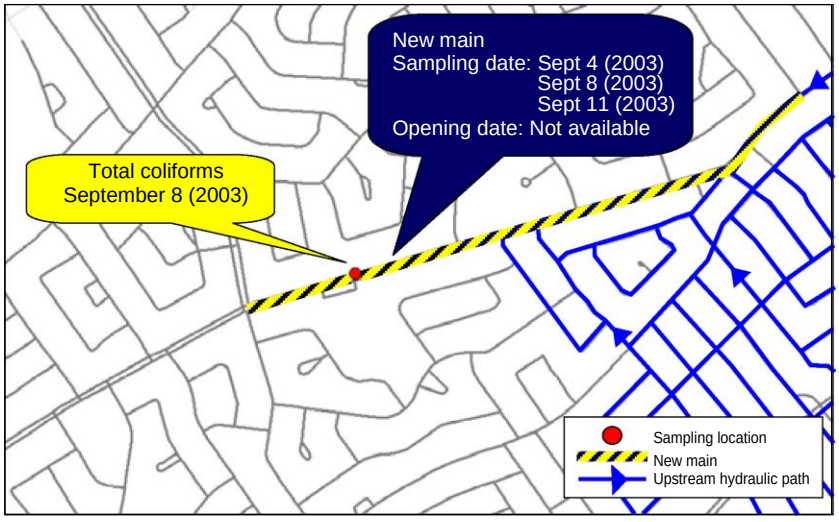

Figure 6 | Positive total coliform sample identified as a non-compliant sanitary release sample collected prior to connecting a new main to the existing system. Source: Friedman *et al.* 2009. © 2009 Water Research Foundation. Reprinted with permission.

It is hypothesized that a high degree of genetic relatedness between coliform isolates could suggest a chronic regrowth issue; whereas a low degree of genetic relatedness could suggest isolated intrusion events. However this particular hypothesis would require further study under controlled laboratory conditions.

Substrate utilization profiles are not an appropriate method for microbial source tracking

Substrate utilization profiles were generated by using a limited number of substrates to identify an isolate based on its ability to metabolize a particular set of substrates under controlled laboratory settings. These individual substrate profiles were then used to indicate the degree of relatedness among isolates. It should, however, be noted that substrate utilization profiles obtained under most standard experimental conditions may not necessarily be reflective of the substrates utilized by the organism in the distribution system (*in situ*) due to differences in the nutrient content, temperature, presence of competing organisms, etc. (Kirk *et al.* 2004). The use of a more complex substrate utlilisation system (with 96 substrates) was explored as a part of Awwa Research Foundation project 3116 (Friedman *et al.* 2009). In that project, over 80 drinking water isolates were studied and it was found that the API® system provided the most reliable data among the phenotypic systems tested. However in comparison to PFGE, API® provided less specificity.

The type and origin of the isolates used to generate the manufacturer's species identification database may be another limitation of the substrate utilization system. Expansion of species identification databases to represent not only clinical strains but also environmental strains of bacteria would strengthen its position as a water quality monitoring tool. Ultimately, substrate utilization profiles may be more useful for differentiating between rather than identifying environmental species (Truu *et al.* 1999).

In spite of these limitations the substrate utilization profile is a fast, easy to use tool when examining bacterial isolates from a coliform event with the goal of determining whether the coliform strains are pathogenic or benign. Substrate utilization can also be used to confirm the identification of positive coliform results obtained from tests such as Colilert® (Chao 2006). For example, using the Colilert® method an ONPG positive and MUG negative isolate would be presumed to be a total coliform but not an *E. coli*. However, this total coliform isolate may in fact be a ß-D-Glucuronidase activity negative (MUG negative) *E. coli* strain. These MUG negative *E. coli* strains represent a small but important fraction of the *E. coli* population (Chang *et al.* 1989; Coyne & Schuler 1994). A substrate utilization profile could then be used to characterize the total coliform isolate and identify it as *E. coli*, in spite of the lack of MUG utilization.

Coliform events and maintenance activities

The data analysis showed that four of the investigated coliform isolates (for which a cause could be identified) were in fact taken from non-compliant sanitary release samples collected prior to the connection of the new/rehabilitated mains to the existing system. These coliform positive and therefore non-compliant sanitary release samples may have resulted from the introduction of soil/groundwater or other debris prior/during construction into the new/rehabilitated main, inadequate flushing and disinfection of the main, etc. Although the rate of sanitary release failures is not recorded by the water utility, such a situation is not unique to the city of Laval. The Philadelphia Water Department reported a sanitary release failure rate of 30% for the 1991–1993 period with reasons for failure including high turbidity, high HPC, total coliforms and background growth in coliform test (Burlingame & Neukrug 1993). Pierson *et al.* (2001) reported a release to service failure rate for new/rehabilitated mains for North West Water of 7% over a 1-month period. Water quality criteria included total/faecal coliforms, turbidity, colour and taste and odor. In Laval, each time a sanitary release sample is collected, the utility also collects a water sample in the adjacent existing distribution system for comparison. Total coliforms were never detected in those samples, indicating a problem with the new/rehabilitated infrastructure. In one of the investigated cases, ten water samples were collected over a 3-month period before the microbial water quality became acceptable to allow for the connection of the new main to the existing system. However, this was an atypical case for the water utility.

Recently serviced mains may also result in water quality problems in distributed water. The five coliform positive water samples collected within a 12-day period at a routine sampling location were likely related to the servicing of a new main supplying the area. In a previous study conducted in the same distribution system, Besner *et al.* (2005) found that of 88 coliform positive events taking place during the 1997–2000 period, 15 of the positive coliform occurrences had a high probability of association with distribution system maintenance work (pipe flushing, low pressure events, valve and hydrant operation). New or rehabilitated water mains did not come out as a likely cause of coliforms at that time as only routine water quality samples were included in that study.

Maintenance activities such as repairs, rehabilitation, and installation of new mains may present several opportunities for distribution system contamination because of the loss of physical integrity. Recently, Nygård *et al.* (2007) found that water consumers exposed to low pressure episodes had significantly increased risk of gastrointestinal illness (risk ration 1.58) compared to unexposed consumers. These results were based on the study of exposed and unexposed households during 88 low pressure episodes. Mains breaks or leaks accounted for 63% of the registered episodes; change of equipment accounted for 26% of the episodes; various causes such as pipe cleaning, adjacent construction work and defective valves accounted for the remaining 11% of the episodes. In the Nygård (2007) study, water samples were only available from 18 of the episodes, and only one sample was found to be positive for *E. coli*. Such results stress the importance of pipe flushing, disinfection (according to AWWA Standard C651) and water quality testing, which are the recommended methods for controlling contamination following repair/rehabilitation/installation of water mains (Pierson *et al.* 2001).

CONCLUSIONS

Results from this study showed that a genotyping method such as PFGE was useful in obtaining a clearer indication of relatedness amongst isolates after the initial tentative special identification using substrate utilization profiles. Based on the degree of genetic relatedness, the coliforms sampled in the Laval distribution system were deemed to be unrelated with the exception two strains from the same site. Integration of the MST methods and the utility data mining approach led to the conclusion that the occurrence of coliforms in the Laval system could attributed to repair and maintenance events. These repair and maintenance events can lead to point-source introductions of coliforms through the physical breaches in distribution system integrity.

The combination of MST and data integration tools can ultimately be used to enhance quality control strategies that are implemented by water treatment utilities. It is suspected that an improved understanding of how microbial events are linked will refine or even redefine preventative and/or corrective actions taken by utilities to address coliform occurrences.

ACKNOWLEDGEMENTS

This project has been funded by the American Water Works Association Research Foundation (Awwa Research Foundation Project 3116) and was made possible by the cooperation of the City of Laval. The authors would like to acknowledge Jeff Rosen and other members of the AwwaRF project team.

REFERENCES

APHA & AWWA 1998 *Standard Methods for the Examination of Water and Wastewater*, 20th edition. American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC.

Australian Government 2003 *Review of Coliforms as Microbial Indicators of Drinking Water Quality*. National Health and Medical Research Council (NHMRC) Publications, Australia.

Besner, M.-C., Gauthier, V., Trépanier, M., Leclair, M. & Prévost, M. 2005 Interactive analyzer for understanding water quality problems in distribution systems. *Urban Water J.* **2**(2), 93–105.

Besner, M. C., Gauthier, V., Trépanier, M., Martel, K. & Prévost, M. 2007 Assessing the effect of distribution system O&M on water quality. *J. AWWA* **99**(11), 77–91.

Burlingame, G. A. & Neukrug, H. M. 1993 Developing proper sanitation requirements and procedures for water main disinfection. *American Water Works Association-Annual Conference, San Antonio, Texas*.

Chang, G. W., Brill, J. & Lum, R. 1989 Proportion of ß-D-Glucuronidase-negative *Escherichia coli* in human fecal samples. *Appl. Environ. Microbiol.* **55**(2), 335–339.

Chao, W. L. 2006 Evaluation of Colilert-18 for the detection of coliforms and *Escherichia coli* in tropical fresh water. *Lett. Appl. Microbiol.* **42**, 115–120.

Chauret, C., Volk, C., Creason, R., Jarosh, J., Robinson, J. & Warnes, C. 2001 Detection of *Aeromonas hydrophila* in a drinking-water distribution system: a field and pilot study. *Can. J. Microbiol.* **47**(8), 782–786.

Coyne, M. S. & Schuler, J. C. 1994 Frequency of MUG negative *Escherichia coli* in Kentucky groundwater samples. *J. Environ. Qual.* **23**(1), 126–129.

Edberg, S. C., Patterson, J. E. & Smith, D. B. 1994 Differentiation of distribution systems, source water, and clinical coliforms by DNA analysis. *J. Clin. Microbiol.* **32**(1), 139–142.

Edberg, S. C., Rice, E. W., Karlin, R. J. & Allen, M. J. 2000 *Escherichia coli*: the best biological drinking water indicator for public health protection. *J. Appl. Microbiol.* **88**, 106S–116S.

Fang, M., Lin, L. & Principe, M. A. 1997 Development of an enhanced water quality monitoring program for coliform control. *American Water Works Association-Water Quality Technology Conference, Denver, Colorado, USA*.

Field, K. G., Chern, E. C., Dick, L. K., Fuhrman, J., Griffith, J., Holden, P. A., LaMontagne, M. G., Le, J., Olson, B. & Simonich, M. T. 2003 A comparative study of culture-independent library-independent genotypic methods of faecal source tracking. *J. Water Health* **1**(4), 181–194.

Fricker, C. R., Pryor, M., Powell, R. M., Mooren, D. & Warden, P. 2006 Investigating the source of coliforms in water distribution systems. *American Water Works Association-Annual Conference, San Antonio, Texas*.

Friedman, M., Hanson, A., Dewis, K., Kirmeyer, G., Gagnon, G., Truelstrup Hansen, L., Krentz, C., Mosher, M., Payne, S. J., Rosen, J., Hargy, T., Sobrinho, J., Besner, M. & Prévost, M. 2009 *Strategies to Manage and Respond to Total Coliforms and E. coli in Drinking Water Distribution Systems*. American Water Works Association Research Foundation, Denver, Colorado.

Giammanco, G. M., Pignato, S., Grimont, F., Grimont, P. A. D., Caprioli, A., Morabito, S. & Giammanco, G. 2002 Characterization of Shiga toxin-producing *Escherichia coli* O157:H7 isolated in Italy and France. *J. Clin. Microbiol.* **40**(12), 4619–4624.

Harwood, V. J., Whitlock, J. & Withington, V. H. 2000 Classification of the antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of faecal contamination in subtropical Florida waters. *Appl. Environ. Microbiol.* **66**(9), 3698–3704.

Health Canada 2006 *Guidelines for Canadian Drinking Water Quality-Summary Table*. http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum_guide-res_recom/micro_e.html#31

Ho, P.-L., Cheung, C., Mak, G. C., Tse, C. W. S., Ng, T.-K., Cheung, C. H. Y., Que, T.-L., Lam, R., Lai, R. W. M., Yung, R. W. H. & Yuen, K.-Y. 2007 Molecular epidemiology and household transmission of community-associated methicillin-resistant *Staphylococcus aureus* in Hong Kong. *Diagn. Microbiol. Infect. Dis.* **57**(2), 145–151.

Hrudey, S. E., Payment, P., Huck, P. M., Gillham, R. W. & Hrudey, E. J. 2003 A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. *Water Sci. Technol.* **47**(3), 7–14.

Janda, M. J. & Abbott, S. L. 2006 *The Enterobacteria*. ASM Press, Washington, DC.

Kelsey, R. H., Webster, L. F., Kenny, D. J., Stewart, J. R. & Scott, G. I. 2008 Spatial and temporal variability of ribotyping results at a small watershed in South Carolina. *Water Res.* **42**(8–9), 2220–2228.

Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H. & Trevors, J. T. 2004 Methods of studying soil microbial diversity. *J. Microbiol. Methods* **58**(2), 169–188.

Kool, J. L., Buchholz, U., Peterson, C., Brown, E. W., Benson, R. F., Pruckler, J. M., Fields, B. S., Sturgeon, J., Lehnkering, E., Cordova, R., Mascola, L. M. & Butler, J. C. 2000 Strengths and limitations of molecular subtyping in a community outbreak of Legionnaires' disease. *Epidemiol. Infect.* **125**(3), 599–608.

Martel, K., Hanson, A., Kirmeyer, G. J., Besner, M. C., Carrière, A., Prévost, M., Lynggaard-Jensen, A. & Bazzurro, N. 2005 *Data Integration for Water Quality Management*. American Water Works Association Research Foundation, Denver, Colorado.

McMath, S. M. & Casey, R. J. 2000 The use of network modelling techniques to resolve water quality issues in distribution systems. *American Water Works Association-Water Quality Technology Conference, Salt Lake City, Utah*.

Meays, C. L., Broersma, K., Nordin, R. & Mazumder, A. 2004 Source tracking fecal bacteria in water: a critical review of current methods. *J. Environ. Manage.* **73**(1), 71–79.

Moussa, S. H. & Massengale, R. D. 2008 Identification of the sources of *Escherichia coli* in a watershed using carbon-utilization patterns and composite data sets. *J. Water Health* **6**(2), 197–208.

Moyer, N. P., Martinetti Luccini, G., Holcomb, L. A., Hall, N. H. & Altwegg, M. 1992 Application of Ribotyping for differentiating *Aeromonads* isolated from clinical and environmental sources. *Appl. Environ. Microbiol.* **58**(6), 1940–1944.

Mullane, N. R., Whyte, P., Wall, P. G., Quinn, T. & Fanning, S. 2007 Application of pulsed-field gel electrophoresis to characterize and trace the prevalence of *Enterobacter sakazakii* in an infant formula processing facility. *Int. J. Food Microbiol.* **116**(1), 73–81.

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

Noble, R., Griffith, J., Blackwood, A., Fuhrman, J., Gregory, J., Hernandez, X., Liang, X., Bera, A. & Schiff, K. 2006 Multitiered approach using quantitative PCR to track sources of fecal pollution affecting Santa Monica Bay, California. *Appl. Environ. Microbiol.* **72**(2), 1604–1612.

Nygård, K., Wahl, E., Krogh, T., Tveit, O. A., Bøhleng, E., Tverdal, A. & Aavitsland, P. 2007 Breaks and maintenance work in the water distribution systems and gastrointestinal illness: a cohort study. *Int. J. Epidemiol.* **36**(4), 873–880.

O'Hara, C. M. 2005 Manual and automated instrumentation for identification of *Enterobacteriaceae* and other aerobic gram-negative bacilli. *Clin. Microbiol. Rev.* **18**(1), 147–162.

Pierson, G., Martel, K., Hill, A., Burlingame, G. & Godfree, A. 2001 *Pratices to Prevent Microbiological Contamination of Water Mains*. American Water Works Association Research Foundation, Denver, Colorado.

Pryor, M., Springthorpe, S., Riffard, S., Brooks, T., Huo, Y., Davis, G. & Sattar, S. A. 2004 Investigation of opportunistic pathogens in municipal drinking water under different supply and treatment regimes. *Water Sci. Technol.* **50**(1), 83–90.

Rossman, L. A. 2000 *EPANET 2. Users Manual*. United States Environmental Protection Agency (USEPA), Cincinnati, Ohio, USA.

Simpson, J., Santo Domingo, J. & Reasoner, D. 2002 Microbial source tracking: state of the science. *Environ. Sci. Technol.* **36**(24), 5279–5288.

Truu, J., Talpsep, E., Heinaru, E., Stottmeister, U., Wand, H. & Heinaru, A. 1999 Comparison of API 20NE and Biolog GN identification systems assessed by techniques of multivariate analyses. *J. Microbiol. Methods* **36**(3), 193–201.

USEPA 1989 *Total coliforms. Final rule*. Federal Register.

Ville de Laval 2007 *Usines de production d'eau potable*. www.ville.laval.qc.ca

First received 3 March 2009; accepted in revised form 27 January 2010