Bacterial composition in a metropolitan drinking water distribution system utilizing different source waters

Vicente Gomez-Alvarez, Ben W. Humrighouse, Randy P. Revetta and Jorge W. Santo Domingo

ABSTRACT

We investigated the bacterial composition of water samples from two service areas within a drinking water distribution system (DWDS), each associated with a different primary source of water (groundwater, GW; surface water, SW) and different treatment process. Community analysis based on 16S rRNA gene clone libraries indicated that Actinobacteria (*Mycobacterium* spp.) and α-Proteobacteria represented nearly 43 and 38% of the total sequences, respectively. Sequences closely related to *Legionella*, *Pseudomonas*, and *Vibrio* spp. were also identified. In spite of the high number of sequences (71%) shared in both areas, multivariable analysis revealed significant differences between the GW and SW areas. While the dominant phylotypes where not significantly contributing in the ordination of samples, the populations associated with the core of phylotypes (1–10% in each sample) significantly contributed to the differences between both service areas. Diversity indices indicate that the microbial community inhabiting the SW area is more diverse and contains more distantly related species coexisting with local assemblages as compared with the GW area. The bacterial community structure of SW and GW service areas were dissimilar, suggesting that their respective source water and/or water quality parameters shaped by the treatment processes may contribute to the differences in community structure observed.

Key words | bacterial diversity, drinking water bacteria, 16S rRNA gene sequencing

Vicente Gomez-Alvarez
Randy P. Revetta
Jorge W. Santo Domingo (corresponding author)
United States Environmental Protection Agency,
Cincinnati,
Ohio,
USA
E-mail: *santodomingo.jorge@epa.gov*

Ben W. Humrighouse
Department of Civil and Environmental
   Engineering,
University of Cincinnati,
Cincinnati,
Ohio,
USA
and
US Centers for Disease Control and Prevention,
1600 Clifton Road, Mail Stop D11,
Atlanta,
GA 30333,
USA

INTRODUCTION

Clear evidence of a link between drinking water quality and waterborne outbreaks was first documented during the London cholera outbreak more than 150 years ago (Snow 1849). Despite considerable improvements in treatment processes and disinfection practices, outbreaks associated with drinking water have been consistently reported in the United States (Craun *et al.* 2010). As a result, effective monitoring of microbial contamination in drinking water distribution systems (DWDS) is critical to reduce health risks, particularly for those populations with immunocompromised systems, such as children and the elderly (Figueras & Borrego 2010). Culture-based methods are primarily used to assess the microbial quality of drinking water, but these assays are selective in nature, providing a limited view of relevant issues, such as: (1) intrinsic

microbial diversity within DWDS (Eichler *et al.* 2006); (2) potential interactions between the different members that might enhance the survival of pathogens, for example, amoeba and Legionella (Lau & Ashbolt 2009); (3) presence of antibiotic resistance genes (Gomez-Alvarez *et al.* 2012a); and (4) bacteria responsible for deterioration of distribution system infrastructure (Gomez-Alvarez *et al.* 2012b).

The implementation of molecular tools (e.g., 16S rRNA gene-based methods) to study the bacterial community structure has provided a better understanding of the phylogenetic affiliation and overall microbial diversity in DWDS (Eichler *et al.* 2006; Hong *et al.* 2010; Pinto *et al.* 2012). More recently, advanced metagenomic surveys have documented that DWDS support a complex microbial network (Gomez-Alvarez *et al.* 2012a). Therefore, a more complete

characterization of DWDS microbial community structure using molecular tools is critical to address public health research questions (e.g., conditions promoting the emergence of pathogens), which may be more difficult to answer using traditional methods.

This study reports on the characterization of the bulk phase water microbial community, based on molecular analyses of 16S rRNA gene sequences, from multiple sampling sites within the Greater Cincinnati Water Works (GCWW) distribution system. The GCWW service area is divided in two main service areas (Figure 1), each associated with their respective source of water (i.e., groundwater, GW and surface water, SW) and treatment facility (Figure 2). Since water treatment processes (e.g., filtration) play an important role in shaping the bacterial community in the distribution system (Pinto *et al.* 2012), we hypothesize that the bulk water at each service area will harbor distinct and diverse bacterial communities as well as unique water chemical properties. This study provides much needed information regarding bacterial community diversity influenced by different raw water sources and different water treatment technologies. It also provides a window into the bacterial diversity at different points within the distribution systems.

MATERIALS AND METHODS

Study sites and sample collection

The GCWW distribution system obtains raw water from the Ohio River and the Great Miami Aquifer. SW from the Ohio River is treated at the Miller treatment plant and supplies about 88% of drinking water, while the Bolton treatment plant treats groundwater from 12 wells in the Great Miami Aquifer and supplies about 12% of drinking water (GCWW 2006). The treatment processes at these plants are similar concerning conventional disinfectant treatment (i.e., chlorination), although the Bolton treatment plant (i.e., GW) utilizes a softening process (Figure 2(a)) and the Miller treatment plant (i.e., SW) uses a granular activated carbon filtration step prior to disinfection (Figure 2(b)).
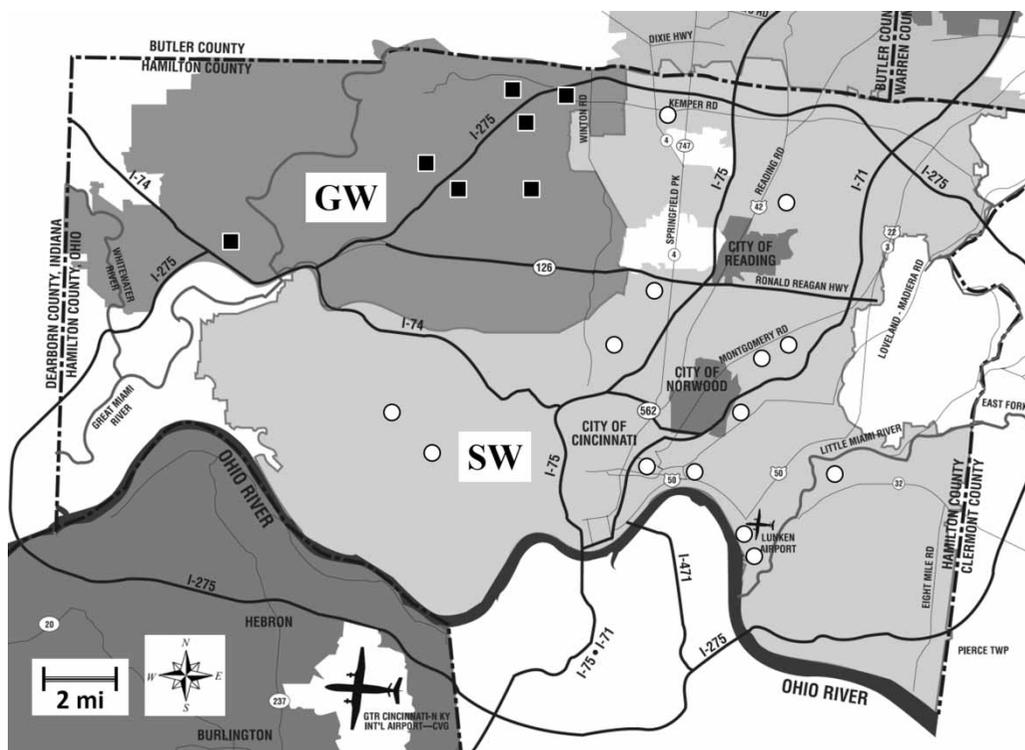

Figure 1 | A simplified service area map shows the two service areas associated with groundwater (GW) and surface water (SW) within the distribution system in Hamilton County (Ohio, USA). Figures represent sampling sites at each service area: GW (■) and SW (○). Modified map from GCWW (2006).
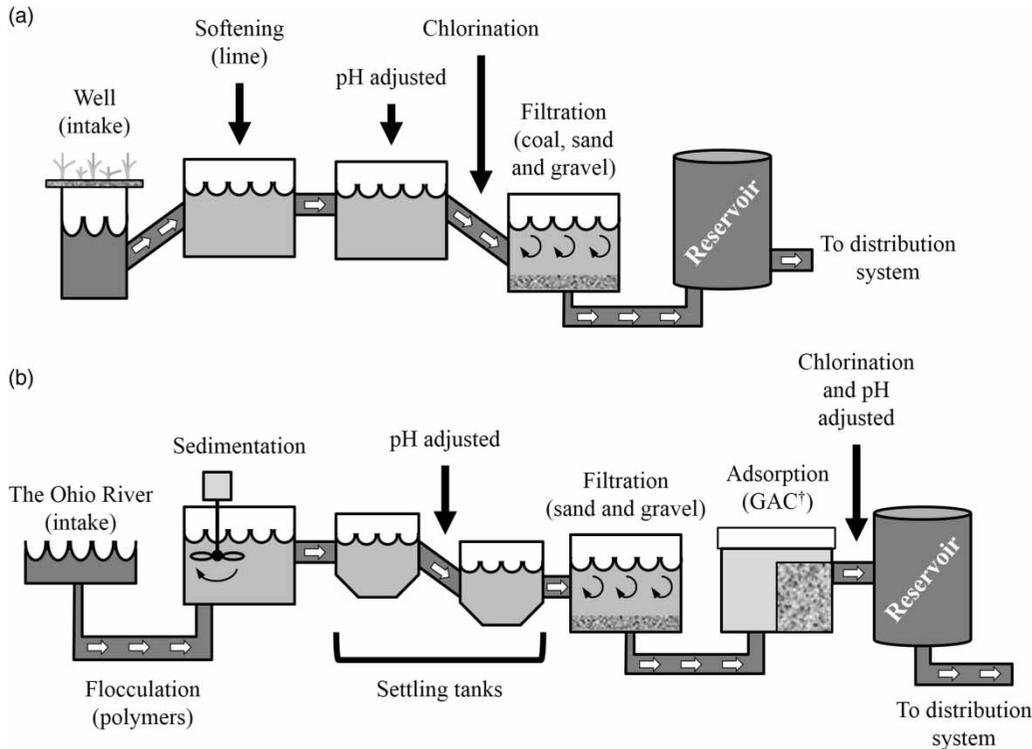
Figure 2 | Water treatment process at the (a) Bolton and (b) Miller treatment plant (adapted from GCWW (2006). †GAC, granular activated carbon.

Water samples ($n = 30$) were obtained directly from faucet heads located in 19 field sites and two houses within the DWDS of the Cincinnati metropolitan area (Figure 1). Several sites were collected more than once to determine how variable community composition could be over time at the same location. Water quality data, including total and free chlorine residual, pH, and temperature, were obtained for each sampling site (Table 1). Faucet heads were sprayed with 70% ethanol prior to sampling and run with the cold valve completely open for 5 minutes. Water samples (5 L) were collected from each sample location using sterile polypropylene bottles (Nalgene, Rochester, NY), transported to the laboratory on ice, and filtered through polycarbonate membranes (47 mm diameter, 0.22 μm pore size; GE Osmonics, Minnetonka, MN) within 3 hours of collection. Membranes were aseptically

Table 1 | Water quality values (±SE) associated with the samples

Characteristics	GW ($n = 8$)	SW ($n = 21$)	P-value
System			
Disinfectant	Chlorine	Chlorine	–
Water source	Groundwater	Surface	–
Parameters[a]			
pH	9.25 (0.04)	8.70 (0.02)	< 0.0001
Temperature [°C]	15.71 (2.32)	20.64 (1.14)	NS
Cl_2 residual, total [ppm]	1.22 (0.08)	0.88 (0.05)	< 0.01
Cl_2 residual, free [ppm]	1.14 (0.09)	0.82 (0.05)	< 0.01

[a]Average results from measurements taken August 2006 to January 2007 from the Cincinnati distribution system.
NS, not significant.

folded in half, placed in 2 mL tubes, and stored at $-80\,°C$ until DNA extractions were performed.

DNA extraction, clone libraries, and sequence analysis

Total DNA was extracted from each membrane using Ultra-Clean Soil DNA kit following the manufacturer's instructions (MoBio Laboratories, Solana Beach, CA). Primers Eub-8f (5′-AGAGTTTGATCCTGGCTCAG-3′) and 787r (5′-GGACTACCAGGGTATCTAAT-3′) were used to amplify the 16S rRNA gene as described by Santo Domingo et al. (2011). Specifically, the PCR assays consisted of 50 μL and contained 5 U of Ex Taq DNA polymerase (TaKaRa Mirus Bio Corp., Madison, WI), 10× Buffer (5 mL), dNTP mix (4 μL), primers (75 picomoles each), DNA template (2 μL of DNA extract), and Ultra Pure water (32.75 μL). The thermal cycler conditions used were an initial denaturation step of 3 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 56 °C, and 1 min at 72 °C, and a final extension step of 7 min at 72 °C. Aliquots of the PCR products were screened using 1% agarose gel electrophoresis (Fisher Scientific, Pittsburgh, PA). GelStarTM was used as the nucleic acid staining dye (BioWhittaker Molecular Applications, Rockland, ME). Replicate PCR products were cleaned with a QIAquick PCR purification kit (QIAGEN, Valencia, CA) and cloned into the pCR4.1 TOPO TA vector following the manufacturer's instructions (Invitrogen™, Carlsbad, CA). Random colonies were screened for PCR products using M13 primers and PCR products of expected size were sequenced in both directions on an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). Sequence assembly was performed using Sequencher software (Gene Codes Corp., Ann Arbor, MI). Chimeric sequences were identified using UCHIME (Edgar et al. 2011) and removed from further analyses. Sequences were deposited in GenBank with the following accession numbers: JX284426 to JX286453.

Phylogenetic analysis and taxonomic classification

Prior to analysis, clone libraries were normalized by randomization to the smallest dataset (i.e., 71 clones). Sequences were aligned and clustered with 97% sequence identity as the cut-off point for each operational taxonomic unit (OTU) using the software MOTHUR v1.22.2 (Schloss et al. 2009; http://www.mothur.org). Taxonomic identification of sequences was performed using the Classifier tool (Ribosomal Database Project II release 10.28) (Cole et al. 2009). Phylogenetic trees were constructed using the maximum likelihood method with the software MEGA v5.03 (Tamura et al. 2011).

Bacterial community diversity

Libraries of randomly selected clones were used to calculate the rarefaction and phylogenetic diversity (PD) curves, the species richness (S), species richness estimates ACE (S_{ACE}) and ChaoI, and the diversity and evenness indices of Shannon (H) and Simpson ($1/D$). Tree construction for PD analysis and index calculations was performed using the software MOTHUR (Schloss et al. 2009).

Statistical analysis of bacterial assemblages

Non-metric multidimensional scaling (nMDS) analysis based on the Bray–Curtis similarity coefficient of the transformed data ($\log[x+1]$) was used to describe the relationships among bulk water communities based on the relative distribution of OTU groups. The robustness of the nMDS ordination was evaluated using the Shepard diagram (i.e., stress test), which measures the goodness of fit of an nMDS plot, for example, <0.1 indicates a good fitting solution (Wickelmaier 2003). nMDS ordination plot was generated using the software PAST v2.14 (Hammer et al. 2001). A one-way analysis of similarities (ANOSIM) was used to identify with significance determination ($p < 0.05$) differences in the structure of community assemblages between GW and SW samples. R values near 0 indicate no difference between groups, whereas those greater than 0 indicate dissimilarities between groups (Clarke 1993). Similarity percentage (SIMPER) analysis was used to determine which species were most responsible for the differences observed between the two distribution areas (Clarke 1993). Ordination and SIMPER analyses were performed with the software PRIMER 6 (PRIMER-E Ltd, Plymouth, UK).

The phylogenetic community structure of bulk water was compared using the net relatedness index (NRI), which calculated the mean pairwise phylogenetic distance

among OTUs in each sample (Webb *et al.* 2002). A positive NRI value indicates phylogenetically clustered communities or assemblages of closely related taxa, while a negative NRI value reveals phylogenetic over-dispersed communities of distantly related species (Horner-Devine & Bohannan 2006). The NRI was calculated using the PHYLOCOM v4.2 software (http://www.phylodiversity.net/phylocom) with abundance data, null model #2 (i.e., OTUs in each sample become random draws from the phylogeny pool without replacement) and 999 permutations (Webb *et al.* 2008).

RESULTS AND DISCUSSION

Water chemistry

The water chemistry analysis indicated that the GW samples were significantly more alkaline and contained higher levels of total and free chlorine residual than the SW samples (*t*-test, $p < 0.05$) (Table 1). Differences in pH may be attributed to different unit processes and/or the point of pH adjustment along the treatment processes at each plant. For example, the alkalinity is 1.1 times higher in the GW samples and is probably due to the addition of lime as softening agent during the treatment process at the Bolton plant (Figure 2(a)), while differences in chlorine residual concentrations observed may be the result of the relationship of chlorine decay and retention time of the

water (Al-Jasser 2007). The average water temperature was not significantly different between the GW and SW samples. In general, the chemical properties at each service area may influence the bacterial composition in the DWDS.

Richness and diversity of bacterial communities

Thirty 16S rRNA libraries from 21 sites in the distribution system were constructed, representing 2,772 bacterial clones with 127 sequences identified as chimeric and removed from further analyses. Clone libraries were normalized by randomization to the smallest dataset resulting in 2,130 sequences for analysis. Taxonomic analysis revealed that the majority of the sequences in all the clone library samples were associated with the phyla Proteobacteria (45%) and Actinobacteria (42%) with additional sequences closely related to Bacteroidetes, Deinococcus-Thermus, Firmicutes, Planctomycetes and Verrucomicrobia phylum detected to a much lesser extent ($\leq 3\%$) (Figure 3(a)). These findings are consistent with previous drinking water studies performed in different geographic locations (Szewzyk *et al.* 2000; Vaerewijck *et al.* 2005; Eichler *et al.* 2006; Hong *et al.* 2010; Henne *et al.* 2012; Pinto *et al.* 2012). A total of 324 OTU ($\geq 97\%$ sequence identity cutoff) were identified, with an average of 15 and 22 phylotypes detected in GW and SW clone libraries, respectively (Table 2). In addition, 93 and 204 phylotypes were found exclusively in GW and SW libraries, respectively (Figure 3(b)). Only 27
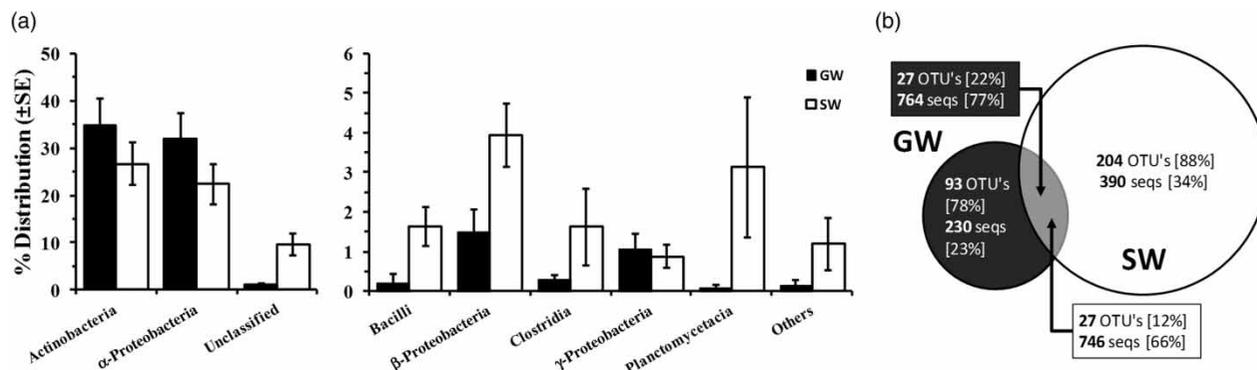

Figure 3 | (a) Distribution of bacterial classes in 16S rRNA gene clone libraries and (b) occurrence of homologous phylotypes (OTUs at ≥97% similarity) associated with groundwater (GW) and surface water (SW). The size of the circles is proportional to the number of OTUs in each treatment. Note the change in scale in the taxonomic distribution. Overlap area is not to scale. The representation of unique and share OTUs/sequences between treatments are indicated in circles and boxes, respectively. The number in brackets represents the percentage of OTUs or sequences.

Table 2 | Community diversity estimates (±SE) of the domain Bacteria. For comparison, the species frequencies were normalized to the smallest library (71 clones)

Source	Sites (libraries)	S^a	Richness estimators		Diversity		Evenness	
			S_{ACE}	ChaoI	Shannon (H)	Simpson ($1/D$)	Shannon (E_H)	Simpson (E_D)
GW	7 (14)	15 (1)	89 (22)	37 (10)	1.64 (0.12)	3.36 (0.28)	0.62 (0.03)	0.24 (0.02)
SW	14 (16)	22 (3)	146 (47)	65 (13)	2.15 (0.21)	8.35 (1.82)	0.69 (0.05)	0.33 (0.04)

[a]Richness is the total number of species in the community (i.e., the total number of independent sequences within the clone library; defined as aligned sequences with ≥97.0% similarity).

OTUs (8% of the total diversity) were shared by both service areas and these represented 71% of the sequences obtained in this study (Figure 3(b)). The majority of the shared OTUs (26 of 27) represents the dominant population in both samples and were identified as either members of the phylum Actinobacteria or Proteobacteria (Table 3).

Based on the richness estimator values (S_{ACE} and ChaoI), a more exhaustive sampling would be necessary

to obtain complete coverage of the bacterial composition in the DWDS (Table 2). The Shannon–Wiener (H) and inverse Simpson's ($1/D$) indices showed a higher diversity for the SW samples (Table 2). In addition, rarefaction curves did not level off, suggesting that further sampling would have revealed more bacterial phylotypes. The use of next generation sequencing methods will be useful at elucidating the identity of the less dominant populations,

Table 3 | Taxonomic affiliation of the top 20 most abundant Bacteria domain representatives in bulk water

GW (*n* = 14)

SW (*n* = 16)

%	OUT's I.D.	Classification	Rank	%	OTU's I.D.	Classification	Rank
35.0 (7.1)	002	*Mycobacterium*	Genus	23.2 (6.5)	002	*Mycobacterium*	Genus
12.0 (5.1)	004	α-Proteobacteria	Class	23.0 (6.3)	004	α-Proteobacteria	Class
10.1 (5.4)	043	*Methylobacterium*	Genus	8.8 (2.4)	006	*Mycobacterium*	Genus
5.6 (3.8)	018	*Microbacterium*[a]	Genus	2.6 (1.6)	156	Planctomycetaceae	Family
5.3 (2.3)	001	*Porphyrobacter*[a]	Genus	2.5 (1.1)	184	Bacteria[a]	Domain
3.6 (3.0)	006	*Mycobacterium*	Genus	2.0 (0.9)	219	β-Proteobacteria[a]	Class
3.5 (1.3)	020	*Blastomonas*[a]	Genus	1.8 (0.6)	005	*Mycobacterium*	Genus
2.1 (1.2)	044	*Sphingomonas*[a]	Genus	1.4 (0.8)	033	Hyphomicrobiaceae	Family
1.6 (0.7)	027	*Delftia*	Genus	1.1 (1.1)	007	*Mycobacterium*	Genus
1.2 (0.7)	050	*Pseudonocardia*[a]	Genus	1.1 (0.4)	132	*Bacillus*[a]	Genus
0.9 (0.5)	051	*Brevundimonas*[a]	Genus	1.1 (0.6)	009	*Sphingomonas*	Genus
0.8 (0.8)	110	α-Proteobacteria	Class	1.1 (0.4)	125	Bacteria[a]	Domain
0.8 (0.5)	005	*Mycobacterium*	Genus	0.9 (0.9)	138	*Thermus*[a]	Genus
0.6 (0.2)	009	*Sphingomonas*	Genus	0.8 (0.4)	027	*Delftia*	Genus
0.6 (0.2)	014	*Mycobacterium*	Genus	0.7 (0.4)	110	α-Proteobacteria	Class
0.6 (0.5)	003	α-Proteobacteria[a]	Class	0.6 (0.6)	214	Bacteria	Domain
0.6 (0.3)	022	Bacteria[a]	domain	0.6 (0.3)	113	*Acidovorax*	Genus
0.5 (0.3)	007	*Mycobacterium*	Genus	0.5 (0.5)	139	β-Proteobacteria[a]	Class
0.5 (0.3)	048	*Sphingobium*[a]	Genus	0.5 (0.3)	032	Sphingomonadaceae	Family
0.5 (0.2)	041	*Sphingobium*[a]	Genus	0.4 (0.3)	136	Comamonadaceae	Family

Percentage (±SE) represents the average of replicates (*n*).
[a]Represents OTUs that are differentially represented between samples from two water sources (METASTATS, $p < 0.05$).

including rare members. However, the slope of the curve created for the GW sample slightly levels off, indicating that the most predominant bacterial groups were likely identified (Figure 4(a)). PD curves show a similar trend of increasing diversity with additional sampling (Figure 4(b)). Both patterns indicate a less diverse community with closely related taxa in the GW samples as compared with SW samples.

Comparative analysis of bacterial communities

The resulting nMDS ordination plot highlighted marked bacterial community differences, revealing a noticeable variability of the bacterial core communities in the bulk water between GW and SW samples (Figure 5). The clustering of samples in the plot was further confirmed by one-way ANOSIM (Global R = 0.258, $p < 0.0001$). Twelve (out of 324) phylotypes explained ≈65% of the dissimilarity (SIMPER analysis, $p < 0.05$) within both service areas (Figure 5). Overall, the GCWW distribution system is mainly composed of two dominant phylotypes, a *Mycobacterium* related species (29%) and Rhizobiales (18%) a member of the α-Proteobacteria phylum (Figure 6). Statistical analysis indicated no significant contribution by these two dominant phylotypes in the ordination of samples (METASTATS, $p < 0.05$). Rather, it is the populations associated with the core phylotypes representing 10 to 1% of their respective communities that significantly contribute to the differences between both service areas (METASTATS, $p < 0.05$) (Figure 5 and Table 3).

The assemblage of a particular microbial community structure may be explained via three primary factors: (1) the inhabitant community in the water source, (2) the subsequent treatment processes, and (3) the creation of microhabitats within the pipe system. In this study, we did not obtain samples from the source water nor the finished water. Nonetheless, Pinto *et al.* (2012) reported that even though the source water seeded the drinking water microbiome, it is the treatment processes (i.e., filter colonization) that played a primary role in the assemblage of the DWDS bacterial community. Microhabitats (i.e., niches) may play an important role in shaping the community structure along their respective service areas by creating an opportunity for distantly related species (e.g., invasive species) to coexist with local assemblages (Webb *et al.* 2002). For example, significant differences (*t*-test, $p < 0.01$) and evidence of distantly related species coexisting with local assemblages were observed for the SW service area (Figure 7) which covered three-quarters of the entire DWDS (Figure 1). Furthermore, it may provide favorable conditions for potential opportunistic pathogens (e.g., *Pseudomonas*, *Legionella*) to establish in the DWDS.

In general, each drinking water system is unique and it is likely that predominant factors influencing downstream microbial communities will also be unique to each situation. Therefore, it is necessary for future studies to analyze the microbial community at different points within the DWDS, as well as source water and finished water prior to discharge to the DWDS. Nevertheless, the results obtained in this study demonstrate the possibility
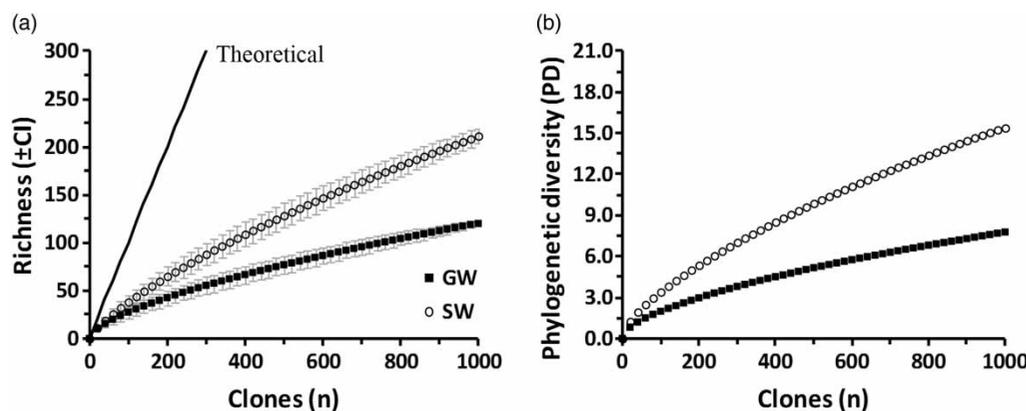

Figure 4 | Rarefaction analysis of (a) observed phylotypes (±CI) and (b) PD. Each curve represents the cumulative value of 1,000 clones from each water source. GW, groundwater; SW, surface water.
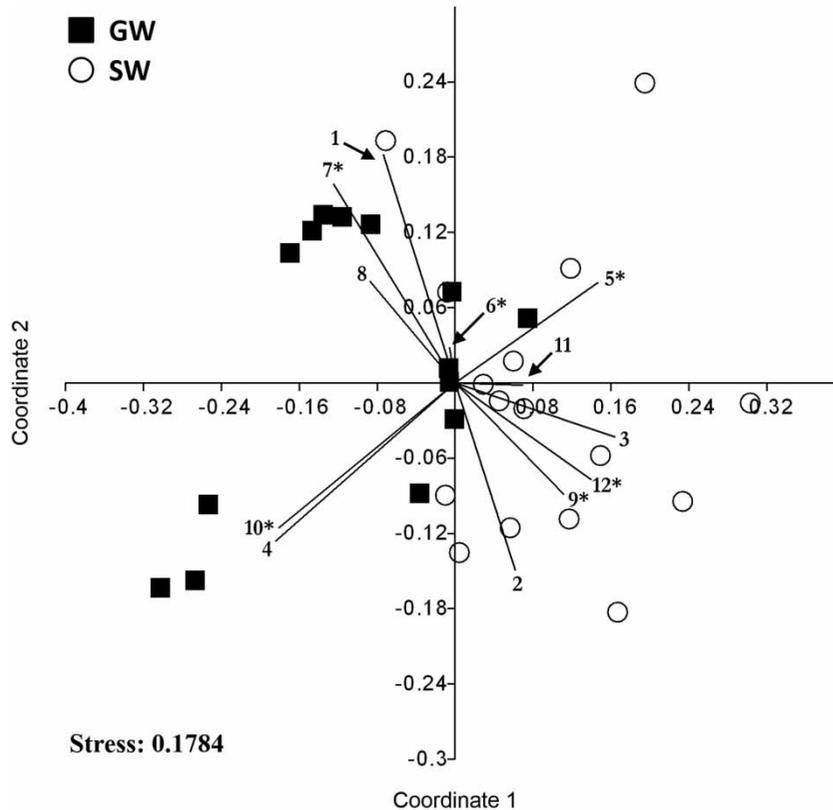
Figure 5 │ nMDS ordination plot of microbial communities ($n = 30$) from two water sources. The analysis was based on Bray–Curtis similarity coefficients calculated from the relative distribution of OTUs. Contribution of OTUs that explained ≈65% (SIMPER analysis) of the dissimilarity within all samples (one-way ANOSIM, $R = 0.258$, $p = 0.0001$) are represented by the size and direction of vectors. Taxonomic classification (all genus, except noted): 1, OTU002 *Mycobacterium*; 2, OTU004 α-Proteobacteria [class]; 3, OTU006 *Mycobacterium*; 4, OTU043 *Methylobacterium*; 5, OTU018 *Microbacterium*; 6, OTU001 *Porphyrobacter*; 7, OTU020 *Blastomonas*; 8, OTU156 Planctomycetaceae [family]; 9, OTU184 Bacteria [unclassified]; 10, OTU044 *Sphingomonas*; 11, OTU005 *Mycobacterium*; and 12, OTU219 β-Proteobacteria [class]. Source: groundwater (■), surface water (○). *Represents OTUs with significant difference in their distribution between sources (METASTAT, $p < 0.05$).

of harboring different microbial communities at different service areas within the same DWDS. These results demonstrate the importance of selecting multiple sampling sites when describing the overall bacterial community structure of distribution systems and when studying factors that may impact the survival/persistence of microorganisms of public health relevance (i.e., pathogens and nitrifiers) as well as those involved in processes that compromise the integrity of the DWDS pipes (i.e., corrosion). The frequency at which sampling should be performed to represent seasonal trends must also be addressed to develop comprehensive sampling protocols.

Bacterial taxonomic identification

Similar to this study, the classes Actinobacteria and α-Proteobacteria were identified as the numerically dominant

groups in clone and metagenome libraries of DWDS from the GCWW (Santo Domingo *et al.* 2003; Williams *et al.* 2004, 2005; Gomez-Alvarez *et al.* 2012a). Many of the dominant species associated with the α-Proteobacteria class recovered in this study were detected in those studies, including those closely related to members of the species *Porphyrobacter, Blastomonas, Methylobacterium, Sphingomonas* and unclassified species of the order Rhizobiales (Figure 6(a)). The relative abundance of α-Proteobacteria (37%) is consistent with a previous study of a DWDS in Michigan (Pinto *et al.* 2012). In contrast, Pinto *et al.* (2012) identified only 0.5% of sequences to be closely related to mycobacterial species compared with nearly 38% recovered in this study. Overall, these studies suggest that different communities might develop in different distribution systems, while in many cases the composition of some of the dominant taxa is shared, albeit not at the same ratio.
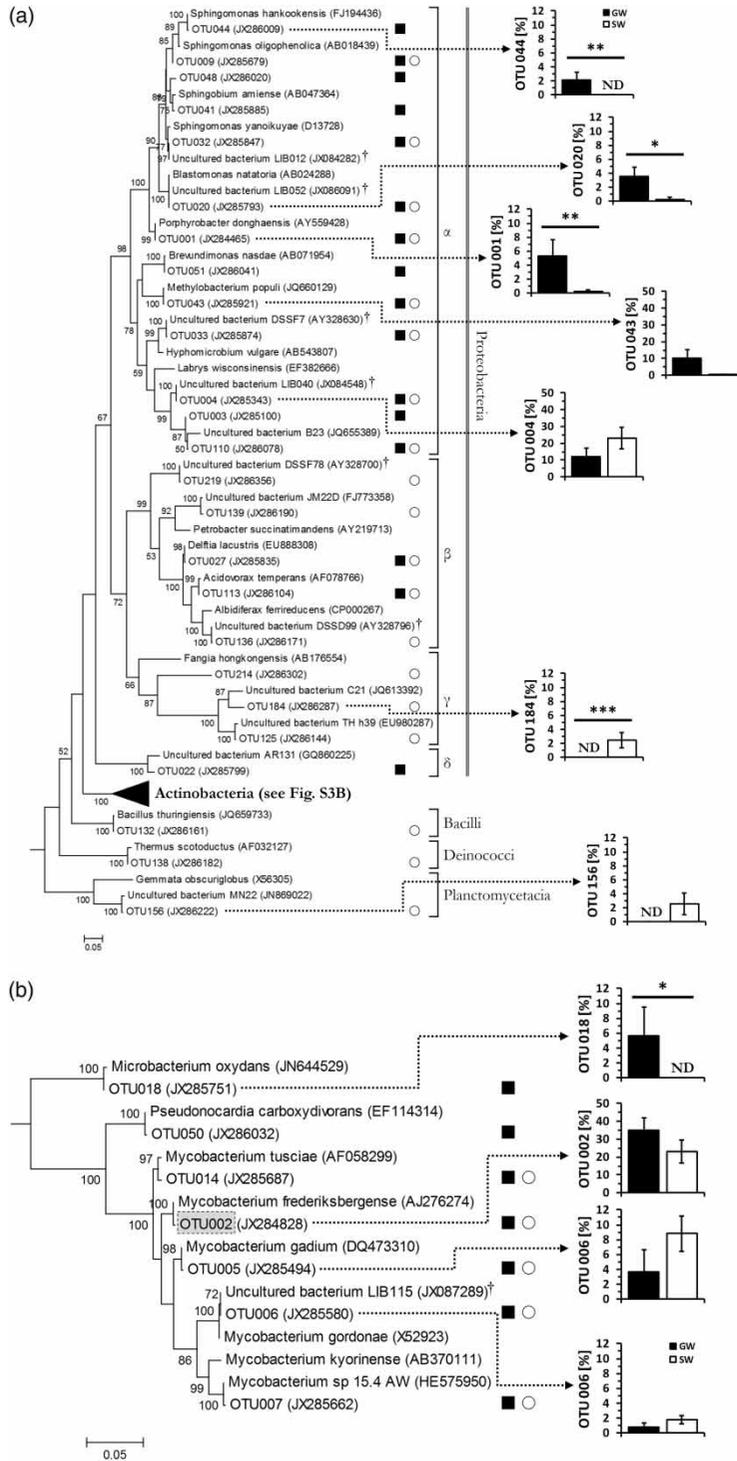
Figure 6 │ (a) Phylogenetic relationships among the 20 most abundant phylotypes from each bulk water communities in groundwater (GW, ■) and surface water (SW, ○). (b) The collapsed group in the main tree is detailed in one subtree for the class Actinobacteria. The tree was inferred from a maximum likelihood analysis of aligned 16S rRNA gene sequence (≈800 bp); nodes with a bootstrap value ≥50% of 500 replicates are identified. The scale-bar represents evolutionary distance and is based on nucleotide substitutions per site. *Methanobacterium congolense* (AF233586) was used as outgroup. The most abundant OTU from each source is framed. †Clones detected in previous studies of drinking water systems. Bar chart shows the relative abundance [%] of selected OTUs that explained ≈65% (SIMPER analysis) of the dissimilarity within all samples (one-way ANOSIM, $R = 0.258$, $p = 0.0001$). Error bars represent the standard error of the mean. Note change in scale. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. ND, not detected.
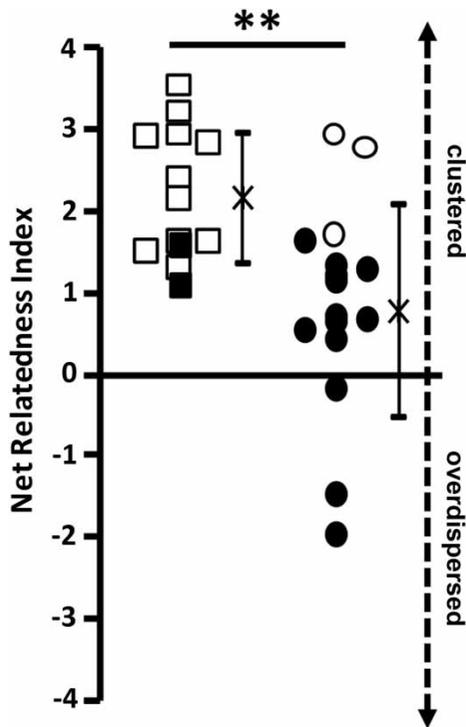
Figure 7 | Variation in NRI for bulk water communities from groundwater (GW, square) and surface water (SW, circle). The NRI reveal a significant increase in clustering for the GW community and a more overdispersed community for the SW samples (**$p < 0.01$). Positive and negative index values indicate phylogenetically clustered and overdispersed communities, respectively. Open symbols represents community phylogenetic structures unlikely to arise by chance ($p < 0.05$). Bars represent the standard deviation with mean ($\times$) of replicates.

Taxonomic classification showed that a significant number of sequences obtained in this study were related to the mycobacterial species *M. tusciae*, *M. frederiksbergense*, *M. gadium*, *M. gordonae*, and *M. kyorinense* (Figure 6(b)). The relative distribution of several *Mycobacterium* species varied between GW and SW samples, but not significantly (Figure 6(b)). Our study further suggests that factors within the DWDS may selectively favor the proliferation of mycobacterial species including: (1) the amount of assimilable organic carbon (AOC), as previous studies have shown that the density of mycobacteria increases with the increase of AOC (Falkinham *et al.* 2001); (2) water retention time and turbidity (Falkinham *et al.* 2001); and (3) hydrogeochemical characteristics of the system (Iivanainen *et al.* 1993).

Numerous studies have identified and documented the occurrence of environmental nontuberculous mycobacteria (NTM) in DWDS treated with chlorine (Vaerewijck *et al.* 2005). Covert *et al.* (1999) detected *Mycobacterium*-related

species in water samples containing detectable levels of free and total chlorine (0.7 ppm and 1.0 ppm, respectively), levels very similar to the water collected during this study (Table 1). The ability of nontuberculous mycobacteria to resist disinfectants has also been described by Bardouniotis *et al.* (2003) in which *M. fortuitum* demonstrated increased resistance to commercially available disinfectants via formation of biofilms, as compared to its planktonic state. In general, our findings suggest a high relative distribution of mycobacteria and that the different proportions of some members of this bacterial group are significant. While the molecular evidence presented in this DWDS study confirmed the presence of *Mycobacterium* spp., there is no evidence that the identified strains are indeed pathogenic.

Our study sheds more light on the fact that the majority of current culture-based techniques for DWDS can underestimate the composition and diversity of microbial communities, including the detection of possible nonmycobacteria opportunistic pathogens. For example, we were able to identify *Legionella*, *Pseudomonas*, and *Vibrio*-like sequences in the DWDS samples. As with the mycobacteria identified, the detection of these sequences does not imply public health risk as DWDS isolates closely related to some pathogenic members of the aforementioned genera seem to have little or no clinical relevance (Vaerewijck *et al.* 2005). Nevertheless, these results are interesting because they suggest that these bacteria can survive the harsh environmental conditions and disinfectant treatments in the drinking water system. One possible mechanism of survival is the ability of *Legionella* and other bacteria to live as intracellular symbionts of environmental protozoa (Lau & Ashbolt 2009). This bacterial–protozoan association can provide chlorine-sensitive bacteria with increased resistance to free chlorine (King *et al.* 1988), and this survival mechanism could have public health relevance as it may increase the persistence of potentially pathogenic bacteria in drinking water systems (Szewzyk *et al.* 2000).

In this study, we did not remove extracellular DNA or dead cells from the original samples, therefore it is possible some sequences are associated with both active/live cells and dead bacterial populations. It should be noted that in a study based on RNA as the target used to generate 16S rRNA gene clone libraries, Revetta *et al.* (2011) demonstrated that Proteobacteria and Actinobacteria were also the dominant members within the 'active' fraction of the community

of a DWDS in Cincinnati. However, in the same study, differences in community structure were obtained between RNA- and DNA-based clone libraries, suggesting that indeed there are some bacteria that are more resistant to the conditions present in DWDS. Using RNA as the PCR template and DNA-binding agents to remove free associated DNA or dead cells will be useful in further identifying active bacterial populations in DWDS and to determine if such bacterial populations are different between and within each distribution system area. Identification of the culturable members can also shed light on some of the populations that are indeed active, even when many of the drinking water bacteria may not grow on conventional microbiological media.

CONCLUSIONS

DWDS are engineered environments that contain distinctive bacterial communities. For example, previous studies have reported β-Proteobacteria as the dominant OTUs in their respective DWDS (Berry *et al.* 2006; Eichler *et al.* 2006; Pinto *et al.* 2012). In contrast, the results from this study are dominated by *Mycobacterium*-related species and members of the Rhizobiales order (α-Proteobacteria). These differences may be attributed to abiotic factors, such as nutrient loads and available organic carbon (Rubin & Leff 2007), disinfectant residual (Mathieu *et al.* 2009), or the treatment processes (Pinto *et al.* 2012). On the other hand, our phylogenetic analysis of the microbial community has shown many similarities with previous studies of DWDS in the Cincinnati area (Santo Domingo *et al.* 2003; Williams *et al.* 2004, 2005; Gomez-Alvarez *et al.* 2012a). In addition, we identified identical sequences within all sample sites suggesting that these bacterial groups may be considered part of the core microbiota of this drinking water system. Finally, many sequences matched organisms that have yet to be cultured, suggesting that many bacteria have specific growth requirements that have not been met by present microbiological media formulations.

ACKNOWLEDGEMENTS

REFERENCES

Al-Jasser, A. O. 2007 Chlorine decay in drinking-water transmission and distribution systems: pipe service age effect. *Water Res.* **41** (2), 387–396.

Bardouniotis, E., Ceri, H. & Olson, M. E. 2003 Biofilm formation and biocide susceptibility testing of *Mycobacterium fortuitum* and *Mycobacterium marinum*. *Curr. Microbiol.* **46** (1), 28–32.

Berry, D., Xi, C. & Raskin, L. 2006 Microbial ecology of drinking water distribution systems. *Curr. Opin. Biotechnol.* **17** (3), 297–302.

Clarke, K. R. 1993 Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18** (1), 117–143.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M. & Tiedje, J. M. 2009 The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37** (1), 141–145.

Covert, T. C., Rodgers, M. R., Reyes, A. L. & Stelma Jr., G. N. 1999 Occurrence of nontuberculous mycobacteria in environmental samples. *Appl. Environ. Microbiol.* **65** (6), 2492–2496.

Craun, G. F., Brunkard, J. M., Yoder, J. S., Roberts, V. A., Carpenter, J., Wade, T., Calderon, R. L., Roberts, J. M., Beach, M. J. & Roy, S. L. 2010 Causes of outbreaks associated with drinking water in the United States from 1971 to 2006. *Clin. Microbiol. Rev.* **23** (3), 507–528.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. 2011 UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27** (16), 2194–2200.

Eichler, S., Christen, R., Holtje, C., Westphal, P., Botel, J., Brettar, I., Mehling, A. & Höfle, M. G. 2006 Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Appl. Environ. Microbiol.* **72** (3), 1858–1872.

Falkinham III, J. O., Norton, C. D. & LeChevallier, M. W. 2001 Factors influencing numbers of *Mycobacterium avium*, *Mycobacterium intracellulare*, and other mycobacteria in drinking water distribution systems. *Appl. Environ. Microbiol.* **67** (3), 1225–1231.

Figueras, M. J. & Borrego, J. J. 2010 New perspectives in monitoring drinking water microbial quality. *Int. J. Environ. Res. Public Health* **7** (12), 4179–4202.

Gomez-Alvarez, V., Revetta, R. P. & Santo Domingo, J. W. 2012a Metagenomic analyses of drinking water receiving different disinfection treatments. *Appl. Environ. Microbiol.* **78** (17), 6095–6102.

Gomez-Alvarez, V., Revetta, R. P. & Santo Domingo, J. W. 2012b Metagenome analyses of corroded concrete wastewater pipe biofilms reveal a complex microbial system. *BMC Microbiol* **12** (1), 122.

Greater Cincinnati Water Works (GCWW). 2006 Safe drinking water report. Available at: http://www.cincinnati-oh.gov/water/downloads/water_pdf15656.pdf, 1–7.

Hammer, Ø., Harper, D. A. T. & Ryan, P. D. 2001 PAST: paleontological statistics software package for evolution and data analysis. *Palaeontol. Electron.* **4** (1), 1–9.

Henne, K., Kahlisch, L., Brettar, I. & Höfle, M. G. 2012 Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany. *Appl. Environ. Microbiol.* **78** (10), 3530–3538.

Hong, P. Y., Hwang, C., Ling, F., Andersen, G. L., LeChevallier, M. W. & Liu, W. T. 2010 Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Appl. Environ. Microbiol.* **76** (16), 5631–5635.

Horner-Devine, M. C. & Bohannan, B. J. M. 2006 Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**(**7 Suppl**), S100–S108.

Iivanainen, E. K., Martikainen, P. J., Väänänen, P. K. & Katila, M. L. 1993 Environmental factors affecting the occurrence of mycobacteria in Brook Waters. *Appl. Environ. Microbiol.* **59** (2), 398–404.

King, C. H., Shotts Jr., E. B., Wooley, R. E. & Porter, K. G. 1988 Survival of coliforms and bacterial pathogens within protozoa during chlorination. *Appl. Environ. Microbiol.* **54** (12), 3023–3033.

Lau, H. Y. & Ashbolt, N. J. 2009 The role of biofilms and protozoa in *Legionella* pathogenesis: implications for drinking water. *J. Appl. Microbiol.* **107** (2), 368–378.

Mathieu, L., Bouteleux, C., Fass, S., Angel, E. & Block, J. C. 2009 Reversible shift in the α-, β- and γ-Proteobacteria populations of drinking water biofilms during discontinuous chlorination. *Water Res.* **43** (14), 3375–3386.

Pinto, A. J., Xi, C. & Raskin, L. 2012 Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environ. Sci. Technol.* **46** (16), 8851–8859.

Revetta, R. P., Matlib, R. S. & Santo Domingo, J. W. 2011 16S rRNA gene sequence analysis of drinking water using RNA and DNA extracts as targets for clone library development. *Curr. Microbiol.* **63** (1), 50–59.

Rubin, M. A. & Leff, L. G. 2007 Nutrients and other abiotic factors affecting bacterial communities in an Ohio River (USA). *Microb. Ecol.* **54** (2), 374–383.

Santo Domingo, J. W., Meckes, M. C., Simpson, J. M., Sloss, B. & Reasoner, D. J. 2003 Molecular characterization of bacteria inhabiting a water distribution system simulator. *Water Sci. Technol.* **47** (5), 149–154.

Santo Domingo, J. W., Revetta, R. P., Iker, B., Gomez-Alvarez, V., Garcia, J., Sullivan, J. & Weast, J. 2011 Molecular survey of concrete sewer biofilm microbial communities. *Biofouling* **27** (9), 993–1001.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009 Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75** (23), 7537–7541.

Snow, J. 1849 *On the mode of communication of cholera*. August/September 1849, Pamphlet, London, UK, pp. 1–31.

Szewzyk, U., Szewzyk, R., Manz, W. & Schleifer, K.-H. 2000 Microbiological safety of drinking water. *Ann. Rev. Microbiol.* **54**, 81–127.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011 MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28** (10), 2731–2739.

Vaerewijck, M. J., Huys, G., Palomino, J. C., Swings, J. & Portaels, F. 2005 Mycobacteria in drinking water distribution systems: ecology and significance for human health. *FEMS Microbiol. Rev.* **29** (5), 911–934.

Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. 2002 Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505.

Webb, C. O., Ackerly, D. D. & Kembel, S. W. 2008 Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24** (18), 2098–2100.

Wickelmaier, F. 2003 An Introduction to MDS: Reports from the Sound Quality Research Unit (SQRU) No. 7. Available at: http://homepages.uni-tuebingen.de/florian.wickelmaier/pubs/Wickelmaier2003SQRU.pdf, 1–26.

Williams, M. M., Santo Domingo, J. W., Meckes, M. C., Kelty, C. A. & Rochon, H. S. 2004 Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *J. Appl. Microbiol.* **96** (5), 954–964.

Williams, M. M., Santo Domingo, J. W. & Meckes, M. C. 2005 Population diversity in model potable water biofilms receiving chlorine or chloramine residual. *Biofouling* **21** (5–6), 279–288.