MONTANA STATE UNIVERSITY LIBRARY

Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system

Authors: Fangqiong Ling, Chiachi Hwang, Mark W LeChevallier, Gary L Andersen, & Wen-Tso Liu

NOTICE: This is a postprint of an article that originally appeared in ISME Journal on August 2015. DOI: http://dx.doi.org/10.1038/ismej.2015.136.

Ling F, Hwang C, LeChevallier MW, Andersen GL, Liu WT, "Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system," ISME Journal 10, no. 3 (August 7, 2015): 582–595. doi:10.1038/ismej.2015.136.

Made available through Montana State University's ScholarWorks
scholarworks.montana.edu

Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system

Fangqiong Ling[1], Chiachi Hwang[1,4], Mark W LeChevallier[2], Gary L Andersen[3] and Wen-Tso Liu[1]

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [2]American Water, Voorhees, NJ, USA and [3]Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract

Drinking water distribution systems (DWDSs) harbor the microorganisms in biofilms and suspended communities, yet the diversity and spatiotemporal distribution have been studied mainly in the suspended communities. This study examined the diversity of biofilms in an urban DWDS, its relationship with suspended communities and its dynamics. The studied DWDS in Urbana, Illinois received conventionally treated and disinfected water sourced from the groundwater. Over a 2-year span, biomass were sampled from household water meters (n = 213) and tap water (n = 20) to represent biofilm and suspended communities, respectively. A positive correlation between operational taxonomic unit (OTU) abundance and occupancy was observed. Examined under a 'core-satellite' model, the biofilm community comprised 31 core populations that encompassed 76.7% of total 16 S rRNA gene pyrosequences. The biofilm communities shared with the suspended community highly abundant and prevalent OTUs, which related to methano-/methylotrophs (i.e., Methylophilaceae and Methylococcaceae) and aerobic heterotrophs (Sphingomonadaceae and Comamonadaceae), yet differed by specific core populations and lower diversity and evenness. Multivariate tests indicated seasonality as the main contributor to community structure variation. This pattern was resilient to annual change and correlated to the cyclic fluctuations of core populations. The findings of a distinctive biofilm community assemblage and methano-/methyltrophic primary production provide critical insights for developing more targeted water quality monitoring programs and treatment strategies for groundwater-sourced drinking water systems.

The ISME Journal advance online publication, 7 August 2015; doi:10.1038/ismej.2015.136

Introduction

Water supplied through piped distribution systems supports the majority of population in the developed world. Since the construction of centralized water supplies in the United States dates back to the late 1800 s, most water treatment and distribution systems are facing the problem of aging infrastruc-ture (National Research Council, 2006). Owing to various chemical and biological processes during distribution, including corrosion (biotic and abiotic) and disinfectant depletion, biofilm growth is a commonly occurring phenomenon in the distribu-tion networks. Distribution system biofilms can cause undesirable water quality changes and violations of public health regulations. Biofilms can act as natural harbors for some opportunistic patho-gens (e.g., *Mycobacterium avium* and *Legionella pneumophila*) that affect immunocompromised populations, allow invasive pathogens to attach when intrusion events occur and remain as a component of waterborne disease risk that is hard to predict (USEPA, 2002). Conversely, biofilms may have positive roles in water quality in ways such as inhibiting pipeline corrosion (Kip and van Veen, 2015) and degrading toxic disinfectant byproducts (Tung and Xie, 2009). To manage properly the risks and exploiting the opportunities, there is a need to better understand biofilms in distribution systems.

Most analyses of the microbiota in drinking water distribution systems (DWDSs) have focused on the suspended community (Lautenschlager et al., 2013; Pinto et al., 2014). Investigations of biofilms are difficult because of limited access and high cost involved in sampling. However, valuable information has been obtained from the laboratory or pilot

reactors that simulate DWDSs. Examining community dynamics in an oligotrophic, disinfectant-free model system revealed that biofilm development on surfaces in contact with drinking water is slow and involves community succession over a multiyear span (Martiny *et al.*, 2003). This study highlights the need for temporal replication for DWDS biofilm studies. Laboratory studies with culture-based or sequencing methods have shown that biofilms can persist after long-term chlorine disinfection, and a distinctive community composition can be selected (LeChevallier *et al.*, 1988; Ling and Liu, 2013). This suggests the need to study DWDS microbiota in countries with different water treatment practices—the residual disinfectant approach in the North America and disinfectant-residual-free approach in certain parts of Europe. Only a few studies have sampled the whole community of biofilms in full-scale systems. Diverse composition (Henne *et al.*, 2012; Kelly *et al.*, 2014) and temporal variation in biomass and dominant species have been observed (Kelly *et al.*, 2014). However, these studies typically used pipeline sections as the source of biofilms, and because of the feasibility of acquiring multiple pipe sections, spatial replication can be limited. Studies adopting spatiotemporal designs, which are crucial for a more complete representation of the community assemblage (Zarraonaindia *et al.*, 2013), have seldom been conducted. Therefore, little is known about many important aspects of the DWDS biofilm ecology, such as species abundance distribution and occupancy-abundance relationship, as well as long-term community dynamics.

In a spatially expansive ecosystem such as DWDSs, populations are influenced by local processes, as well as regional dispersal; therefore, it is important to look into the occupancy of a population (i.e., percentage of sites covered) in addition to its local abundance, as exemplified previously (Pinto *et al.*, 2014). Studies in macroecology have long observed a positive correlation between abundance and regional distribution and using 'core-satellite' hypothesis (Hanski, 1982) to partition species into widely distributed, abundant core populations and rare satellites. Such partition is useful in understanding many phenomena, including species abundance distribution (Magurran and Henderson, 2003; Ulrich and Zalewski, 2006; Van der Gast *et al.*, 2011), yet one of the fundamental benefits is to provide a conceptual tool to identify important taxa in a spatially expansive ecosystem and to facilitate an understanding of how the ecosystem functions (Hanski, 1982).

In this study, we obtained biofilm samples from household water meters, which allowed examination of DWDS biofilms with temporal and spatial replications. The biofilms retrieved from the interior surfaces of water meters developed on the same substratum material, under the same flow regime and orientation and were approximately the same age. In addition, water meters also represent sites where the municipal and household water supply meet, and therefore are relevant to end-users' concerns (Hong *et al.*, 2010). In total, 213 water meter biofilm samples were collected during a 2-year period. Suspended communities were collected in the same DWDS. The 454 pyrosequencing was conducted on the 16 S rRNA gene to examine the abundance and occupancy distribution patterns. The core-satellite model and multivariate analysis of whole community were used in parallel to investigate: (i) what was the extent of community commonness between biofilm and water communities?; (ii) is there any seasonal variation existing in microbial communities on a system scale and how important is it compared with other possible influences?; and (iii) were core populations linked to the dynamics of the whole community?

Materials and methods

Water treatment processes

The DWDS in this study received groundwater treated through convention treatment processes. In the treatment plant, raw water from two aquifers was combined before entry into the treatment basins. The combined water is softened by the addition of lime and sedimentation, recarbonated to lower pH, chlorinated before entering a dual media filtration bed, passed through a clear well and then delivered through the DWDS (Figure 1). After such stringent treatment processes, the product water quality showed stable chemical composition (Supplementary Table S1). To maintain water quality integrity in the DWDS, the water utility applies residual disinfectants in accordance with compliance standards. The water treatment processes were described in detail previously (Hwang *et al.*, 2012b).

Sampling schemes

We worked with Illinois American Water during a program that replaced household water meters in service for ~15 years. The water meters were located at the end of service lines (Figure 1). Each water meter comprised a lead-free bronze alloy main case and a proprietary polymer flowing chamber (Neptune Technology Group, Tallassee, AL, USA). Upon removal, the water meter was plugged at both ends with sterile rubber stoppers to prevent contamination and maintain moisture during transport. Water meter collection was conducted at six different sampling periods and covered a range of water age (the traveling time for treated water to reach the customer) in the DWDS. Sampling took place between May 2009 and February 2011 at selected time periods in summer, fall and winter. The exact day and location of meter collection was affected by the presence of residents at the collection sites. As a result, most but not all of the samples met the designed scheme. The actual times are shown in

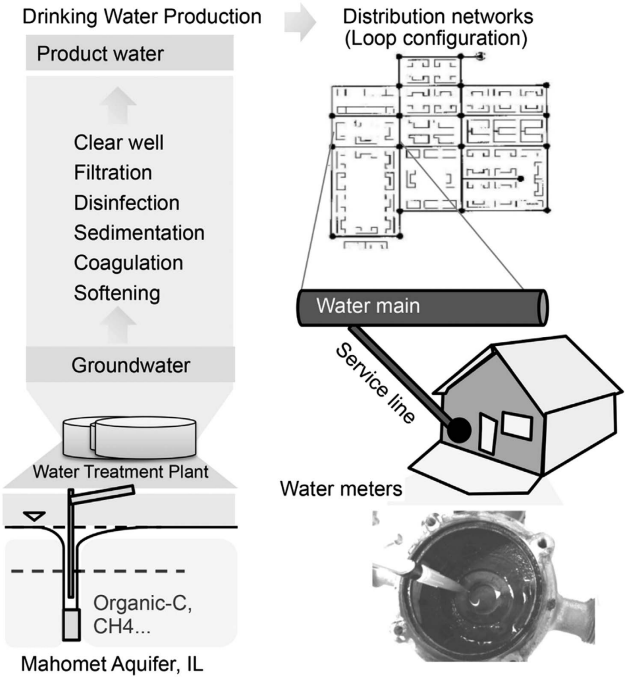


Figure 1 Schematics of a full-scale DWDS. Distribution systems are placed downstream to water source and treatment, receive product water from treatment and deliver it to end-users. Water mains are pipelines for municipal water supply, and service lines connect water mains to households. Water meters are placed at the end of service lines.

Supplementary Figure S1 A for summer (22 May–8 June) and fall (3–21 September) of 2009, early winter (19 January–22 March), summer (3 May–24 June) and fall (7 September–8 October) of 2010 and early winter (11 January–10 February) of 2011. Suspended communities were collected between winter 2010 and winter 2011 (5 March 2010, 10 May 2010, 8 October 2010 and 14 January 2011) from five different sites covering water ages of 0–24 h. A summary of the water meter and water sample information is provided in Supplementary Table S2.

Accessing historical data post hoc

DWDSs are ever-evolving systems where network construction and water distribution span several decades as the service area expands. Ideally, construction years and pipe materials should be included into experimental designs *a priori*, yet it was impractical in this study because construction of water mains was specific to streets, and that of service lines was specific to households. Such data were documented in paper copy by year of construction, where a search engine of a digital database could not be applied. We adopted an *ad hoc* approach where we accessed the historical project archive at Illinois American Water, and retrieved the year of construction, materials used and size of the pipelines (water mains and service lines) specific to the sampled sites. Documentation from the water utility was available since the 1920 s, but we also accessed the archives of local newspapers to search the start of water service for pipeline age estimation.

Other data retrieval from public databases

Temperature data were retrieved from Illinois State Water Survey (http://www.isws.illinois.edu/atmos/statecli/cuweather/index.htm). The ambient temperature shows the same trend as water temperature (Supplementary Figure S1B). In water distribution system design practices, water mains are usually placed underground along paths of streets; hence, pairwise route distances between water meter sites were used to approximate the pipe length between sites. The data were retrieved from Google Map API with a python script written for this study. The script is made available on Github (https://github.com/gaoce/dist_matrix). The distance calculation results are provided in Supplementary Table S3.

DNA extraction

All samples were processed within 6 h of collection. Biofilms on water meter surfaces were obtained with sterile cotton swabs (VWR, Radnor, PA, USA). The swabs were then suspended in 1× phosphate-buffered saline and vortexed vigorously to dislodge the collected biomass. Biomass pellets were obtained by centrifugation and stored at −80 °C until DNA extraction. DNA was extracted using the FastDNA Spin Kits (MPBio, Santa Ana, CA, USA) according to the manufacturer's instructions and stored at −80 °C until further processing. The biomass from suspended communities was collected by filtering tap water and collecting the retentate with 0.22 μm nitrocellulose membrane filters (Millipore, Billerica, MA, USA). The DNA was extracted with a protocol involving enzymatic digestion, bead beating and phenol–chloroform–IAA extraction. The criterion for selection of DNA extraction protocol was reported in a previous publication that evaluated different protocols for DNA extraction of DWDS water and biofilm samples (Hwang *et al.*, 2012a).

16 S rRNA pyrosequencing and sequence processing

The pyrosequencing reaction was performed according to procedures previously described (Tamaki *et al.*, 2011). Universal primers, forward 515 F (5′-Fusion A-Barcode-CA linker-GTGYCAGCMGCCGCGGTA-3′) and reverse 907 R (5′-Fusion B-TC linker-CCCCGYCAATTCMTTRAGT-3′), were used for PCR amplification. PCR products were gel-purified according to the manufacturer's instructions (Promega, Madison, WI, USA). The 454 pyrosequencing was performed on 454 Life Science Genome Sequencer GS FLX (Roche, Branford, CT, USA) at the Roy J Carver Biotechnology Center of University of Illinois Urbana-Champaign (Champaign, IL, USA). Tag sequences were sorted and quality filtered with Quantitative Insights Into

Microbial Ecology (v1.6.0) pipeline (Caporaso *et al*., 2010). Chimeric sequences were identified with Chimera Slayer and removed from downstream analyses (Haas *et al*., 2011). Sequences were aligned to SILVA bacteria alignment trimmed to the amplicon size by the 'pcr.seqs' command in Mothur v.1.33 (Schloss *et al*., 2009). Also in Mothur, distances between sequences were calculated and operational taxonomic units (OTUs) were defined at 97% similarity. The taxonomy of unique sequences was classified with the least common ancestor method against most recent SILVA at an identity score of 0.8. The taxonomy classification results and nearest neighbors are provided in Supplementary Table S4. Phylogenetic trees for 16S rRNA gene pyrosequences and previously reported sequences were constructed using the ARB program based on the neighbor-joining algorithm. Pyrosequences were inserted using the parsimony insertion tool of the ARB program (Saitou and Nei, 1987). The topology of the trees was estimated by 1000 bootstrap replicates (Felsenstein, 1985). Sequencing data have been submitted to NCBI Sequence Read Archive under the bioproject accession number PRJNA279206.

Diversity analysis

Rarefaction curves were generated using Mothur (Schloss *et al*., 2009). The sequences were subsampled to lowest read depth among all samples (858 seqs per sample) for diversity analyses. Box plots were generated with Origin 7.0 (OriginLab Corporation, Northampton, MA, USA). Comparison of α-diversity indices between biofilm and suspended samples were conducted with one-tailed permutation Welsh's *t*-test using the R package 'Deducer' (Fellows, 2012). Community similarity between samples was calculated as Bray–Curtis similarity after square-root transformation. Multivariate tests were performed with distance-based methods such as analysis of similarities (Clarke, 1993; Chapman and Underwood, 1999), PERMANOVA (permutational analysis of variance; Anderson, 2001) and DistLM (distance-based linear models; McArdle and Anderson, 2001). PERMDISP (permutational analysis of multivariate dispersion; Anderson, 2006) was performed to check multivariate dispersions. These tests were performed in the PRIMER-6.0 package (PRIMER-E, Plymouth, UK) according to the author's manual (Anderson and Gorley, 2008). For the multivariate tests on water meter data, temporal binning was applied to reduce possible biases from the unequal sampling season. In applying the temporal binning, each sampling window in the multivariate analyses covered the same length of 20 days, and a more even sample size for each group was achieved. Sampling windows with extremely small sample size (⩽3 samples per 20-day window) were excluded from the analysis to reduce bias. This results in exclusion of two sampling windows (W10-4 and W11-2), including four samples in total from downstream analysis. Sampling in winter 2010 and summer 2010 extended more than 40 days, and the dates not overlapping with the other year (W10-3, W10-4 and S10-1) were labeled as 'transitional'. A summary on the sample size is provided in Table 1. The temporal binning was also conducted at 10-day to test if the length would influence the ecological interpretation. The binning assignment for each sample is provided in Supplementary Table S2. Other details about statistical tests are provided in Supplementary Information.

Results

Surveys on pipeline age, material, configuration and water chemistry of the distribution network

The drinking water system in Champaign-Urbana dates back to 1885 (Urbana Municipal Ordinances, 1885). At the start, water was supplied only during certain hours of a day, and in 1899, a continuous supply began (Urbana archive document, source unclear). The distribution pipelines, 'water mains', developed along with city expansion. Among the 213 sites studied, 29 were supplied by water mains

Table 1 Definition of sampling windows

20-Day window summary	*Season*	*Day 1*	*Day 20*	*Number of samples*
1	S09	22/5/2009	11/6/2009	19
2	F09	3/9/2009	23/9/2009	13
3	W10-1	19/1/2010	8/2/2010	27
4	W10-2	8/2/2010	28/2/2010	34
5	W10-3 (transitional)	28/2/2010	20/3/2010	9
6	W10-4 (transitional)	22/3/2010	11/4/2010	1*
7	S10-1(transitional)	3/2010	23/5/2010	33
8	S10-2	25/2010	14/6/2010	8
9	S10-3	15/6/2010	5/7/2010	9
10	F10-1	7/9/2010	27/9/2010	16
11	F10-2	27/9/2010	17/10/2010	15
12	W11-1	11/1/2011	31/1/2011	26
13	W11-2	3/2/2011	23/2/2011	3*

* indicate sampling windows containing less or equal to three samples.

built before the record. Their ages were approximated using the median between 1899 and 1927 (i.e., 1913). The overall distribution of water main ages is provided in Supplementary Figure S2A. Pipeline construction over time also resulted in mixed materials in the current DWDS. Most are cast iron and ductile iron (Supplementary Figure S2C). Cast iron was the primary material from 1927 to mid-1970 s, and then gradually ductile iron was used for better mechanical properties (Supplementary Figure S2D). Service lines (pipeline connecting municipal supplies to households) had similar distribution of age (Supplementary Figure S2B). The material was mostly copper (93% of all sites in this study), with the rest being galvanized steel or ductile iron. The size of water mains ranged between 1.5 and 12 in based on their designed capacity. The service lines were mostly 0.75–1.25 in wide, with a few exceptions of 2–4 in lines.

The overall configuration of the DWDS was in a 'loop' layout to enhance circulation and minimize stagnation (Swamee and Sharma, 2008). Most of the sites were supplied by loops (74.6%), with a few exceptions supplied directly by one water main (12.2%) or a dead-end connecting from a regional loop (12.7%) (Supplementary Figure S2E). In a loop configuration, the direction of water flow is not fixed, but changes according to the relative transient pressure (Swamee and Sharma, 2008). Thus, water age in this study was estimated through hydraulic models and varied from 0.7 to 89 h. (Supplementary Figure S2F).

Owing to a robust water treatment process, the product water supplied into the DWDS generally maintained stable water chemistry (Hwang *et al.*, 2012b). Water quality monitoring across multiple sites in the distribution system also showed stable pH and turbidity (Table 2). Disinfectants were applied using free chlorine during most of the sampling time, except F09 and W10 when partial monochloramine disinfection was used (Supplementary Figure S1).

'Core-satellite' model in DWDS biofilm and planktonic communities

Two hundred and thirteen water meter biofilms and 20 suspended communities were collected and analyzed, which resulted in 3639 and 1189 OTUs at a sequence similarity of 97% in each category. After rarified to even depths, 199 914 sequences were retained. That yielded 1647 and 507 OTUs in biofilm and suspended communities, respectively. The rarefaction curves did not reach a plateau (Supplementary Figure S3A), yet rank-abundance and occupancy curves suggest that dominant and prevalent species in the community were captured (Figures 2a and b). The abundance and occupancy distribution also showed highly skewed communities in both biofilm and suspension, with a presence of few highly dominant or prevalent OTUs and a tail of low abundance or rare OTUs. The biofilm community has a lower slope in the early stage of rank-abundance curve compared with the suspended community, indicating a difference in community structure. This difference was further demonstrated through a mean comparison of α-diversity indices. Biofilm samples exhibited significantly lower means in Chao 1 diversity and Simpson index compared with suspended samples (Figures 2c and d), suggesting lower diversity and evenness in the biofilm community. Other α-diversity indices showed similar trends (Supplementary Figure S4).

The highly skewed rank-abundance and occupancy curves (Figures 2a and b) led us to examine the correlation between local abundance and regional prevalence. Positive correlation was observed in both biofilm and suspended communities (Pearson's

Table 2 Sample information

	Range
Spatiotemporal	
Year	2009–2011
Season	Summer, fall and winter
Distance between samples	0–11.63 km
Distribution systems	
Range of water age	0.68–89 h
Water main installaion (years)	1927–1997
Service line installation (years)	1897–2010
Materials of service line	Copper
Materials of water mains	Cast iron, ductile iron and copper (extremely few), not recorded (before 1927)
Configuration of adjacent pipeline	Loop, dead ends, direct intake from large water main
Water quality	
Disinfectant type	Free chlorine, monochloramine
Disinfectant concentration	2.69-3.29 mg l^{-1}
pH	8.36–8.86
Turbidity	0.087–0.13

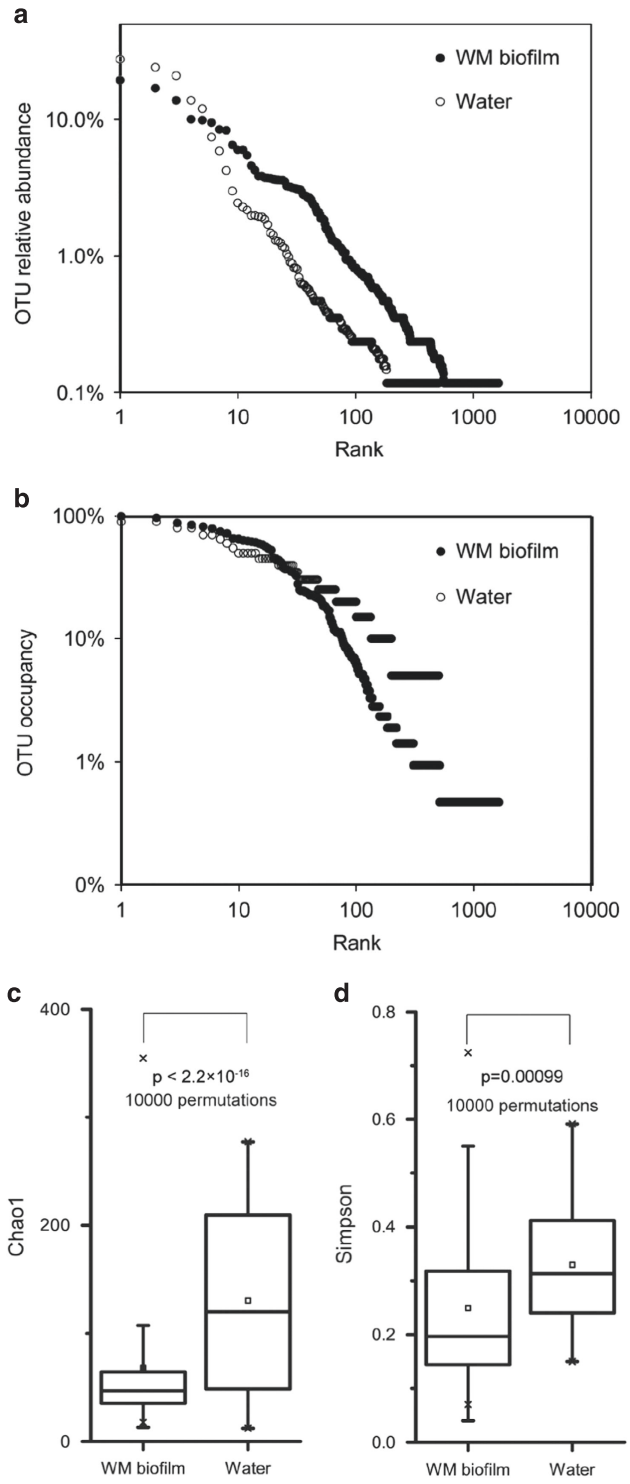


Figure 2 Diversity in biofilm (i.e., WM biofilm) and suspended (i.e., water) communities. (**a** and **b**) The rank-abundance distributions (close symbols: biofilm; open symbols: suspension). Dominance of high abundance and high occupancy samples are observed in both categories of samples. (**c** and **d**) Compare the Chao 1 (**c**) and Simpson's indices (**d**) between biofilm and suspended communities. Biofilms have lower mean values for both indices (*P*-values from one-sided Welch's *t*-test are provided on the plot). In box plots, box represents 25 to 75 percentiles and 'x' represents 1 and 99 percentiles.

$R = 0.695$ for biofilms and 0.484 for suspended communities; Figures 3a and b, respectively), suggesting that the 'core-satellite' model (Hanski, 1982) could be applied conceptually. For comparison purposes, the biofilm community was subsampled by the minimum occupancy in the suspended community (5%) to account for sample size difference. The number of core populations depended on the cutoff occupancy level (Figure 3c) and followed a log-decay function at occupancy between 10% and 90%. To facilitate the discussion, we operationally defined a core community at 30% occupancy, and this generated 31 OTUs in biofilms and 47 in suspended community (Figure 3d). Such defined core populations covered 86.3% of the total sequences, suggesting that such operational definition was useful to capture the majority of diversity.

As the core populations of biofilm and suspended communities overlap, 'shared core OTUs' were further defined, as opposed to those only detected in biofilm or suspended community (Figure 3d). The shared populations were low in number (14 out of 1963 OTUs) but high in weight (62.2% of all the sequences; Figure 3d). When compared across communities, the abundance of biofilm-only core OTUs in suspended communities (green dots in Figure 3e) showed a smaller standard deviation compared with the abundance of suspension-only core population in biofilms (s.d. = 0.0031 vs 0.0156). This difference indicates that the dispersion of the abundance data from biofilm-core populations had a smaller variation in the suspended communities and suggests that certain biofilm populations were independent from the suspended communities.

Core community composition

The core community in biofilms comprised α-, β- and γ-*Proteobacteria*, *Actinobacteria*, *Verrocomicrobia*, *Bacteroidetes* and *Cyanobacteria*. The core-suspended community covers similar divisions, and also includes sequences related to the *Candidatus* phylum Melainabacteria that were genomically predicted as non-photosynthetic bacteria (Di Rienzi *et al.*, 2013). The suspended community also harbored '*Candidatus* Omnitrophica'-related sequences at low abundance. Noticeably, methylotrophy-related OTUs in α-, β- and γ-*Proteobacteria* are abundant and prevalent in both biofilm and suspended communities (Figure 4a, blue symbols). Methylophilaceae- and *Methylococcaceae*-related OTUs are present in the shared core populations. Further phylogenetic analysis suggested their similarity to *Methylotenera* (OTU-1), *Methylomicrobium* (OTU-2), *Methylobacter* (OTU-7) and *Clonothrix* (OTU-12). In addition, a *Methylocystis*-related OTU (OTU-8) was detected in the biofilm-core community, and *Methylobacterium* (OTU-29), *Hyphomicrobium* (OTU-76) and an uncultured cluster close to *Methylococcaceae* and *Crenothrix* were detected in the core-suspended community.

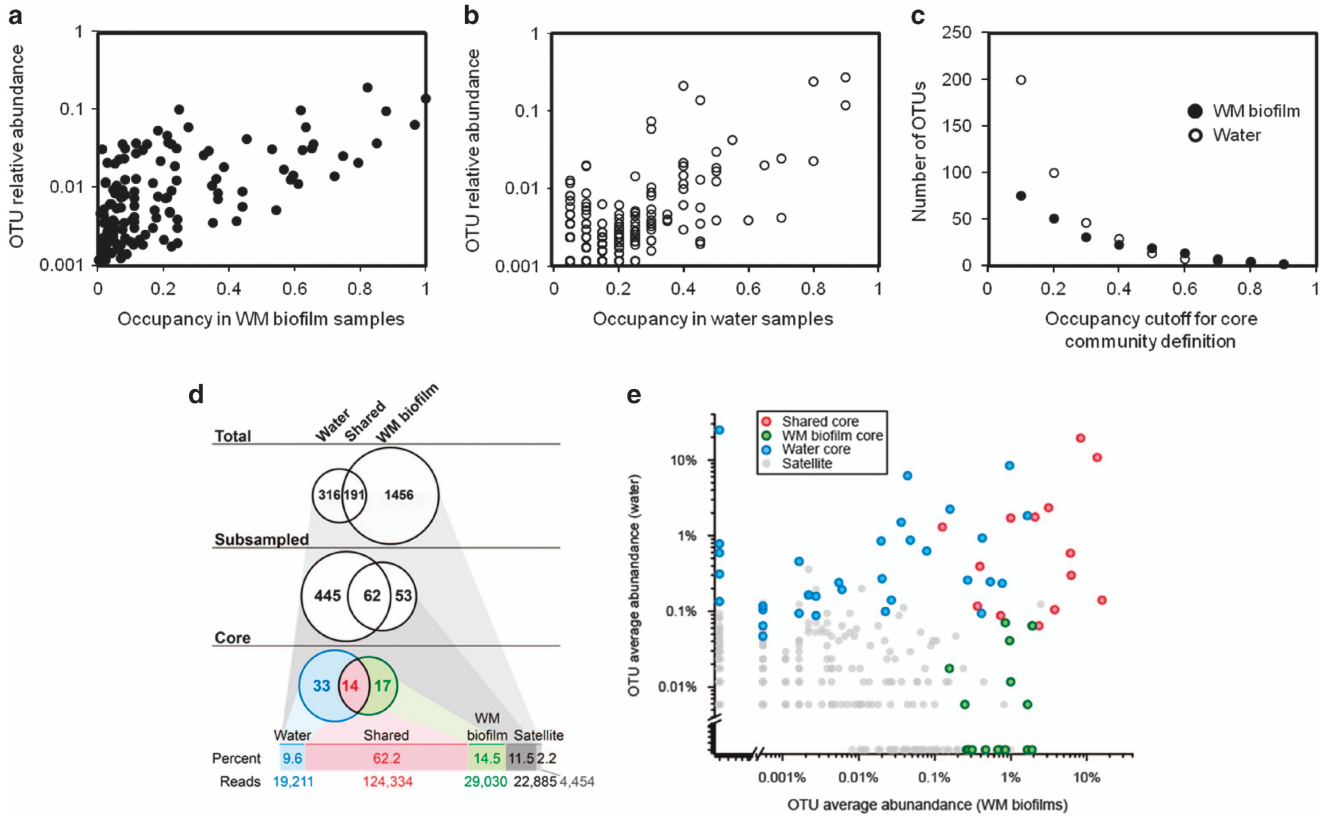
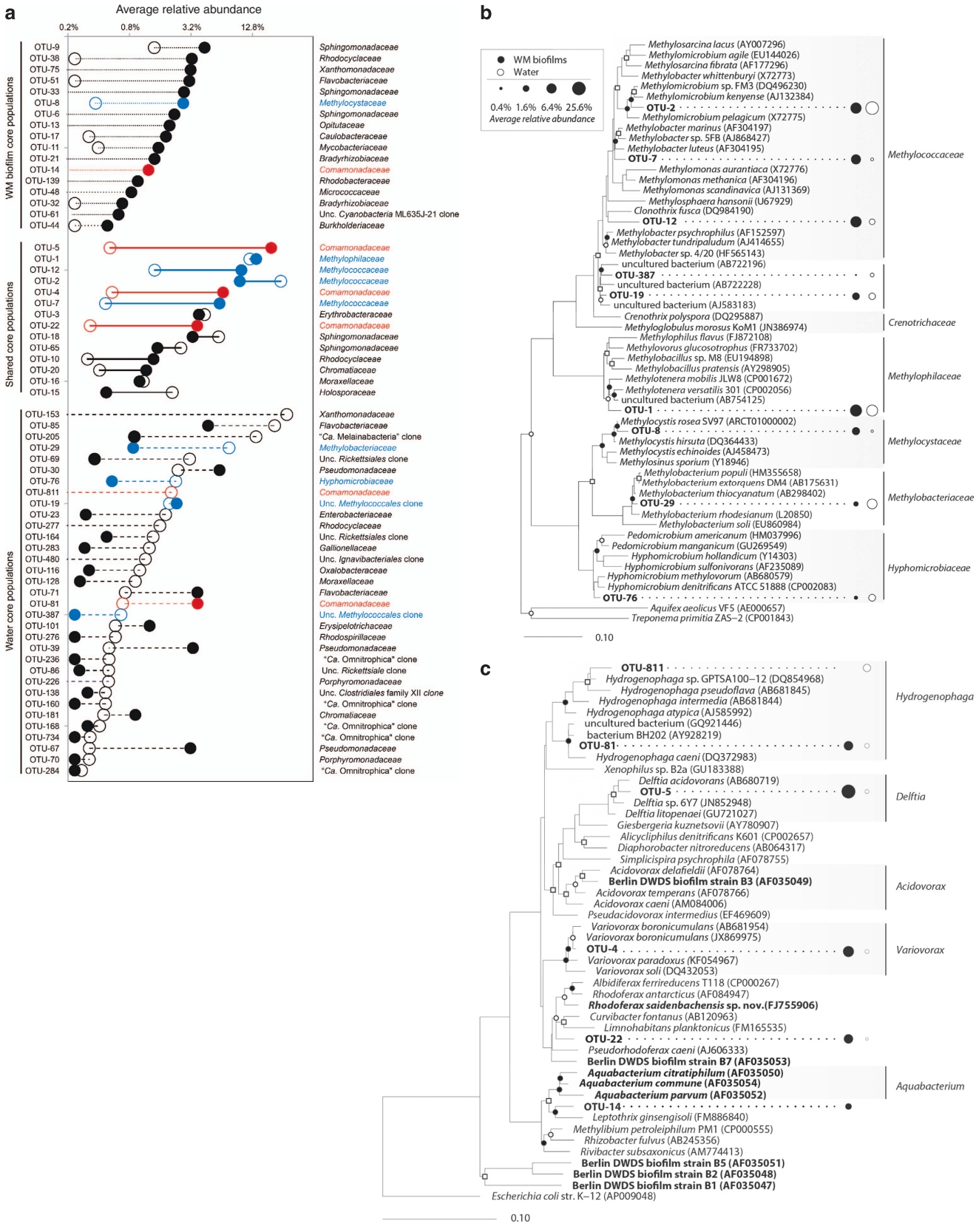


Figure 3 Definition of core populations. Positive correlation of OTU relative abundance and occupancy in biofilm (**a**) and suspended communities (**b**) supports the use of core-satellite model. Operational definition of core communities given at different occupancy level resulted in different core community size (**c**). Number of core communities and their correlating reads are provided (**d**). Biofilm samples were subsampled by occupancy to compare with suspended communities. Average local abundance of shared (red), biofilm-only (green) and suspension-only (blue) core communities in contrast to satellite populations (gray) are provided in (**e**).

Among these taxa, *Methylomicrobium*, *Methylobacter* and *Methylocystis* are methanotrophs, and isolated strains have been reported of using methane as the sole or preferred carbon source (Kalyuzhnaya *et al.*, 2008; Bowman *et al.*, 1993; Belova *et al.*, 2013). For methanol utilization, *Methylotenera* are capable of using methanol as the sole carbon source (Kalyuzhnaya *et al.*, 2012), and *Hyphomicrobium* and *Methylobacterium* are known as facultative methylotrophs (Chistoserdova *et al.*, 2009). Other facultative methylotrophy-related OTUs are also present in the core communities, such as *Mycobacterium* and *Pseudomonas*. A full enumeration is not provided here because recent findings suggest that facultative methylotrophy may be widespread (Chistoserdova *et al.*, 2009).

The DWDS ecosystem harbors many aerobic chemoorganotrophy-related taxa. Five biofilm or shared core OTUs were classified to the family *Sphingomonadaceae* (OTU-66, -33, -65, -9 and -18 in Figure 4a). Bacteria in this family are known as strictly aerobic, able to form biofilms and often observed in oligotrophic environments (Glaeser and Kampfer, 2014). They have long been observed in the chlorinated water environment (Szewzyk *et al.*, 2000; Koskinen *et al.*, 2000; Vaz-Moreira *et al.*, 2011). Six OTUs were classified to the family *Comamonadaceae* (Figure 4a, red symbols), a diverse family that has been observed in many aquatic environments. Further examination on the phylogeny of shared or biofilm-core OTUs (Figure 4c) indicated classification to the genera *Delftia* (OTU-5, shared), *Variovorax* (OTU-4, shared) and *Leptothrix* (OTU-14, biofilm-only). The sequence of a shared core OTU formed a basal branch close to *Limnohabitans* and *Pseudorhodoferax* (OTU-22). An OTU (OTU-811) in the core-suspended community was classified to *Hydrogenophaga*. The phylogenetic diversity of the *Comamonadaceae*-related OTUs in this study overlapped with a previous culture-based study in the DWDS of Berlin, Germany (Kalmbach *et al.*, 1999).

The phylogenetic analysis results suggest other possible energy sources and ecosystem dynamics. Some OTUs were classified to freshwater iron-oxidizing bacteria, including *Gallionella* (OTU-283) and *Leptothrix* (OTU-14). *Gallionella* isolates were the first freshwater iron-oxidizing bacteria described. Cultivated strains exhibit potentials for chemolithoautotrophy and mixotrophy (Emerson *et al.*, 2010). All cultivated species of *Leptothrix* described so far are able to oxidize iron and manganese in ways of chemolithoautotrophy or chemoorganotrophy, and are known for forming sheathes where iron/


a
Average relative abundance
0.2%
0.8%
3.2%
12.8%
WM biofilm core populations
Shared core populations
Water core populations
OTU-9 Sphingomonadaceae
OTU-38 Rhodocyclaceae
OTU-75 Xanthomonadaceae
OTU-51 Flavobacteriaceae
OTU-33 Sphingomonadaceae
OTU-8 Methylocystaceae
OTU-6 Sphingomonadaceae
OTU-13 Opitutaceae
OTU-17 Caulobacteraceae
OTU-11 Mycobacteriaceae
OTU-21 Bradyrhizobiaceae
OTU-14 Comamonadaceae
OTU-139 Rhodobacteraceae
OTU-48 Micrococcaceae
OTU-32 Bradyrhizobiaceae
OTU-61 Unc. Cyanobacteria ML635J-21 clone
OTU-44 Burkholderiaceae
OTU-5 Comamonadaceae
OTU-1 Methylophilaceae
OTU-12 Methylococcaceae
OTU-2 Methylococcaceae
OTU-4 Comamonadaceae
OTU-7 Methylococcaceae
OTU-3 Erythrobacteraceae
OTU-22 Comamonadaceae
OTU-18 Sphingomonadaceae
OTU-65 Sphingomonadaceae
OTU-10 Rhodocyclaceae
OTU-20 Chromatiaceae
OTU-16 Moraxellaceae
OTU-15 Holosporaceae
OTU-153 Xanthomonadaceae
OTU-85 Flavobacteriaceae
OTU-205 "Ca. Melainabacteria" clone
OTU-29 Methylobacteriaceae
OTU-69 Unc. Rickettsiales clone
OTU-30 Pseudomonadaceae
OTU-76 Hyphomicrobiaceae
OTU-811 Comamonadaceae
OTU-19 Unc. Methylococcales clone
OTU-23 Enterobacteriaceae
OTU-277 Rhodocyclaceae
OTU-164 Unc. Rickettsiales clone
OTU-283 Gallionellaceae
OTU-480 Unc. Ignavibacteriales clone
OTU-116 Oxalobacteraceae
OTU-128 Moraxellaceae
OTU-71 Flavobacteriaceae
OTU-81 Comamonadaceae
OTU-387 Unc. Methylococcales clone
OTU-101 Erysipelotrichaceae
OTU-276 Rhodospirillaceae
OTU-39 Pseudomonadaceae
OTU-236 "Ca. Omnitrophica" clone
OTU-86 Unc. Rickettsiale clone
OTU-226 Porphyromonadaceae
OTU-138 Unc. Clostridiales family XII clone
OTU-160 "Ca. Omnitrophica" clone
OTU-181 Chromatiaceae
OTU-168 "Ca. Omnitrophica" clone
OTU-734 "Ca. Omnitrophica" clone
OTU-67 Pseudomonadaceae
OTU-70 Porphyromonadaceae
OTU-284 "Ca. Omnitrophica" clone
b
WM biofilms
Water
0.4% 1.6% 6.4% 25.6%
Average relative abundance
Methylosarcina lacus (AY007296)
Methylomicrobium agile (EU144026)
Methylosarcina fibrata (AF177296)
Methylobacter whittenburyi (X72773)
Methylomicrobium sp. FM3 (DQ496230)
Methylomicrobium kenyense (AJ132384)
OTU-2
Methylomicrobium pelagicum (X72775)
Methylobacter marinus (AF304197)
Methylobacter sp. 5FB (AJ868427)
Methylobacter luteus (AF304195)
OTU-7
Methylomonas aurantiaca (X72776)
Methylomonas methanica (AF304196)
Methylomonas scandinavica (AJ131369)
Methylosphaera hansonii (U67929)
Clonothrix fusca (DQ984190)
OTU-12
Methylobacter psychrophilus (AF152597)
Methylobacter tundripaludum (AJ414655)
Methylobacter sp. 4/20 (HF565143)
uncultured bacterium (AB722196)
OTU-387
uncultured bacterium (AB722228)
OTU-19
uncultured bacterium (AJ583183)
Methylococcaceae
Crenothrix polyspora (DQ295887)
Methyloglobulus morosus KoM1 (JN386974)
Crenotrichaceae
Methylophilus flavus (FJ872108)
Methylovorus glucosotrophus (FR733702)
Methylobacillus sp. M8 (EU194898)
Methylobacillus pratensis (AY298905)
Methylotenera mobilis JLW8 (CP001672)
Methylotenera versatilis 301 (CP002056)
uncultured bacterium (AB754125)
OTU-1
Methylophilaceae
Methylocystis rosea SV97 (ARCT01000002)
OTU-8
Methylocystis hirsuta (DQ364433)
Methylocystis echinoides (AJ458473)
Methylosinus sporium (Y18946)
Methylocystaceae
Methylobacterium populi (HM355658)
Methylobacterium extorquens DM4 (AB175631)
Methylobacterium thiocyanatum (AB298402)
OTU-29
Methylobacterium rhodesianum (L20850)
Methylobacterium soli (EU860984)
Methylobacteriaceae
Pedomicrobium americanum (HM037996)
Pedomicrobium manganicum (GU269549)
Hyphomicrobium hollandicum (Y14303)
Hyphomicrobium sulfonivorans (AF235089)
Hyphomicrobium methylovorum (AB680579)
Hyphomicrobium denitrificans ATCC 51888 (CP002083)
OTU-76
Hyphomicrobiaceae
Aquifex aeolicus VF5 (AE000657)
Treponema primitia ZAS–2 (CP001843)
0.10
c
OTU-811
Hydrogenophaga sp. GPTSA100–12 (DQ854968)
Hydrogenophaga pseudoflava (AB681845)
Hydrogenophaga intermedia (AB681844)
Hydrogenophaga atypica (AJ585992)
uncultured bacterium (GQ921446)
bacterium BH202 (AY928219)
OTU-81
Hydrogenophaga caeni (DQ372983)
Xenophilus sp. B2a (GU183388)
Hydrogenophaga
Delftia acidovorans (AB680719)
OTU-5
Delftia sp. 6Y7 (JN852948)
Delftia litopenaei (GU721027)
Delftia
Giesbergeria kuznetsovii (AY780907)
Alicycliphilus denitrificans K601 (CP002657)
Diaphorobacter nitroreducens (AB064317)
Simplicispira psychrophila (AF078755)
Acidovorax delafieldii (AF078764)
Berlin DWDS biofilm strain B3 (AF035049)
Acidovorax temperans (AF078766)
Acidovorax caeni (AM084006)
Acidovorax
Pseudacidovorax intermedius (EF469609)
Variovorax boronicumulans (AB681954)
Variovorax boronicumulans (JX869975)
OTU-4
Variovorax paradoxus (KF054967)
Variovorax soli (DQ432053)
Variovorax
Albidiferax ferrireducens T118 (CP000267)
Rhodoferax antarcticus (AF084947)
Rhodoferax saidenbachensis sp. nov.(FJ755906)
Curvibacter fontanus (AB120963)
Limnohabitans planktonicus (FM165535)
OTU-22
Pseudorhodoferax caeni (AJ606333)
Berlin DWDS biofilm strain B7 (AF035053)
Aquabacterium citratiphilum (AF035050)
Aquabacterium commune (AF035054)
Aquabacterium parvum (AF035052)
Aquabacterium
OTU-14
Leptothrix ginsengisoli (FM886840)
Methylibium petroleiphilum PM1 (CP000555)
Rhizobacter fulvus (AB245356)
Rivibacter subsaxonicus (AM774413)
Berlin DWDS biofilm strain B5 (AF035051)
Berlin DWDS biofilm strain B2 (AF035048)
Berlin DWDS biofilm strain B1 (AF035047)
Escherichia coli str. K–12 (AP009048)
0.10

manganese deposits accumulate (Emerson *et al.*, 2010). Given that the DWDS water mains in this study were built with cast iron or ductile iron (Table 2 and Supplementary Figure S2), these OTUs likely came from corroded pipe scales. OTUs phylogenetically related to parasitic or mutualistic symbionts were detected at low abundance, including the order *Rickettsiales* (OTU-69, -86, -15 and -164). Previous detection of *Rickettisales* in drinking water-related environments suggest its connection to free-living amoeba (Fritsche *et al.*, 1999; Winiecka-Krusnell and Linder, 2001). Their presence indicates the possibility of grazing activities in the DWDS ecosystem.

Factors influencing overall community structure

One-way analysis of similarities analysis suggested significant temporal variation in the biofilm community ($R = 0.242$, $P < 0.001$). Other factors, including distribution system characteristics, disinfectant type and service line or water main ages, did not explain a comparable proportion of community variation ($R < 0.05$) despite significant *P*-values for some variables (Table 3 and Supplementary Figure S5). Numeric variables tested by DistLM also showed low *R* values (Table 3). The correlation between pairwise Bray–Curtis distance and site distance was examined to seek potential influence from the dispersal process, yet no consistent results were observed (Supplementary Figure S6).

To further examine the temporal change in biofilm communities, PERMANOVA and PERMDISP were used to compare locations of centroids and dispersion. The overall test results showed significant difference in centroid positions ($R = 0.18$, $P < 0.001$). Pairwise comparisons of all combinations of sampling windows are provided in Supplementary Table S5A. The results could be grouped into three categories: significant difference ($P = 0.001$), borderline significance ($0.001 < P < 0.01$) and nonsignificance ($0.01 \leqslant P < 1$). Neighboring sampling windows were mostly similar (S10-2 vs S10-3, $P = 0.9$; F10-1 vs F10-2, $P = 0.882$; W10-1 vs W10-2, $P = 0.011$; S10-1 vs S10-2, $P = 0.01$), except for the transitional W10-3. Borderline or nonsignificant difference were shown for the majority of pairs between winter 2010 and fall 2010 (W10-1 vs F10-1, $P = 0.014$; W10-1 vs F10-2, $P = 0.006$; W10-2 vs F10-1, $P = 0.004$), fall 2010 and winter 2011 (F10-1 vs W11, $P = 0.011$; F10-2 vs W11, $P = 0.01$) and summer 2009 and summer 2010 (S09 vs S10-2, 0.005; S09 vs S10-3, 0.008). The results support the hypothesis of seasonal variation and suggest a repeated pattern between years. Between-centroid distance showed distinctive seasonal clustering, and a back-and-forth swing by seasons between the two clusters (Figure 5a). The same seasonal trend emerged from the analysis based on 10-day binning (Supplementary Figure S7A), suggesting that the bin size did not affect the seasonal variation observed here. Pairwise PERMDISP tests (Supplementary Table S5B) suggest that most groups were not significantly different in their dispersion and support the observed difference in centroids positions, although exceptions exist for pairs involving the F09 window ($P < 0.01$).

The biofilm communities have seasonal variation in diversity. The richness (observed OTUs) shows a fluctuation with time, and a decrease from summer to winter in the second year. The diversity in nonparametric Shannon index descended from summer to winters in both years (Supplementary Figure S8).

Key OTUs explaining seasonal change

From the multiple statistical tests above, seasonality explained most of the variation in biofilm community structure. Therefore, constrained coordination against seasons was conducted through canonical analysis of principle coordinates (CAPs). The result shows that the first axis strongly correlated to season, with fall and winter in the positive side and summer in the negative side (Supplementary Figure S7B). This agreed with analysis on centroids in Figure 5a. OTUs explaining the repeated seasonal pattern were revealed through Spearman's correlation to the first CAP axis (Figure 5b, $r > 0.4$ or < -0.4). Interestingly, these OTUs were all categorized as the 'shared core community' (defined in Figure 3). Among them, OTU-1 related to *Methylotenera* and showed negative correlation with CAP1, indicating summer abundance (Figure 5c). Another two OTUs related to *Comamonadaceae* (OTU-5, *Delfia*; OTU-22, Unc. *Comamonadaceae*) and showed higher abundance in fall and winter (Figure 5d). The rest of OTUs in the core community showed weaker ($0.2 < |r| < 0.4$) or no correlation ($|r| < 0.2$) to CAP1, meaning alternative seasonal pattern or seasonal coherence. The results of Spearman's correlation factors are provided in Supplementary Table S6.

Figure 4 Phylogeny and abundance of core OTUs in biofilm and suspended communities. (**a**) The taxonomy classification of core OTUs to family level and average relative abundance of each OTU in biofilm (closed) and suspended (open) communities. Lines indicate shared (solid), biofilm-only (dotted) and suspension-only (dashed) core populations. Highlighted OTUs are methylotroph- (blue) and *Comamonadaceae*-related (red). (**b** and **c**) Distance matrix tree of 16S rRNA gene sequences assigned to known methylotrphs (**b**) and the family *Comamondaceae* (**c**) based on the neighbor-joining method. Boldface indicates the sequences obtained in this study or other drinking water systems. The 16S rRNA gene sequences of *Aquifex aeolicus* VF5 (AE000657), *Treponema prmitia* ZAS-2 (CP001843) and *Escherichia coli* str. K-12 (AP009048) were used as outgroups. The bar indicates 10% base substitution. Branching points supported probabilities >95%, >75% and >50% by bootstrap analyses (based on 1000 replicates) are indicated by solid circle, open circles and open square, respectively.

Table 3 Test statistics of Bray–Curtis similarity matrix vs explanatory variables

	Method: ANOSIM	
	Global R	*Significant level of sample statistics*
Spatiotemporal		
20-Day sampling windows	0.242	0.001
Distribution system characteristics		
Service line age quartiles	0.02	0.014
Water main age quartiles	0.024	0.001
Service line size groups	0.017	0.162
Water main size groups	0.016	0.323
Distribution system configuration	0.014	0.371
Water main materials	−0.002	0.504
Water quality		
Disinfectant types	0.029	0.151
	Method: DistLM	
	P-*value*	*Proportion of variance explained* (%)
Variable		
Total chlorine	0.001	3.90
Temperature	0.001	4.70
Service line age	0.055	0.80
Water main age	0.007	1.20
Water age	0.003	1.50
	Method: PERMANOVA	
	P-*value (permutation)*	*Sq.root of estimates of component of variation*
Variable		
20-Day sampling windows	0.001	18.87
Residual		44.98

Abbreviations: ANOSIM, analysis of similarities; DistLM, distance-based linear models; PERMANOVA, permutational analysis of variance.

Discussion

By sampling an urban DWDS temporally and spatially, this study reveals novel aspects of the DWDS microbiota, including the presence of highly prevalent core populations and occupancy distribution of the biofilm and suspended communities. Using the 'core-satellite' model allowed us to characterize the ecosystem ecology through identifying abundant and prevalent core populations. Specifically, prevalence of obligate methanotrophs and methylotrophs suggests the presence of methane and methanol as a chemolithotrophic energy source in the distribution system. One may speculate that source water-derived methane (methanogenic Mahomet aquifer; Hackley *et al.*, 2010; Kirk *et al.*, 2004) may be carried through into the distribution system as gas is not stripped during treatment. This is supported by the detection of methane in tap water from all of our water sampling sites in the distribution system (~0.2 mM; Supplementary Information) and reports of taxa associated with methanogensis in the aquifer (Flynn *et al.*, 2013). Thus, methane oxidation may yield methanol as a byproduct for downstream methyltrophic metabolism. Further, such methano- and methylotrophic primary production may generate organic carbon that supports growth of the observed heterotrophs in this oligotrophic ecosystem. Several core populations were classified to genera that were reported as facultative methylotrophs and also related to pathogenic species (i.e., *Acinetobacter* spp., *Methylobacterium* spp., *Mycobacterium avium* and *Pseudomonas aeruginosa*) (Szewzyk *et al.*, 2000); thus, groundwater-sourced drinking water systems with dissolved methane may expose to risks associated with certain methane- or methanol-utilizing opportunistic pathogens.

Among the heterotrophic bacteria potentially supported by the aforementioned primary production, we identified diverse, abundant and prevalent OTUs classified to the family *Comamonadaceae* (Figures 4a and c). Previous studies identify related organisms predominating in DWDS-associated biofilms (Kalmbach *et al.*, 1999; Kalmbach, 2000), faucet biofilms (Liu *et al.*, 2012) and tap water

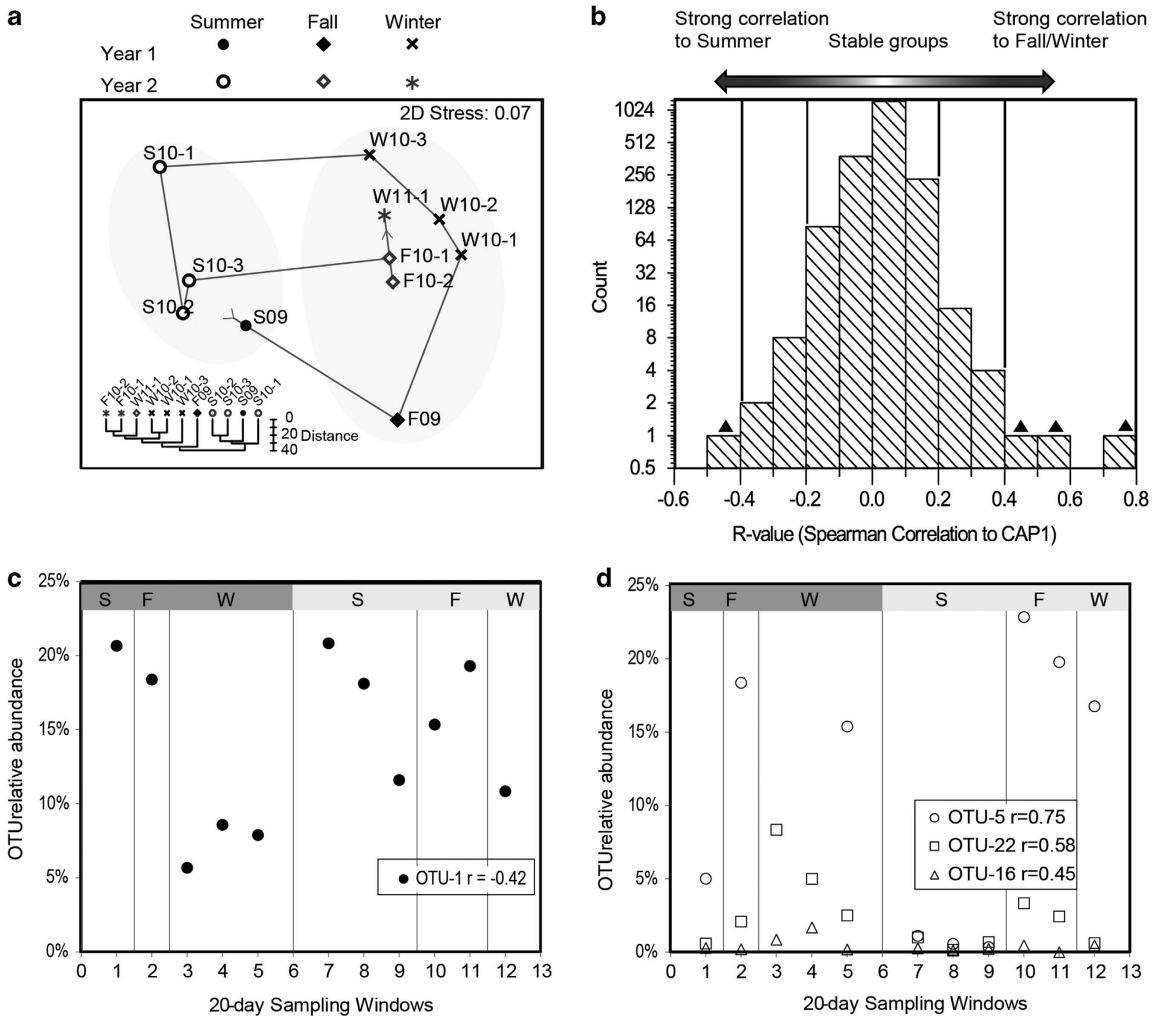


Figure 5 Seasonality of biofilm community structure. (**a**) Non-metrical multidimensional scaling for centroids of 20-day sampling windows. Gray eclipses indicate clusters at 40% Bray–Curtis distance. The cluster analysis result is shown as an insert. (**b**) Distribution of seasonal variation represented by Spearman's coefficient between OTUs and the first canonical principle coordinate constrained by seasonal variation (result of CAPs is shown in Supplementary Figure S7B). The OTUs strongly correlated to the first CAP axis (Spearman's correlation $r > 0.4$ or < -0.4) are plotted against sampling time in (**c**) (negative correlation) and (**d**) (positive correlation).

(Lautenschlager *et al.*, 2013; Pinto *et al.*, 2014). Furthermore, it is interesting to note that members of *Comamonadaceae* have been observed as dominant groups in biofilm and suspended communities of other fast-flowing, low-temperature and oligotrophic aquatic ecosystems including glacier-fed streams and headwater streams (Wilhelm *et al.*, 2013; Besemer *et al.*, 2012), suggesting that DWDS ecosystem shares similarities with other freshwater ecosystems in nature despite its engineered properties.

We observed a distinctive biofilm community from the suspended community (i.e., tap water) in the DWDS. As for specific core populations, methanotrophic *Methylocystis* was biofilm-specific and *Hydrogenophaga* from the family *Comamonadaceae* was water-specific. The biofilms also exhibited lower diversity and evenness compared with the suspended community. These observations indicate that even though water meter biofilms and tap water were present in the same distribution system, their differences in physical properties have yielded different metacommunities. Similarly, the suspended community presented a system with continuous dispersal, from both the fluids and the biofilms. In contrast, biofilm communities were assembled through species sorting on the time-scale of decades, where species fit for biofilm growth were selected. Such differentiation agrees with previous reports on community assemblage processes during stream biofilm formation (Besemer *et al.*, 2012; Jackson *et al.*, 2001).

This study also observed seasonal variation as a key factor to drive the overall variation in the biofilm community (Figures 3a and b) and influence the dynamics of several core populations. Similar observations were reported in DWDS pipe section biofilms (Henne *et al.*, 2012). We hypothesize that seasonal water temperature fluctuation is the likely reason (Supplementary Figure S1B); however, further

studies are needed for better elucidation. Other engineering factors examined in this study did not appear to affect the biofilm communities as significantly, although disinfectant type had been reported previously to affect the structure of suspended communities in the DWDS ecosystem (Hwang *et al*., 2012b; Wang *et al*., 2014). A plausible explanation for this observation is that the biofilms in DWDS, because of the long-term community assemblage discussed above and reduced disinfectant penetration yielded by the polymeric biofilm matrix (Chen and Stewart, 1996), are likely resistant to certain perturbations. This study also observed that spatial effects were not pronounced by water age or by between-site distance (Table 3). This is could be explained by the configuration of the DWDS studied here (i.e., loop shape), which was designed to enhance flow and reduce stagnation, and thus could reduce possible species sorting caused by stagnation. Meanwhile, flow in water meters generally occurs in one direction (from water mains to households), which could further reduce chances of microbial dispersal between sites.

In summary, our findings provide novel insights for bioinformed engineering and pathogen surveillance in drinking water supply. In groundwater-sourced DWDS, methanotrophic populations could serve as the primary producers and likely support the overall growth of biofilms, and of methylotrophic or heterotrophic opportunistic pathogens. To address this, processes that can reduce methane in source water can be explored and used in the treatment of groundwater, and methylotrophic opportunistic pathogens can be better identified and targeted for routine monitoring in groundwater-sourced DWDS. In addition, the presence of a distinctive core biofilm community suggests that monitoring of biological quality in the tap water cannot fully represent the risks in a DWDS. Thus, we recommend monitoring biofilms, especially as a precautionary measure in high-risk water supply systems that have experienced severe contamination, such as many distribution systems reported of sewage contamination after Hurricane Sandy (Redlener *et al*., 2012). Monitoring biofilms in such systems can provide early preventative detection of disease-causing agents. Furthermore, the insights on the ecology of water meters from this study can be translated into a new ecology-inspired monitoring and waterborne disease prevention framework that involves a biofilm sampling device network with a similar spatial resolution as household water meters.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the staff at Illinois American Water for providing water meter samples, water quality data and construction history. We thank Ce Gao for help with the python code for pairwise distance retrieval, Masaru Nobu for discussion on data visualization and Drs Rachel Whitaker, James O'Dwyer and Vern Snoeyink for meaningful discussions. The study is funded by Water Research Foundation and the Academic Excellence Alliance program from King Abdullah University of Science and Technology. The authors declare no conflict of interest.

References

Anderson MJ. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**: 32–46.

Anderson MJ. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**: 245–253.

Anderson MJ, Gorley RN. (2008). *PERMANOVA for PRIMER: Guide to Software and Statistical Methods*. PRIMER-E Ltd: Plymouth, UK.

Belova SE, Kulichevskaya IS, Bodelier PLE, Dedysh SN. (2013). Methylocystis bryophila sp. nov., a facultatively methanotrophic bacterium from acidic Sphagnum peat, and emended description of the genus Methylocystis (ex Whittenbury et al. 1970) Bowman *et al*. 1993. *Int J Syst Evol Microbiol* **63**: 1096–1104.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ *et al.* (2012). Unraveling assembly of stream biofilm communities. *ISME J* **6**: 1459–1468.

Bowman JP, Sly LI, Nichols PD, Hayward AC. (1993). Revised taxonomy of the methanotrophs: description of *Methylobacter* gen. nov., emendation of Methylococcus, validation of *Methylosinus* and *Methylocystis* species, and a proposal that the family Methylococcaceae includes only the group I Methanotrophs. *Int J Syst Bacteriol* **43**: 735–753.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Chapman MG, Underwood AJ. (1999). Ecological patterns in multivariate assemblages : information and interpretation of negative values in ANOSIM tests. *Mar Ecol Prog Ser* **180**: 257–265.

Chen X, Stewart PS. (1996). Chlorine penetration into artificial biofilm is limited by a reaction – diffusion interaction. *Environ Sci Technol* **30**: 2078–2083.

Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. (2009). The expanding world of methylotrophic metabolism. *Annu Rev Microbiol* **63**: 477–499.

Clarke KR. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecol* **18**: 117–143.

Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC *et al.* (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife* **2**: e01102.

Emerson D, Fleming EJ, McBeth JM. (2010). Iron-oxidizing bacteria: an environmental and genomic perspective. *Annu Rev Microbiol* **64**: 561–583.

Fellows I. (2012). Deducer: a data analysis GUI for R. *J Stat Softw* **49**: 1–15.

Felsenstein J Confidence limits on phylogenies: an approach using the bootstrap. (1985). *Evolution (NY)*. **39**: 783–791.

Flynn TM, Sanford RA, Ryu H, Bethke CM, Levine AD, Ashbolt NJ *et al.* (2013). Functional microbial diversity explains groundwater chemistry in a pristine aquifer. *BMC Microbiol* **13**: 146.

Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer K-H, Wagner M. (1999). *In situ* detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order rickettsiales. *Appl Envir Microbiol* **65**: 206–212.

Van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW *et al.* (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* **5**: 780–791.

Glaeser SP, Kampfer P. (2014). *The Prokaryotes*. SpringerBerlin, Heidelberg, Germany.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward D V, Giannoukos G *et al.* (2011). Chimeric 16 S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**: 494–504.

Hackley KC, Panno S V, Anderson TF. (2010). Chemical and isotopic indicators of groundwater evolution in the basal sands of a buried bedrock valley in the midwestern United States: Implications for recharge, rock-water interactions, and mixing. *Geol Soc Am Bull* **122**: 1047–1066.

Hanski I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* **38**: 210–221.

Henne K, Kahlisch L, Brettar I, Höfle MG. (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**: 3530–3538.

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

Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. (2012a). Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. *Microbes Environ* **27**: 9–18.

Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. (2012b). Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl Environ Microbiol* **78**: 7856–7865.

Jackson CR, Churchill PF, Roden EE. (2001). Successional changes in bacterial assemblage structure during epilithic biofilm development. *Ecology* **82**: 555–566.

Kalmbach S. (2000). *In situ* probing reveals *Aquabacterium commune* as a widespread and highly abundant bacterial species in drinking water biofilms. *Water Res* **34**: 575–581.

Kalmbach S, Manz W, Wecke J, Szewzyk U. (1999). *Aquabacterium* gen. nov., with description of *Aquabacterium citratiphilum* sp. nov., *Aquabacterium parvum* sp. nov. and *Aquabacterium commune* sp. nov., three *in situ* dominant bacterial species from the Berlin drinking water system. *Int J Syst Bacteriol* **49**: 769–777.

Kalyuzhnaya MG, Beck DAC, Vorobev A, Smalley N, Kunkel DD, Lidstrom ME *et al.* (2012). Novel methylotrophic isolates from lake sediment, description of *Methylotenera versatilis* sp. nov. and emended description of the genus Methylotenera. *Int J Syst Evol Microbiol* **62**: 106–111.

Kalyuzhnaya MG, Khmelenina V, Eshinimaev B, Sorokin D, Fuse H, Lidstrom M *et al.* (2008). Classification of halo(alkali)philic and halo(alkali)tolerant methanotrophs provisionally assigned to the genera *Methylomicrobium* and *Methylobacter* and emended description of the genus Methylomicrobium. *Int J Syst Evol Microbiol* **58**: 591–596.

Kelly J, Minalt N, Culotti A, Pryor M, Packman A. (2014). Temporal variations in the abundance and composition of biofilm communities colonizing drinking water distribution pipes. *PLoS One* **9**: e98542.

Kip N, van Veen JA. (2015). The dual role of microbes in corrosion. *ISME J* **9**: 542–551.

Kirk MF, Holm TR, Park J, Jin Q, Sanford RA, Fouke BW *et al.* (2004). Bacterial sulfate reduction limits natural arsenic contamination in groundwater. *Geology* **32**: 953.

Koskinen R, Ali-Vehmas T, Kampfer P, Laurikkala M, Tsitko I, Kostyal E *et al.* (2000). Characterization of Sphingomonas isolates from Finnish and Swedish drinking water distribution systems. *J Appl Microbiol* **89**: 687–696.

Lautenschlager K, Hwang C, Liu W-T, Boon N, Köster O, Vrouwenvelder H *et al.* (2013). A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. *Water Res* **47**: 3015–3025.

LeChevallier MW, Cawthon CD, Lee RG. (1988). Factors promoting survival of bacteria in chlorinated water supplies. *Appl Environ Microbiol* **54**: 649–654.

Ling F, Liu W-T. (2013). Impact of chloramination on the development of laboratory-grown biofilms fed with filter-pretreated groundwater. *Microbes Environ* **28**: 50–57.

Liu R, Yu Z, Guo H, Liu M, Zhang H, Yang M. (2012). Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. *Sci Total Environ* **435-436**: 124–131.

Magurran AE, Henderson PA. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.

Martiny AC, Jørgensen TM, Albrechtsen H-J, Arvin E, Molin S. (2003). Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system. *Appl Environ Microbiol* **69**: 6899–6907.

McArdle BH, Anderson MJ. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**: 290–297.

National Research Council (2006). *Drinking Water Distribution Systems: Assessing and Reducing Risks*. National Academies Press. Available at: https://books.google.com/books?id = 0Jz2fgfBLIgC&pgis = 1 (last accessed 11 December 2014).

Pinto AJ, Schroeder J, Lunn M, Sloan W, Raskin L. (2014). Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. *MBio* **5**: e01135–14.

Redlener I, Reilly MJ. (2012). Lessons from Sandy–preparing health systems for future disasters. *N Engl J Med* **367**: 2269–2271.

Saitou N, Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing

microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
Swamee PK, Sharma AK. (2008). *Design of Water Supply Pipe Networks*. Wiley. Available at: https://books.google.com/books?id=F5dz28bt5RkC&pgis=1 (last accessed 11 December 2014).
Szewzyk U, Szewzyk R, Manz W, Schleifer K-H. (2000). Microbiological safety of drinking water. *Annu Rev Microbiol* **54**: 81–127.
Tamaki H, Wright CL, Li X, Lin Q, Hwang C, Wang S *et al.* (2011). Analysis of 16 S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. *PLoS One* **6**: e25263.
Tung H-H, Xie YF. (2009). Association between haloacetic acid degradation and heterotrophic bacteria in water distribution systems. *Water Res* **43**: 971–978.
Ulrich W, Zalewski M. (2006). Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos* **114**: 338–348.
USEPA. (2002). Health risks from microbial growth and biofilms in drinking water distribution systems. Report of water office, United State Environmental Protection Agency. Available online at http://www.epa.gov/ogwdw/disinfection/tcr/pdfs/whitepaper_tcr_biofilms.pdf (last accessed 28 July 2015).
Vaz-Moreira I, Nunes OC, Manaia CM. (2011). Diversity and antibiotic resistance patterns of Sphingomonadaceae isolates from drinking water. *Appl Environ Microbiol* **77**: 5697–5706.
Wang H, Proctor CR, Edwards MA, Pryor M, Santo Domingo JW, Ryu H *et al.* (2014). Microbial community response to chlorine conversion in a chloraminated drinking water distribution system. *Environ Sci Technol* **48**: 10624–10633.
Wilhelm L, Singer GA, Fasching C, Battin TJ, Besemer K. (2013). Microbial biodiversity in glacier-fed streams. *ISME J* **7**: 1651–1660.
Winiecka-Krusnell J, Linder E. (2001). Bacterial infections of free-living amoebae. *Res Microbiol* **152**: 613–619.
Zarraonaindia I, Smith DP, Gilbert JA. (2013). Beyond the genome: community-level analysis of the microbial world. *Biol Philos* **28**: 261–282.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)