UNDERSTANDING THE MICROBIOTA IN DRINKING WATER DISTRIBUTION NETWORKS – FROM MUNICIPAL TO INDOOR WATER SUPPLY SYSTEMS

BY

FANGQIONG LING

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Wen-Tso Liu, Chair
Doctor Mark W. LeChevallier, American Water
Professor Benito J. Mariñas
Professor Thanh H. Nguyen
Professor Rachel J. Whitaker

ABSTRACT

A drinking water distribution system can be viewed as an ecosystem that is comprised of diverse microorganisms interacting with each other and the environment. These microorganisms are central to water quality integrity during distribution. This study investigated the diversity, structure, and spatio-temporal variation of microbial communities in the municipal distribution system and indoor plumbing. The Champaign-Urbana water distribution system, which received conventionally treated and disinfected groundwater, was used as a model system. High throughput sequencing (454 pyrosequencing and Illumina Mi-seq sequencing) was used to study the diversity, and flow cytometry was used to quantify microbial abundance.

For the municipal distribution system, the study focused on the biofilm component. Microbial communities were sampled from household water meters (n=213). Tap water communities (n=20) were also sampled for comparisons. A positive correlation between OTU abundance and occupancy was observed. Highly abundant and prevalent OTUs were observed and defined as "core populations" in the biofilm and suspended communities. The biofilm core population overlapped with the suspended community and formed a "shared core population," including taxa related to methano-/methylotrophy and aerobic heterotrophy. Despite that, the biofilm community differed from the suspended community by specific core populations and lower

diversity and evenness. Multivariate tests indicated seasonality as the main contributor to community structure variation.

Indoor plumbing has smaller pipelines and is prone to stagnation, therefore biological growth is expected. However, the magnitude of biological growth and the possible community composition change is not clear. Thus, we examined the impact of stagnation on the community composition in the tap water community. We treated three dormitory buildings in Champaign-Urbana as natural laboratories and conducted a controlled stagnation test. Our results showed that the microbial abundance increased from $<10^3$ cells/mL to $\sim10^5$ cells/mL after a week-long stagnation. The community structure of post-stagnation water significantly differed from the pre-stagnation water. Multivariate analysis showed a significant difference between stagnant and fresh tap water communities. The building, floor, and faucet of sample collection were also shown as significant sources of variation, yet to a lesser degree than stagnation. Temporal variation did not significantly influence the community structure. The post-stagnation communities further exhibited differentiation by flow volumes, which again indicate the influence from the pipeline structure. Methylotrophy and aerobic heterotrophy-related taxa were observed in the post-stagnation communities.

Overall, this study has demonstrated that spatiotemporal experiments combined with hypothesis testing can lead to new understanding of drinking water supply systems. Source water community, seasons, and water use (stagnation) were shown to profoundly influence microbial

communities in the distribution system. Our findings further showed taxa indicative of a certain carbon source and cell count gradients indicative of stagnation. These findings suggest that the microbiota in the distribution system is a valuable source of information within the distribution system and can be harnessed to complement current monitoring.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Professor Wen-Tso Liu, for guidance and support for my dissertation work and giving me much freedom in my research. I have always felt grateful to have the opportunity to work on drinking water microbiology. I have learned a lot from discussions with my dissertation committee members. Dr Mark LeChevallier has provided insights from water quality engineering and directed me to important literature that could easily be skipped by the search engines. He has also given many comments on my manuscripts and asked questions that made me think. I learned a lot from Professor Rachel Whitaker on experimental design. It was in her class that I learned to ask ecological questions. I want to thank Professor Helen Nguyen and Professor Benito Mariñas for giving me support, and for discussions on transport phenomena, which I always want to keep in mind when I look at an environmental process. I have also learned much from Professor James O'dwyer on ecological models and from Professor Snoeyink on drinking water treatment.

The research work in this dissertation has received much support from Illinois American Water. Ms. Elizabeth Doellman worked with me since my start as a Master's student. We went through sample collection, water quality data acquisition, and project history retrieval. Mr. Charles Andrew McCarrey, Andy, navigated me through the piles of Illinois American Water project documentation since the 1920s. By reading these documents I learned how the pipelines in

Champaign-Urbana came into their current shape. I cannot describe how exciting it was to see a real “blueprint” for the first time in my life, and to flip through the project correspondence of engineers in the 1920s.

The experiment on indoor water supply has received tremendous support from University of Illinois Housing. I want to thank John E. Collins, Housing Director 1998-2013, who talked to me when I was making a cold visit to University Housing. John put me in contact with David Boehm, Assistant director for housing maintenance and operations. It was through his coordination that my experimental design of “treating buildings as reactors” finally came true. There are many others to thank at U of I Housing: the inspectors (Otto, Joshua, Scott, and many others) walked through the buildings with me during all my sampling and preparations to insure my safety; the cleaning crew postponed their tasks in order to make my sampling schedules work.

I want to thank University Facilities and Services for their assistance on building materials and documentation. Richard Drew, project manager, helped me with the acquisition of a 90-year old pipe for my reactor design during an update project. Ms. Beth Leitz helped me to access the plumbing drawings and project specifications for the buildings where I collected water samples.

Drinking water contains a very low number of bacteria, therefore in order to make the community analysis happen, massive amounts of samples had to be collected. I want to thank Dr Chiachi Hwang for working with me on the sampling and molecular experiments of water meters. In the indoor water study project, more than 1334 liters of samples were collected, and

transporting the samples was made possible by the efforts from many friends in the EES program. I want to thank Aimee Gall, Priscilia Tunas, Johnathan Moor, John Jurevis, Shengkun Dong, Ruiqing Lu, Lin Ye, Yifan He, Min Yin, Jinyong Liu, Yin Wang, Hanting Wang, and so many others for lending a hand. Most of the buildings we sampled were built before the Americans with Disabilities Act compliance was required, which means there were no elevators; therefore, they had to carry the 10-liter carboys and walk up and down stairs multiple times. Johnathan Moor designed a rubber handle padding to reduce the pressure of the very narrow carboy handles on their fingers.

Many thanks are extended to Dr Shaoying Qi for always being helpful for facilities in the Environmental Engineering Science laboratory and also for sharing his experience in water chemistry research and scientific writing with me.

During my doctoral study I have been a member of "CASE", which stands for Cultural Awareness and Speech Enhancement, hosted by Professor Snoeyink. It has been the place where I practice presentations and communication skills. I have also learned a lot about other students' backgrounds and their research through CASE.

I have benefited from working with students and researchers in Professor Wen-Tso Liu's research group. Many of the students and postdoctoral researchers have talked to me and given critiques on my work. I want to thank Masaru Nobu for our discussions on data visualization and Camilla Carlos, Na Kyung Kim, Michael Siegert, and Ran Mei for many valuable discussions.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Motivation

This thesis work is about microorganisms in the public supplied drinking water, especially the distribution infrastructures. It tries to address the following questions: what microbes are there in the water supplies? Is there any pattern for their spatial and temporal distributions? And what are the processes underlying these patterns? These are microbial ecology questions, yet asking these questions is also important for water engineers because the course of exploration will involve adapting the latest technologies and concepts in biology to the monitoring of water, and that will eventually help us to better define "safe" drinking water for the public and develop more up-to-date criteria for process innovations.

Historically, the breakthroughs in microbiology have profoundly impacted water treatment technologies and water quality monitoring. It is now well known that proper treatment of drinking water is important in preventing epidemics of cholera and typhoid, yet the knowledge did not come from nowhere. The birth of "germ theory" in bacteriology and the plate count technique provided the very basis for monitoring the biological quality of water (Szewzyk *et al.*, 2000; Payment *et al.*, 2003). After Robert Koch's lab developed the plate count technique, British sanitarian scientist Parcy Frankland received training on the technique from Koch's lab and applied it on assessing filtration efficiency after he went back to Britain (Hamlin, 1990). He found "strikingly" fewer microbes in Thames River water after filtration, to a degree of 95%-99% removal. Robert Koch himself was the first to link filter efficiency to the prevention of cholera. During his investigation of the 1892 cholera outbreak along the river Elbe in Germany, Koch

found that in a small town, Altona, escaped the outbreak, although their drinking water was drawn from the river impacted by sewage of cholera-hit cities. However, the river water was filtered before being supplied to the public, and the filters were properly operated. Using his plate count techniques, Koch further established a water treatment goal for cholera prevention: "filtered water containing more than 100 germs capable of development per cubic centimetre must not be allowed to enter the pure-water-reservoir" (Koch *et al.*, 1895; Payment *et al.*, 2003). This treatment goal has a long lasting impact on the way drinking water is treated. It can be viewed as the prototype of heterotrophic plate count, a widely used water quality indicator today. The value of 100 colony forming units per milliliter is still used in many countries, though the culturing media and conditions have been refined over the years (Exner et al., 2003).

Interestingly, at the time when plate count was first introduced to water engineering, many of the debates centered on the limits of the culture technique in representing the abundance and diversity of microorganisms in their natural environment. The following text is an excerpt from the 1908 version of "Elements of Water Bacteriology with Special Reference to Sanitary Water Analysis":

"That the customary methods for determining the number of bacteria do not reveal the total bacterial content, but only a very small fraction of it, becomes apparent when we consider the large number of organisms, nitrifying bacteria, cellulose-fermenting bacteria, strict anaerobes, etc., which refuse to grow, or grow only very slowly in ordinary culture media, and which, therefore, escape our notice" (Prescott and Winslow, 1908; Payment et al., 2003).

Today, answering "what microbes are there" no longer relies on the "ordinary culture media" criticized by the sanitarians in the early 1900s. The advent of ribosomal RNA sequence analysis

revolutionized microbial ecology; biologists can take a glimpse of the microbial diversity of an environment by extracting the community DNA, amplifying the ribosomal RNA genes, and comparing with databases built from previous studies. High throughput sequencing technologies further make this process much faster and less costly. Despite that, current drinking water monitoring still relies on culturing certain indicator microorganisms. This is understandable, because much needs to be learned before we can develop new molecular techniques that can be used for drinking water monitoring. We currently know very little about what microorganisms are present in drinking water supplies, what their baseline community structure is, and what environmental conditions or engineering factors would cause community variation. In this thesis, we aim to apply high throughput sequencing technologies to investigate the microbial communities in a drinking water supply, and leverage on spatio-temporal sampling to further understand the variations.

1.2 Biofilms and water quality problems in distribution systems

In modern public water supply systems, water quality is managed by using qualified water sources, meeting treatment goals, and maintaining pressure and disinfectants during distribution. While producing water of regulated quality at water treatment plants is a well-achieved goal with conventional treatment and disinfection, maintaining water quality stability during distribution still remains challenging. Various processes, including pipeline corrosion, pressure transients, and loss of disinfectants in dead-ends, could occur in distribution systems and cause adverse effects on water quality. The situation is further worsened by climate change: under extreme climate events, power outage usually leads to the interruption of water service, the loss of water

pressure, and consequently sewage infiltration into the distribution system (National Research Council, 2006).

Diverse microorganisms have been observed in the water distribution systems. Among heterotrophic plate count bacteria, non-photosynthetic pigmented bacteria of yellow and orange were frequently detected in systems carrying residual disinfectants (Reasoner *et al.*, 1989). Pink pigmented bacteria that were methylotrophic have also been isolated (Hiraishi *et al.*, 1995). In addition to the suspended communities, water distribution systems provide vast solid-liquid interfaces for microorganisms to attach, grow, and form biofilms. Early microscopic studies on pipe sections showed the presence of diverse morphologies on the pipeline corrosion product -- "tubercles" and the iron-oxidizing bacteria *Gallionella* spp. (Ridgway *et al.*, 1981). Heterotrophic plate counts have been cultured from pipe surfaces, indicating the presence of heterotrophic bacteria in pipeline biofilms (LeChevallier *et al.*, 1987).

Biofilms in distribution systems are central to several compliance issues and public health risks. Certain opportunistic pathogens have a natural reservoir in the environment (e.g., water or soil) and may be able to grow once they enter distribution systems and attach to biofilms, including *Mycobacterium* spp., *Legionella* spp., *Aeromonas* spp., and *Pseudomonas* spp. (Szewzyk *et al.*, 2000). Among these bacteria, *Mycobacterium* spp. and *Legionella* spp. are on the USEPA Candidate Contaminate List for public water systems. Biofilms are also considered a source for "coliform regrowth" -- product water that meets the Total Coliform Rule at the treatment plant experiences coliform count increase in distribution systems, which often causes violations. Through biochemical profiling of coliform bacteria in a distribution system, it was shown that the coliforms in "regrowth" were related to those in distribution system biofilms (LeChevallier *et al.*, 1987).

Due to the water quality problems related to biofilms in distribution systems, there is much engineering literature on how to control their growth. Several engineering factors were shown to correlate to less regrowth, including filtration, limiting assimilable organic carbon, maintaining residual disinfectants, and applying corrosion control (LeChevallier *et al.*, 1996). However, there are other factors not as easily controllable. For example, it was shown that water temperature over 15ºC correlated to biological regrowth, suggesting seasonal temperature fluctuation may affect the biological growth in biofilms (LeChevallier *et al.*, 1996). In a surface water sourced distribution system in New Jersey, it was shown that rainfalls correlated to higher regrowth, suggesting that source water microbes may persist in distribution systems despite treatment and disinfection (LeChevallier *et al.*, 1987). Thus, it is important to understand the distribution system biofilms as an ecosystem, and identify how source water biogeography, local climate and engineering factors interact to influence the microbial community assemblages.

1.3 Biological water quality in premise plumbing systems

Premise plumbing is the last part of water distribution that impacts the water quality before the product water from municipal water treatment reaches the point of use. It branches from the network and supplies water to various properties, such as schools, hospitals, and public and private housing. It is estimated that the distribution mains throughout the United States measures at approximately 1 million miles, and the premise plumbing portion of the network is more than 5 million miles in length (National Research Council, 2006). The design of building pipelines is based on water demand; therefore, premise plumbing systems usually have smaller pipeline sizes than distribution systems. In addition, the water use pattern of people in buildings dictates the

inactivity of water use of varying lengths. Thus, premise plumbing systems can be considered as a kind of "dead-end" where water stagnates and is likely to experience complete depletion of disinfectants and elevated levels of bacterial growth (LeChevallier *et al.*, 1996).

In the United States, the Safe Drinking Water Act (SDWA) is the federal law that ensures the quality of public water supplies. Under the SDWA, the United States Environmental Protection Agency (EPA) regulates water quality and oversees the implementation of such regulations. Among the current regulations, most of the water quality parameters are regulated before distribution. For example, the National Primary Drinking Water Regulations (http://water.epa.gov/drink/contaminants/index.cfm#List) cover the source water and the treatment process and include 51 organic compounds, 16 inorganic chemicals, 4 radionuclides, indicator microorganisms, and disinfectants. Among these, only the rules for disinfectants, disinfectant byproducts (Disinfectant/Disinfection Byproduct Rule [D/DBPR]), total coliform (Total Coliform Rule [TCR]), and lead and copper (Lead and Copper Rule [LCR]) govern quality changes that might occur after treatment. The TCR and the D/DBPR target the distribution system and require water samples to be taken after flushing the tap, and the LCR is the only rule that purposefully monitors premise plumbing. Without routine monitoring, very little is known about the biological water quality change in premise plumbing.

1.4 Previous work on extraction of DNA from drinking water

This thesis work used high throughput sequencing to characterize microbial communities. This method has circumvented the bias associated with culturing microbes in the lab, but the molecular methods are not perfect either. Our view of the microbial communities can be

distorted by steps in the molecular experiments and the bioinformatics. The first step towards community characterization with sequencing is to disrupt or lyse cells and to separate DNA from the environmental matrix. In this step, incomplete extraction could underestimate the diversity, extraction efficiency could vary for microbes having different cell wall compositions, and the environmental matrix might protect the cells from the lysis reagents (Frostegard *et al.*, 1999; Martin-Laurent *et al.*, 2001). Natural and built environments present unlimited possibilities of chemical compositions and microbial diversity; therefore, it is virtually impossible to develop a standardized DNA extraction method that fits all. Thus researchers have been working to understand the possible DNA extraction bias with respect to the environment to be studied. While many evaluations have been conducted for natural environments, especially soil, there have been limited reports on the drinking water environment.

Prior to conducting this thesis work, we had evaluated the possible limitations of several widely used DNA extraction protocols or commercial kits in characterizing the drinking water biofilms (Hwang *et al.*, 2012). The protocols we examined are physical disruption, chemical lysis, as well as a combination of physical disruption and chemical and enzymatic lysis, herein referred to as Miller's, Zhou's, and Schmidt's methods (Table 1). These methods all include a final step using organic solvents (i.e., phenol and chloroform) to separate DNA from cell debris, which requires careful handling and generates hazardous wastes. Commercial kits using silica bead cartridges for DNA adsorption can circumvent the use of chloroform. These methods also render additional convenience through prepackaged solutions and supplies. Thus we also included two commercial kits, the FastDNA spin kit and the MoBio PowerSoil kit, in our evaluation. We evaluated these methods based on yield, DNA integrity, and possible biases.

To evaluate these procedures, we first tested them on bacterial monocultures of strains with different cell wall characteristics. These include Gram-negative bacteria *Aeromonas caviae, Aquabacterium parvum,* and *Sphingomonas* sp. RO2; the Gram-positive bacteria *Bacillus subtilis* and *Gordonia hirsuta*; the acid-fast bacterium *Mycobacterium smegmatis*; and *Escherichia coli* as a positive reference strain. In addition, these strains were selected because their close relatives were isolated from drinking water environments or reported as biofilm formers. Our results showed that the laboratory protocols achieved 4-5 times more DNA yield than the commercial kits tested. The DNA extracts from the FastDNA spin kit, Zhou's protocol, and Schmidt's protocol preserved better integrity. When the yield was normalized to the *E.coli* DNA yield, the FastDNA spin kit, Miller's protocol, and Schmidt's protocol produced more consistent yields across different strains.

Based on the monoculture test results, we further tested the yield and possible biases from the FastDNA spin kit, Zhou's protocol, and Schmidt's protocol on a kind of community samples representative of the drinking water systems, the biofilms from household water meters. The community variation was characterized with community fingerprinting (T-RFLP) and pyrosequencing. Similar to the results from the monoculture tests, the laboratory protocols produced higher yields. The different methods caused variation in community fingerprints and composition, yet the community differentiation between water meters, the main biological signal in this case, still emerged as the main source of variation in both analyses.

1.5 Thesis structure

This thesis work used the Champaign-Urbana water supply as a model system to study the microbiota inhabiting drinking water supplies. Because of the important role biofilms have in the post-treatment stability of drinking water, we first examined the microbial communities in distribution system biofilms. In Chapter 2, we adopted household water meters as our biofilm sampling device and asked the following questions: i) what is the extent of community commonness between the biofilm and water communities? ; and ii) what are the environmental conditions that shape the variation in the biofilm communities?

Following the direction of water flow, we then looked into the indoor plumbing. As discussed previously in section 1.3, this is the part of the water supply where water stagnation is inevitable, and microbial growth is expected. However, the magnitude of biological growth and the possible community variation associated with it is not clear. In Chapter 3, we aimed to answer these questions by treating buildings as our natural laboratory and conducting a controlled stagnation experiment.

Lastly, after we made the observations about community variations, we wanted to ask i) what are the biological processes behind the observed patterns?; and ii) how does the pipeline design interact with these processes? In Chapter 4, we synthesized the data from Chapters 2 and 3, and modeled the dispersal and demographic processes in indoor plumbing under the framework of the unified theory of biodiversity.

1.6 Reference

Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, *et al.* (1999). Quantification of Bias Related to the Extraction of DNA Directly from Soils. *Appl Envir Microbiol* **65**:5409–5420.

Hamlin C. (1990). A Science of Impurity: Water Analysis in Nineteenth Century Britain. University of California Press: Berkeley http://ark.cdlib.org/ark:/13030/ft667nb43t/.

Hiraishi A, Furuhata K, Matsumoto A, Koike K, Fukuyama M, Tabuchi K. (1995). Phenotypic and genetic diversity of chlorine-resistant Methylobacterium strains isolated from various environments. *Appl Envir Microbiol* **61**:2099–2107.

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

Koch R, Duncan G, Gairdner S. (1895). Professor Koch on the Bacteriological Diagnosis of Cholera, Water-filtration and Cholera, and the Cholera in Germany During the Winter of 1892-93.

LeChevallier MW, Babcock TM, Lee RG. (1987). Examination and characterization of distribution system biofilms. *Appl Envir Microbiol* **53**:2714–2724.

LeChevallier MW, Welch NJ, Smith DB. (1996). Full-scale studies of factors related to coliform regrowth in drinking water. *Appl Environ Microbiol* **62**:2201–11.

Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, *et al.* (2001). DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* **67**:2354–9.

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

Payment P, Sartory DP, Reasoner DJ. (2003). The history and use of HPC in drinking-water quality management. In:*Heterotrophic plate counts and drinking-water safety*. https://books.google.com/books?hl=en&lr=&id=HQx8a_KtBbAC&oi=fnd&pg=PA20&dq=The+history+and+use+of+HPC+in++drinking-water+quality+management+&ots=QNfZNv6PGA&sig=zRdznTSjEN1tUdahA5Pvl9MRnc0 (Accessed July 7, 2015).

Prescott SC, Winslow C-EA. (1908). Elements of Water Bacteriology: With Special Reference to Sanitary Water Analysis. J. Wiley https://books.google.com/books?id=mvkKAQAAIAAJ&pgis=1 (Accessed July 7, 2015).

Reasoner DJ, Blannon JC, Geldreich EE, Barnick J. (1989). Nonphotosynthetic pigmented bacteria in a potable water treatment and distribution system. *Appl Environ Microbiol* **55**:912–21.

Ridgway HF, Means EG, Olson BH. (1981). Iron Bacteria in Drinking-Water Distribution Systems: Elemental Analysis of Gallionella Stalks, Using X-Ray Energy-Dispersive Microanalysis. *Appl Envir Microbiol* **41**:288–297.

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

CHAPTER 2 CORE-SATELLITE POPULATIONS AND SEASONALITY OF WATER METER BIOFILMS IN A METROPOLITAN DRINKING WATER DISTRIBUTION SYSTEM[1]

2.1 Abstract

Drinking water distribution systems (DWDS) harbor microorganisms in biofilm and suspension, yet the diversity and spatiotemporal distribution have been studied mainly in the suspended community. This study examined the diversity of biofilms in an urban DWDS, its relationship to suspended communities, and its dynamics. The studied DWDS in Urbana, Illinois received conventionally treated and disinfected water sourced from groundwater. Over a two-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 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 16S 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

[1] The results described in this chapter have been compiled into a journal paper and accepted. The citation is as follows: Fangqiong Ling, Chiachi Hwang, Mark W. LeChevallier, Gary L. Andersen, Wen-Tso Liu. Core-satellite populations and seasonality of water meter biofilms in a metropolitan drinking water distribution system. The ISME Journal. Accepted.

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 cyclic fluctuations of core populations. The findings of a distinctive biofilm community assemblage and methano-/methyl-trophic primary production provide critical insights for developing more targeted water quality monitoring programs and treatment strategies for groundwater-sourced drinking water systems.

2.2 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 1800s, most water treatment and distribution systems are facing the problem of aging infrastructure (National Research Council, 2006). Due 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 distribution 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 pathogens (e.g., *Mycobacerium 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 play 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 properly manage 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 (DWDS) have focused on the suspended community (Lautenschlager *et al.*, 2013; Pinto *et al.*, 2014). Investigations of biofilms are difficult due to limited access and high cost involved in sampling. However, valuable information has been obtained from laboratory or pilot reactors that simulate DWDS. 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 multi-year 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 necessity of studying DWDS microbiota in countries of 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 utilized pipeline sections as the source of biofilms, and due to 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, occupancy-abundance relationship, as well as long-term community dynamics.

In a spatially expansive ecosystem like a DWDS, 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 positive correlation between abundance and regional distribution and using "core-satellite" hypothesis (Hanski, 1982) 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, 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. 454 pyrosequencing was conducted on the 16S 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 to other possible influences?; and iii) were core populations linked to the dynamics of the whole community?

2.3 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 prior to entry into the treatment basins. The combined water is softened by lime

addition 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 2.1). After such stringent treatment processes, the product water quality showed stable chemical composition (Table A.2). To maintain water quality integrity in the DWDS, the water utility applies residual disinfectants in accordance with compliance standards. A detailed description of the operation at the water treatment plant was described previously (Hwang *et al.*, 2012b).

Sampling schemes We worked with Illinois American Water (ILAW) during a program that replaced household water meters in service for approximately 15 years. The water meters were located at the end of service lines (Figure 2.1). Each water meter comprised a lead-free bronze alloy main case and a proprietary polymer flowing chamber (Neptune, 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 A.1A for summer (May 22 to June 8) and fall (September 3-21) of 2009, early winter (January 19-March 22), summer (May 3 – June 24), and fall (September 7-October 8) of 2010, and early winter (January

11 to February 10) of 2011. Suspended communities were collected between winter 2010 and winter 2011 (3/5/2010; 5/10/2010; 10/8/ 2010; 1/14/2011) from five different sites covering water ages of 0-24 hrs.

Accessing historical data *post hoc* DWDS 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 ILAW, 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 1920s, 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 was 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 (Figure A.1B). 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).

DNA extraction All samples were processed within six hours of collection. Biofilms on water meter surfaces were obtained with sterile cotton swabs (VWR, Radnor, PA). The swabs were then suspended in 1X PBS 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) according to manufacturer's instructions and stored at -80°C until further processing. Biomass from suspended communities was collected by filtering tap water and collecting the retentate with 0.22 μm nitrocellulose membrane filters (Millipore, Billerica, MA). 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).

16S rRNA pyrosequencing and sequence processing The pyrosequencing reaction was performed according to procedures previously described (Tamaki *et al.*, 2011). Universal primers forward 515F (5'-Fusion A-Barcode-CA linker-GTGYCAGCMGCCGCGGTA-3') and reverse 907R (5'-Fusion B-TC linker-CCCCGYCAATTCMTTRAGT-3') were used for PCR

amplification. PCR products were gel-purified according to manufacturer's instructions (Promega, Madison, WI). 454 pyrosequencing was performed on 454 Life Science Genome Sequencer GS FLX (Roche, Branford, CT) at the Roy J. Carver Biotechnology Center of University of Illinois Urbana-Champaign. 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 OTUs were defined at 97% similarity. The taxonomy of unique sequences were classified using a naive Bayesian algorithm (Wang *et al.*, 2007) against most recent SILVA at an identity score of 0.8. The taxonomy classification results and nearest neighbors are provided in Table A.3. 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 1,000 bootstrap replicates (Felsenstein, 1985). Sequencing data has been submitted to NCBI Sequence Read Archive under the bioproject accession 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/sample) for

diversity analyses. Box plots were generated with Origin 7.0 (OriginLab Corporation, Northampton, MA, USA). Comparison of alpha 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 analysis of similarities (ANOSIM), (Clarke, 1993; Chapman and Underwood, 1999), PERMANOVA (Anderson, 2001), and DistLM (McArdle and Anderson, 2001). PERMDISP (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 authors' 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 ($\leq$3 samples per 20-day window) were excluded from the analysis to reduce bias. This results in exclusion of 2 sampling windows (W10-4 and W11-2) including 4 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 2.3. The temporal binning was also conducted at 10-day to test if the length would

influence the ecological interpretation. Other details about statistical tests are provided in supplementary information.

2.4 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 built prior to 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 Figure A.2A. Pipeline construction over time also resulted in mixed materials in the current DWDS. Most are cast iron and ductile iron (Figure A.2C). Cast iron was the primary material from 1927 to mid-1970s; then gradually ductile iron was used for better mechanical properties (Figure A.2D). Service lines (pipeline connecting municipal supplies to households) had similar distribution of age (Figure A.2B). 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 between1.5 and 12 inches based on their designed capacity. The service lines were mostly 0.75-1.25 inches wide, with a few exceptions of 2-4 inch 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%) (Figure A.2E). 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 hours. (Figure A.2F).

Due 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 2.1). Disinfectants were applied using free chlorine during most of the sampling time, except F09 and W10 when partial monochloramine disinfection was used (Figure A.1).

"Core-satellite" model in DWDS biofilm and planktonic communities 213 water meter biofilms and 20 suspended communities were collected and analyzed, which resulted in 3,639 and 1,189 OTUs at a sequence similarity of 97% in each category. After rarified to even depths, 199,914 sequences were retained. That yielded 647 and 507 OTUs in biofilm and suspended communities, respectively. The rarefaction curves did not reach a plateau (Figure A.3A), yet rank-abundance and occupancy curves suggest that dominant and prevalent species in the community were captured (Figure 2.2A and 2.2B). 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 than the suspended community, indicating a difference in community structure. This difference was further demonstrated through a mean comparison of alpha diversity indices. Biofilm samples exhibited significantly lower means in Chao1 diversity and Simpson index than suspended samples (Figure 2C and 2D), suggesting lower diversity and evenness in the biofilm community. Other alpha-diversity indices showed similar trends (Figure A.4).

The highly skewed rank abundance and occupancy curves (Figure 2.2A and 2.2B) led us to examine the correlation between local abundance and regional prevalence. Positive correlation was observed in both biofilm and suspended communities (Pearson's R = 0.695 for biofilms and 0.484 for suspended communities; Figure 2.3A and 2.3B, 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 cut-off occupancy level (Figure 2.3C) 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 2.3D). 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 2.3D). The shared populations were low in number (14 out of 1963 OTUs) but high in weight (62.2% of all the sequences, Figure 2.3D). When compared across communities, the abundance of biofilm-only core OTUs in suspended communities (green dots in Figure 2.3E) showed a smaller standard deviation than the abundance of suspension-only core population in biofilms (standard deviation = 0.0031 vs. standard deviation = 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, but 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 2.4A, 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.

Among these taxa, *Methylomicrobium*, *Methylobacter* and *Methylocystis* are methanotrophs, and *Methylomicrobium* and *Methylobacter* are known for obligate methanotrophy (Kalyuzhnaya *et al.*, 2008; Bowman *et al.*, 1993; Belova *et al.*, 2013). For methanol utilization, *Methylotenera* are capable of utilizing methanol as 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 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 2.4A). 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 2.4A, 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 2.4C) indicated classification to the genera *Delftia* (OTU-5, shared), *Variovorax* (OTU-4, shared), and *Leptothrix* (OTU-14, biofilm-only). The sequence of a shred 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/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 2.1, Figure A.2), 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 ANOSIM 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 2.2, Figure A.5). Numeric variables tested by DistLM also showed low R values (Table 2.2). 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 (Figure A.6).

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 Table A.4A. The results could be grouped into three categories: significant difference ($p=0.001$), borderline significance ($0.001<p<0.01$),

and non-significance ($0.01 \leq p < 1$). 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 non-significant 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 2.5A). The same seasonal trend emerged from the analysis based on 10-day binning (Figure A.7A), suggesting that the bin size did not affect the seasonal variation observed here. Pairwise PERMDISP tests (Table A.4B) suggest that most groups were not significantly different in their dispersion and support the observed difference in centroids positions, though 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 non-parametric Shannon index descended from summer to winters in both years (Figure A.8).

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 (CAP). The result shows the first axis strongly correlated to season, with fall and winter in the positive side and summer in the negative side (Figure A.7B). This agreed with analysis on centroids in Figure 2.5A. OTUs explaining the repeated seasonal pattern were revealed through Spearman correlation to the first CAP axis (Figure 2.5B, $r>0.4$ or <-0.4). Interestingly, these OTUs were all categorized as the "shared core community" (defined in Figure 2.3). Among them, OTU-1 related to *Methylotenera* and showed negative correlation with CAP1, indicating summer abundance (Figure 2.5C). Another two OTUs related to *Comamonadaceae* (OTU-5, *Delfia*; OTU-22, Unc. *Comamonadaceae*) and showed higher abundance in fall and winter (Figure 2.5D). 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.

2.5 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. Employing 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 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 methylo-trophic 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.

Amongst the heterotrophic bacteria potentially supported by the aforementioned primary production, we identified diverse, abundant, and prevalent OTUs classified to the family *Comamonadaceae* (Figure 2.4A and 2.4C). 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 (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 than the suspended community. These observations indicate that even though water meter biofilms and tap water were present in the same distribution system, their difference in physical properties have yielded different meta-communities. Likely, 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 (Figure 2.3A and 2.3B) 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 (Figure A.1B), 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, due to 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 2.2). 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, 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 utilized in 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.

2.6 Tables and Figures

Table 2.1 Sample Information

Spatio-Temporal	Range
Year	2009-2011
Season	Summer, fall and winter
Distance between samples	0-11.63km
Distribution systems	
Range of water age	0.68-89 hrs
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 (prior to 1927)
Configuration of adjacent pipelines	loop, dead-ends, direct intake from large water main
Water Quality	
Disinfectant type	Free chlorine, monochloramine
Disinfectant concentration	2.69-3.29 mg/L
pH	8.36-8.86
Turbidity	0.087-0.13

Table 2.2 Test statistics of Bray-Curtis similarity matrix versus explanatory variables

Method: ANOSIM		
Spatio-temporal	Global R	Significant level of sample statistics
20-day sampling windows	0.242	0.001
Distribution system characteristics		
Service line age quartiles	0.020	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		
Variable	p	Proportion of variance explained
Total Chlorine	0.001	3.9%
Temperature	0.001	4.7%
Service Line Age	0.055	0.8%
Water Main Age	0.007	1.2%
Water Age	0.003	1.5%
Method: PERMANOVA		
Variable	p(permutation)	Sq.root of estimates of component of variation
20-day sampling windows	0.001	18.87
Residual		44.98

Table 2.3 Definition of sampling windows (asterisks indicate sampling windows containing less or equal to three samples)

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

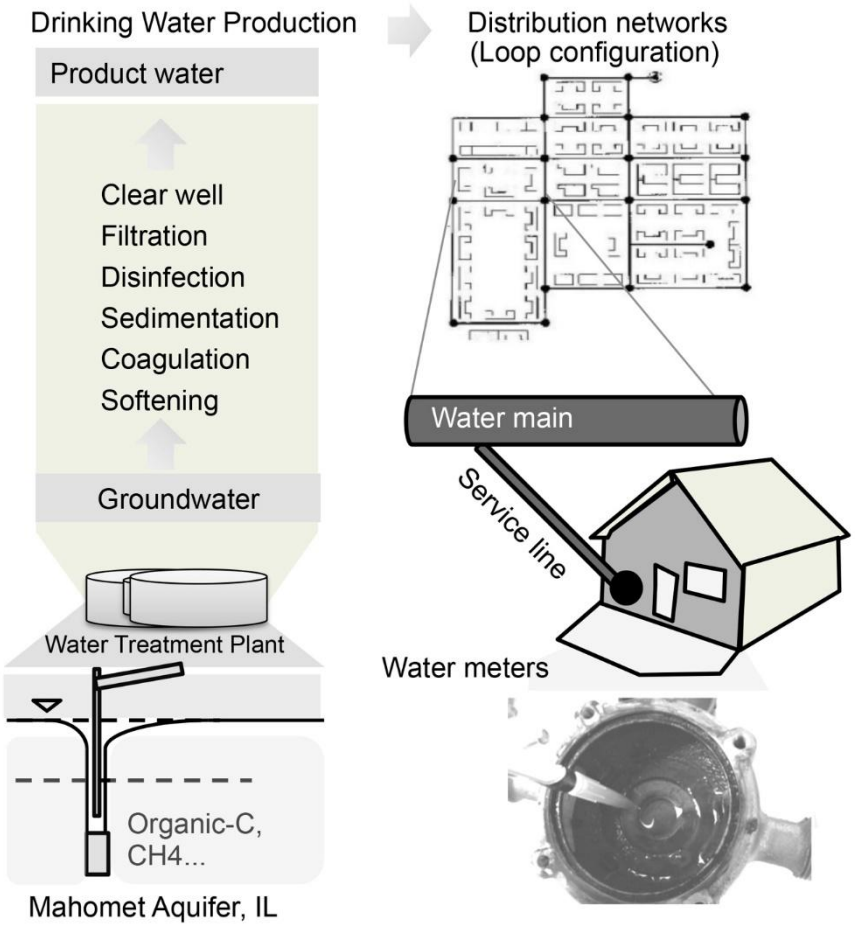


Figure 2.1 Schematics of a full-scale drinking water distribution system. Distribution systems are placed downstream to water source and treatment, receive product water from treatment, and deliver it to end users. Water mains are pipelines for municipal water supply, and service lines connect water mains to households. Water meters are placed at the end of service lines.

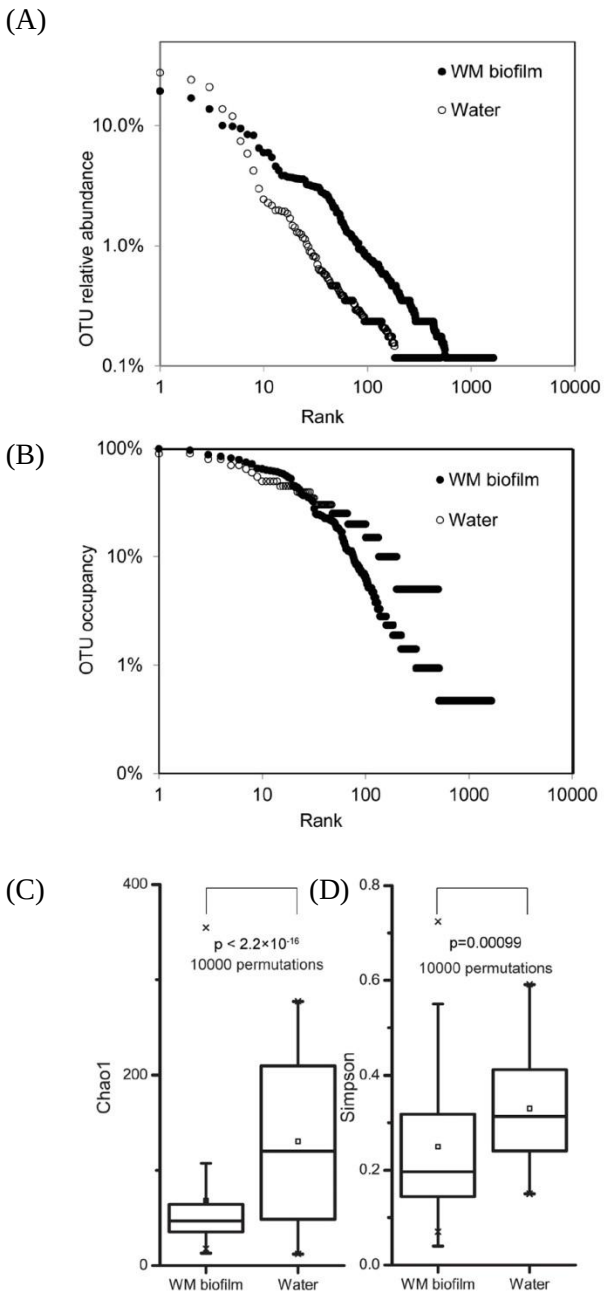


Figure 2.2 Diversity in biofilm (*i.e.* WM biofilm) and suspended (*i.e.* water) communities. Panels A and B show the rank abundance distributions (close symbols: biofilm; open symbols: suspension). Dominance of high abundance and high occupancy samples are observed in both categories of samples. Panels C and D compare the Chao 1 (C) and Simpson's indices (D) between biofilm and suspended communities. Biofilm 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.

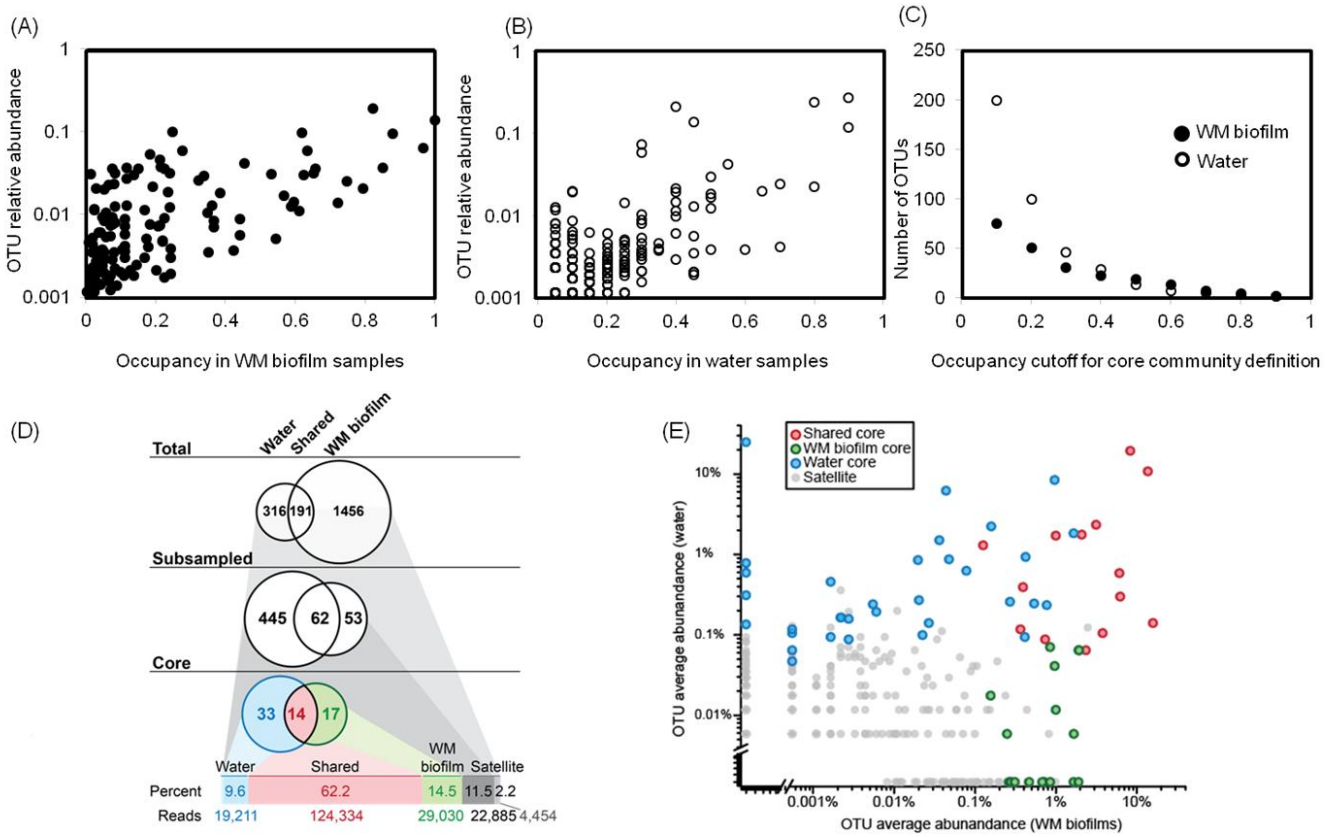


Figure 2.3 Definition of core populations. Positive correlation of OTU relative abundance and occupancy in biofilm (A) and suspended communities (B) supports the use of core-satellite model. Operational definition of core communities given at different occupancy level resulted in different core community 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 (grey) are provided in panel E.

(A)

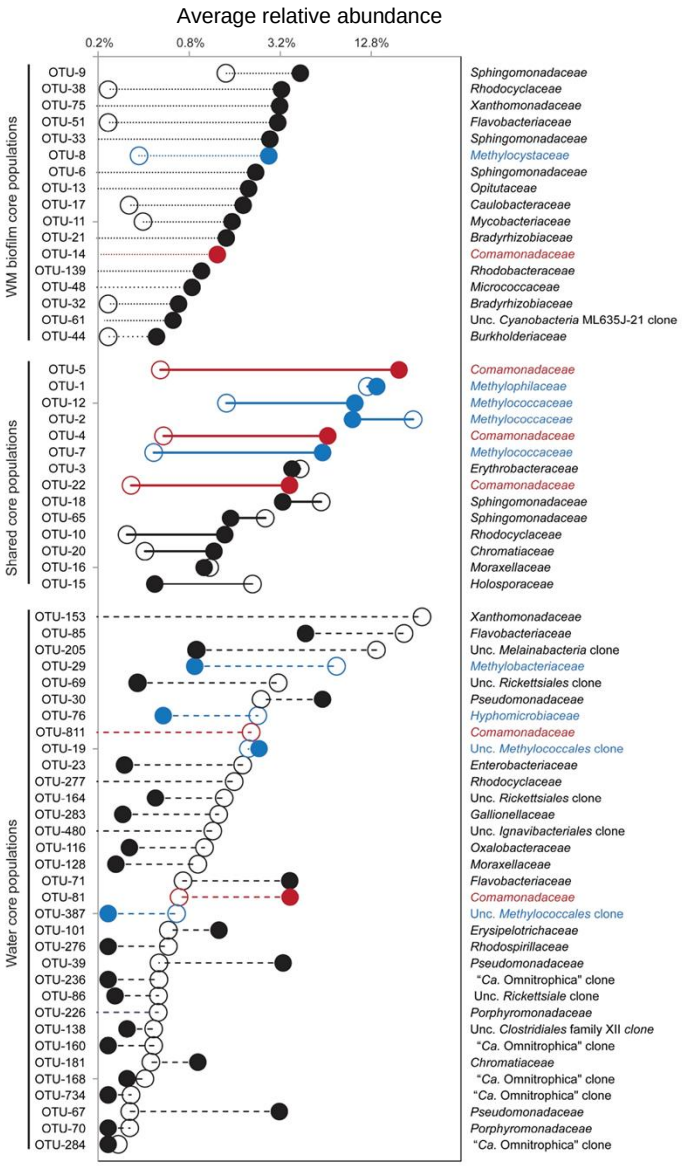


Figure 2.4 Phylogeny and abundance of core OTUs in biofilm and suspended communities. Panel A shows the taxonomy classification of core OTUs to family level and average relative abundance of each OTU in biofilm (closed) and suspended (open) communities. Lines indicate shared (solid), biofilm-only (dotted), and suspension-only (dashed) core populations. Highlighted OTUs are methylotroph- (blue) and *Comamonadaceae*-related (red). Panels B and C are 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 1,000 replicates) are indicated by solid circle, open circles, and open square, respectively.

(B)

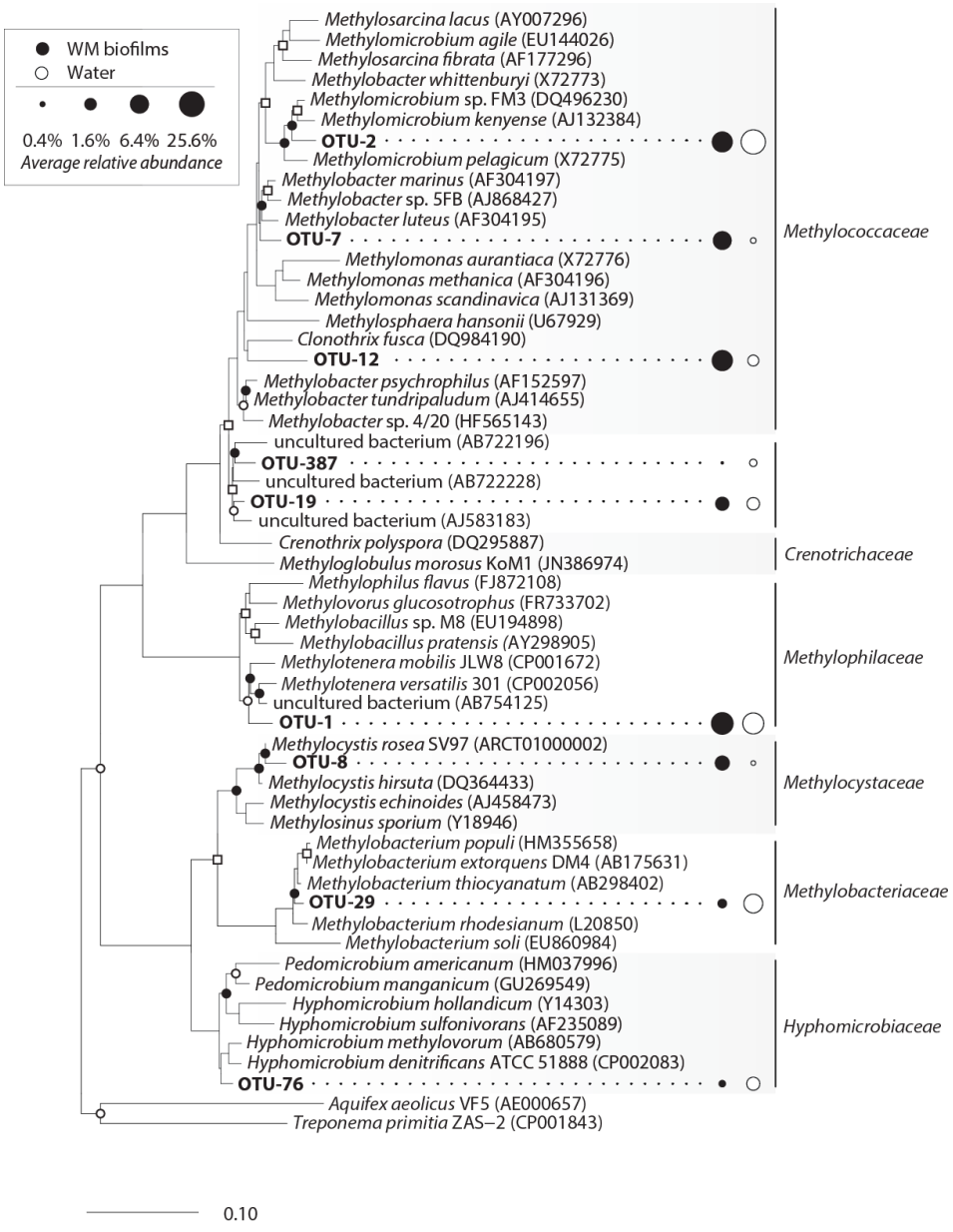


Figure 2.4 (Cont.)

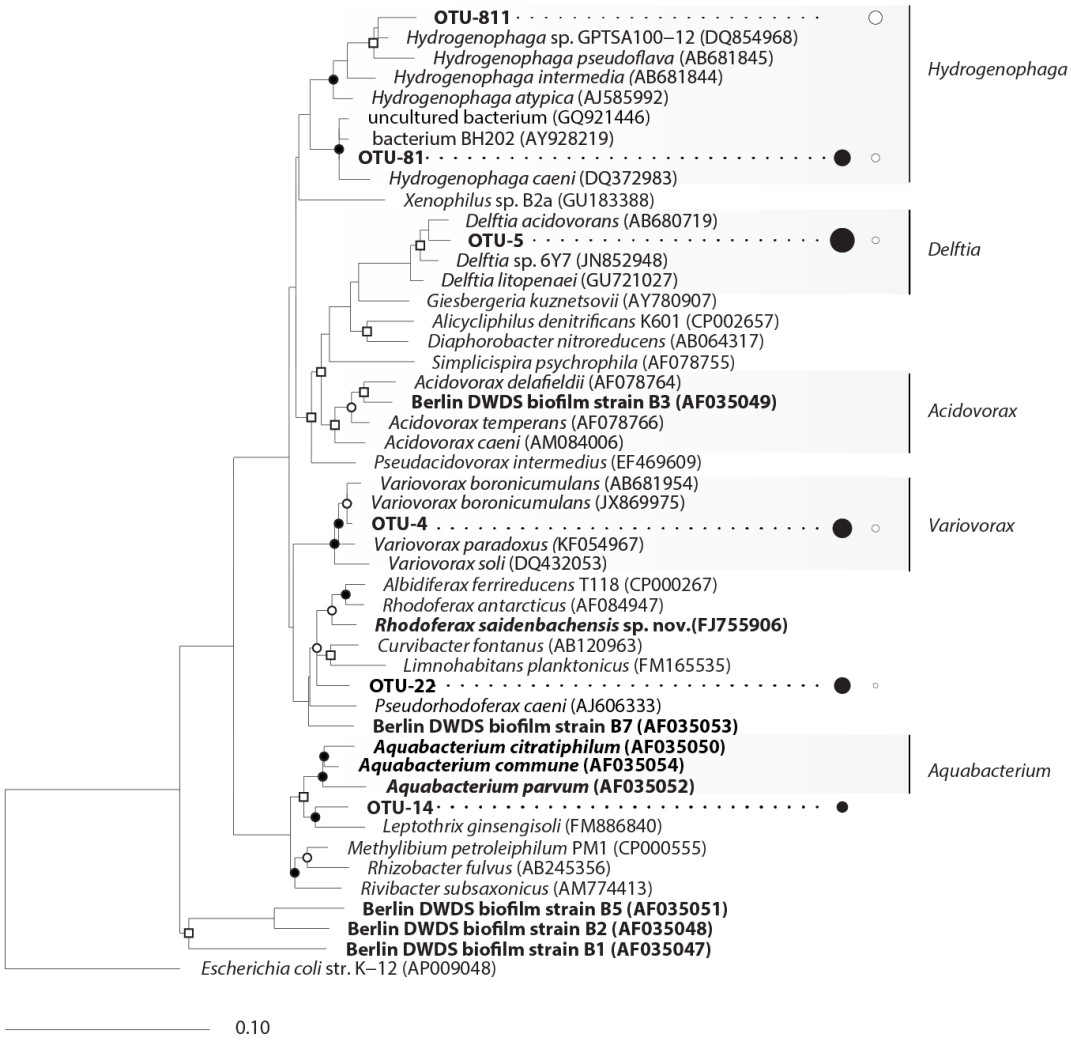


Figure 2.4 (Cont.)

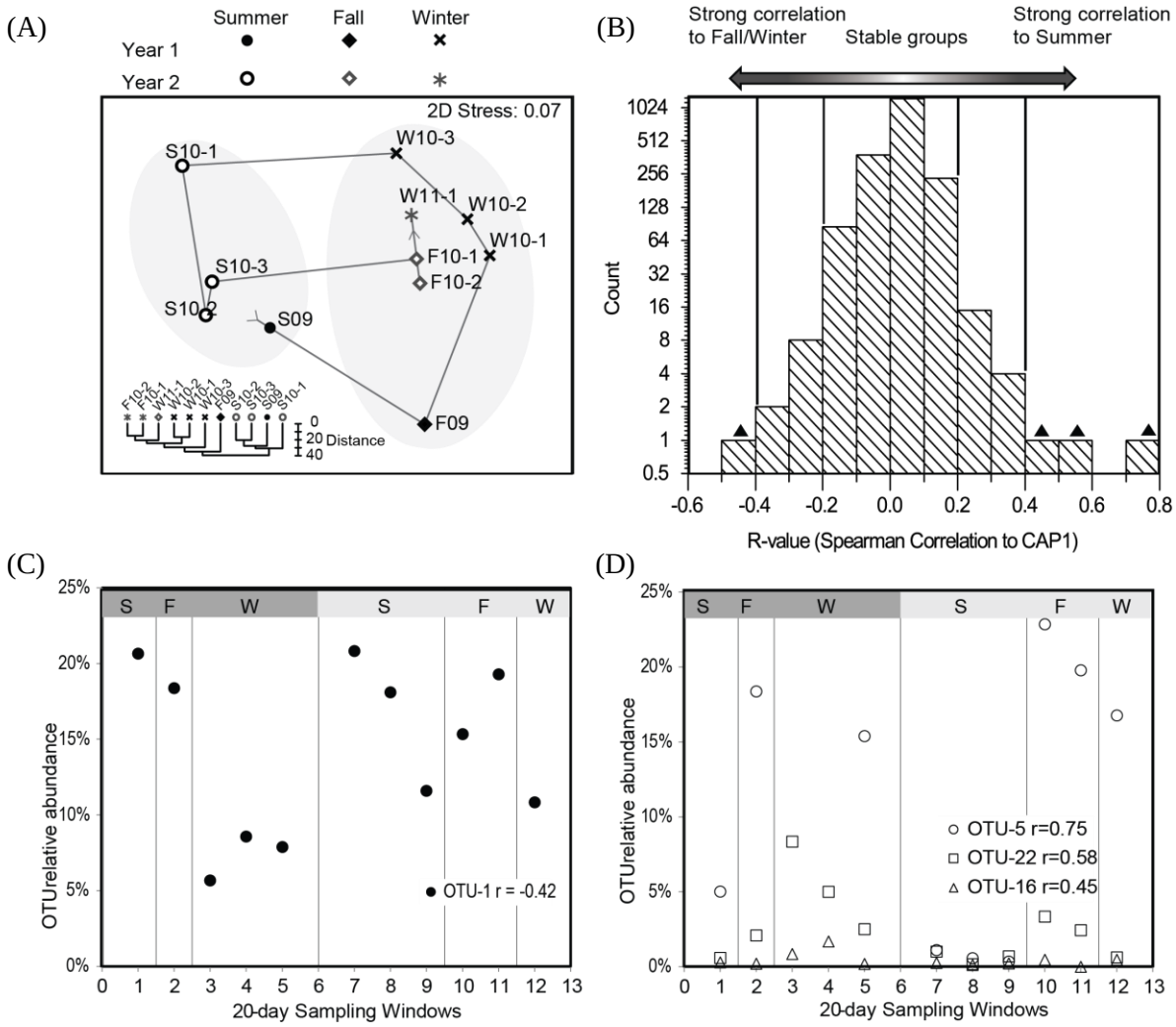


Figure 2.5 Seasonality of biofilm community structure. Panel A shows non-metrical multidimensional scaling for centroids of 20-day sampling windows. Grey eclipses indicate clusters at 40% Bray-Curtis distance. The cluster analysis result is shown as an insert. Panel B shows distribution of seasonal variation represented by spearman coefficient between OTUs and the first canonical principle coordinate constrained by seasonal variation (result of canonical analysis of principle coordinates is shown in Figure S7B). The OTUs strongly correlated to the first CAP axis (Spearman correlation r>0.4 or <-0.4) are plotted against sampling time in panel C (negative correlation) and D (positive correlation).

2.7 Reference

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**:245–53.

Anderson MJ, Gorley RN. (2008). PERMANOVA for PRIMER: guide to software and statistical methods. PRIMER-E Ltd,.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ, *et al.* (2012). Unraveling assembly of stream biofilm communities. *ISME J* **6**:1459–68.

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**:335–6.

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**:477–99.

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

Emerson D, Fleming EJ, McBeth JM. (2010). Iron-oxidizing bacteria: an environmental and genomic perspective. *Annu Rev Microbiol* **64**:561–83.

Fellows I. (2012). Deducer: A Data Analysis GUI for R. Journal of Statistical Software. *J Stat Softw* **49**:1–15.

Felsenstein J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap Author(s): Joseph Felsenstein Source: *Evolution (N Y)* **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**:780–91.

Glaeser SP, Kampfer P. (2014). The Prokaryotes. Springer Berlin Heidelberg: Berlin, Heidelberg doi:10.1007/978-3-642-30197-1.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward D V, Giannoukos G, *et al.* (2011). Chimeric 16S 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**:3530–8.

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**:5631–5.

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**:7856–65.

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 Pt 2**:769–77.

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**:106–11.

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**:591–6.

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**:542–51.

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**:3015–25.

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

Ling F, Liu W-T. (2013). Impact of chloramination on the development of laboratory-grown biofilms fed with filter-pretreated groundwater. *Microbes Environ* **28**:50–7.

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**:124–31.

Magurran AE, Henderson PA. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**:714–6.

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**:6899–907.

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 https://books.google.com/books?id=0Jz2fgfBLIgC&pgis=1 (Accessed December 11, 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.

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.

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**:7537–41.

Swamee PK, Sharma AK. (2008). Design of water supply pipe networks. John Wiley & Sons https://books.google.com/books?id=F5dz28bt5RkC&pgis=1 (Accessed December 11, 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 16S 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**:971–8.

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.

Vaz-Moreira I, Nunes OC, Manaia CM. (2011). Diversity and antibiotic resistance patterns of Sphingomonadaceae isolates from drinking water. *Appl Environ Microbiol* **77**:5697–706.

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**:10624–33.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**:5261–7.

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.

CHAPTER 3 BIOLOGICAL GROWTH AND COMMUNITY STRUCTURE SHIFTS IN TAP WATER DURING STAGNATION IN INDOOR PLUMBING

3.1 Abstract

Indoor water supply systems support the daily water use of humans in built environments and have a direct impact on microbes in contact with humans through water use. This part of water supplies are more prone to frequent and prolonged water stagnation than the municipal water supplies. Currently there is limited understanding on the impact of stagnation in indoor plumbing on microbial community in drinking water. In this study we conducted a stagnation experiment in water supply systems of dormitory buildings and examined how stagnation would interact with the design of plumbing pipelines to influence the composition of microbial communities in tap water. Our results showed that the microbial abundance increased from $<10^3$ cells/mL to $\sim10^5$ cells/mL after week-long stagnation. The cell count significantly depended on the volume of water flowing out of the faucet, indicating an influence from the pipeline structure. The biological growth during stagnation was associated with drastic change in community structure (ANOSIM R= 0.9, p=0.001). Furthermore, within the stagnant water, we observed a spatial differentiation in community composition that correlated to the configuration of water supply pipelines (R=0.508, p=0.001). Other spatial factors did not cause a comparable difference, including building (R=0.039, p=0.03) and floor (R=0.028, p=0.049). Temporal variation did not

significantly influence the community structure (R=0.011, p=0.182), suggesting the communities in the stagnant water, although ephemeral, was composed in a pattern resilient to perturbations from water use.

3.2 Introduction

In modern societies, access to drinking water relies on water treatment and distribution networks. In U.S. and many other parts of the world, residual disinfectants are applied to rigorously treated water in order to protect its microbiological quality during distribution. However, in distribution systems, reactions in the water column and on surfaces can lead to partial or complete decay of residual disinfectants (Biswas *et al.*, 1993), which is associated with elevated growth of heterotrophic microorganisms and microbiological risks (LeChevallier *et al.*, 1996; Carter *et al.*, 2000). Indoor plumbing is the last component of water supply pipelines and can influence the quality of water in contact with humans through drinking, showering, and direct contact. Compared to municipal distribution systems, indoor pipelines are more prone to disinfectant decay and biological growth due to smaller pipelines and more often and extended periods of water stagnation (National Research Council, 2006). Meanwhile, routine monitoring of biological quality in distribution systems is stipulated to collect samples after thoroughly flushing the tap water through indoor plumbing. Thus, although biological growth in indoor plumbing is expected, much is unknown about the magnitudes, neither the effects on community composition.

Although public health risk indicators have been the focus of routine microbiological monitoring in drinking water, the presence of diverse microbial communities has been recognized since standard microbiological techniques and media was used in water quality monitoring (Geldreich *et al.*, 1972). It has been recognized that different stages of treatment and distribution are drastically different in chemistry and flow regimes by design, and thus may present different habitats for microorganisms (Proctor and Hammes, 2015). In indoor water supplies, it is not uncommon to have water stay unused for more than a day. For example, in schools, offices, and public buildings, vacation times may lead to week-long water stagnation in pipelines, which can alter the habitat profoundly through leaching of pipe materials, consumption of disinfectants, and formation of disinfectant byproducts (Proctor and Hammes, 2015). However, there is limited knowledge about the diversity of microbial communities in indoor pipelines, perhaps because these pipelines are usually private properties where rigorous experiments are difficult to conduct. Shower head biofilms have been used as a source of samples, and an elevated abundance of non-tuberculous mycobacteria than municipal drinking water was observed (Feazel *et al.*, 2009). A few studies utilized "first-draw" samples, which were one-liter samples of tap water that had stood motionless in the plumbing pipes for at least 6 hours without flushing the tap, and revealed community structure differentiation from distribution system water with fingerprinting techniques (i.e T-RFLP or DGGE) (Wang *et al.*, 2012; Lautenschlager *et al.*, 2010). There are still knowledge gaps about phylogeny and the spatial variation in the microbial community after stagnation in indoor pipelines.

In this study, we used the cold water lines that supplied for hand-washing basins as a model system to study the effect of stagnation on drinking water microbiota. Three dormitory buildings in University of Illinois, Urbana-Champaign were used for the study. Through controlled access

to these buildings, we obtained post-stagnation samples after a week-long pause of activities. Fresh distribution system water samples were collected at the same faucets before stagnation. Spatial replication was included between and within buildings, and the experiment was conducted twice for temporal replication. Illumina pair-end sequencing was used to profile microbial community and flow cytometry to quantify total cell count. We ask the following two questions: i) how do the stagnant microbial communities in indoor pipelines differentiate from the fresh municipal water in abundance, community composition and structure? ii) Does the configuration of the water supply system influence the microbial community structure?

3.3 Materials and Methods

Study location We conducted the experiment in three 4-storey buildings in Champaign, Illinois. The water supply in the area sourced from groundwater aquifers and treated with a conventional treatment process. The treatment processes, and the diversity of the suspended and biofilm communities in this system has been studied (Hwang *et al.*, 2012b). The system carried stable free chlorine disinfection during the time of this study.

Three residence halls T, V, and S were chosen for their design and availability. In terms of design, these buildings adopted direct cold water supply system, where all potable fittings were supplied with cold water direct from the water main (no water tank involved). It is the most basic type of indoor water supply design and we chose it to eliminate interference from building architectural complexity. These buildings were managed by University Housing at University of Illinois, Urbana-Champaign (UIUC). Through collaboration with University Housing, the access

to water use devices was controlled for none or minimal disruptions to study sites during designed stagnation periods. The T Hall and V Hall were built in 1955 and S Hall in 1963 as part of the on-campus residence at UIUC. T Hall and V Hall were located adjacently and received water supply through the same municipal water distribution pipeline (water main). S Hall was located 1.1 mile away. In each building, we controlled access of the second, third and fourth floor during the experiment.

Water supply system design In a typical building water supply design, there are vertical lines (*i.e.* risers) that delivers tap water upward to different floors, and horizontal lines in each floor, then device-associated pipes for each hand-washing basins. These pipelines are referred to as Branch Level I, II, and II in this study (Figure 3.1A). The sizing of these pipelines was designed based on the water demand as determined by the number of people served. Thus, the diameter decreased from Branch Level I, II, and III; within the Branch Level I, the diameter decreased as the elevation ascended. The difference in diameter further determines the difference in other important physical properties, i.e. the volume of water held in each section of pipelines and the surface area per volume. We accessed the original blueprints and project manual for the studied buildings to acquire design data. Certain details were not included in the project manual due to common knowledge in the trade, and were supplemented by estimation from architects at University Housing. An example calculation on the cold water line of the S Hall is provided in Table 3.1. Level I branches were designed as 2" to 1" from Floor 1 to 4, connected to Level II branches of 3/4"- 1/2", and then Level III branches of 3/8". The specific area in the pipe sections increased as the branch deepens, from 1.05-1.57 cm^{-1} in Level I to 4.2 cm^{-1} in Level III lines. The volume of water in each section can also be estimated based on branch diameter and length. Common to all the floors, the Level III branch (3/8") could hold 21 mL water, and the Level II

branch 1212 mL. The capacity of Level III branches differed by floor and ranged 1389 mL to 3125 mL from 4th floor to 2nd floor.

Water sample design and collection methods We treated the buildings as reactors to study the effect of stagnation on the microbial communities in water. The samples were collected from cold water faucets at the handwashing basins in the dormitory bathrooms. Before the experiment, faucet accessory structure, aerators were removed to eliminate possible interference on the community profiles (Cross et al., 1966). Prior to the designed stagnation, tap water in the studied pipelines were replenished by running tap water for 30 minutes, and then the *pre-stagnation* samples were collected on each floor to represent tap water without the influence of stagnation. After that, the sites were controlled for 5-6 days with controlled access (Figure 3.1B). At the end of the stagnation period, the *post-stagnation* samples were collected at different flow volume by letting tap water flow out at minimum flow rate. Samples were taken as every 100 mL for the first liter and every 500 mL in the 2100-3000 mL for free chlorine and total cell count measurement, and were pooled as first 100 mL, 200 -1000 mL, 1100 -2000 mL, and 2100 -3000 mL for community analysis (hence referred to as S1, S2, S3, and S4). Based on the aforementioned volume calculation, the post-stagnation samples at different volume were used to represent branch levels (Table 3.1). Biological replicates of tap and stagnant samples were taken spatially as one faucet per floor from three floors in each building, and then repeated once for temporal replication. Another replication for inter-faucet difference was designed and taken as three faucets per floor from three floors in one building.

Biomass collection and DNA extraction Water samples were collected at the sampling site with steam-sterilized polypropylene carboys or bottles (Fisher scientific), and transferred to the laboratory immediately, then concentrated by 0.22 μm filter retention in a biological safety hood

prepared by standard sterilization with UV and ethanol wiping. The filters used in this study were prepackaged sterilized polyethersulfone filter units (Millipore). Because of the need for other tests, the filtration volume for pre-stagnation was 40 liter per sample, and the volume of stagnant water samples were 80 mL for S1, 880 mL for S2, and 980 mL for S3 and S4 samples. Filters were removed from the filter unit with sterile scalpels and preserved in sterile falcon tubes -80°C prior to DNA extraction. DNA was extracted using the Schmidt's protocol (Schmidt *et al.*, 1991) and purified with Promega Wizard DNA cleanup system (Promega) and was stored at -80°C. The protocol was selected based on a previous publication that evaluated the different protocols for DNA extraction of drinking water samples from distribution systems (Hwang *et al.*, 2012a).

Illumina Sequencing Sequencing analysis was conducted with Illumina dual-index strategy. The extracted DNA was amplified with the following bacterial specific forward 515F and 806R (Read1-TATGGTAATTGTGTGCCAGCMGCCGCGGTAA; Read2-AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT; Linker -ATTAGAWACCCBDGTAGTCCGGCTGACTGACT) (Kozich *et al.*, 2013). The amplified products were purified with Promega PCR clean-up systems. Sequencing was performed at the W.M. Keck Center, part of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign.

Diversity analysis The pair-end sequences were assembled and quality trimmed in Mothur (Schloss *et al.*, 2009). Following that, unique sequences were picked and aligned to SILVA Gold bacteria alignment. Distances between sequences were then calculated and OTUs were defined at 97% similarity. The taxonomy of unique sequences were classified using a naive Bayesian algorithm (Wang *et al.*, 2007) against most recent SILVA taxonomy at an identity score of 0.8.

Rarefaction curves were generated using Mothur. The sequences were subsampled to the lowest read depth among all samples for diversity analyses. Descriptive analyses on alpha diversity were generated with Origin 7.0 (OriginLab Corporation, Northampton, MA, USA). Community similarity between samples was calculated as Bray-Curtis similarity. Multivariate tests were performed with distance-based methods analysis of similarities (ANOSIM), (Clarke, 1993; Chapman and Underwood, 1999). These tests were performed in the PRIMER-6.0 package (PRIMER-E, Plymouth, UK) according to the authors' manual (Anderson *et al.,* 2008).

Flow cytometry The samples for flow cytometry were collected in gamma-irradiation sterilized falcon tubes (Fisher) and dechlorinated on site with filter-sterilized 1% sodium thiosulfate. The staining and flow cytometry measurement was conducted according to a published method for drinking water samples (Berney *et al.*, 2007; Hammes *et al.*, 2008). The samples were transported to lab within 1 hour of collection, stored in 4ºC refrigerator temporarily, and measured within 8 hours after collection. Bacterial cells were stained with 10 µL/mL SYBR Green (Invitrogen) and incubated in dark at room temperature for 20 minutes. Flow cytometry was conducted with a Partec CyFlow space flow cytometer equipped with a 50mW solid-state laser (488nm). The fluorescence signals were collected with a green fluorescence channel FL1=536mm and was also set as the trigger for the signals. Electronic gating for cell signals was performed with the Flowmax software (Partec). Cell signals were separated from noise by forward fluorescence signals and side scattering (Berney *et al.*, 2007; Hammes *et al.*, 2008).

Water quality measurement Water samples were tested for time-sensitive parameters on-site, including temperature and free chlorine. The free chlorine is measured using the N, N-diethyl-p-phynylenediamine (DPD) chlorine test kit (Hach, CO), a method approved by EPA

for water utilities to monitor chlorine disinfectants. Other water chemistry measured at the laboratory as 0.22-µm-filtrate of the first liter. The measured parameters include inductively coupled plasma (ICP) metals, orthophosphate, anions (nitrate, nitrite, and sulfate), ammonia, non-volatile organic carbon (NVOC), and alkalinity. The chemical analyses were conducted at the Illinois State Water Survey in accordance with USEPA methods.

3.4 Results

Water Chemistry Chemical profiles of pre-stagnation and post-stagnation tap water We tested water quality before and after stagnation (Table 3.2). The pre-stagnation samples showed free chlorine (sodium hypochloride) at 2.2 ± 0.3 mg-Cl/L, which was close to the free chlorine level expected in the distribution system. The level and consistency shows that the replenishing process was able to bring the municipal water into the plumbing lines and set a common starting point for the stagnation test. We observed several changes in the water chemistry profile after stagnation, including significantly lower free chlorine concentrations with partial or complete depletion (0.3 ± 0.3 mg-Cl/L, $p<0.001$), and detection of trace zinc and copper, which were components of brass plumbing materials and fittings immediately connected to the faucets. We also observed increase in water temperature from 19.4 °C in pre-stagnation samples to 29.2 °C in post-stagnation samples. These results are expected for water stagnation and indicate the controlled stagnation test was successful. Copper and lead, the regulation targets for in-building water quality change were below regulated levels (Lead and Copper Rule, lead <15 ppb and copper <1.5 ppm). Other chemical parameters including dissolved carbon, hardness

(sum of Ca and Mg concentrations), and alkalinity (the acid neutralization capacity) showed stability. The concentrations of nitrite, nitrate and orthophosphate were below or close to detection limits.

Plumbing system structure influences on microbial abundance and disinfectant residuals

For samples collected prior to stagnation, we were not able to detect meaningful signals with flow cytometry (detection limit: 1000 cells/mL). In post-stagnation samples, we observed total cell count of 10^4 -10^6 cells/mL (Figure 3.2). Further examination of the post-stagnation samples revealed that cell count depended significantly on the sampling volume, which reflected the influence from the plumbing structure (Figure 3.3A). The average for total cell count decreased by two orders of magnitudes from the first 100 mL (10^6 cells/mL) to the 3100th mL (10^4 cells/mL). When the cell count data from each building was subjected to forward selection of explanatory variables including flow volume, floor, and building in a general linear model, flow volume was the first to choose among all the tested ($p < 2.2\times10^{-16}$ as opposed to $p>0.001$ for other factors tested). Other variables of intermediate significance were floor and building ($p=0.0025$ and $p=0.0031$). Temporal replication was not significant ($p=0.83$), suggesting consistency between the two experimental runs. A linear model with volume as the sole explanatory factor was then constructed, and shown to capture the majority of variation in the full model (adjusted R-square for simplified models versus full model were building-S: 0.79/0.82; Building-T: 0.58/0.81; Building-T: 0.72/0.79).

Along with the observation on microbial abundance, we detected the dependence of disinfectant concentration on plumbing structure (Figure 3.3B). Similarly, flow volume was the predictor that explained the highest variation for free chlorine data, although the pattern is not identical to the cell count data.

In-plumbing stagnation influences on tap water microbiota The 16S rRNA gene Illumina sequencing yielded 3,652,429 reads which were subsequently grouped into 3008 OTUs (329 940 reads and 1784 OTUs after a subsampling to even depths across all samples at 3666 seqs/sample, Figure 3.4A). Both pre-stagnation and post-stagnation samples showed long tails in rank-abundance curves, indicating the presence of rare species in both categories (Figure 3.4B). An overview of the taxonomy classification for dominant taxa (over 1%) is shown in Figure 3.5. Overall, *Proteobacteria* (86%), *Cyanobacteria* (4.4%), *Bacteroidetes* (4.1%), *Planctomycetes* (1.6%), *Actinobacteria* (1.3%) and *Firmicutes* (1.0%) were detected at over 1% in relative abundance. Most of the *Cyanobacteria*-related sequences were classified to "*Candidatus"* Melainabacteria. While these phyla were present in both pre-stagnation and post-stagnation samples, the composition of post-stagnation samples showed dominance by Alpha-, Beta- and Gamma-*Proteobacteria* (Figure 3.5A and 3.5B).

Comparison of community similarity shows that whether the sample was collected before or after stagnation had more influence on the community structure than the building, floor, or experimental run they were collected (Table 3.3). Sample types (before or after stagnation and different volumes after stagnation) had significant and strong influences on the community structure (ANOSIM R=0.508, p=0.001). Building and floor also showed significance in the test (p=0.0049 for floor and p=0.03 for buildings), yet the amount of variance explained was lower in magnitude compared to sample type (R=0.028 for floor and R=0.039 for building). Temporal replication did not significantly affect the community structure (p=0.182), suggesting the post-stagnation community was resilient to the disturbance introduced by water use. In a separate experiment carried out in Building S to examine the inter-faucet differences, the sample type

(pre-stagnation and different collections of the post-stagnation samples) also showed significant difference (p=0.002), whereas the faucet difference was not (p=0.21).

Further examination on the sample type individually showed that pre-stagnation and post-stagnation samples differed significantly (R=0.9 and p=0.001, Figure 3.6A). Within the post-stagnation samples, samples collected at the first 100 mL (S1) which represented mixture from Branch Level II and III, significantly differed from the rest, as shown in Figure 3.6B and D (p=0.001 for all pairwise comparisons), and to a larger extent when compared against samples representing Branch Level I (R>0.9 for S1 vs. S3 and S1vs. S4, Figure 3.6B) than to Branch Level II (R=0.264, Figure 3.6D). The samples represented Branch Level II also significantly differed from those representing Level III (p=0.001 for S2 vs. S3 and S2 vs. S4, Figure 3.6C), but to a lesser extent (S2 vs. S3, R=0.271; S2 vs. S4, R=0.333). The samples collected within Branch Level I were not significantly different (S3 vs. S4 p=0.857 and R<0, Figure 3.6E).

In addition to the differentiation in community similarity described above, pre-stagnation and post-stagnation samples also exhibited differences in alpha diversity. Pre-stagnation samples showed higher evenness than post-stagnation samples (Table B.2). Within post-stagnation samples, samples representing the distal ends (Stagnant-1 and Stagnant-2) exhibited higher evenness than the main lines (Stagnant-3 and Stagnant-4). Diversity indices based on dominance including Berger-Parker index, non-parametric Shannon index, and inverse Simpson index showed similar trends (Figure 3.7, Table 3.4).

OTUs associated with the pre-stagnation community and post-stagnation communities from different branch levels Based on the differentiation of community structure described above, we further examined correlation between OTUs and stagnation. By applying Spearman

correlation to the constrained canonical coordinates based on sample types, we grouped OTUs by having correlation with pre-stagnation, branch level I (S3 and S4), or branch level II/III (S1). The OTUs exhibited varying levels of correlation. We provided the level of correlation for abundant OTUs (>1% relative abundance in at least one category) together with their taxonomy classification in Table 3.5. Results are visualized as by the size and partition sequences in Figure 3.8.

Amongst OTUs that correlated to the pre-stagnation condition, the more abundant ones related to Melainabacteria (OTU-25), *Hyphomicrobium* (OTU-29), and *Sphingomonas* (OTU-9). Many other OTUs also correlated to the pre-stagnation condition and were classified to *Nitrosomonas* (OTU-6), *Pseudomonas* (OTU-7 and OTU-10), *Comamonadaceae* (OTU-3, -28, and -8), *Cytophagaceae* (OTU-11), and uncultured *Planctomycetes* (OTU-12 and OTU-121).

For post-stagnation conditions, OTU-5 related to *Methylocystis* and did not show strong correlation to branch levels. Other OTUs showed correlation to different branch levels. Among the more abundant ones, OTU-2 related to *Porphyrobacter* and correlated to branch level I. More diverse OTUs showed correlation to branch level II/III, including those classified to *Dechloromonas* (OTU-1), *Methylophilus* (OTU-4), *Acinetobacter* (OTU-13), and *Denitromonas* (OTU-47).

3.5 Discussion

By conducting controlled water stagnation test in situ, this study has shown that microbial communities distinctive from municipal drinking water formed in the habitat provided by indoor plumbing systems. Within stagnant communities, variation in microbial abundance and community structure correlated to branching of the plumbing systems. The pattern was robust to spatial and temporal replication.

Similar to our previous studies in the municipal water distribution systems, we observed taxa related to methanotrophy, methylotrophy and aerobic heterotrophy in the building plumbing systems. In specific, we detected an increase in relative abundance after stagnation for taxa related to obligate methanotrophy (*Methylocystis* spp.), obligate or restricted facultative methylotrophy (*Methylophilus* spp.), and facultative methylotrophy (*Acinetobacter* spp.) (Table 3.5). These taxa were detected in our previous study as the core populations in the biofilm or suspended communities in the local municipal distribution system which sourced from a methanogenic aquifer. Our new results suggest that the source water influence on microbial community in distribution can persist to the indoor water supply.

Despite this similarity, our results suggest that community composition in the indoor plumbing cannot be extrapolated from distribution systems. In the indoor plumbing, the community composition strongly correlated to the deepening of the pipeline branches (Figure 3.6B). For the distribution system, it has been shown that the composition of microbial communities tied closely to source water chemistry and treatment processes (Hwang *et al.*, 2012b; Lautenschlager *et al.*, 2013; Pinto *et al.*, 2012), but the spatial effect, which has been explicitly examined in two

independent studies, has not shown statistical meaningful patterns (Hwang *et al.*, 2012b; Pinto *et al.*, 2012).

Considering the gradient in chlorine along the flow of stagnant water (Figure 3.3), it is tempting to conclude that this variation is driven by selection. However, the community evenness shows an increase towards the faucet end, thus, community assembly driven by selection is not likely. It is possible that contact with biofilms has contributed to the diversity in stagnant community, though further studies are needed to confirm this influence.

Our findings hold implications for drinking water treatment technology and management in multiple ways. As mentioned previously in the introduction, currently there is no monitoring program targeting at the change of biological quality in water after entering buildings. Our results showed that stagnation caused microbial abundance and community structure to change profoundly from the municipal water. It highlights the need to avoid long-term stagnation in buildings. One consideration is that in green buildings, water age is usually longer than regular buildings and stagnation is promoted by water saving features, thus the impact of long-term stagnation would probably be amplified. Our repeated observation of the microbial abundance suggests that extent of growth in indoor plumbing may be predictable given certain size and pipe age information. Upon further studies, this could be valuable for predicting managing risks and aid regulation for microbiological quality change in indoor water supplies.

3.6 Tables and Figures

Table 3.1 Pipeline diameter, volume and specific area calculation based on Building S

Branch Type	Internal diameter		Volume	Specific Area	Material
Level I	4F	1"	1389 mL	1.57 cm-1	Galvanized steel
	3F	1-1/4"	2170 mL	1.26 cm-1	
	2F	1-1/2"	3125 mL	1.05 cm-1	
Level II	1/2", 3/4",1"		1212 mL*	1.95 cm-1**	
Level III	3/8"		21 mL	4.2 cm-1	Brass

* total volume including all connecting lines

** average for all connecting lines

Table 3.2 Physical and chemical property of water samples

	Fresh	Stagnant
Sensitive parameters		
Temperature (°C)	19.4 ± 0.8	29.2 ± 1.0
ClO^- (mg/L)	2.2 ± 0.3	0.3 ± 0.3
Cu (mg/L) *	n.d. ~ 0.005	0.018 ± .006
Zn (mg/L) *	n.d.	0.022 ± 0.008
Stable parameter		
Ca (mg/L)	13.3 ± 0.6	
Mg (mg/L)	12.6 ± 0.8	
pH	8.5 ± 0.1	
Alkalinity (mg/L)	134.8 ± 3.0	
o-PO4 (mg/L)	0.019 ± 0.009	
Nitrate (mg-N/L)	<0.07	
Nitrite (mg-N/L)	<0.02	
NVOC(mg/L)	1.74 ± 0.51	
Pb (ug/L)	<2	

* Detection limit for Cu and Zn are 0.8 ug/L and 9.7 ug/L repectively

Table 3.3 Community structure differentiation tested by ANOSIM

ANOSIM		
Factor	R	p
Test 1: Building T, V, S N=90		
Sample Type	0.508	0.001
Floor	0.028	0.049
Experimental Run	0.011	0.182
Building	0.039	0.03
Test 2: Building S N=39		
Faucet	0.044	0.21
Sample type	0.382	0.002

Table 3.4 Variation of Evenness and diversity estimators among pre-stagnation and post-stagnation samples

	Simpson index-based measure of evenness	Shannon index-based measure of evenness	Heip's metric of community evenness	Inverse Berger-Parker index	Non-parametric Shannon index	Inverse Simpson index
Fresh	0.104	0.660	0.221	4.06	3.27	14.13
Stagnant-1	0.113	0.513	0.140	3.00	2.05	5.63
Stagnant-2	0.084	0.417	0.100	2.13	1.66	4.10
Stagnant-3	0.043	0.301	0.050	1.45	1.28	2.35
Stagnant-4	0.027	0.291	0.035	1.40	1.35	2.00

Table 3.5 Classification of OTUs over 1% and their spearman correlation to sample types

	Conditions	Spearman Correlation (++++ >0.8; +++ >0.6; ++ >0.4; + >0.2)
Post-stagnation		
OTU-5	*Methylocystis*	+
	Branch Level II and III (S1)	
OTU-1	*Dechloromonas*	++++
OTU-4	*Methylophilus*	+++
OTU-13	*Acinetobacter*	
OTU-21	*Blastomonas*	+++
OTU-31	*Pseudorhodoferax*	+
OTU-38	*Lacibacter*	+++
OTU-47	*Denitratisoma*	+++
OTU-53	*Ohtaekwangia*	++
	Branch Level I (S3 and S4)	
OTU-2	*Porphyrobacter*	++++
OTU-22	*Candidatus Melainabacteria*	+
OTU-78	*Stenotrophomonas*	+
Pre-stagnation		
OTU-3	Unc. Comamonadaceae	++
OTU-6	*Nitrosomonas*	++++
OTU-9	*Sphingomonas*	+
OTU-10	*Pseudomonas*	+++
OTU-11	Unc. *Cytophagaceae*	+++
OTU-12	Unc. *Planctomycetes*	+++
OTU-15	*Rhodanobacter*	++++
OTU-25	*"Candidatus* Melainabacteria"	+++
OTU-28	*Schlegelella*	
OTU-29	*Hyphomicrobium*	++
OTU-121	Unc. *Phycisphaeraceae*	++
OTU-7	*Pseudomonas*	+
OTU-8	Unc. *Comamonadaceae*	+

(A) (B)

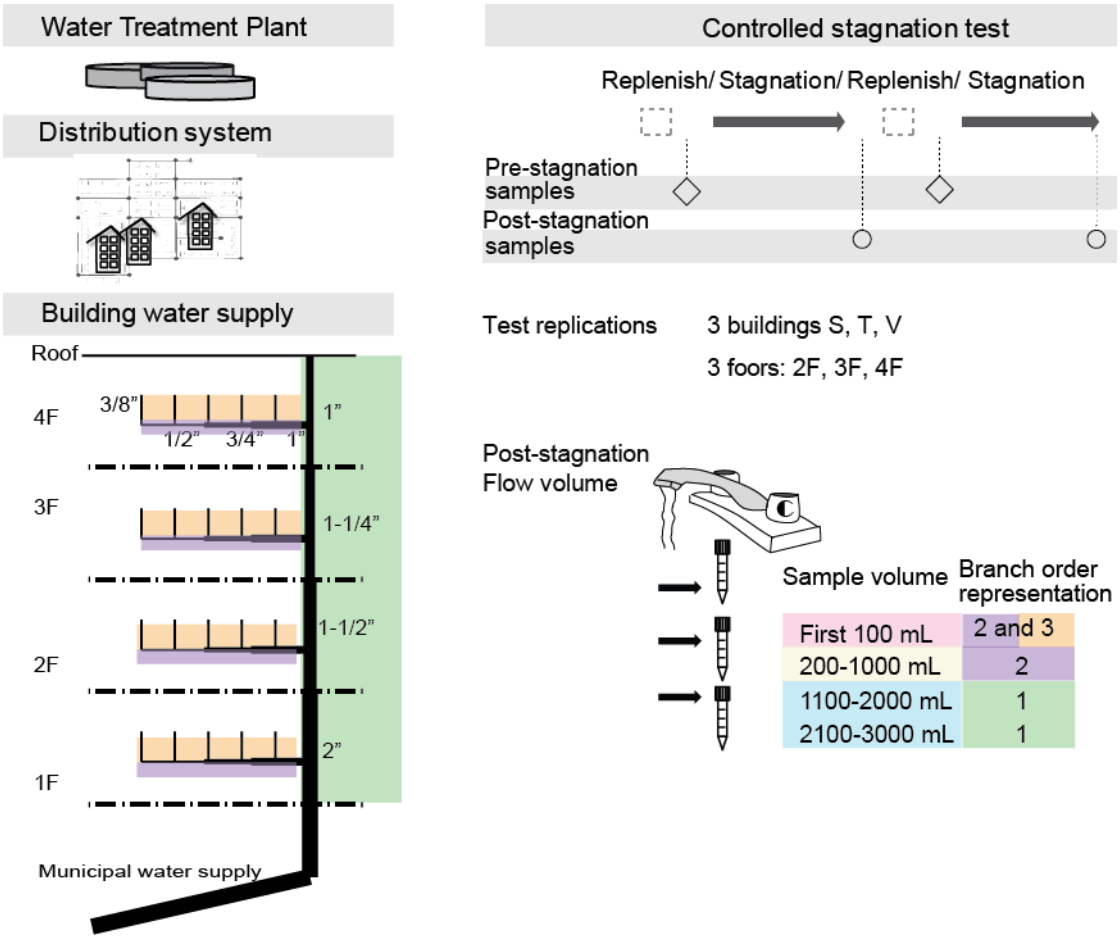


Figure 3.1 Building water supply system schematics and experimental design. Indoor water supply systems are usually designed as a branched structure with a vertical line delivering water upward (green), horizontal lines in each floor (purple) and device-connecting lines (orange) in the end (referred to as branch 1, 2 and 3 in the table associated). These branches are different in size because difference in number of people supplied. This difference in diameter determines the volume held in each level of branches as well as an increase in specific area as it branches deeper (A). Samples were collected before and after stagnation from 3 buildings and 3 floors with temporal replication (B). At each site, cell count and biomass were sampled at different flow volume representing the branch levels.

(A) (B)

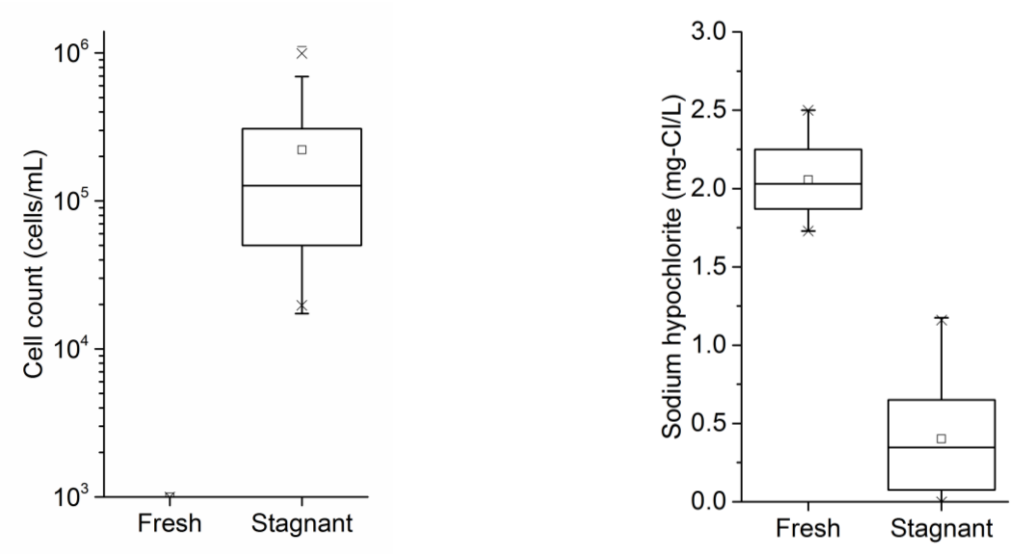
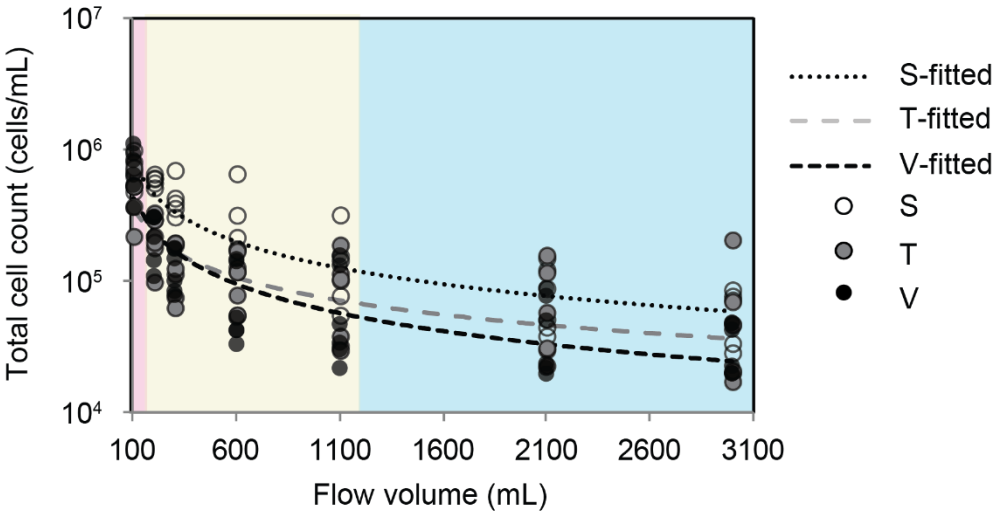


Figure 3.2 Comparison of microbial abundance (A) and disinfectant concentrations (B) in before and after stagnation. $n_{fresh} = 18$, $n_{stagnant} = 162$ (chlorine) or 126 (cell count).

(A)



(B)

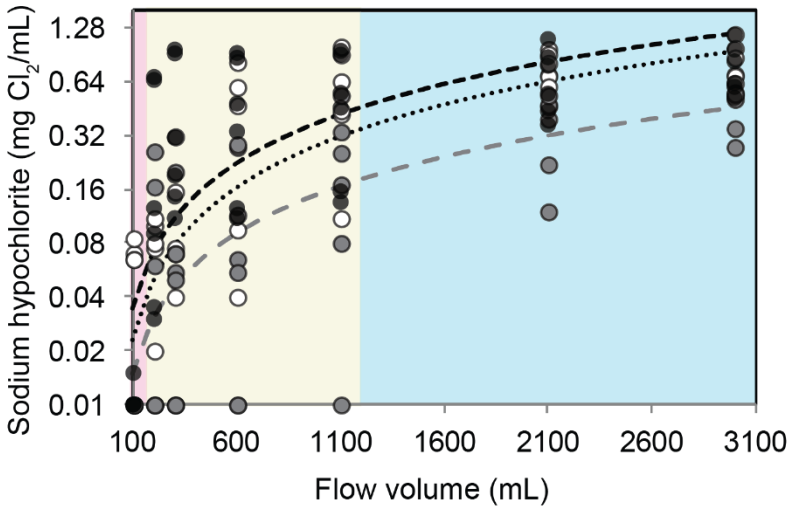
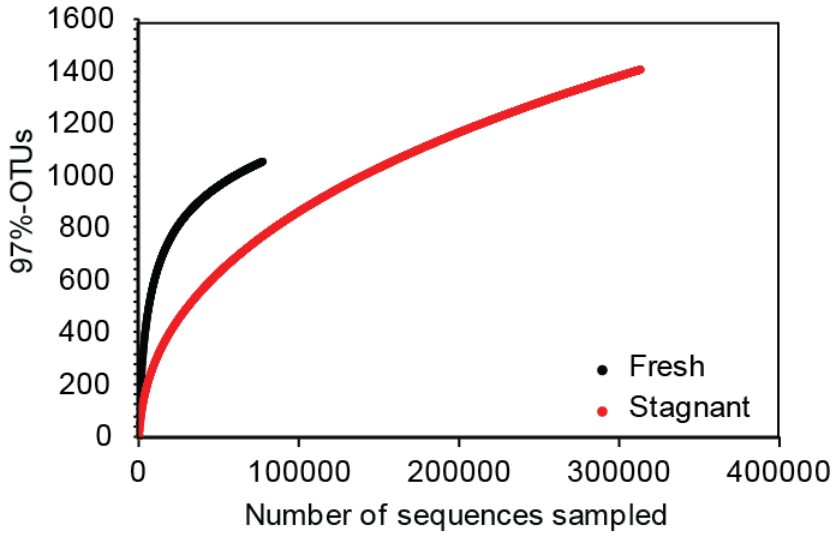


Figure 3.3 Dependence of residual disinfectants (A) and microbial abundance (B) on flow volume.

(A)



(B)

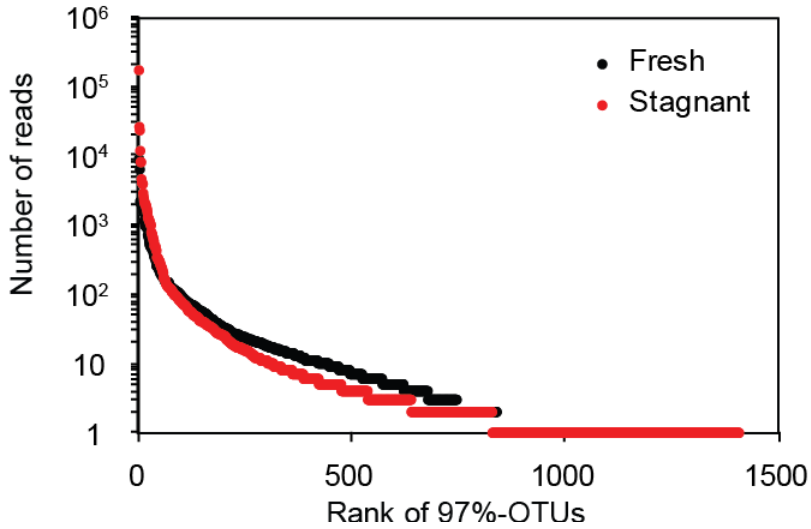
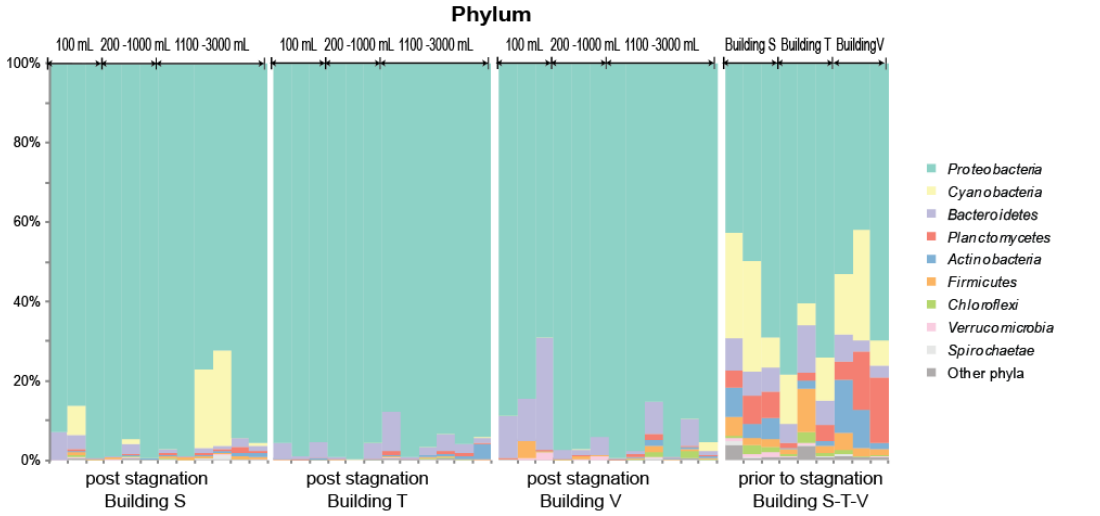


Figure 3.4 Rarefaction curves (A) and relative abundance (B) of pre-stagnation and post-stagnation communities.

(A)



(B)

(C)

Figure 3.5 Community composition comparison between post-stagnation and pre-stagnation samples. Phyla over 0.1% (A) and classes (B) and orders (C) over 1% are shown.

(A)

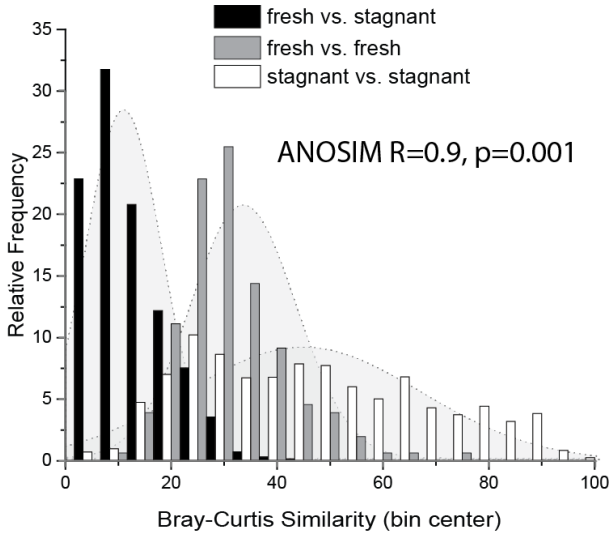


Figure 3.6 Differentiation of community structures before and after stagnation and within stagnation samples. (A) Histograms of Bray-Curtis similarities (bars) and regression to normal distribution (shades) within and between pre-stagnation and post-stagnation samples. (B) - (E) Distribution of community similarities in the first-100 mL (S1), 200-1000 mL (S2), 1100-2000 mL (S3) and 2100-3000 mL (S4) and between-group distances. ANOSIM test results indicate significant difference between pre-stagnation and post-stagnation samples. Within post-stagnation samples, test results indicates significant difference between branch level 3 (S2 and S3) and branch level 1 and 2 (S1 and S2), moderate difference between S1 and S2, but non-significant within branch level 3.

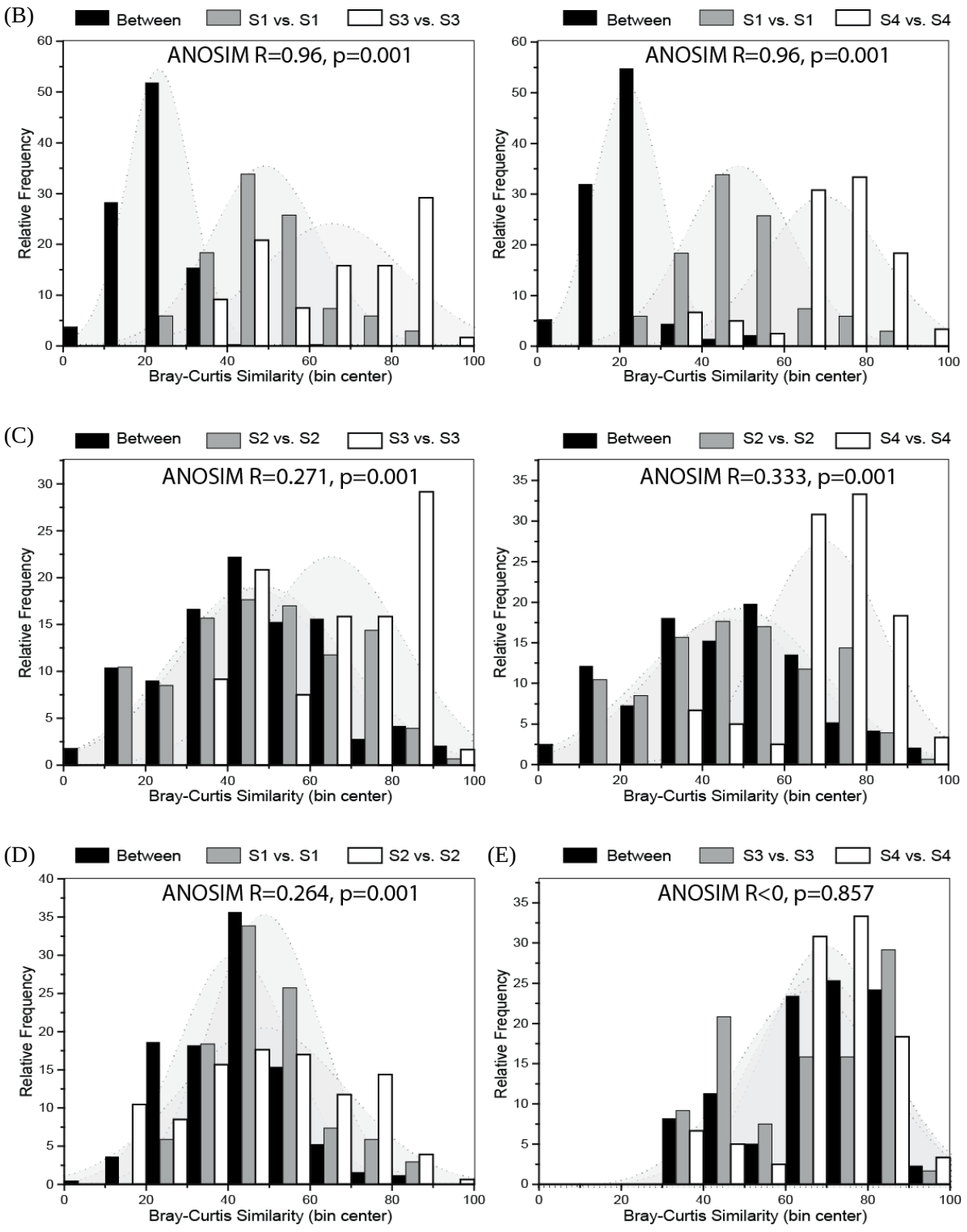


Figure 3.6 (Cont.)

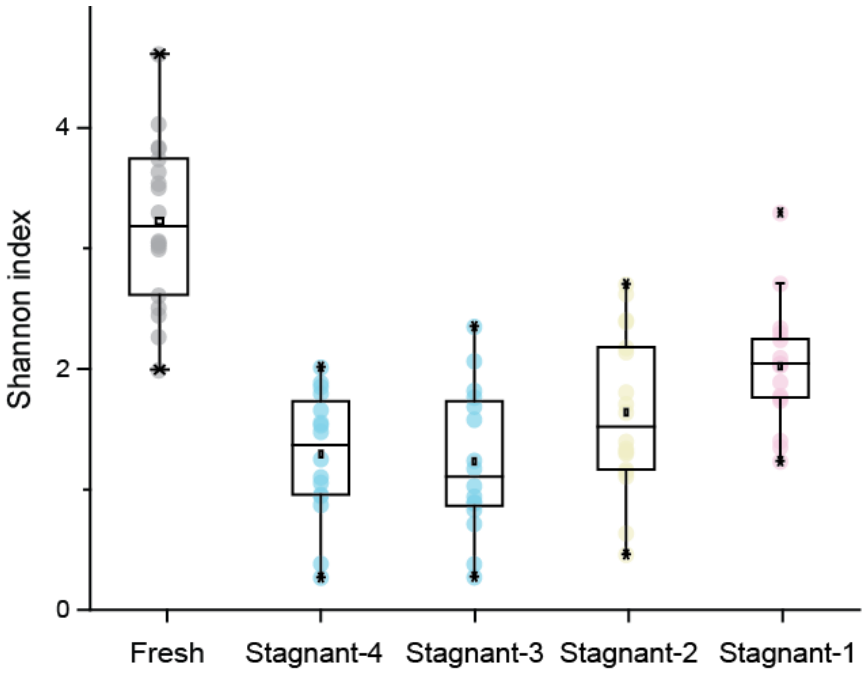


Figure 3.7 Comparison of community evenness for samples before and after stagnation, and between different stagnant samples. Stagnant-1: first-100 mL; Stagnant-2: 200-1000 mL ; Stagnant-3: 1100-2000 mL; Stagnant-4: 2100-3000 mL.

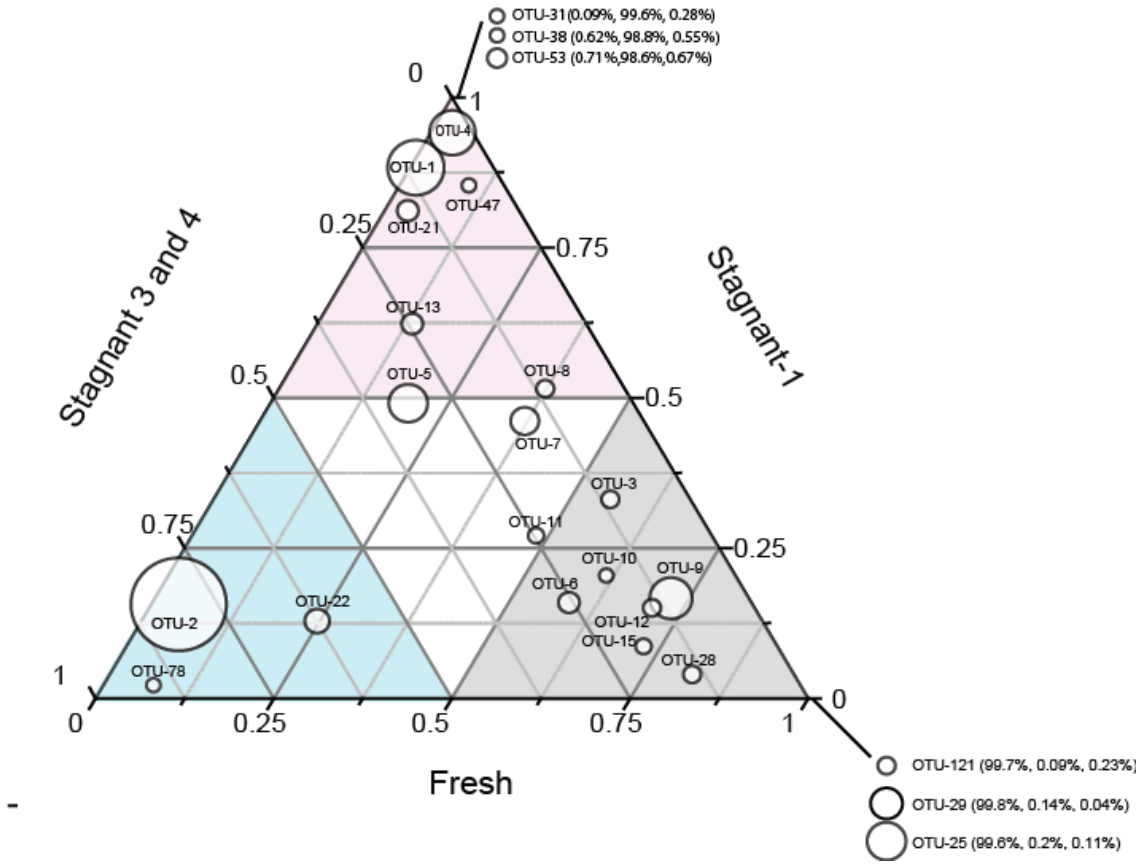


Figure 3.8 Characteristics OTUs before and after stagnation. Stagnant-1: first-100 mL; Stagnant 3 and 4: pooled community from 1100-2000 mL and 2100-3000 mL.

3.7 Reference

Anderson MJ, Gorley RN, Clarke KR. (2008). PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. *2008* 214.

Berney M, Hammes F, Bosshard F, Weilenmann H-U, Egli T. (2007). Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. *Appl Environ Microbiol* **73**:3283–90.

Biswas P, Lu C, Clark RM. (1993). A model for chlorine concentration decay in pipes. *Water Res* **27**:1715–1724.

Carter JT, Rice EW, Buchberger SG, Lee Y. (2000). Relationships between levels of heterotrophic bacteria and water quality parameters in a drinking water distribution system. *Water Res* **34**:1495–1502.

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.

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

Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. (2009). Opportunistic pathogens enriched in showerhead biofilms. *Proc Natl Acad Sci* **106**:16393–16399.

Geldreich EE, Nash HD, Reasoner DJ, Taylor RH. (1972). Necessity of controlling bacterial populations in potable waters: community water supply. *Amer Water Work Ass J* **64**:596–602.

Hammes F, Berney M, Wang Y, Vital M, Köster O, Egli T. (2008). Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Res* **42**:269–77.

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**:7856–65.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**:5112–20.

Lautenschlager K, Boon N, Wang Y, Egli T, Hammes F. (2010). Overnight stagnation of drinking water in household taps induces microbial growth and changes in community composition. *Water Res* **44**:4868–77.

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**:3015–25.

LeChevallier MW, Welch NJ, Smith DB. (1996). Full-scale studies of factors related to coliform regrowth in drinking water. *Appl Environ Microbiol* **62**:2201–11.

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

Pinto AJ, Xi C, Raskin L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environ Sci Technol* **46**:8851–9.

Proctor CR, Hammes F. (2015). Drinking water microbiology—from measurement to management. *Curr Opin Biotechnol* **33**:87–94.

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**:7537–41.

Schmidt TM, DeLong EF, Pace NR. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* **173**:4371–8.

Wang H, Edwards M, Falkinham JO, Pruden A. (2012). Molecular survey of the occurrence of Legionella spp., Mycobacterium spp., Pseudomonas aeruginosa, and amoeba hosts in two chloraminated drinking water distribution systems. *Appl Environ Microbiol* **78**:6285–94.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**:5261–7.

CHAPTER 4 UNDERSTANDING THE ROLES OF BIOFILMS AND MUNICIPAL TAP WATER IN THE ASSEMBLY OF MICROBIAL COMMUNITIES IN INDOOR WATER SUPPLIES

4.1 Introduction

In drinking water treatment process design, microorganisms are usually considered as a kind of contaminants to be eliminated by disinfection. To control biological growth during water distribution, applying certain levels of residual disinfectants to distribution systems is common (USEPA, 2016). In the previous chapter, we have reported that in the case of paused water use in buildings, residual disinfectants in tap water would be partially or completed consumed, which was associated with biological growth. The stagnant water formed a community whose composition was distinctive from the fresh municipal water. Meanwhile, our experiments showed that community structure within the stagnant water communities correlated to the configuration of the water supply pipelines. These observations were reproducible in temporal and spatial replicates, suggesting it may be possible to track and manage the communities in building water supplies. To do that, a better understanding on the mechanisms shaping the communities is crucial.

The microbial communities in a building water supply can be viewed as a kind of "mainland-island" system (MacArthur and Wilson, 1967; Hubbell, 2001). Two kinds of

"mainland" – i) the fresh tap water distributed by the utility and ii) the biofilms already present in the inner surfaces of the pipelines, are possible sources contributing to the "islands" -- the community formed after stagnation. When tap water is drawn from the faucet, fresh municipal water flows into the building; along with the fluid, microbes in the suspension of municipal water can be brought in to the buildings. When there is no water use, water flow is stopped, which makes time for the microbes from the pipeline biofilms to migrate into the suspended community. With an extended stagnation time, the individuals in the island can undergo demographic changes, i.e., random births and deaths. These processes can contribute to the community diversity in the islands.

In a system where dispersal and random births and deaths are the main contributors to diversity, the species present on the island may follow the prediction from the Unified Neutral Theory of Biodiversity (UNTB, Hubbell, 2001), where trophically similar groups of species are assumed to be demographically equivalent to birth, death, immigration and speciation. Under this assumption, the more abundant groups in the mainland are more likely to be detected in the island (Sloan *et al.*, 2006). Although it is difficult for this assumption to be fully met in natural environments, UNTB provides a null hypothesis to test against. It has also provided excellent fitting in a variety of macro- and microbial communities (Burbrink *et al.*, 2015; Venkataraman *et al.*, 2015; Burns *et al.*, 2015).

In the present study, we examined the ability of UNTB in explaining the community assembly in indoor water supply. We then examined the link between the pipeline configuration and the contribution of the two mainland communities, the biofilms and fresh municipal water. The stagnation experiment in Chapter 3 provides an interesting data set to test on. In the experiment, we created an extreme condition where the tap water in building water supplies were thoroughly replenished. That can be used to approximate the mainland community in fresh municipal water. Following there was no water use in the faucets where experiments were conducted, which would allow migration from biofilms to the suspended community to happen. Due to the difficulty in extracting fresh biofilms from a live building, the biofilm data from the water meters in the same distribution system (described in Chapter 2) were used as a surrogate for the mainland community in biofilms.

4.2 Materials and Methods

Data To characterize the microbial communities developed in indoor plumbing pipelines, we used previously generated 16S rRNA gene sequences from a week-long *in situ* stagnation experiment performed at dormitory buildings in University of Illinois, Urbana-Champaign. By allowing continuous flow of the stagnant water from a faucet and sampling in fractions (this included the first 100mL, 200-1000 mL, 1100-2000 mL, and 2100-3000 mL), we sought to collect samples that represent the spatial gradient present in pipeline sections at different

distances to the faucet. The sample set was replicated in 3 different floors and buildings, as well as two sampling time points. We also used fresh tap water samples from the same buildings. Biofilms in direct contact with suspended communities in indoor plumbing are pipelines inside building walls that were not practical to sample, thus we used water meter biofilms as a proxy.

Sequence processing All sequences were curated with Mothur v1.33.3 (Schloss et al., 2009). The sequences from the dormitory experiment have been assembled into contigs assembled from Illumina pair-ended sequences targeting the V4 region (515F/806R). The water meter biofilm sequences were previously denoised sequences targeting V4-V5 region (515F/907R) and were trimmed (using pcr.seqs command in Mothur) to the same size as the dormitory samples. Following that, unique sequences were picked and aligned to SILVA Gold bacteria alignment (Quast *et al.*, 2013). Unique sequences were used for OTU picking at 97% similarity. Taxonomy classification was conducted using the SILVA online server. The sequences were subsampled to the lowest depth (1067 seqs/sample) to achieve an even sampling depth for diversity analysis.

Testing the effects of biofilm and fresh tap sources Community dissimilarity was calculated in Mothur using Bray-Curtis dissimilarity. The distance matrices were used for hypothesis testing with ANOSIM (Chapman and Underwood, 1999). Right-tailed P-values for each ANOSIM test were computed using 1000 iterations.

The Sloan version of neutral model was used to explore the role of neutral processes in the assembly of indoor plumbing water communities (Sloan *et al.*, 2006). We assumed the mainland

community to be stable (no speciation), because the experiment time was relatively short. The stagnant water communities from each volume fraction were treated as the local communities, and the biofilm or fresh tap water communities as the metacommunities, or the source communities. The fitting was performed with a non-linear regression package, Minpack.lm, in R (Elzhov *et al.*, 2013; Burns *et al.*, 2015). The goodness of fit was examined with R-squared. The fitting of the neutral model was compared with binomial distribution using Akaike Information Criterion (Sloan *et al.*, 2007; Spiess and Neumeyer, 2010). Confidence intervals were calculated with the Hmisc package in R (Harrell, 2015).

SourceTracker (Knights *et al.*, 2011) was used to identify sources of OTUs in each indoor plumbing sample. Each stagnant sample was designated as a "sink", while the biofilm communities and fresh tap water samples were used as potential sources. An unknown source was also included by default in the SourceTracker program. The probability for biofilm or fresh tap source was calculated for the communities and each OTU.

4.3 Results

Composition of biofilm and suspended communities Combining the data from previous sequencing efforts yielded 3,569,318 reads in total. We subsampled the sequences to an even depth (1,067 sequences/sample), and these sequences were classified to 1,466 OTUs at 97%

sequence similarity. To examine the overall composition of biofilms, fresh and stagnant water, we first treated each category as one sample and subsampled to the same depth (11,737 sequences/category). The results showed that these three categories largely overlapped (Figure 4.1A), and shared more than 50% of the sequence reads in common (Figure 4.1B). We then compared individual samples across categories using the Bray-Curtis similarity. The samples from biofilms, fresh and stagnant water formed distinctive clusters in nMDS (Figure 4.2) and were significantly different in ANOSIM tests ($p<0.001$, R=-0.3-0.94). A closer examination on the subcategories of stagnant samples showed that the samples collected closer to the distal ends of the pipeline where pipe diameters were smaller (i.e. the first 100 mL and the 200-1000th mL), clustered closer to the biofilm communities. This trend was also shown in the ANOSIM test, where the R values of the pairwise tests increased as the samples represented locations further away from the faucets (Table 4.1), indicating higher degrees of differences.

Contribution of the neutral processes to stagnant water communities First, we considered the fresh tap water as a potential metacommunity for the stagnant communities. The fresh tap water community represents the starting point of the stagnation process, and can shape the stagnant communities through random growths or deaths within the community members. With the goodness-of-fit test (R^2), it was shown that the neutral processes from fresh tap water can well explain the frequency distribution of taxa in stagnant communities (Figure 4.3A to 4.3D). However, the fit of the neutral model varied in stagnant communities from different

locations in the pipelines, and showed an overall decrease as samples approached the faucet (Figure 4.4A).

We then considered the biofilm communities as another metacommunity. Biofilms on pipe walls surround the suspended communities in indoor plumbing. This community can shape the stagnant communities first through dispersal into the suspended phase, and then random births and deaths. The models with the biofilm community as the metacommunity provides fitting comparable to the fresh tap water only in the stagnant communities representing locations close to the faucets (the first 100 mL shown in Figure 4.3E and the 200-1000th mL shown in Figure 4.3F), but not the samples representing the pipeline farther away (1100-2000th mL and 2100-3000th mL, shown in Figure 4.3G and 4.3H). Note that we used the biofilm community data from water meters as a surrogate for pipeline communities, hence the estimated contribution of biofilms is likely to be on the conservative side.

In both scenarios, the neutral model provides better fit than binomial distribution (Figure 4.4C), suggesting that the neutral processes explains more of the community composition than random sampling of the source communities.

Whole-community source tracking as a sensor for stagnation level Is it possible to apply the knowledge from community assembly to develop water quality monitoring technologies? We applied a machine-learning based method, SourceTracker, to demonstrate an application. This method allows source tracking on whole-community level as well as for each taxon (Figure

4.3A). Each OTU was assigned a source community at a probability (Figure 4.3D), and the community-wide source assignment was achieved through the accumulation of individual OTUs. With fresh tap water and biofilm samples as the potential sources, the communities from all the stagnant samples showed mixed source assignment. The results showed certain degree of variation, however, the overall trend suggests fresh tap water as a primary source for the stagnant water samples representing the main lines (subcategories 1100-2000th mL and 2100-3000th mL), and the biofilm community a source for samples from the distal ends (Figure 4.3B and 4.3C).

4.4 Discussion

At the start of this chapter, we asked what biological processes might have shaped the communities in stagnant tap water. Our results showed that the taxa frequency distribution in stagnant water communities was well explained by dispersal and random births/deaths from biofilms and fresh tap water. Such processes can be easily ignored if only community similarity was looked at. For example, in the samples of main plumbing lines (1100-2000th mL and 2100-3000th mL samples), the pattern in community composition was significantly different from the fresh tap water (ANOSIM R= 0.88-0.99, $p<0.001$), but the neutral processes from fresh tap water explained high percentage of frequency distribution (R^2=0.41-0.48).

Our results suggest that the relative contribution of neutral processes from fresh municipal water and biofilms correlated to the pipeline configuration. The pipeline in indoor water supplies are usually designed in a tree structure where the size of the diameters are scaled down as the branches deepens. This change in diameter can affect both the growth and the dispersal of microbes. First, the pipeline diameter strongly influences the decay of disinfectants in pipelines: the smaller the pipeline diameter is, the faster the consumption on the pipe wall would be (Hallam *et al.*, 2002; Hua *et al.*, 1999; Rossman *et al.*, 1994). Hence, communities at a smaller diameter would be more likely to experience a longer exposure to sub-detrimental chlorine levels. Second, pipeline diameter can also influence the probability of dispersal from biofilms to the suspended community. The smaller the pipeline is, the higher biofilm-covered pipeline area it is for unit volume of water, and the more likely it is for dispersal from biofilms to the suspended community to happen. This is consistent with our result that the fresh municipal water provided better fit for the communities from the main branches, while the biofilm community provides better fit for the communities close to the faucet.

Disentangling these two sources provides insights on both monitoring and control of water quality change in premise plumbing. It first highlights the role of biofilms in the altering the microbial communities in suspended communities. Therefore, remedies on problematic systems can consider the replacing or thorough cleaning of pipelines with heavy biofilm growth. Further, the whole-community source tracking in our analysis shows a possibility to use the community

information for decision making: for a controlled length of stagnation, the relative contribution of biofilm can be used as a criterion to determine if certain pipelines should be retired. Although further validation is needed before responsibly applying this method, the methods used here provides an approach for such investigations.

4.5 Tables and Figures

Table 4.1 ANOSIM for community composition between stagnant water and fresh tap water or biofilms

Comparisons	R-value	P-value
Fresh tap water v.s. WM biofilms	0.76	<0.001*
First 100 mL v.s. Fresh tap water	0.94	<0.001*
200-1000 th mL v.s. Fresh tap water	0.87	<0.001*
1100-2000 th mL v.s. Fresh tap water	0.88	<0.001*
2100-3000 th mL v.s. Fresh tap water	0.89	<0.001*
First 100 mL v.s. WM Biofilms	0.30	<0.001*
200-1000 th mL v.s. WM Biofilms	0.53	<0.001*
1100-2000 th mL v.s. WM Biofilms	0.78	<0.001*
2100-3000 th mL v.s. WM Biofilms	0.80	<0.001*

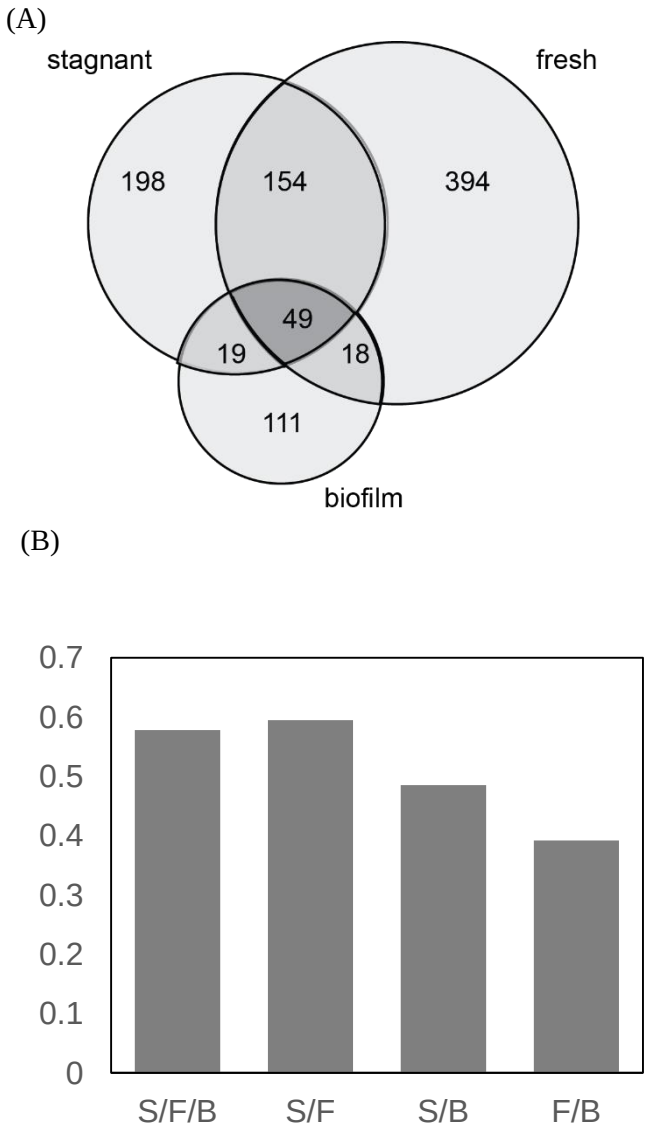


Figure 4.1 Venn diagram of OTUs in fresh municipal water, stagnant water, and water meter biofilms. Panel A represents the number of shared and distinctive OTUs by area. Panel B shows the proportion of reads shared (S-stagnant, F-fresh, and B-biofilms).

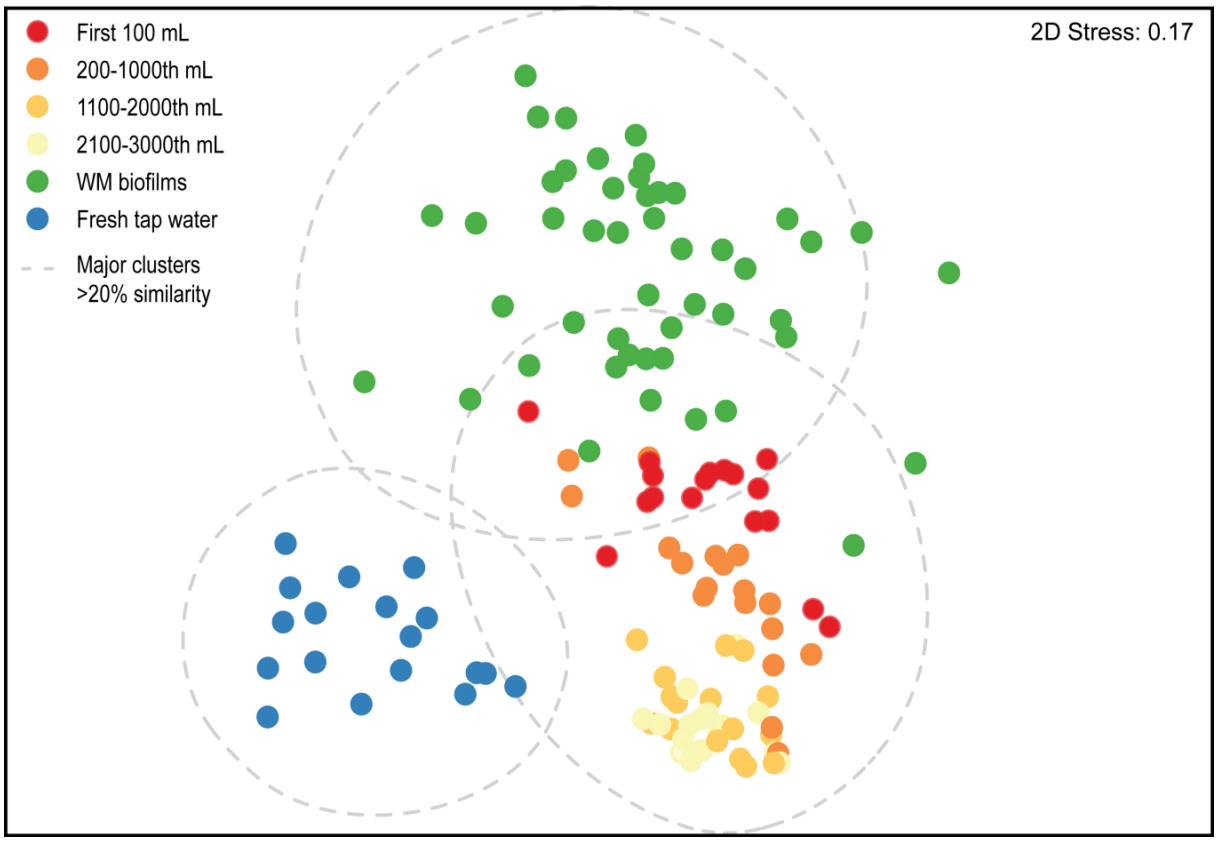


Figure 4.2 Non-metrical multidimensional scaling plot representing community similarities between samples.

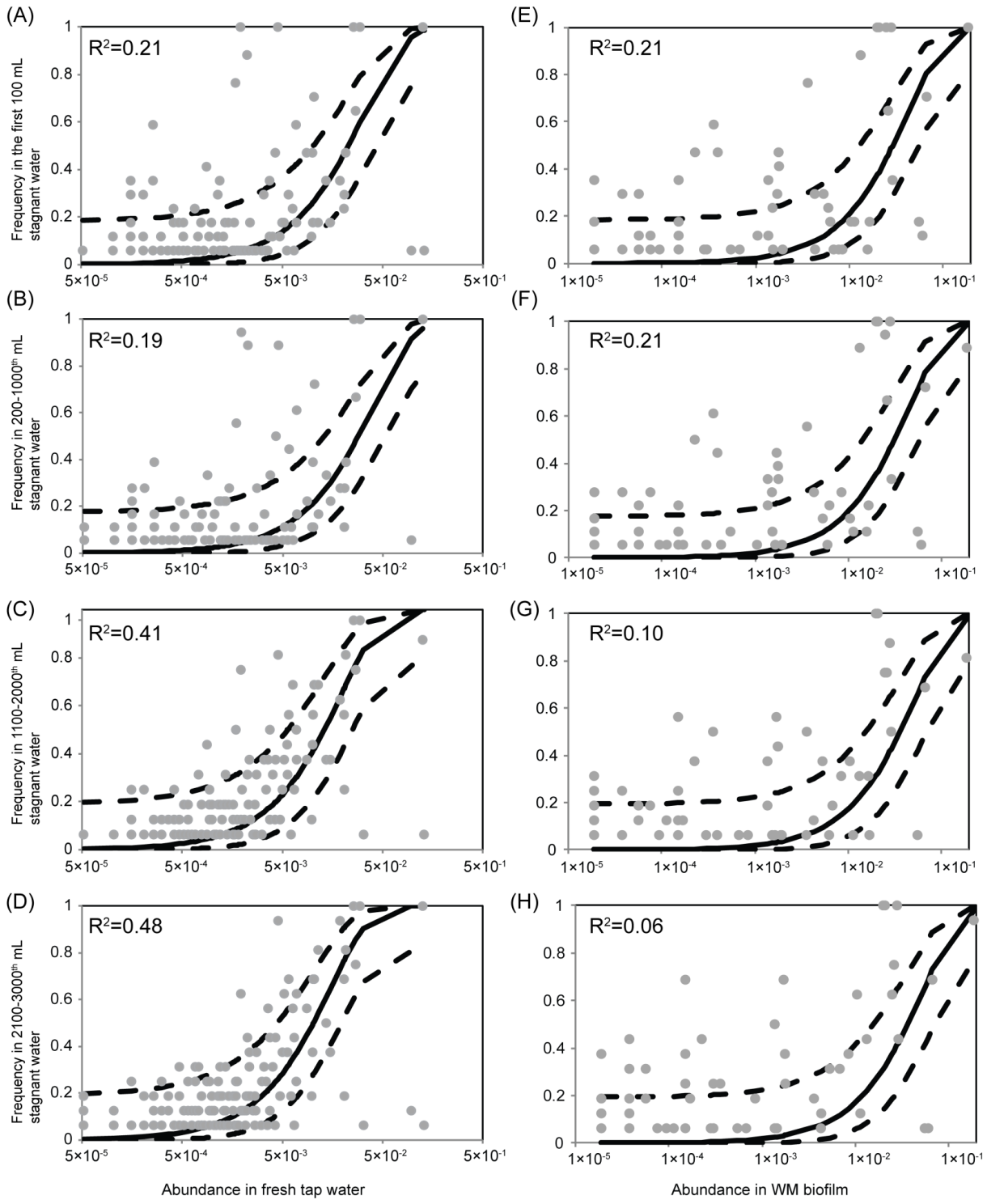


Figure 4.3 Results of neutral model fitting with fresh tap water (A-D) and biofilms (E-H) as the potential sources. A and E, the first 100 mL stagnant water; B and F, the 200th-1000th mL stagnant water; C and G, the 1100-2000th mL stagnant water; D and H, the 2100-3000th mL stagnant water.

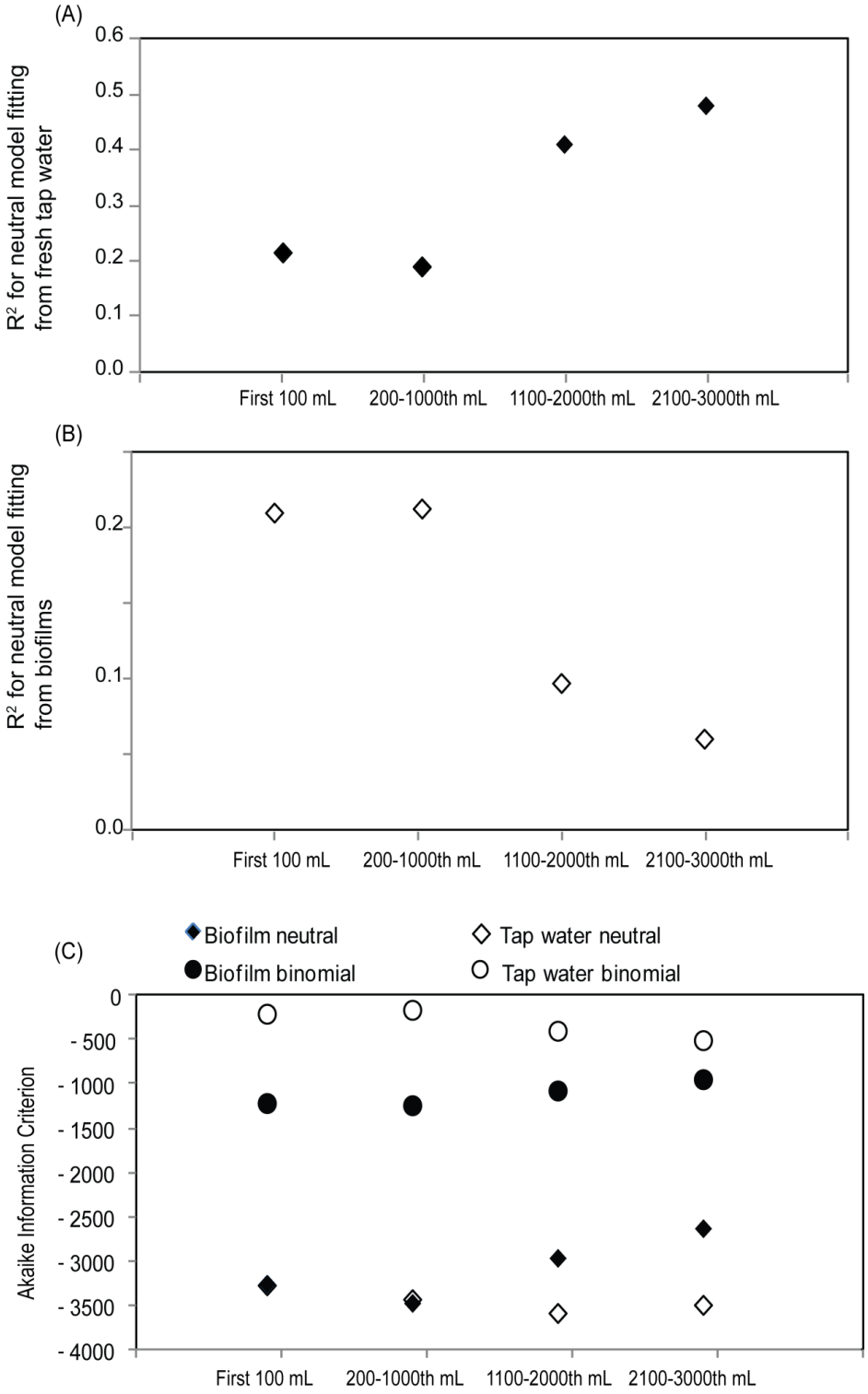


Figure 4.4 Variation of the relative contribution from fresh tap water (A) and biofilms (B) and the comparison of model fit to binomial distribution (C).

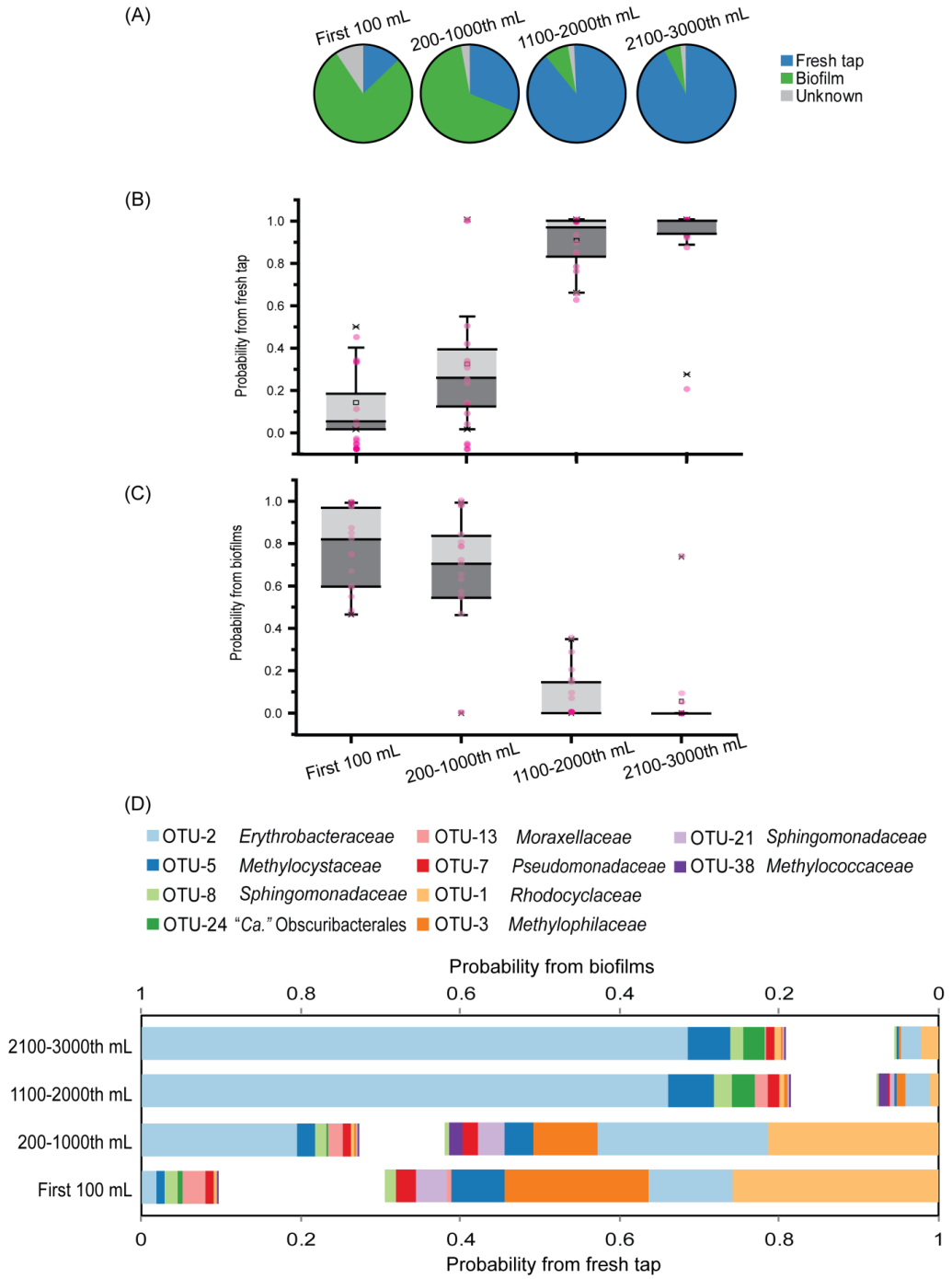


Figure 4.5 Whole-community source tracking in stagnant water. Panel A shows the average source tracking in fractions of stagnant water. Panel B and C shows the box plots representing probabilities for individual samples contributed by fresh tap water (B) and biofilms (C). Panel D shows the source tracking for dominant OTUs (top 10 abundant in stagnant communities).

4.6 Reference

Burbrink FT, McKelvy AD, Pyron RA, Myers EA. (2015). Predicting community structure in snakes on Eastern Nearctic islands using ecological neutral theory and phylogenetic methods. *Proc Biol Sci* **282**:20151700.

Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, *et al.* (2015). Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J*. doi:10.1038/ismej.2015.142.

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.

Elzhov T V., Mullen KM, Spiess A-N, Bolker B. (2013). Minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. https://cran.r-project.org/web/packages/minpack.lm/index.html (Accessed January 7, 2016).

Hallam N., West J., Forster C., Powell J., Spencer I. (2002). The decay of chlorine associated with the pipe wall in water distribution systems. *Water Res* **36**:3479–3488.

Harrell F. (2015). Hmisc: Harrell Miscellaneous.
https://cran.r-project.org/web/packages/Hmisc/index.html (Accessed January 7, 2016).

Hua F, West J., Barker R., Forster C. (1999). Modelling of chlorine decay in municipal water supplies. *Water Res* **33**:2735–2746.

Hubbell SP. (2001). The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). Princeton University Press
http://books.google.com/books?hl=en&lr=&id=EIQpFBu84NoC&pgis=1 (Accessed December 11, 2014).

Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, *et al.* (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* **8**:761–3.

MacArthur RH, Wilson EO. (1967). The Theory of Island Biogeography. Princeton University Press.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2013). The SILVA

ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**:D590–6.

Rossman LA, Clark RM, Grayman WM. (1994). Modeling Chlorine Residuals in Drinking Water Distribution Systems. *J Environ Eng* **120**:803–820.

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**:7537–41.

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**:732–40.

Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. (2007). Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microb Ecol* **53**:443–55.

Spiess A-N, Neumeyer N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol* **10**:6.

USEPA. Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules. http://www.epa.gov/dwreginfo/stage-1-and-stage-2-disinfectants-and-disinfection-byproducts-rules (Accessed January 7, 2016).

Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, *et al.* (2015). Application of a neutral community model to assess structuring of the human lung microbiome. *MBio* **6**:e02284–14.

CHAPTER 5 SUMMARY AND ENGINEERING IMPLICATIONS

Drinking water supply systems comprises of stages of infrastructure drastically different in chemistry and flow regimes by design, and therefore may present different habitats for microorganisms. The diversity of the microbiota and its response to the environment has been an interest for the drinking water microbiology field for decades, but limited by available tools and samples (Geldreich *et al.*, 1972; Proctor and Hammes, 2015). This thesis work used the Champaign-Urbana drinking water distribution system as a model system, investigated the microbial communities inhabiting the distribution system spatio-temporally with high throughput sequencing, and provided new understanding on the phylogeny and community structure in the distribution systems. In specific, this work went beyond the heavily monitored suspended community in the municipal water supply system, and focused on the biofilm component in the municipal water network and the post-stagnation community in indoor water supply. The findings revealed profoundly different community composition and structure in these two habitats from the distribution system water. The key findings are:

- Occupancy-abundance correlation and presence of abundant and prevalent core populations were observed in the water meter biofilm communities.
- The core populations in the biofilm communities were distinctive from the suspended communities despite certain shared populations.

- Resilient pattern of seasonal fluctuation was observed in the biofilm community.
- Drastic abundance and community composition/structure change was observed in the tap water community after week-long stagnation in buildings.
- Microbial abundance and community structure in post-stagnation indoor water correlated with the residual chlorine gradient, which was determined by pipeline structure.

These findings hold implication for water treatment and monitoring:

- Certain components of the biofilm core populations were not detected as the core populations in suspended communities, indicating that there are bacteria that are more adapted to the biofilm habitat. These bacteria would unlikely to be represented in routine monitoring of the suspended community. In distribution systems recovering from a contamination event, it may be advisable to monitor the biofilm communities in addition to routine monitoring, and the water meters could serve as a potential sampling device for their easy accessibility and ample spatial replication.
- In groundwater sourced water supplies, methane may act as a gaseous carbon source that can support microbial growth in distribution systems. Thus, treatment targeting methane can be used to control overall biological growth and certain methane or methanol utilizing bacteria.
- Drastic variation of microbial abundance and community structure from the distribution system water can occur in indoor plumbing, given a week-long stagnation time. This

highlights the need to develop biological quality risk control programs for premise plumbing, especially for green buildings where water stagnation is worsened by water saving features. Our results showed the dependence of microbial abundance and community on the plumbing structure, which suggests the sampling program for premise plumbing can be designed based on the demand served by the plumbing system.

APPENDIX A SUPPLEMENTAL MATERIALS FOR CHAPTER 2

I. Methods on multivariate analysis

All multivariate analyses were performed in PRIMER 6. Service line material, water pH, and water turbidity were ruled out from the analysis, because of their overall consistency. Distribution system configurations were grouped into "loops" and "dead-ends" when the configuration of the pipe construction map at the utility showed the particular type supplied the site, or "water mains" when a site directly drew water from a major water main. Water main materials were grouped as "cast iron", "ductile iron", and "prior to records".

For numeric variables, a distance-based linear model (DistLM) was used to test the relationship between the resemblance matrix and each factor. These factors were also tested in ANOSIM (Chapman and Underwood) after grouped into quartiles. Service line size was grouped as "0.75 inch", "1 inch" and "over 1 inch". Water main size was grouped as "less than 6-inch", "6-inch", "8-inch" and "over 8-inch". Water age, water main age and service line age were grouped into quartiles. Missing values were marked as not available as one category.

Because PERMANOVA is more robust in dealing with heterogeneity and unbalanced designs (Anderson and Walsh, 2013), we conducted PERMANOVA and PERMDISP (Anderson and Gorley, 2008) to distinguish differences in centroids and dispersion for factors shown to be significant in ANOSIM. For one-way design, unrestricted permutation of the raw data at 999

permutations was used. Canonical analysis of principal coordinates (CAP) was performed to sampling windows (Anderson and Willis, 2003).

Reference:

Anderson MJ, Gorley RN. (2008). PERMANOVA for PRIMER: guide to software and statistical methods. PRIMER-E Ltd,.

Anderson MJ, Walsh DCI. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol Monogr* **83**:557–574.

Anderson MJ, Willis TJ. (2003). Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84**:511–525.

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

II. Detection of methane in DWDS water sampling sites

To confirm presence of methane in the distribution system as suggested by the detection of methanotrophic bacteria as core populations (Figure 4A), we sampled tap water from the water sampling sites described in Method (Hwang *et al.*, 2012) and measured for methane. Thoroughly flushed tap water samples were collected by slowly filling in serum bottles and sealed with butyl stoppers on-site. Headspace methane was measured the same day with a GC-2014 Gas Chromatograph (Shimadzu Scientific Instruments) (Narihiro *et al.*, 2014) and then converted to aqueous concentration (Rettich *et al.*, 1981).

Table A.1 Results of methane measurements at water sampling sites

Sites	Methane (mM)	Water temperature (°C)
N2	0.2	11.5
S2	0.2	13.5
C0	0.2	13.0
N1	0.2	13.5
S1	0.3	14.0

Reference:

Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. (2012). Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl Environ Microbiol* **78**:7856–65.

Narihiro T, Nobu MK, Kim N-K, Kamagata Y, Liu W-T. (2014). The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ Microbiol*. doi:10.1111/1462-2920.12616.

Rettich TR, Handa YP, Battino R, Wilhelm E. (1981). Solubility of gases in liquids. 13. High-precision determination of Henry's constants for methane and ethane in liquid water at 275 to 328 K. *J Phys Chem* **85**:3230–3237.

Table A.2 Distribution system water chemistry*

Variables (mg/L)	Sampling Sites				
	N2	N1	C0	S1	S2
NVOC	1.565 ± 0.294	1.66 ± 0.36	1.77 ± 0.44	1.38 ± 0.22	1.60 ± 0.26
Alkalinity	141.71 ± 11.05	142.15 ± 13.37	142.78 ± 10.49	132.71 ± 9.45	138.15 ± 10.95
Ca	12.73 ± 1.33	12.74 ± 1.33	12.06 ± 0.73	13.36 ± 1.10	12.76 ± 1.11
Mg	12.64 ± 1.76	12.61 ± 1.77	12.89 ± 1.83	11.25 ± 0.89	11.20 ± 0.69
K	2.28 ± 0.52	2.48 ± 0.18	2.43 ± 0.12	2.51 ± 0.14	2.49 ± 0.16
Na	35.25 ± 2.98	35.33 ± 4.13	38.86 ± 6.24	32.36 ± 1.58	38.38 ± 9.32
Si	4.54 ± 0.43	4.58 ± 0.32	4.556± 0.43	4.62 ± 0.40	4.42 ± 0.35
SO_4^{2-}	0.45 ± 0.18	0.38 ± 0.13	0.36 ± 0.01	0.51 ± 0.24	0.46 ± 0.23
NO_3^-	0.03 ± 0.04	0.04 ± 0.04	0.04 ± 0.04	0.03 ± 0.05	0.02 ± 0.03
NO_2^-	<0.025	<0.025	<0.025	<0.025	<0.025
**NH_3	0.24 ± 0.41	0.27 ± 0.41	0.19 ± 0.42	0.32 ± 0.71	0.30 ± 0.67
**Total P	<0.05	<0.05	<0.05	<0.05	<0.05
**TDS	171.4 ± 10.16	169.80 ± 11.43	179.0 ± 18.38	158.6 ± 7.09	176.8 ± 32.24

*This table is adapted from Table S1 in Hwang 2012b.

**Indicate values averaged from 5 time points as these parameters were not measured in the winter 2010 water samples. NH_3 measurement was conducted to seasons with chloramine disinfection.

Table A.3 SILVA Taxonomy classification for core OTUs

OTU_id	Phylum	Class	Order	Family	Nearest neighbor
OTU-1	*Proteobacteria*	*Betaproteobacteria*	*Methylophilales*	*Methylophilaceae*	CP002056
OTU-2	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*	*Methylococcaceae*	AJ132384
OTU-3	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Erythrobacteraceae*	AB630719
OTU-4	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	JX869975
OTU-5	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	AB680719
OTU-6	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	FJ772064
OTU-7*	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*	*Methylococcaceae*	AF304195
OTU-8	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	DQ364433
OTU-9	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	AB681243
OTU-10	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	AF035048
OTU-11	*Actinobacteria*	*Actinobacteria*	*Corynebacteriales*	*Mycobacteriaceae*	GU574063
OTU-12	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*	*Methylococcaceae*	DQ984190
OTU-13	*Verrucomicrobia*	*Opitutae*	*Opitutales*	*Opitutaceae*	KJ721192
OTU-14	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	FM886840
OTU-15	*Proteobacteria*	*Alphaproteobacteria*	*Rickettsiales*	*Holosporaceae*	AB630425
OTU-16	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	EF103560
OTU-17	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	FR691411
OTU-18	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	KJ667149
OTU-19	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*		AB722228
OTU-20	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Chromatiaceae*	AY701891
OTU-21	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	AY599912
OTU-22	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	FM165535
OTU-23	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*	GU272379
OTU-29	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylobacteriaceae*	AB298402
OTU-30	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	AB733397
OTU-32	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	AF508803
OTU-33	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	EU286535
OTU-38	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	AF479766

Table A.3 (Cont.)

OTU_id	Phylum	Class	Order	Family	Nearest neighbor
OTU-39	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	KC844779
OTU-44	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae*	KC757053
OTU-48	*Actinobacteria*	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	FJ546064
OTU-51	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	HM149209
OTU-61	*Cyanobacteria*	ML635J-21			EF589990
OTU-65*	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	AY328556
OTU-67	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	AY303329
OTU-69**	*Proteobacteria*	*Alphaproteobacteria*	*Rickettsiales*		AY957940
OTU-70	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*	AB078842
OTU-71	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	HM031972
OTU-75	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	FR682714
OTU-76	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	AB680579
OTU-81	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	AY928219
OTU-85	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	FJ497452
OTU-86**	*Proteobacteria*	*Alphaproteobacteria*	*Rickettsiales*		HQ218746
OTU-101	*Firmicutes*	*Erysipelotrichia*	*Erysipelotrichales*	*Erysipelotrichaceae*	FJ382108
OTU-116	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	AY728038
OTU-128	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	KF722521
OTU-138	*Firmicutes*	*Clostridia*	*Clostridiales*	Family XII	AB476700
OTU-139	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	AB122032
OTU-153	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	EF623862
OTU-160	Candidate division OP3				AB858608
OTU-164**	*Proteobacteria*	*Alphaproteobacteria*	*Rickettsiales*		FJ437938
OTU-168	Candidate division OP3				AB722208
OTU-181	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Chromatiaceae*	HQ324860
OTU-205	*Cyanobacteria*	*Melainabacteria*	*Obscuribacterales*		AY328558
OTU-226	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*	AB355049
OTU-236	Candidate division OP3				AB364884

Table A.3 (Cont.)

OTU_id	Phylum	Class	Order	Family	Nearest neighbor
OTU-276	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*	Y17389
OTU-277	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	DQ664239
OTU-283	*Proteobacteria*	*Betaproteobacteria*	*Nitrosomonadales*	*Gallionellaceae*	AB600388
OTU-284	Candidate division OP3				AB232821
OTU-387	*Proteobacteria*	*Gammaproteobacteria*	*Methylococcales*		AB722196
OTU-480	*Chlorobi*	*Ignavibacteria*	*Ignavibacteriales*		AB661522
OTU-734	Candidate division OP3				AB232821
OTU-811	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*	DQ8854968

* The nearest neighbors of OTU-7 and OTU-65 sequences from SILVA alignment are sequences from uncultured bacteria, thus a SILVA taxonomy was not provided automactially from the online servie. The classification were inferred from further analysis of the OTUs and and their neareast neighbors with ARB.

** OTU-69, -86, and -164 were classified as Rickettisales incertae sedis in SILVA.

Table A.4A Pairwise comparison

Sampling windows	t	P(perm)	Unique permutations
1, 2	1.8675	0.002	997
1, 3	2.2881	0.001	999
1, 4	2.4223	0.001	998
1, 5	1.806	0.001	999
1, 10	1.6255	0.001	999
1, 11	1.6374	0.001	997
1, 12	2.0719	0.001	999
1, 9	1.3903	0.008	999
1, 8	1.4247	0.005	999
1, 7	2.0411	0.001	998
2, 3	1.8403	0.001	997
2, 4	2.2429	0.001	999
2, 5	1.9879	0.001	998
2, 10	1.8252	0.001	999
2, 11	1.8663	0.001	999
2, 12	2.2823	0.001	998
2, 9	2.3068	0.001	999
2, 8	2.1929	0.001	993
2, 7	3.0234	0.001	997
3, 4	1.4669	0.011	999
3, 5	1.6733	0.002	998
3, 10	1.4159	0.014	999
3, 11	1.5589	0.006	998
3, 12	2.0228	0.001	998
3, 9	2.1809	0.001	997
3, 8	2.1655	0.001	996
3, 7	3.4561	0.001	998
4, 5	1.533	0.003	999
4, 10	1.713	0.004	998

Table A.4A (Cont.)

Sampling windows	t	P(perm)	Unique permutations
4, 11	1.8028	0.001	999
4, 12	2.2342	0.001	998
4, 9	2.2198	0.001	998
4, 8	2.1937	0.001	996
4, 7	3.4552	0.001	999
5, 10	1.6178	0.001	999
5, 11	1.5321	0.001	997
5, 12	1.5139	0.001	999
5, 9	1.6945	0.001	973
5, 8	1.6132	0.001	979
5, 7	1.9023	0.001	998
10, 11	0.8687	0.822	998
10, 12	1.3968	0.011	999
10, 9	1.4853	0.003	998
10, 8	1.5216	0.001	997
10, 7	2.5142	0.001	999
11, 12	1.381	0.01	998
11, 9	1.5653	0.001	998
11, 8	1.5752	0.001	997
11, 7	2.4278	0.001	998
12, 9	1.7085	0.001	998
12, 8	1.5866	0.002	997
12, 7	2.8066	0.001	998
9, 8	0.80356	0.9	980
9, 7	1.6292	0.001	999
8, 7	1.4801	0.01	996

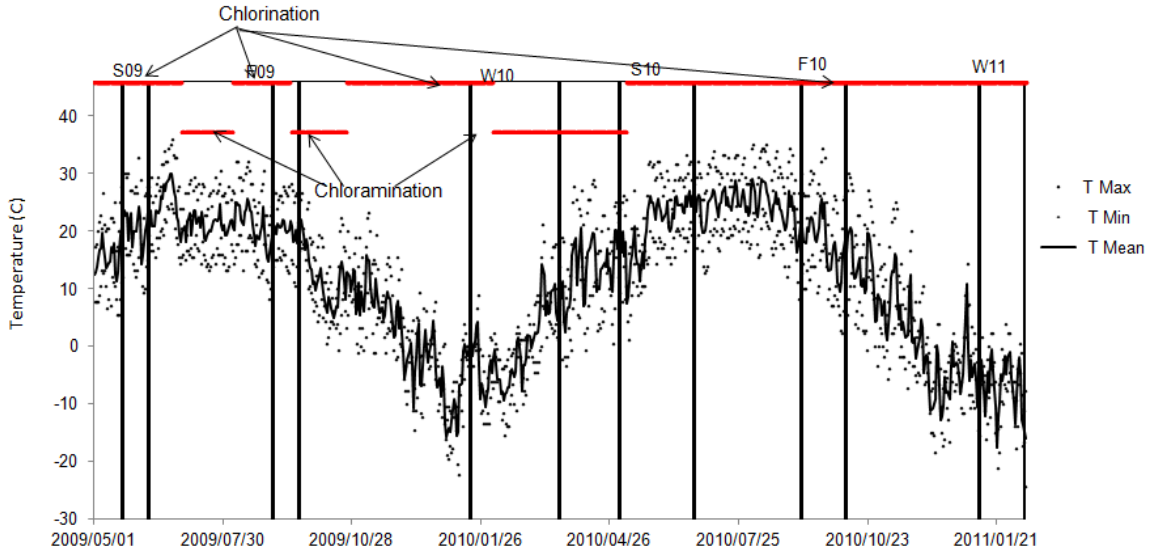
Table A.4B Pairwise PERMDISP

Sampling windows	t	P(perm)
(1,2)	2.5429	2.40E-02
(1,3)	0.29209	0.786
(1,4)	1.3563	0.247
(1,5)	0.71353	0.567
(1,10)	0.51246	0.632
(1,11)	0.44865	0.706
(1,12)	1.6093	0.129
(1,9)	0.91047	0.461
(1,8)	0.79645	0.484
(1,7)	1.9146	8.70E-02
(2,3)	2.9897	7.00E-03
(2,4)	4.6848	1.00E-03
(2,5)	2.8485	2.00E-02
(2,10)	2.4868	3.20E-02
(2,11)	2.8071	5.00E-03
(2,12)	4.545	1.00E-03
(2,9)	3.6735	2.00E-03
(2,8)	3.7622	1.00E-03
(2,7)	5.0698	1.00E-03
(3,4)	1.1209	0.299
(3,5)	0.56386	0.622
(3,10)	0.85748	0.425
(3,11)	0.23192	0.848
(3,12)	1.4498	0.175
(3,9)	0.74255	0.529
(3,8)	0.61733	0.594
(3,7)	1.7606	0.102
(4,5)	8.38E-02	0.947
(4,10)	2.101	7.60E-02

Table A.4B (Cont.)

Sampling windows	t	P(perm)
(4,11)	0.63954	0.585
(4,12)	0.55237	0.608
(4,9)	6.65E-02	0.961
(4,8)	6.69E-02	0.957
(4,7)	0.79836	0.46
(5,10)	1.1915	0.281
(5,11)	0.30899	0.782
(5,12)	0.38945	0.753
(5,9)	9.55E-02	0.958
(5,8)	1.44E-02	0.99
(5,7)	0.53555	0.676
(10,11)	0.94972	0.396
(10,12)	2.2882	5.00E-02
(10,9)	1.5354	0.176
(10,8)	1.4426	0.251
(10,7)	2.6461	2.10E-02
(11,12)	0.94128	0.419
(11,9)	0.45453	0.751
(11,8)	0.35644	0.806
(11,7)	1.166	0.325
(12,9)	0.28892	0.804
(12,8)	0.4002	0.771
(12,7)	0.16756	0.869
(9,8)	0.10005	0.935
(9,7)	0.43325	0.735
(8,7)	0.55064	0.701

(A)



(B)

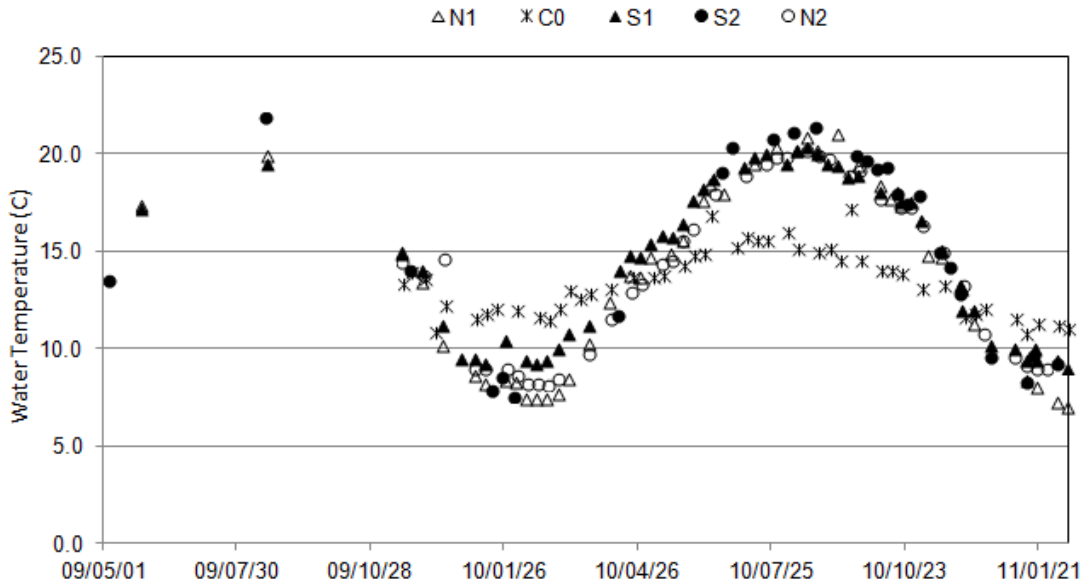
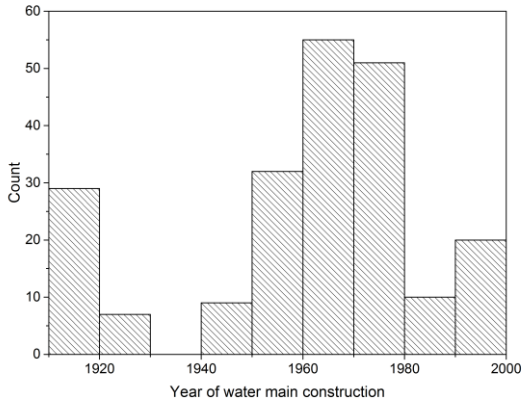
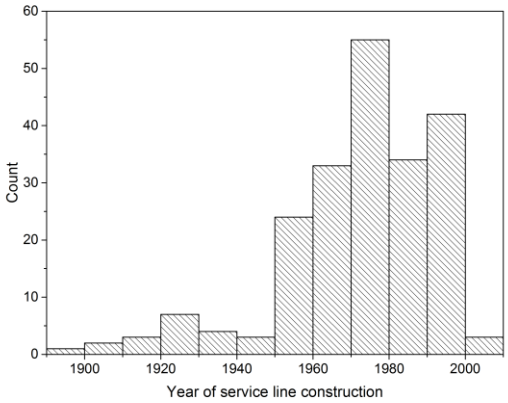


Figure A.1 Ambient temperature and disinfectant types of sampling periods and water temperature. In panel A, red lines indicate periods of stable disinfectant application. The types of disinfectant are labeled to the lines. In panel B, water temperature data were retrieved from the monitoring data of Illinois American Water. Symbols indicate monitoring sites N1, C0, S1, S2, and N2 within the distribution system.

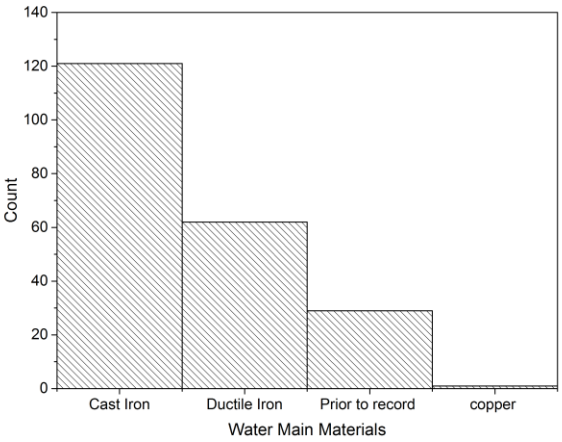
(A)



(B)



(C)



(D)

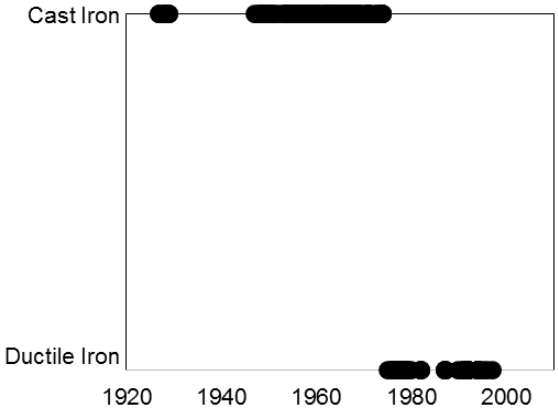
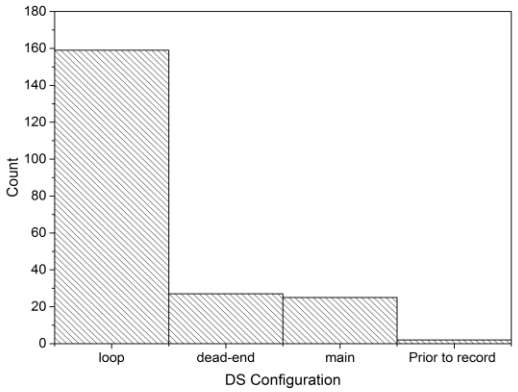


Figure A.2 Description on distribution system conditions associated with water meter sites. (A) Installation Years of water mains that supplies the sample sites. (B) Installation year of service lines connected to the water meters. (C) Water main materials. (D) The correlation of water main material to water main installation year. (E) The distribution system configuration. (F) Water age.

(E)



(F)

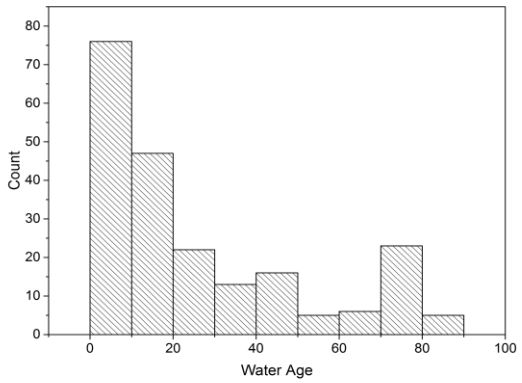
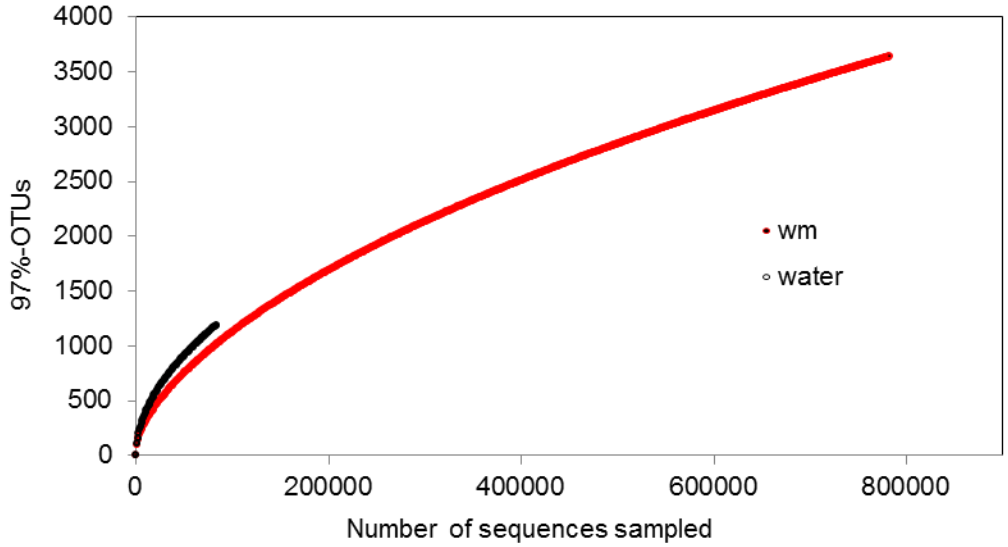


Figure A.2 (Cont.)

(A)



(B)

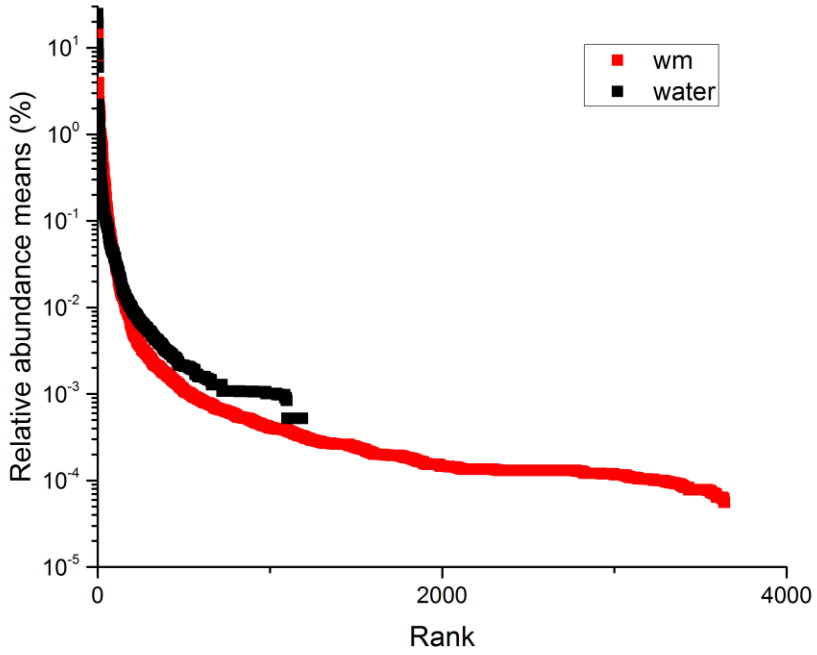


Figure A.3 Rarefaction curves (A) and rank abundance curves (B) in WM biofilms and water samples before sub-sampling.

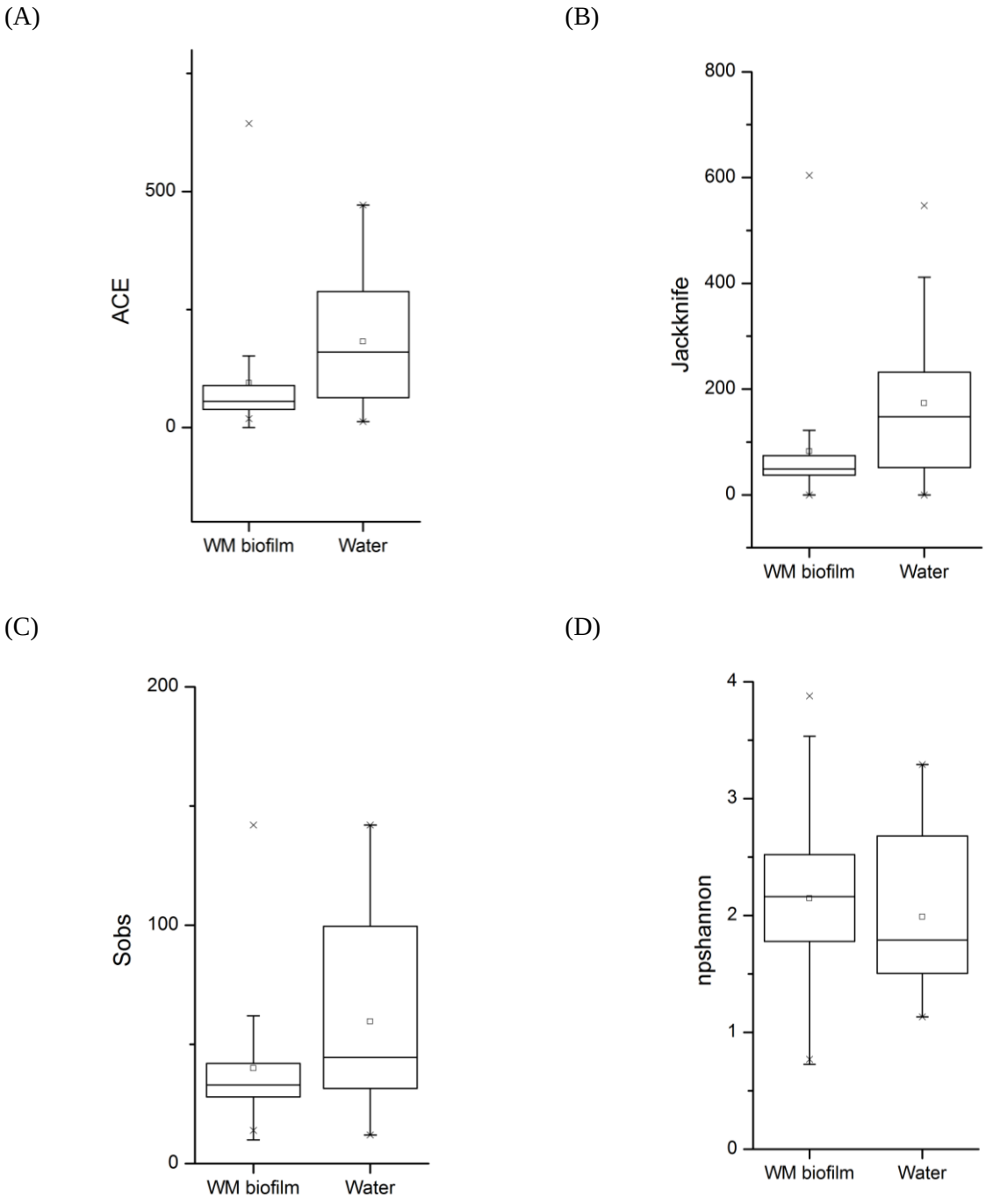


Figure A.4 Alpha diversity comparison between water and water meter samples. The alpha diversity was represented as abundance-based coverage estimators (ACE, Figure A.4.A), Jacknifed richness (B), richness (C), and non-parametric Shannon index (D).

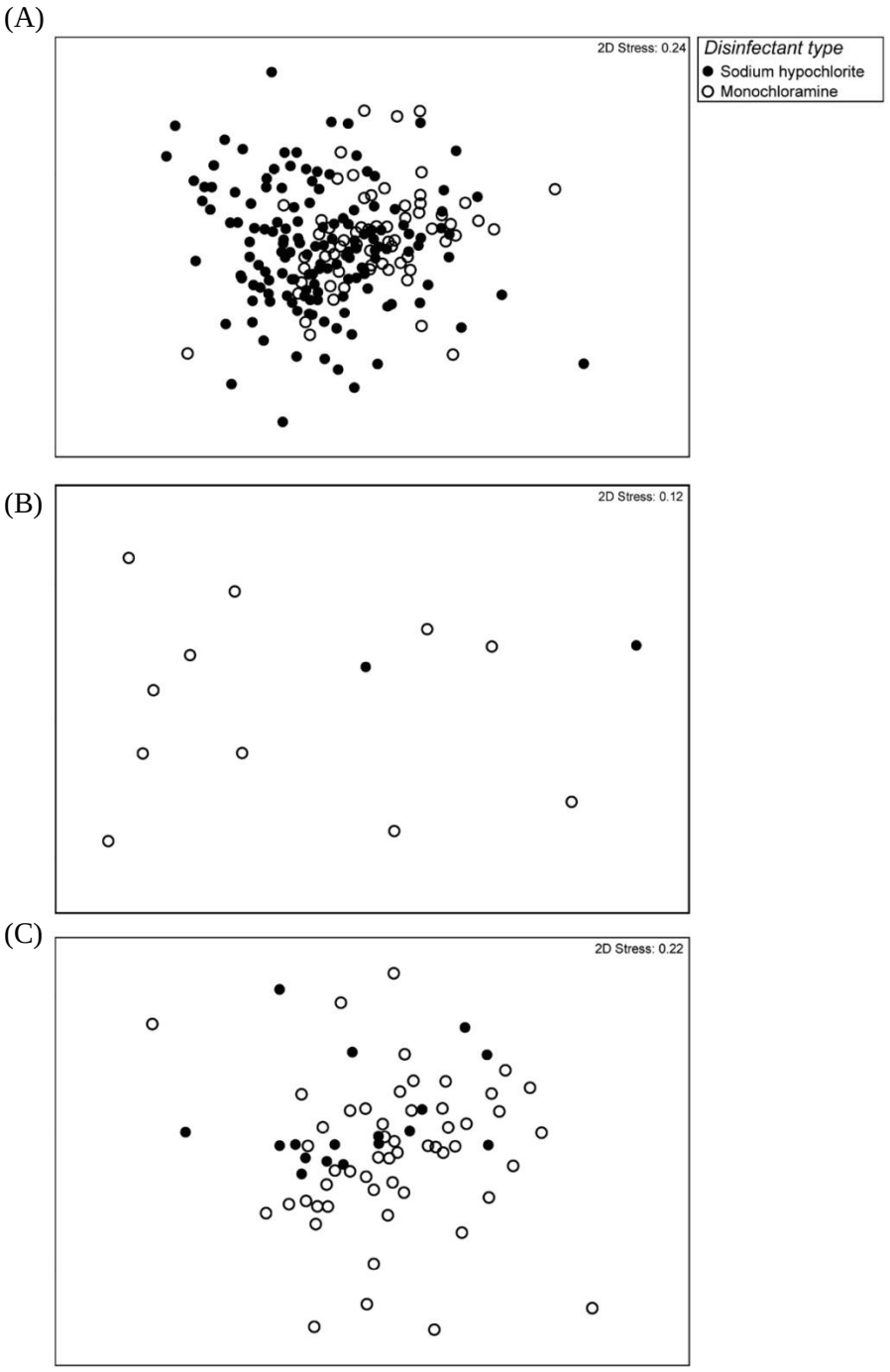


Figure A.5 Multidimensional scaling of community similarity based on Bray-Curtis distances for all the biofilm samples (A) and those from seasons receiving partial free chlorine disinfection (fall 2009 shown in Panel B and winter 2010 shown in Panel (C). Figures supports that community structure were not strongly affected by disinfectant type.

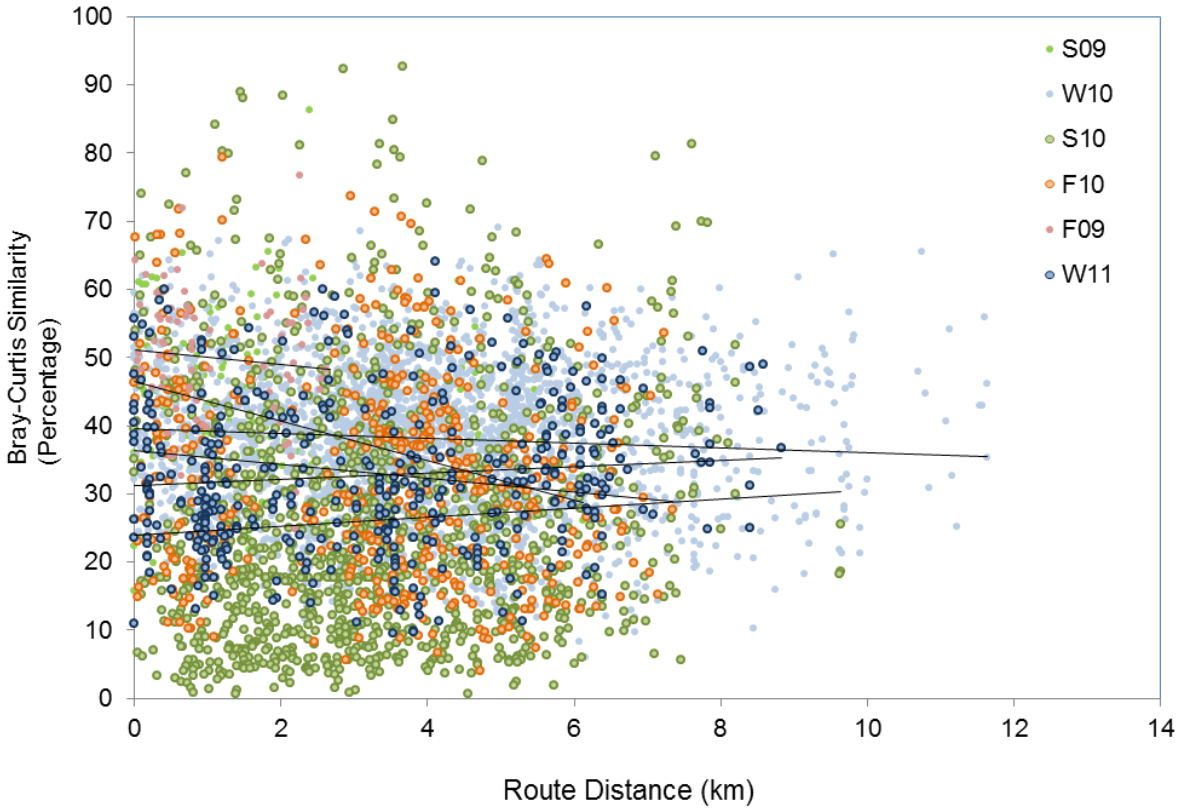


Figure A.6 Absence of correlation between Bray-Curtis similarity and route distance. Black lines indicate the regression results.

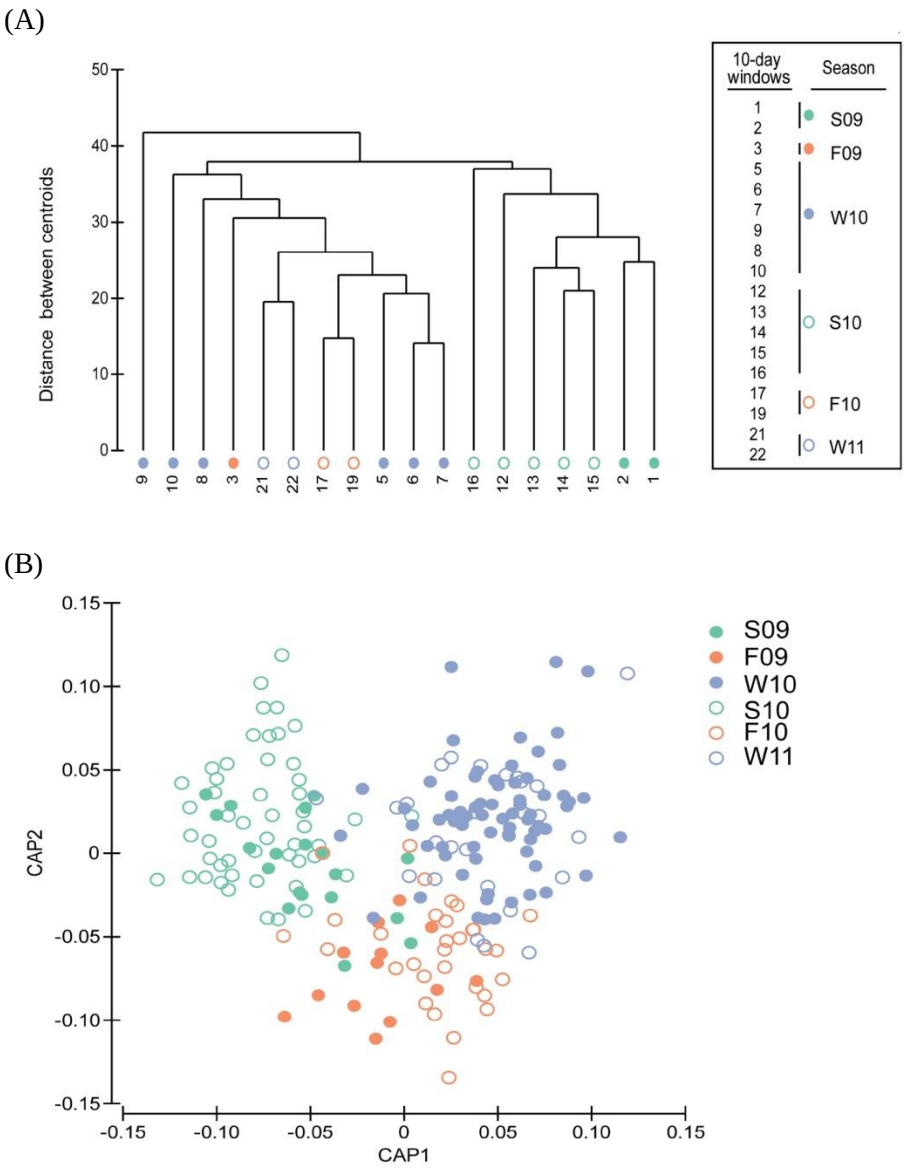


Figure A.7 (A) Cluster analysis on the group centroids of 10-day temporal bins. Bins containing 3 or less samples were excluded from the analysis. Results showed similar trend as the cluster analysis from 20-day bins. (B) Canonical analysis of principle coordinates for Bray-Curtis distance between water meter biofilm samples, with seasons (summer, fall and winter) as the constraint variables. The analysis was done as discriminant analysis and the number of PCO axes was chosen to maximize a leave-one-out allocation success to groups. The proportion of variation explained by the chosen principle coordinates is 92.6%; the size of the first squared canonical correlation is 0.743, and the second is 0.466. Percentage of allocation success is 80.86%.

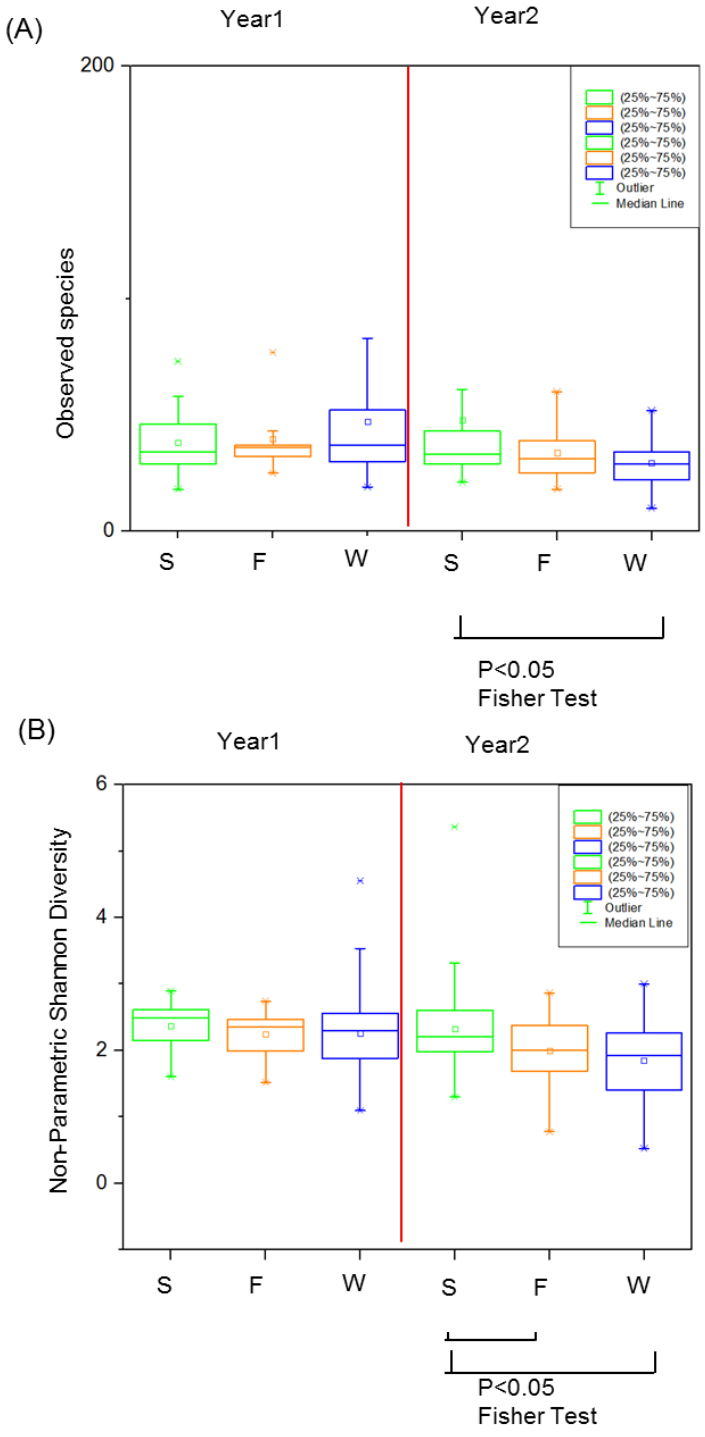


Figure A.8 Seasonal fluctuation of alpha diversity in WM biofilms. The alpha diversity was represented as richness (A) and non-parametric shannon diversity (B). ANOVA on both groups showed significant between-group differences. Groups with significant difference ($p<0.05$) in Fisher's test are marked on the plot.

APPENDIX B SUPPLEMENTAL MATERIALS FOR CHAPTER 3

Table B.1 Community structure analysis with PERMANOVA and PERMDISP

(A)

Factors	**Levels**
Sample type	F, S1, S2, S3, S4 *
Floor	2F, 3F, 4F
Experimental run	I, II
Building	S, V, T

(B)

PERMANOVA			
Groups	t	P (perm)	Unique permutations
F, S1	2.06	0.017	999
F, S2	2.0977	0.02	999
F, S3	2.3194	0.008	998
F, S4	2.3755	0.01	998
S1, S2	1.5983	0.069	998
S1, S3	2.2768	0.008	999
S1, S4	2.5743	0.004	998
S2, S3	1.6393	0.06	999
S2, S4	1.7152	0.05	999
S3, S4	0.93408	0.625	997

Table B.1 (Cont.)

(C)

PERMDISP	
Groups	P
F,S1	1.00E-03
F,S2	1.10E-02
F,S3	1.00E-03
F,S4	1.00E-03
S1,S2	0.99
S1,S3	1.70E-02
S1,S4	1.00E-03
S2,S3	4.90E-02
S2,S4	5.00E-03
S3,S4	0.578

Table B.2 Comparison of alpha diversity estimators for pre-stagnation and post-stagnation samples

Sample Type	**Alpha-diversity estimators**					
	Sobs	Chao1	ACE	Jacknifed	Shannon	Simpson
Before stagnation	138.222	173.475	180.678	208.649	3.22459	0.1121
After stagnation	71.6111	112.113	144.364	396.227	1.67836	0.38898
One-tail p	1.06E-06	0.00176	0.09914	0.24017	1.78E-09	2.48E-13

Note: Equal Variance NOT Assumed (Welch Correction)

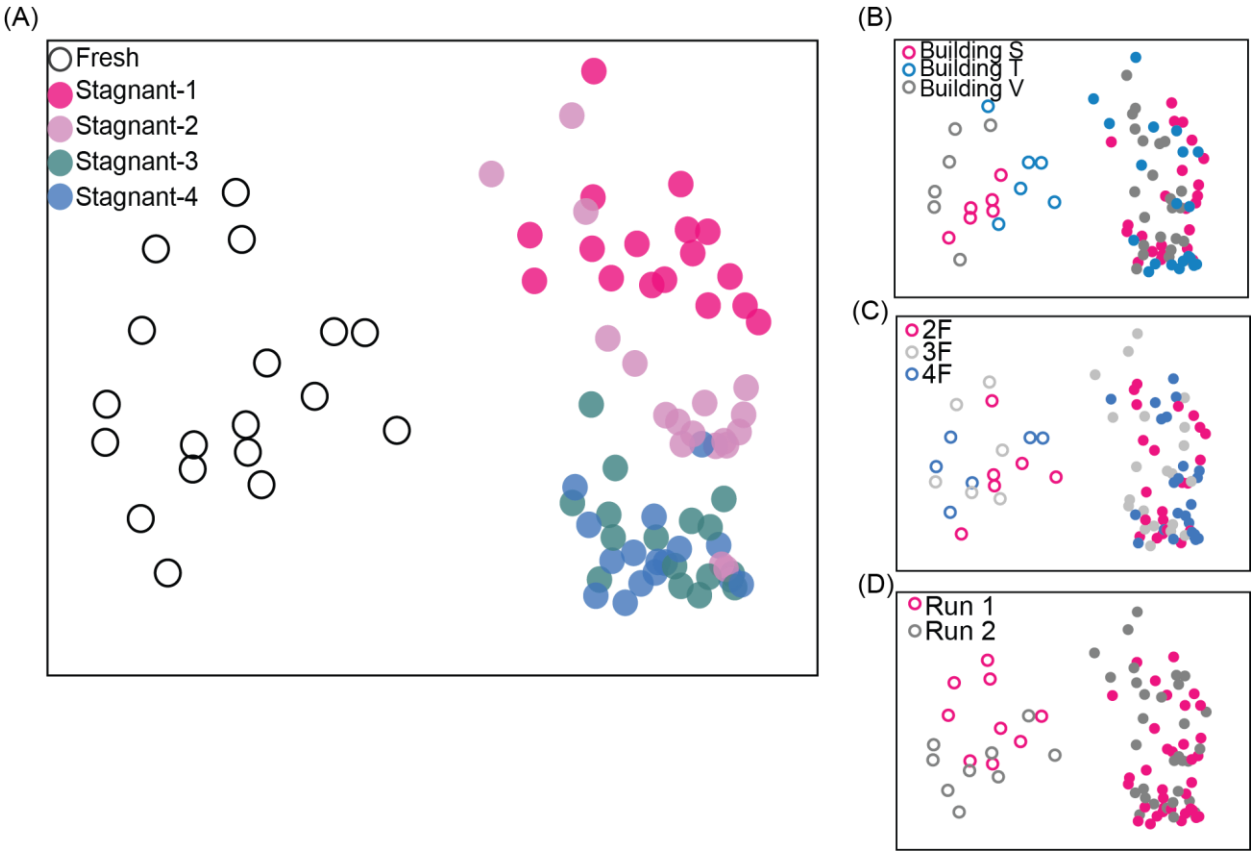


Figure B.1 Non-metric multidimensional scaling for pre-stagnation and post-stagnation samples labeled by different categories. The groups by different volumes (A) showed clear distinction, while those by buildings (B), floors (C), and experimental run (D) did not.