Drinking Water Microbiome Project: Is it Time?

Natalie M. Hull¹, Fangqiong Ling², Ameet J. Pinto³, Mads Albertsen⁴, H. Grace Jang⁵, Pei-Ying Hong⁶, Konstantinos T. Konstantinidis⁷, Mark LeChevallier⁸, Rita R. Colwell⁹, and Wen-Tso Liu¹⁰*

¹Civil, Environmental, and Geodetic Engineering, The Ohio State University, Columbus, OH 43210, USA.
²Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA.
³Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA.
⁴Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark.
⁵The Water Research Foundation, Denver, CO 80235, USA.
⁶Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia.
⁷School of Civil and Environmental Engineering and School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA.
⁸Dr. Water Consulting, LLC, Morrison, CO 80465, USA.
⁹Center for Bioinformatics and Computational Biology, University of Maryland Institute for Applied Computer Studies, College Park, Maryland, and CosmosID, Rockville, MD 20742, USA.
¹⁰Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

*Correspondence: wtliu@illinois.edu (W.-T. Liu)

Keywords
Microbiome, Meta-omics, Drinking water, Microbial ecology
Glossary

- **Meta-Omics, Omics**: Incorporates more than one of the following -omics approaches: metagenomics (study of genomic contents in a sample), metatranscriptomics (study of transcribed genes in a sample), metabolomics (study of metabolites in a sample), metaproteomics (study of proteins in a sample), and interactomics (study of interactions between multiple molecule types in a sample).

- **Drinking Water**: water that has been treated to be potable, or derives from a protected source.

- **Community Water Systems**: public systems supplying water for human consumption.

- **Microbial Ecology**: the relationships of microorganisms (including bacteria, archaea, fungi, protists, and viruses) with each other and with their environment [1].

- **Microbiome**: the collection of microorganisms, their activities, and all other biotic and abiotic factors in their environment [1].

- **Biogeography**: distribution of biological communities over space and time.

- **Biogeochemistry**: chemical, physical, geological, and biological processes and reactions affecting biological systems.

- **Premise plumbing**: portion of potable drinking water distribution systems in buildings.

- **Citizen Science**: public participation in research conducted by professional scientists.
Abstract
Now is an opportune time to fund and foster collaborations across sectors and geographical boundaries to enable development of best practices for drinking water (DW) microbiome research, focusing on accuracy and reproducibility of meta-omic techniques (while learning from past microbiome projects). A large-scale coordinated effort that builds on this foundation will enable the urgently needed comprehensive spatiotemporal understanding and control of DW microbiomes by engineering interventions to protect public health. This opinion paper highlights the need to initiate and conduct a large-scale coordinated DW microbiome project by addressing key knowledge gaps, and recommends a roadmap for this effort.
Justification for a Large-scale Coordinated Drinking Water Microbiome Project (DWMP)

Drinking water distribution systems (DWDSs) throughout the world host diverse microorganisms in biofilms and bulk water suspensions that make up the drinking water (DW) microbiome, resulting in tap water teeming with microbial life, ranging from $10^6$ – $10^9$ microbial cells per liter [2,3]. In the United States alone, 39 billion gallons of water are withdrawn every day from the environment by community water systems [4] and pass through DWDSs comprised of conventional or advanced treatment processes, more than 2 million miles of distribution networks, and premise plumbing in buildings before delivery of DW to taps (Figure 1).

Although water may be microbiologically safe according to regulations and monitoring designed to limit fecal contamination, other microbial risks may be present in treated water or may increase in distribution systems and premise plumbing, where stagnation and decay of disinfectant residual favor growth of microorganisms, especially in biofilms. For example, biofilms and amoebae in DWDSs can harbor opportunistic pathogens (OPs) (such as *Mycobacterium avium*, *Legionella pneumophila*, and *Pseudomonas aeruginosa*) which can cause infections that are increasing in frequency (e.g., by 3.5-fold between 2000-2011 for *Legionella* [5]) with estimated hospitalization costs of $2 billion per year [6]. This is further exemplified by the legionnaire’s disease outbreak during the Flint water crisis, where changes in water quality in distribution systems associated with interrupted corrosion control were linked with high levels of *Legionella* DNA markers in tap waters [7]. While OPs are frequently present in DW microbiomes and various engineering treatments can cause unintended changes in OP relative abundance, factors that affect presence, proliferation, and control of OPs in DW microbiomes are not yet fully understood [8]. Additionally, antibiotic resistant bacteria (ARB) and genes (ARGs) are prevalent in DWDS and tap water [9], and ARGs can be transferred between organisms throughout the DW microbiome continuum, which can escalate illness costs (as has been shown for some OPs [6]) and potential risk of morbidity and mortality among vulnerable and infected individuals.

Despite the huge importance that exposure to DW microbiomes plays in public health, current understanding of DW microbiomes and their link to water quality parameters, engineering interventions, and disease is incomplete. Using the recent advances in nucleic acid sequencing technologies and big-data science, broad multi-agency programs have been developed to study and elucidate microbiomes of the human body [10] and built environments [11] (see Box 1). Microbiome studies in natural aquatic systems and in human-associated systems have also enabled
development of novel diagnostic biomarkers and treatments for disease [10]. In comparison, DW microbiome study significantly lags behind in both sequencing effort invested and knowledge generated (Figure 2). Recent small-scale DNA sequencing surveys have revealed high spatiotemporal heterogeneity and microbial diversity of DW microbiomes [12–15]. These findings and initial models/meta-analyses [14,16,17] suggest it is challenging to generalize DW microbiomes, for example across >55,000 community water systems in the US. Therefore, it remains difficult without a coordinated effort to devise the next advances in DW microbiome diagnostics and treatments to decrease risks to public health.

Understanding of DW microbiomes is also urgently needed to optimize future water infrastructure development and combat increasing global challenges. A 2015 US EPA report estimated a $472.6 billion need for new US DW infrastructure and for rehabilitation, expansion, or replacement of existing DW infrastructure over the following 20 years [18]. When accounting for population growth from 2010-2035, the American Water Works Association estimated a $1 trillion need for US DW infrastructure [19]. To maximize return on this investment, there is a pressing need to understand the impact of microbiomes on the sustainability of these infrastructures (e.g., impacts of nitrification, corrosion, etc.), and the impact of infrastructure design on DW microbiomes relevant to both operational performance and human health.

The US is just one example of impending infrastructure development needed to provide safer drinking water around the world. In other regions, water infrastructure age and designs vary, as do source water quality/quantity and socioeconomic capacity for new and existing infrastructure. Because climate change will pose increasing regional threats on water quality and quantity, the extreme rainfall and flooding in some areas but increasingly arid conditions in other areas will likely lead to DW production relying increasingly on wastewater reuse, desalination, unconventional source waters, and advanced treatment technologies. Existing infrastructure and monitoring programs may not be designed to cope with massive inflow (in the earlier scenario), or to ensure water quality when tapping alternative water sources (in the latter scenario). Rising temperatures may also increase the possibility of waterborne diseases outbreaks [16,20–22]. Thus, developing tools to predict DW microbiome responses to changing treatments, source waters, or water quality is becoming important. Likewise, population changes (e.g., an aging population with increased health vulnerabilities, increased global population, increased urbanization, etc.) will also increase the need to engineer DW microbiomes that are safer for all human exposures.
Knowledge Gaps the DWMP Could Address

The ability to define and successfully engineer safe or healthy DW microbiomes will require the DWMP as a discovery process enabling advanced monitoring, predictive modeling, and optimization of engineering interventions to ensure continued public health protection. By conducting a large-scale coordinated DWMP, we can achieve the required depth and breadth of expertise necessary to address important knowledge gaps and inform DW infrastructure investments while anticipating changes in climate, treatment practices, and exposed populations.

The most important knowledge gap is to understand the DW microbiome under different conditions, prior to an attempt to define and engineer a healthy or safe DW microbiome. To achieve this, high quality supporting data must be collected to overcome inherent stochasticity in biological systems, and to integrate characteristics of the larger water and ecological systems affecting DW microbiomes. Because of the influence of past conditions on current DW microbiomes (especially in stagnation-prone distribution systems and buildings), efforts should focus on well-defined, composite and grab sampling methodologies (for both biofilms and bulk waters), and on thorough characterization of key biogeochemistry and physical parameters. Representative parameters include pipe materials and diameters, flow rates and regimes, temperature, sampling time, disinfectant type and residual, and measures of current or potential microbial activity (e.g., ATP or assimilable organic carbon), disinfection by-products (DBPs), turbidity, inorganics such as metals, and nutrients (particularly C and N). Water usage and storage data should also be collected. Such data are useful to identify outliers in microbiome data, although there may be disconnects in time scales for relevant underlying processes. Comprehensively defined samples and well-documented, high-quality sampling procedures and analyses will be essential to tease out confounding interrelationships between these parameters. In-depth studies are particularly needed to understand the underlying mechanisms behind the formation of DW microbiome constituents in biofilms. Most DW microbiome biomass is known to reside in biofilms (including minority microbial populations), and upstream biofilm communities can contribute to downstream DW microbiome populations through sloughing and seeding [20,23,24]. Biofilm communities can have negative biochemical impacts on water quality, facilitate antibiotic resistance transfer, decrease water biostability, and harbor organisms with increased disinfectant resistance [25]. Through the DWMP discovery process, we can better derive knowledge of complex DW microbiomes, and use the knowledge gained to predict them through measurements
of associated biotic and abiotic parameters. With this concerted effort, the DW microbiome can be better managed.

The Outstanding Questions box lists additional knowledge gaps. One of them is to agree on standard protocol(s) to support DW microbiome studies. Numerous studies have shown that variation in microbiomes of different samples can be caused by bias associated with procedures in sampling, DNA extraction, sequencing, and data analysis [26]. Possible protocols can first be identified from previous and ongoing large-scale microbiome characterization projects (Box 1). Next, participants in the DWMP can participate in round-robin validation tests of various approaches. The use of omics tools alone cannot generate all the required knowledge needed by water utilities to better manage water quality, shape the DW microbiome, and ultimately protect public health. It is also important to evaluate coupling various omics tools (e.g., metatranscriptomics and metagenomics to confirm viable and nonviable cells), with other molecular methods (e.g., qPCR and metagenomics to derive quantitative measurements), or with routinely collected metadata, and to make the underlying methodologies more accessible. These efforts can support the selection of optimal protocols and ensure that samples and data can properly be collected and processed to minimize biases and allow reproducible comparisons among future DW microbiome studies.

Another important gap is the lack of a comprehensive microbial inventory and key biomarkers for rapid monitoring. Currently, the rRNA gene is the most widely used marker for determining microbial diversity, but it offers limited resolution at the species level and cannot differentiate between closely related pathogens and their innocuous commensal counterparts. In addition, viruses and microbial eukaryotes must also be accounted since they are an important but overlooked component of the DW microbiome. Compared to the reference genome databases of bacteria that can provide within-species resolution for some pathogens, reference databases for environmental bacteria, archaea, viruses, and eukaryotes (e.g., fungi and protozoa) are less complete. Omics tools can be applied to establish comprehensive microbial inventories as sequencing costs continue to drastically decrease [27], especially for long-read technologies. Whole genome sequencing, reference database development, and curation of existing databases using ontologies relevant for environmental niches and physiochemical aspects of DW microbiomes will increase efficiency of meta-omics efforts by reducing unclassified/unknown sequences.
Figure 3 depicts the roadmap for a large-scale coordinated DWMP with the goal to involve multi-sector efforts to generate data necessary to link DW microbial ecology with abiotic factors to better understand, monitor, predict, and control DW microbiomes. Phase 1 of the DWMP will start with recruitment of diverse stakeholders across sectors to conduct a comprehensive critical review of existing DW microbiome literature. This could be funded by a workshop/training grant for university labs and other stakeholders including utilities/regulators to work together. The effort would focus on the feasibility and applicability of potential methods, experimental comparisons, pros/cons of various approaches, determining what constitutes value-added microbiome data, and meta-analyses of existing longitudinal full-scale/pilot-system data. The outcome of this review would be used to direct investment required for the large-scale coordinated effort, and to guide essential round-robin testing to standardize DWMP methodology.

Other Phase 1 groundwork would be to devise an appropriate framework for organizing collection, analysis, and storage of samples and data necessary for comprehensive characterization in DWDS under normal operations. This requires leadership by a central lab or network of labs, which might be recruited or self-selected from academia (e.g., the authors of this opinion), industry, or government. The leading labs need to be supported by a network including university, government, and commercial laboratories and entities such as utilities and citizen scientists. The central labs would implement guidance for best practices including sample and data collection and processing, and requirements for the minimum QA/QC procedures and information to accompany samples that were determined in the initial critical review. Guidance from the central leading labs in a top-down approach will enable even those with fewer laboratory resources (e.g., citizen scientists and small utilities) to participate and provide meaningful engagement in a bottom-up support network. This concurrent top-down bottom-up approach will deconstruct the monolithic DWMP into feasible and cohesive projects to advance understanding of DW microbiomes. Finally, the consortium will facilitate sharing data, protocols, and samples.

Phase 2 of the DWMP will focus on comprehensive baseline characterizations in diverse full-scale systems (e.g., urban vs. rural, various treatment practices, etc.) to build a strong foundation. Bench-scale experiments should also be used to test hypotheses that fall out of the initial review and ongoing full-scale sampling. This will enable the consortium to build an understanding of ecological processes and contributing factors affecting the DW microbiome.
through space and time. Although different baselines have been observed for various treatment and distribution processes at different spatiotemporal scales, commonalities and trends that can be modeled and engineered may continue to emerge because finished water quality parameters do not generally fluctuate over a wide range. To illustrate, the water industry is typically required to monitor microbial and physiochemical water quality, and operational data (hydrodynamics, treatment processes, etc.). These valuable long-term continuous data can be leveraged along with epidemiological data to draw connections between nascent microbiome data, functionality of the water treatment and distribution systems, and potentially even with human health outcomes. Additionally, utilities could archive samples obtained during their routine surveillance for microbiome analyses by scientific members of the DWMP. This ongoing effort might be funded through research mechanisms for large centers, such as the NSF ERC (Engineering Research Centers) or GRC (Growing Convergence Research), or the NIEHS (National Institute of Environmental Health Sciences) EHSCC (Environmental Health Sciences Core Centers) in the USA to bring together the biological and chemical aspects of the system. The Phase 1 workshop series will build the coalition and provide the background knowledge and justification necessary for such an endeavor.

Phase 3 of the DWMP will establish effective communication developed by scientists, regulators, and utilities to raise awareness of the importance and value of DW microbiomes and share research findings with the public. The DWMP engagement of citizen scientists will enable natural channels for public communication. Questions relative to omics analyses and how to design studies and interpret data with respect to shortcomings of molecular techniques should be disseminated by the leadership lab(s) among scientific communities. Clearly identifying and communicating the importance of determining DW microbiomes as a discovery process and of managing DW microbiomes to protect public health will incentivize this research. This communication and education component will be integral to DWMP success.

**Concluding Remarks and Future Perspectives**

The time is opportune for a large-scale coordinated drinking water microbiome project that includes cross-directorate funding, involvement of laboratories around the world, and cooperation with utilities and the public. Although we can learn from other microbiome projects, DW microbiomes require a dedicated research focus because of their unique ecosystems and their diverse environmental and engineered parameters. It is critical to determine specific effects of
engineering treatments on various source waters and distribution systems, and parameters affecting changes in DW microbiomes, highlighted in the Outstanding Questions Box. Particularly needed is information on distribution systems and buildings to reduce pathogens, and to explore how DW microbiomes respond to new and/or alternative water sources in responses to climate and population changes. Characterization of planktonic and biofilm communities and their functions will help differentiate baseline and impacted DW microbiomes. Reference database development will be required using ontologies relevant to the DW microbiome. The ultimate goal is to provide safe and healthy drinking water for all.
Box 1: It’s never too late to learn … from other microbiome projects.

Broad coordinated efforts to characterize microbiome dynamics have been conducted across a range of different ecosystems, from host-associated [28–30] to environmental [31] ecosystems. These have taken the form of top-down efforts coordinated across multiple research groups (e.g., Human Microbiome Project, Meta-HIT), to citizen science projects (e.g., American Gut), and/or a hybrid involving dissemination of a common set of standardized protocols adopted by numerous research groups (e.g., Earth Microbiome Project). Despite evidence that differences in sample handling, DNA extraction, sequencing, and data processing workflows adopted by different groups can result in sufficient variation to mask underlying biological signal [26,32], there has been a lack of consensus on standardization efforts primarily due to demonstrable limitations of early standardization efforts. More recently, the focus has shifted from standardizing across studies towards development and adoption of best practices [33] in design and implementation of microbiome studies that may be adopted depending on study design and hypothesis [34]. Nonetheless, the key consensus has emerged from these large-scale microbiome efforts around data recording/reporting standards and minimum quality assurance and control (QA/QC) requirements. For instance, the Genome Standards Consortium encouraged the adoption of the MIMARKS and MIxS reporting standards to allow for better comparison of amplicon sequencing data from multiple studies [35]. Similarly, building on the Minimum Information about Genome Sequence (MIGS) specifications proposed by Field et al. in 2008 [36], Bowers et al. recently proposed the MIMAG and MISAG reporting specifications for metagenome assembled and single amplified genomes, respectively [37]. These reporting standards are crucial towards ensuring data quality and re-usability over the long-term. The second key consensus has been around inclusion of negative controls and mock communities primarily as a means of identifying both contamination artifacts [38] and batch effects [39,40] that may confound both within- and cross-study data analyses.
Figure 1: Sources of microbes and factors affecting biogeography of drinking water microbiomes to which the public is exposed. There are many sources of microorganisms and various physiochemical factors shaping microbial biogeography throughout the process of DW production and delivery to taps. The complex relationships in each niche throughout the process all play a role in determining the DW microbiome to which the public may be exposed. Some of the niches for microbiomes throughout this continuum include bulk water and biofilms in source waters (such as surface water from streams or reservoirs, ground water, desalinated seawater, recycled wastewater, or blends of these sources), various treatment processes (such as biostabilization without disinfectant residual treatment approaches shown in A and conventional treatment with disinfectant residual approaches shown in B) to distribution (including storage, transmission, and travel through or stagnation in water main pipes and premise plumbing). Some of the many parameters potentially contributing to the complexity of DW microbiomes from source to tap include differences in source water type and quality (brought on by climate and population changes), differences in engineering treatment technologies, distribution system design and materials (including changes due to infrastructure rehabilitation and redesign), indoor building design and materials (including green building design which increases water stagnation in pipes), and consumer behaviors (such as water conservation). After delivery to taps, DW microbiomes inoculate and may even alter human microbiomes through ingestion (e.g., gastric infection), inhalation (e.g., respiratory infection), and adsorption (e.g., skin infection). UV/AOP: Ultraviolet light and advanced oxidation processes.
**Figure 2 (Key Figure): Knowledge gap in drinking water microbiomes compared to other microbiomes.** A. The total number of metagenome projects in the JGI IMG database (downloaded Oct 8, 2018) shown in the flow diagram (built on sankeymatic.com) are 2,184; 11,813; and 4,058 for engineered systems; environmental systems; and host-associated systems, respectively. Of the total 18,055 metagenome projects in the database, only 205 study or sample names mention drinking water or water treatment. B. The total basepairs in the NCBI SRA database (downloaded July 2018) depicts similar disparity in 2017 for metagenomic studies of water treatment systems versus host-associated and other environmental and engineered systems (not shown). Longitudinally, the total basepairs in the NSBI SRA database show data for DW microbiomes was unavailable before 2012, and continues to lag behind studies of aquatic environment and human-associated microbiomes. C. and D. Published manuscripts indexed in Scopus (downloaded October 2018) comparing study of drinking water to study of other metagenomes (in C) and microbiomes in general (in D) also demonstrate a large and expanding knowledge gap.
Figure 3: Proposed Drinking Water Microbiome Project (DWMP) framework and roadmap.

The roadmap for operation of the DWMP involves three stages including 1) experimental design, 2) the main effort of the DWMP project dedicated to better understanding DW microbiomes, and 3) ongoing outreach. The experimental design phase will consist of a critical review of existing literature to inform recommendations for best practices for protocols and strategies (which will be validated in round-robin testing by various stakeholders), and establishment and recruitment of participants in the stakeholder framework. The framework will enable the primary effort of the DWMP, which will be to coordinate the collection, analysis, and storage of the vast quantity of samples and data necessary for comprehensive characterization of DW microbiomes in a concurrent top-down and bottom-up approach. From the top-down, the central lab/network will provide guidance and coordination while maintaining databases and a biobank, while the involved stakeholders across sectors will provide sampling, analysis, and funding effort from the bottom-up. This will allow meta-analyses by the central lab or any participants that will enable elucidation of the factors affecting DW microbiomes over space and time in various systems. Throughout the data analysis effort and engagement with these various stakeholders, high quality and frequent communication strategies will be essential for effective outreach and education. Stakeholders should work together to derive guideline or best management practices for their engineering and distribution operations so as to meet the required drinking water quality that can best protect public health.
Acknowledgements

NH, FL, AP, and WL conceived and outlined the opinion. NH drafted the manuscript with contribution from FL, AP, PH, and WL. GJ, AP, PH, KK, MA, RRC, and ML provided critical review and comments on the manuscript. Authors acknowledge support from NSF (NSF) Division of Chemical, Bioengineering, Environmental, and Transport Systems (CBET) under award number 1701641. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of NSF. Because many discussions at an NSF Workshop held March 12-13, 2018 in Denver, CO (Advancing understanding of Microbiomes in Drinking Water Distribution Systems and Premise Plumbing Using Meta-omics Techniques) informed this opinion, authors would like to acknowledge input of all workshop participants.
References


4. Maupin, M.A. et al. (2014) Estimated Use of Water in the United States in 2010,


AWWA Buried No Longer: Confronting America’s Water Infrastructure Challenge. (2012), 1–37


Sinha, R. et al. (2017) Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* DOI: 10.1038/nbt.3981


Yeh, Y.-C. et al. (2018) Taxon Disappearance from Microbiome Analysis Reinforces the
Value of Mock Communities as a Standard in Every Sequencing Run. *mSystems* DOI:

10.1128/mSystems.00023-18