A SOURCE TO TAP INVESTIGATION OF MICROORGANISMS IN MINNESOTA'S GROUNDWATER SUPPLIES USED FOR DRINKING WATER

A THESIS SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

JOHN RANDOLPH GALT

IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

TIMOTHY M. LAPARA AND RAYMOND M. HOZALSKI

OCTOBER 2018

Acknowledgements

I would like to thank the Legislative-Citizen Commission on Minnesota Resources (LCCMR) for recommending this project for funding provided by the Minnesota Environment and Natural Resources Trust Fund (ENRTF).

I would also like to extend warm gratitude to my advisors, Tim LaPara and Ray Hozalski, for their mentorship and for giving me the opportunity to work on this project. Tim, I enjoyed the sampling trips spent at your cabin (so long and thanks for all the fish) – those trips were fine blends of work and leisure. Ray, thank you for hosting your Christmas parties every year. Dinners were delicious and I am sorry you had to hear me butcher songs on karaoke.

Thank you to all of the water utilities that participated in this study. You have helped all Minnesotans learn more about the quality of one of our most precious resources.

Michael Waak and Ben Ma, thank you for your patience and support throughout my analyses. Mike, I sincerely appreciate the time you spent with me helping me through my sequencing data – I have learned a great deal from you. *Tusen takk*. Ben, thank you for spending the time to help me with the solid phase extractions/mass spectrometry data analysis and statistical methods – I have learned a great deal from you too.

Thank you to everyone who has helped me with sample processing and instrumentation in the labs, particularly Emma O'Leary and Ellie Hill. I couldn't have collected and processed all of my samples without you.

Dedication

To all my family and friends who have supported me throughout the duration of this project.

Abstract

Groundwater is often a desirable drinking water source because it is generally free of suspended solids and microbial pathogens and thus requires minimal, if any, treatment prior to distribution. Epidemiological studies have shown, however, that consumption of untreated groundwater increases risk of gastrointestinal illness. Previous work in Wisconsin, USA reported the occurrence of pathogenic viruses in groundwater supplies and resulting health impacts but bacterial pathogens were not investigated. In this study, a high-volume (300 – 1500 L) dead-end ultrafiltration sampling method was used to capture and recover microbes from 16 public groundwater systems throughout the State of Minnesota. The systems were sampled at the wellhead or source, after treatment if any (i.e., two systems did not treat or disinfect before distribution), and from one location in the distribution system. DNA was extracted from the microbes recovered in these samples and used as template for quantitative PCR analyses targeting 14 genes corresponding to pathogenic bacteria, one gene for a DNA virus, and the 16S rRNA gene as a marker for total bacteria. All samples were negative for the targeted genes from *Campylobacter jejuni*, *Shigella* spp., and Adenovirus; *Escherichia coli*-specific genes were only detected in water from a non-potable well with a documented history of contamination. Genes markers for two genera, *Legionella* and *Mycobacteria*, that include species that are opportunistic pathogens, were detected in four of the 16 public groundwater supplies, with *Legionella* levels decreasing in disinfected systems while *Mycobacteria* levels tended to increase. Raw water 16S rRNA gene concentrations ranged from 10^5 – 10^8 gene copies L^{-1}, decreased to background levels after disinfection,

then rebounded at the tap in the majority of cities. There was no significant difference in 16S rRNA gene concentrations from source-to-tap in the two non-disinfecting cities. Raw water samples contained diverse and previously uncharacterized organisms as revealed by DNA sequencing analyses, and beta diversity analyses suggest that community composition is driven by source water and/or disinfection. The results from this study suggest that groundwaters supplying public water systems in Minnesota are largely free of enteric pathogens but may contain opportunistic pathogens.

TABLE OF CONTENTS

List of Tables

Main body tables

Appendix tables

List of Figures

Main body figures

Appendix Figures

1 Introduction

Approximately 2.5 billion people worldwide rely on groundwater for daily use.[1] Groundwater is generally understood to be free of pathogens and safe for consumption with little to no treatment. In fact, many public water systems in the USA and elsewhere supply groundwater with little or no treatment and even without disinfection.[2,3] Nevertheless, groundwaters are susceptible to microbiological contamination from a variety of sources such that consumption of contaminated groundwater without disinfection can sometimes result in waterborne disease.[4,5]

Previous work regarding microbial contamination of groundwater supplies used for drinking water has been limited to pathogenic viruses. Studies conducted in Wisconsin, USA have detected pathogenic viruses in groundwater supplies resulting from leaky septic systems,[6] surface water intrusion events,[7] and vertical migration from agricultural practices.[8] The drinking water can also become contaminated after it enters the drinking water distribution system (DWDS) through negative pressure events,[9] maintenance events,[10] and poorly maintained sanitary sewer infrastructure.[11] Furthermore, epidemiological studies have shown that the consumption of untreated groundwater via private wells or small systems (i.e., < 1,000 people served) increases risk of acute gastro-intestinal illness (AGI).[5,12,13] In the United States between 1971 – 2008, 248 waterborne disease outbreaks involving groundwater systems were reported to the U.S. Center for Disease Control (CDC). From these outbreaks, over 23,000 people became ill resulting in 390 hospitalizations and 13 deaths.[4] Many epidemiological

analyses of groundwater supplies focus on viral pathogens; comparatively little is known about pathogenic bacteria in groundwater systems.

In the United States of America (USA), the legislation governing groundwater disinfection is known as the Groundwater Rule (GWR) administered by the U.S. Environmental Protection Agency (USEPA).[14] Disinfection of groundwater supplies is not mandated as part of the GWR because not all groundwater supplies are susceptible to contamination. In many cases, universal disinfection would result in unnecessary treatment and costs. Rather, the GWR focuses on a risk-targeted approach for ensuring safe drinking water, primarily requiring source water monitoring for indicator organisms. Water samples are collected in compliance with the Total Coliform Rule (TCR), which sets guidelines for corrective action following the occurrence of a coliform-positive drinking water sample. Similar approaches are followed in the European Union (EU), whose Drinking Water Directive (DWD) must be translated into national law by all EU member states.[15] The DWD does not explicitly require disinfection of drinking water, but that drinking water at the tap be free of any concentration of pathogenic microorganisms that may pose a potential risk to human health. Many EU member countries, such as the Netherlands and Austria, prefer high quality protected groundwater sources so that disinfection is not needed.[3,16] Despite the stipulations in the GWR and the EU member state implementation of the DWD, there are growing bodies of evidence that consumption of non-disinfected groundwater increases risk of disease and gastrointestinal illness, even when routinely monitored for indicator organisms.[5,13,17,18]

To mitigate this potential risk of waterborne disease, groundwaters are often disinfected and a residual disinfectant is maintained within the DWDS to suppress microbiological regrowth. Free chlorine ($HOCl/OCl^-$) and chloramines (predominantly monochloramine, NH_2Cl) are common and effective disinfectants used in drinking water treatment. Free chlorine has been used for drinking water disinfection since 1909[19] and historically it has been the most commonly used disinfectant. Free chlorine, however, has been shown to produce halogenated disinfection by-products (DBPs) from reactions with natural organic matter in the water.[20] Some of these DBPs are known or suspected human carcinogens.[20,21] Chloramine use has increased in recent years because chloramines do not form halogenated DBPs,[21] although there is emerging concern about the potential formation of nitrosamines.[22] Another major drawback to the use of chloramine as a residual disinfectant is its instability in the presence of nitrite. Accelerated chloramine decay can occur in the presence of ammonia-oxidizing microorganisms as the nitrite produced by these microorganisms reacts with chloramine.[23] Furthermore, chlorine and chloramines can present aesthetic problems to consumers in the form of undesirable tastes and odors,[24] undermining consumer perception of water quality. Because of these drawbacks, the use of chlorine-based disinfectants is increasingly being called into question.[25]

This research was performed to characterize the microorganisms (primarily bacteria) in groundwater-sourced public water systems in Minnesota, USA from source to tap, including systems with and without disinfection. A total of 16 water systems located in four different regions of the state were analyzed on one to four different occasions

each over a 16-month period from June 2016 to October 2017. Ultrafiltration membrane cartridges were used to capture microbes from 320 – 2454 L of raw groundwater, treated water leaving the treatment facility, and one location at the tap in each system. The use of hollow-fiber ultrafiltration membrane cartridges for microbial capture allowed for much greater sample volumes than is afforded by conventional filter disks (~1 L), thereby increasing sensitivity of the subsequent microbial assays. The quantities of 16 genes (14 genes associated with bacterial pathogens, the 16S ribosomal RNA (rRNA) gene as a measure of total biomass, and one DNA virus) in the concentrates were assessed via quantitative polymerase chain reaction (qPCR). Bacterial community structure was analyzed by sequencing PCR-amplified 16S rRNA gene fragments using the Illumina MiSeq platform. Water samples were also collected for chemical analyses, including indicators of water age (tritium content) and wastewater contamination (carbamazepine).

2 Materials and Methods

2.1 Study Systems

Four general regions in Minnesota were selected for investigation: the Twin Cities Metro area (metro region), the Brainerd Lakes area (central region), the Karst region of southeast Minnesota (southeast region), and the agricultural region of southwest Minnesota (southwest region). Municipal water supplies in the metro region draw water from deep wells (> 200 ft) and disinfect with free chlorine; two of the four systems in the central region subsequently quench with ammonia, leaving a chloramine residual in the distribution system. The central region is characterized by high water tables and numerous lakes. Several small municipalities in this region do not disinfect. The southeast region is unique in Minnesota because of its Karst geology. Karst geology is characterized by numerous cracks and sinkholes, making the underlying groundwater particularly susceptible to surface runoff infiltration and contamination.[26] Systems in the southeast region that participated in this study, however, drew water from deep wells (> 300 ft) that penetrated through the Karst layer and extracted water from the underlying geologic layers. Lastly, the agricultural southwest region is typified by high water tables and susceptibility to high nitrate and nitrite concentrations due to high agricultural land use.

2.2 Quantification of Total Coliforms and *Escherichia coli (E. coli)*

Water samples for culture of total coliforms and *E. coli* were collected in 100 mL plastic bottles containing sufficient sodium thiosulfate for neutralization of up to 15 mg

Cl_2 L^{-1} chlorine. All water samples were incubated at 35°C for 24 hours and analyzed with the IDEXX Colilert® Quanti-Tray test kit (IDEXX, Westbrook, ME) within 48 hours of collection. Measurements for total coliforms and *E. coli* were reported as MPN 100 mL^{-1}.

2.2.1 Dead-end Ultrafiltration

Microorganisms were captured from the raw, treated, and tap water on-site using REXEED 25S ultrafiltration membrane cartridges (Asahi Kasei, Tokyo, Japan) as described by Smith and Hill.[27] Each sample tap was first flame sterilized, then braided PVC tubing was attached and the tap was opened to rinse the tap and tubing for 2 to 3 minutes prior to connecting an ultrafilter cartridge. Then, the flow was turned off and a new ultrafilter cartridge was connected to the braided PVC tubing. After connecting a flow meter to the outlet of the ultrafilter cartridge, the flow was resumed at a target flow rate of 4-6 L min^{-1}. The total volume filtered ranged from 320 – 2454 L with a mean of 922 L ± 339 L (mean ± standard deviation) across all samples. The membrane cartridges were transported on ice to the laboratory for subsequent backflushing and concentration of microorganisms. Method blank ultrafilter samples were collected by backflushing fresh unused ultrafilter cartridges.

Environmental samples and method blank ultrafilters were backflushed using 500 mL autoclave-sterilized surfactant solutions of 0.50% Tween-80, 0.01% sodium hexametaphosphate (NaPP), and 0.001% Y-30 anti-emulsion. The microbial cells were collected from the backflush solution via coagulation with a solution containing 0.2 M

sodium chloride, 8% (w/v) polyethylene glycol, and 1% (w/v) beef extract, settling for 24 hours, and finally centrifugation at 12,000 × g for 45 min. The supernatant was decanted, and the remaining pellet was resuspended using 1- 5 mL of 10X TE buffer. The resulting final concentrated sample volumes (FCSVs) were stored at -20°C prior to DNA extraction. Concentration factors on average were ~10^5-fold. The true average concentration factors was ~10^4-fold, however, because only 0.3 mL of each FCSV was used for DNA extractions (approximately 10% of the FCSV). A total of 109 environmental samples and 7 method blanks were collected.

2.2.2 DNA Extraction & Purification

DNA was extracted from the FCSVs using the FastDNA™ SPIN Kit (MP Biomedicals, Santa Ana, CA). Lysis buffer (5% m/v SDS, 120 mM sodium phosphate buffer, pH 8.0) was added to a 300 µL aliquot of concentrated samples, which were then subjected to three freeze-thaw cycles, followed by a 90-minute incubation at 70 °C. DNA was stored at -20 °C until further use.

2.2.3 Quantitative PCR

Quantitative PCR was performed on each environmental sample targeting 14 genes specific to bacterial pathogens, the 16S rRNA gene for quantifying total biomass, and one gene specific to a viral pathogen (Table 1). Assays were performed using a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA). Final reaction mixtures were 20 µL and consisted of nuclease-free water, 10 µL SsoAdvanced™

Universal Probes Supermix (EvaGreen for the 16S rRNA gene assay), 20 μg bovine serum albinum, 1 μL template DNA, and varying concentrations of primers and probes depending on the assay (Table AP1).

Table 1 - qPCR gene targets, primer and probe sequences, and references.

Taxonomic Target	Target gene name	Gene product	Primer[a] & Probe[b] (5'-3' sequence)	Reference
Campylobacter jenjuni	*cadF*	Fibronectin-binding protein	F: TGC TAT TAA AGG TAT TGA TGT RGG TGA R: GCA GCA TTT GAA AAA TCY TCA T P: UPL 039	Ishii et al. 2013[28]
Campylobacter jenjuni	*ciaB*	Invasion antigen B	F: GCG TTT TGT GAA AAA GAT GAA GAT AG R: GGT GAT TTT ACT TTC ATC CAA GC P: UPL 137	Ishii et al. 2013
Mycobacterium avium complex (MAC)	ITS	Internal transcribed spacer region	F: TTG GGC CCT GAG ACA ACA CT R: GCA ACC ACT ATC CAA TAC TCA AAC AC P: CCG TGT GGA GTC CCT CCA TCT TGG	Rocchetti et al. 2016[29]
E. coli	*ftsZ*	Cell division protein	F: CTG GTG ACC AAT AAG CAG GTT R: CAT CCC ATG CTG CTG GTA G P: UPL 071	Ishii et al. 2013
E. coli	*uidA*	Beta-D-glucuronidase	F: CCC TTA CGC TGA AGA GAT GC R: TTC ATC AAT CAC CAC GAT GC P: UPL 113	Ishii et al. 2013
Enterohemorrhagic *E. coli* (EHEC)	*eaeA*	Intimin	F: GGC GAA TAC TGG CGA GAC TA R: GGC GCT CAT CAT AGT CTT TCT T P: UPL 028	Ishii et al. 2013
Enterohemorrhagic *E. coli* (EHEC)	*stx1*	Shiga toxin 1 subunit A	F: TGT AAT GAC TGC TGA AGA TGT TGA T R: TCC ATG ATA RTC AGG CAG GA P: UPL 060	Ishii et al. 2013
Enterohemorrhagic *E. coli* (EHEC)	*stx2*	Shiga toxin 2 subunit A	F: TCT GGC GTT AAT GGA GTT YAG R: GTG ACA GTG ACA AAA CGC AGA P: UPL 126	Ishii et al. 2013
Shigella spp. and enteroinvasive *E. coli*	*virA*	Secreted VirG-processing protein	F: GGC AAT CTC TTC ACA TCA CG R: TTC GGA CAT AAT TTG GGC ATA P: UPL 006	Ishii et al. 2013
Enterococcus spp.	23S rRNA	Large subunit, ribosomal RNA	F: GAG AAA TTC CAA ACG AAC TTG R: CAG TGC TCT ACC TCC ATC ATT P: TGG TTC TCT CCG AAA TAG CTT TAG GGC TA	Ludwig and Schleifer 2000[30]
Legionella spp.	*ssrA*	Transfer-messenger RNA	F: GGG CGA CCT GGC TTC R: GGT CAT CGT TTG CAT TTA TAT TTA P: ACG TGG GTT GCA A	Benitez and Winchell 2013[31]
Legionella pneumophila	*mip*	Macrophage infectivity potentiator	F: TTG TCT TAT AGC ATT GGT GCC G R: CCA ATT GAG CGC CAC TCA TAG P: CGG AAG CAA TGG CTA AAG GCA TGC A	Benitez and Winchell 2013
Legionella pneumophila serogroup 1	*wzm*	ABC transporter of LPS O-antigen	F: TGC CTC TGG CTT TGC AGT TA R: CAC ACA GGC ACA GCA GAA ACA P: TTT ATT ACT CCA CTC CAG CGA T	Benitez and Winchell 2013
Mycobacteria spp.	*atpE*	ATP synthase C chain AtpE	F: CGG YGC CGG TAT CGG YGA R: CGA AGA CGA ACA RSG CCA T P: ACS GTG ATG AAG AAC GGB GTR AA	Radomski et al. 2013[32]
All Bacteria	16S rRNA	Small subunit, ribosomal RNA	F: ACT CCT ACG GGA GGC AGC AG R: ATT ACC GCG GCT GCT GG -	Muyzer et al 1993[33]
Adenovirus	*hex*	Hexon protein for capsid coat	F: GGA CGC CTC GGA GTA CCT GA R: CGC TGI GAC CIG TCT GTG G P: CAC CGA TAC GTA CTT CAG CCT GGG T	Lambertini et al. 2012[34]

[a]Forward and reverse primer sequences are preceded by the letters 'F' and 'R', respectively.

[b]Probe sequences are preceded by the letter 'P'. Items containing "UPL" followed by a number represent proprietary probe sequences from the Universal ProbeLibrary® (Roche Molecular Systems. Inc, Pleasanton, CA)

Gene copies in each sample were determined through standard curves generated during each qPCR using known quantities of serially diluted gBlocks® Gene Fragments (Integrated DNA Technologies, Skokie, IL). Standard curves consisted of a minimum of five sequential standards that yielded amplification efficiencies from 88% - 105% and R^2 values from 0.995 - 1.00. Background signal was quantified for each assay using template DNA extracted from method blank samples. For each assay, positive detects were defined as samples with C_q values that were two standard deviations below the average C_q value of all negative controls.

2.2.4 Microbiome Analysis

Bacterial community analysis was performed by sequencing the PCR-amplified V3 region of the 16S rRNA gene using the Illumina MiSeq platform[33,35] at the University of Minnesota Genomics Center. Barcode sequences were attached to the 5' end of the forward and reverse primers before amplification, allowing pooling of samples before addition to a single-lane flow cell. DNA samples were sequenced using 2x150 paired-end reads on a MiSeq System instrument (Illumina, Inc., San Diego, CA).

DNA sequences from environmental samples were analyzed using QIIME 2 v2018.2.[36] First, the 338F/518R primers were removed from reads using the 'cutadapt' plugin,[37] then Phred quality scores were calculated and used to trim forward and reverse reads to 127 and 91 bp, respectively. Amplicon errors were resolved using the Divisive Amplicon Denoising Algorithm (DADA2), allowing sequence filtering, de-replication, sample inference, chimeric sequence detection, and merging of paired-end reads.[38]

Amplicon sequence variants (ASVs) were re-checked and filtered to remove chimeras and borderline chimeras using the *de novo* UCHIME algorithm[39] through the VSEARCH plugin.[40] Next, sequences were shortened to the V3 region using the 'feature-classifier' plugin. Non-target ASVs were identified and removed by comparison against the SILVA (release 128) 99% operational taxonomic unit reference sequences,[41] maintaining a minimum 97% alignment and 70% identity. The reference sequences contained in the SILVA and "All-species Living Tree Project" (release 128) taxonomic framework[42] were used to train a naïve Bayesian classifier,[43] which was used to assign taxonomy to the ASVs with at least 90% bootstrap confidence.

2.3 Chemical Analyses

2.3.1 Chlorine Analysis

Free and total chlorine were measured immediately after water sample collection using a Hach® DR900 Multiparameter Portable Colorimeter (Hach, Loveland, CO) using the USEPA DPD Method 10245 for free chlorine (valid range of 0.05 to 4.0 mg L^{-1} Cl_2) and Method 10250 for total chlorine (valid range of 0.05 to 4.0 mg L^{-1} Cl_2). Water samples were collected in clean 1 L high-density polyethylene (HDPE) bottles and returned to the lab on ice for subsequent water quality analyses. Water samples were stored at 4°C and analyzed for nitrite and nitrate within 48 hours of sample collection with the Hach® DR900 colorimeter, using the Diazotization Method 8507, and the Cadmium Reduction Method 8171, respectively.

2.3.2 Chloride and Sulfate Analysis

Water samples for chloride and sulfate concentration measurements were collected from each sampling location in 50 mL vials and stored at 4°C until analysis by ion chromatography (IC). Chloride and sulfate concentrations were measured using a Metrohm 930 Compact IC Flex instrument (Metrohm, Riverview, FL). Each injection drew 15 mL of sample into the instrument at a flow rate of 0.7 mL min^{-1} using a 3.2 mM carbonate/1.0 mM bicarbonate buffer as the eluent and 0.5 M H_2SO_4 as the regenerant. Anion separation was performed using a 150mm x 4.0mm Metrosep A Supp 5 column (Metrohm, Riverview, FL). Chloride and sulfate stock solutions were prepared with reagent-grade salts (Fisher Scientific, Waltham, MA) and diluted with ultrapure water to produce standards ranging from 1 – 100 mg L^{-1} for chloride and 1 – 500 mg L^{-1} for sulfate.

2.3.3 Carbamazepine Analysis

The method for carbamazepine analysis was adapted from Vanderford and Snyder.[44] Water samples for carbamazepine analysis were collected once from each raw water sampling location in each system. Four liters of raw groundwater were collected in 1 L pre-combusted glass bottles and returned to the lab on ice. The samples were processed within 24 hours as described in the Appendix. Samples were filtered using 0.7 μm glass fiber filters, then acidified to pH 2.0 with H_2SO_4. All samples were spiked with a deuterated carbamazepine isotope solution (Cambridge Isotope Laboratories, Tewksbury, MA) to a final concentration of 10 ng L^{-1}. Two of the four samples were

spiked with an unlabeled carbamazepine solution (Sigma Aldrich, St. Louis, MO) to a final concentration of 50 ng L^{-1}. Samples were stored for up to 10 days before performing a solid phase extraction (SPE). The unknown carbamazepine starting concentration was calculated using a previously constructed standard curve. The limit of detection (LOD) and the limit of quantification (LOQ) were 0.018 ng L^{-1} and 0.059 ng L^{-1}, respectively (Table AP2).

2.3.4 Tritium Analysis

Tritium quantities were determined by liquid scintillation counting (LSC) at the Environmental Isotope Laboratory (EIL, University of Waterloo, Ontario, Canada). Water samples were collected in 1 L HDPE bottles, sealed with parafilm over the bottlecap, and shipped to the EIL for analysis. Samples were pre-treated by ion-exchange or azeotropic distillation until conductivity was below 100 μS cm^{-1}. Afterwards, samples were electrolytically enriched 15 times then counted using a liquid scintillation counter. Detection limits were 0.8 tritium units (TU) corresponding to 2.58 picocuries L^{-1}.

2.4 Statistical analyses

Data handling and statistical analyses were performed using R (version 3.4.4).[45] Non-parametric hypothesis testing was performed using *Wilcoxon* rank-sum tests[46] when analyzing differences between two samples from different populations, and *Kruskal-Wallis* one-way analysis of variance tests when comparing samples from three or more groups. *Post hoc* testing was performed with pairwise *Wilcoxon* tests to determine

pairwise significance between groups. All p-values reported from *post hoc* pairwise tests include the unadjusted p-value and a false discovery rate adjusted (FDR) p-value.[47] Statistical analyses on qPCR data were performed on $\log_{10}$-transformed gene concentrations, and non-detect samples were substituted with the detection limit where applicable to assign low ranks to these samples. Statistical significance in all tests were performed at the $\alpha = 0.05$ level.

Between sample diversity (beta diversity) was evaluated using the Bray-Curtis, weighted, and unweighted UniFrac dissimilarity matrices. Dissimilarity matrices were computed with 'beta-group-significance' plugin[48] and exported for principal coordinate analyses using the vegan package in R.[49]

Permutational analysis of variance (PERMANOVA) tests were performed (999 random permutations) on the Bray-Curtis dissimilarity metric using the 'adonis' function in the vegan package in R. Raw water sample community compositions were assessed for the effects of well depth (sample bins were assigned as follows: < 200 ft = 'shallow' wells, 200 – 400 ft = 'medium' wells, and > 400 ft = 'deep' wells) and location (region and systems within each region). The effect of disinfection type (raw water samples were treated as non-disinfected samples) on community composition was assessed including all sample types (raw, treated, and tap samples). *Post hoc* analyses were performed by pair-wise comparisons using the 'pairwise.adonis' package in R.[50] Homogeneities of sample group dispersions (an assumption of PERMANOVA) were computed using the 'betadisper' and 'permutest' (999 permutations) functions in the vegan package in R, and *post hoc* analyses were performed with pairwise *Wilcoxon* tests.

A constrained correspondents analysis (CCA) was conducted on raw water samples using a collapsed genus-rank ASV table. Variance in raw water samples was constrained to well depth, tritium content, and carbamazepine concentration, first individually then in tandem, to determine the fraction of total variance explained by each environmental variable and the combination of all three. Significance of the explained variance by each environmental variable, and the combination of all three, was assessed by randomly permuting the metadata (999 random permutations) and recalculating the CCA each time. The fraction of randomized variances that exceeded the observed variances calculated using the non-permuted metadata was used as the p-value.

3 Results

3.1 Raw water conditions

Well depths, tritium content, and carbamazepine concentration at the raw water sampling locations varied from region to region (Table 2). The central and southwest regions contained the shallowest wells; more than half of the wells in these two regions were less than 100 ft deep. Deeper wells were observed in the metro region (range 219 – 414 ft), but the deepest wells were observed in the southeast region (range 334 – 1204 ft). All wells in the southeast region penetrated through the karst geology, circumventing the possibility of drawing groundwater that may be susceptible to surface contamination. Tritium content varied significantly from region to region (*Kruskal-Wallis,* $p = 0.026$) and correlated with well depth (*Spearman,* $\rho = -0.77$, $p = 5.5 \times 10^{-10}$). All samples from the central region and the majority of samples collected from the southwest region (4/5) contained young water (i.e., tritium content was detectable, median 5.1 and 5.2 TU for the central and southwest regions, respectively). Two of the four samples from the metro region and all samples from the southeast region did not contain detectable levels of tritium (< 0.8 TU), indicating older water.

Table 2- Raw water conditions for all public water systems

Region/system	Well Depth (ft)	^{3}H (TU ± 67% CI)[a]	Carbamazepine (ng L^{-1})
Central			
C1	60	5.0 ± 0.5	< LOD
C2	148	5.2 ± 0.5	< LOD
C3	115	4.2 ± 0.5	5.08
C4	76	7.2 ± 0.7	< LOD
Metro			
M1	262	4.7 ± 0.5	< LOD
M2	219	<0.8 ± 0.2	< LOD
M3	395	1.9 ± 0.3	< LOD
M4	414	<0.8 ± 0.2	< LOD
Southeast			
SE1	904	<0.8 ± 0.2	< LOD
SE2	334	<0.8 ± 0.3	-
SE3	1204	<0.8 ± 0.3	< LOD
SE4	450	<0.8 ± 0.3	-
Southwest			
SW1	298	<0.8 ± 0.3	2.34
SW2	67	5.2 ± 0.5	0.093
SW3	37	5.7 ± 0.6	1.93
SW41[b]	335	5.2 ± 0.5	< LOD
SW42	261	4.5 ± 0.5	< LOQ

[a]Tritium content measured as tritium units (TU), 1 TU ≈ 3.22 picocurries L^{-1}. Reported value of ^{3}H content is within the margin of error with 67% confidence.
[b]Non-potable well

3.2 Water quality

3.2.1 Total coliforms and *E. coli*

Total coliforms and *E. coli* detection was sparse among all samples (Table AP3). The majority (9/10) of total coliform-positive samples were at or below 12.2 MPN 100 mL^{-1}. Two of 45 raw water samples were positive for total coliforms and each contained 1 MPN 100 mL^{-1}; three of 29 treated water samples were positive with a maximum concentration of 3.1 MPN 100 mL^{-1}. At the tap, four of 33 water samples were total-

coliform positive and ranged from 1 – 12.2 MPN 100 mL^{-1}. All samples drawn from potable-use water systems were negative for *E. coli*. The highest total coliform concentration and the only *E. coli* positive sample was drawn from the non-potable well in the southwest region that contained 727 MPN 100 mL^{-1} total coliforms, and 249.5 MPN 100 mL^{-1} *E. coli*.

3.2.2 Water chemistry

Total, free, and combined chlorine concentrations from treated and tap water samples for each system are summarized in Table AP4. Systems with free chlorine concentrations higher than chloramine concentrations at the tap were classified as chlorinated systems, and vice versa for chloraminated systems. Of the 13 systems that perform disinfection as part of regular treatment operations, four were classified as chloraminated (C2, M1, M3, and SW1), and the remaining nine systems were classified as chlorinated. Systems C1 and C3 do not perform disinfection as part of their treatment operations. Total chlorine concentrations ranged from 0.4 – 3.8 mg Cl_2 L^{-1} among all treated water samples, with the highest frequency (46% of samples) containing between 1.0 – 1.5 mg Cl_2 L^{-1}. System C2 on average supplied the highest total chlorine concentration of any system and kept consistent chlorination among replicate samples (3.8 ± 0.2 mg Cl_2 L^{-1}, mean ± standard deviation). System C4 on average supplied the lowest total chlorine concentration (0.4 ± 0.05 mg Cl_2 L^{-1}), which included one replicate containing no measurable chlorine in the treated water. Residual chlorine or chloramine decay (i.e., lower concentration at the tap than the treated water) between the treated and

tap water was observed for the majority (10/13) of the systems that disinfect. The southwest region contained the most stable chlorine residuals and showed little decay between the treated and tap water sampling locations.

Water quality parameters from source to tap for the central, metro, southeast, and southwest regions are compiled in Tables AP5, AP6, AP7, and AP8, respectively. Overall, less than half of all samples contained detectable concentrations of nitrate and nitrite (21/50 and 22/51, respectively). Nitrate and nitrite were either zero or present in low concentrations (< 1.0 mg NO_3^--N L^{-1}) in the central, metro, and southeast regions, and higher nitrate concentrations were measured in the southwest agricultural region (6/20 samples contained nitrate > 2 mg NO_3^--N L^{-1}). Nitrite concentrations in all samples did not exceed 0.02 mg NO_2^--N L^{-1}. Chloride concentrations were low (< 10 mg Cl^- L^{-1}) in samples collected from the metro and southeast regions, but in the central and southwest regions, concentrations were greater and correlated significantly with well depth (*Spearman*, $\rho = -0.57$, $p = 5.7 \times 10^{-3}$). Samples from all regions contained nominal sulfate concentrations (1.1 – 53.4 mg SO_4^{2-} L^{-1}) except system SW1, which contained concentrations ranging from 398 – 471 mg SO_4^{2-} L^{-1}, well in excess of the Secondary Drinking Water Standards (250 mg SO_4^{2-} L^{-1}). The pH of all samples was measured in the neutral to slightly basic range (range 7.0 – 8.0) and temperatures ranged from 12.9°C – 18.8°C.

3.3 Quantification of total bacteria

The total quantity of bacteria (as 16S rRNA gene copies) decreased roughly 100-fold on average between the raw water and treated water sampling locations in systems that disinfect (Figure 1). Total bacteria regrowth was observed in the majority of these systems (7/13). No significant change in total bacteria between the raw and tap water samples was observed in the two non-disinfecting systems (*Wilcoxon,* system C1: p = 0.89, system C3: p = 1.0). Total bacteria concentrations in all regions were consistent over the 1.5 – year long sampling period (Figure 2), although sampling events within the central, southeast, and southwest regions were sparse. There was no significant difference in total bacteria concentration in all regions between different seasons (*Kruskal-Wallis,* p = 0.54, p = 0.79, p = 0.71, and p = 0.65 for the central, metro, southeast, and southwest regions, respectively). Table AP9 provides further detail into this analysis.

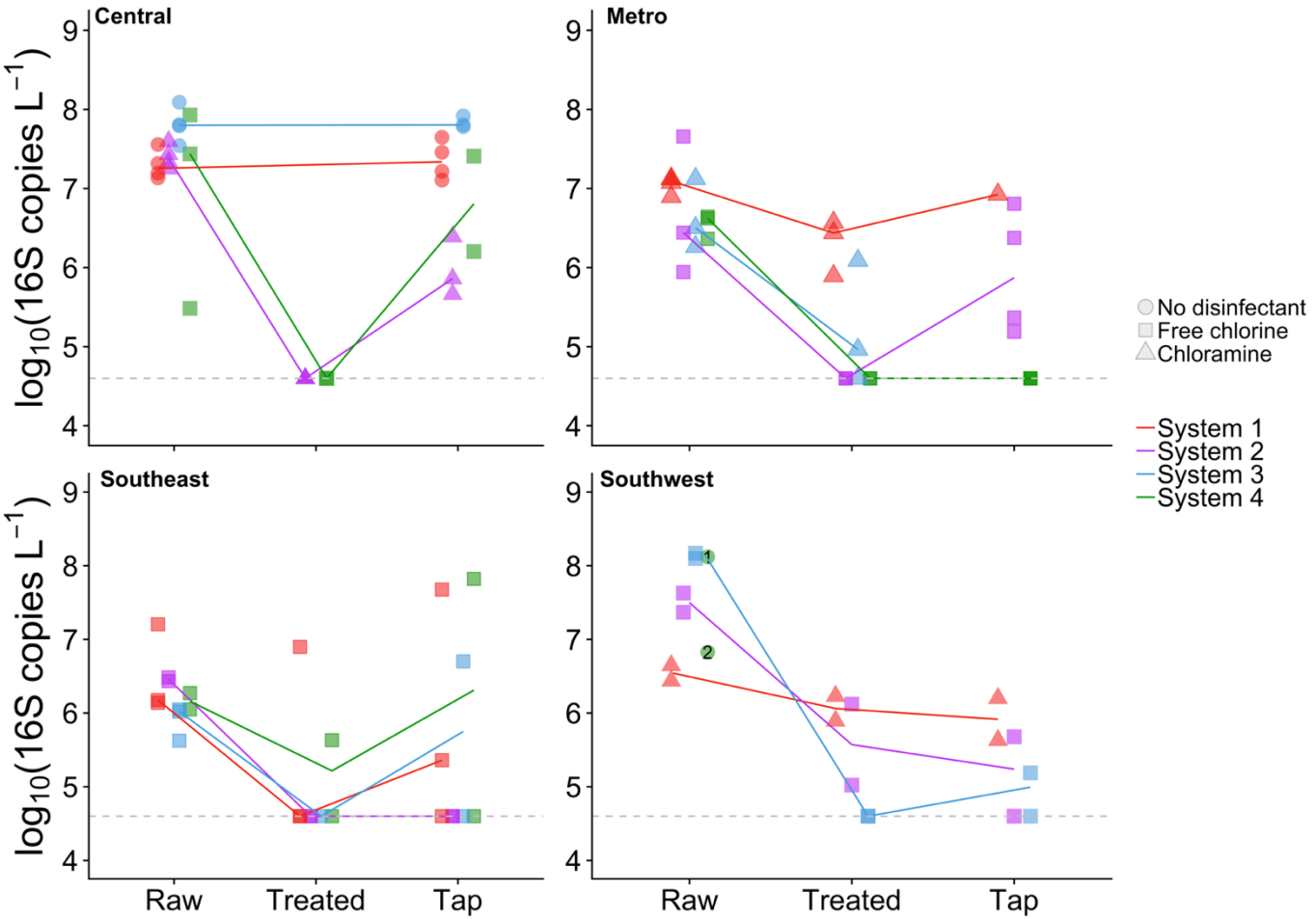

Figure 1 - Real-time qPCR quantification of total Bacteria of all samples. Colors represent separate water systems within each region, and shape denotes disinfectant type. Lines are drawn through the median observation at each location. Samples whose 16S rRNA gene copy number were below the average of the negative controls are located on the dashed line. Green circles in the Southwest plot labeled '1' and '2' represent (1) the non-potable well and (2) the potable well within the SW4 system.

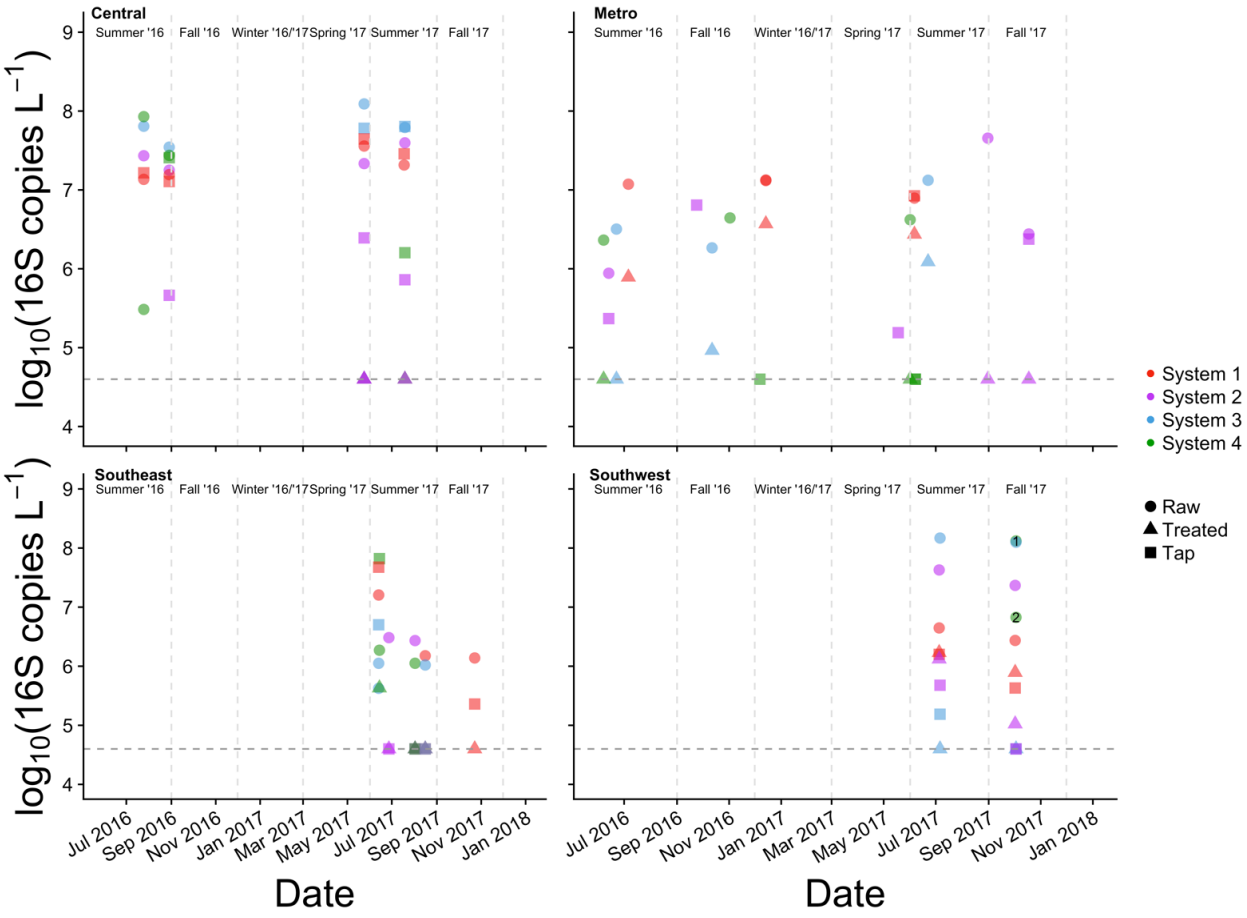

Figure 2 - Temporal variation in total bacteria concentrations as 16S rRNA gene copies within each system. Color denotes the system within each region and shape denotes the sampling location type (raw vs. treated vs. tap). Green circles in the Southwest subplot labeled '1' and '2' represent (1) the non-potable well and (2) the potable well within the SW4 system.

Raw water total bacteria quantities between systems within all regions did not differ significantly (*Kruskal-Wallis,* $p = 0.13$, $p = 0.40$, $p = 0.082$, $p = 0.16$, for the central, metro, southeast, and southwest regions, respectively). Pairwise comparisons of raw water total bacteria quantities between regions revealed significant differences between several regions (Table 3).

Table 3 – Pairwise post hoc comparisons results for 16S rRNA gene concentrations in raw water samples between regions (Pairwise Wilcoxon rank sum test; significant p-values are in bold)

	p	adj. $p_{Wilcoxon}$
Central vs metro	**2.9×10^{-4}**	**8.6×10^{-4}**
Central vs southeast	**1.6×10^{-4}**	**8.6×10^{-4}**
Central vs southwest	0.78	0.78
Metro vs southeast	**0.015**	**0.022**
Metro vs southwest	**0.045**	0.053
Southeast vs Southwest	**5.5×10^{-4}**	**1.1×10^{-3}**

Total bacteria in raw water samples inversely correlated with well depth (*Spearman,* $\rho = -0.81$, $p = 6.5 \times 10^{-11}$) (Figure 3A), positively correlated with tritium content (*Spearman,* $\rho = 0.77$, $p = 8.2 \times 10^{-10}$) (Figure 3B), and positively correlated with carbamazepine concentration (*Spearman*, $\rho = 0.45$, $p = 3.1 \times 10^{-3}$) (Figure 3C).

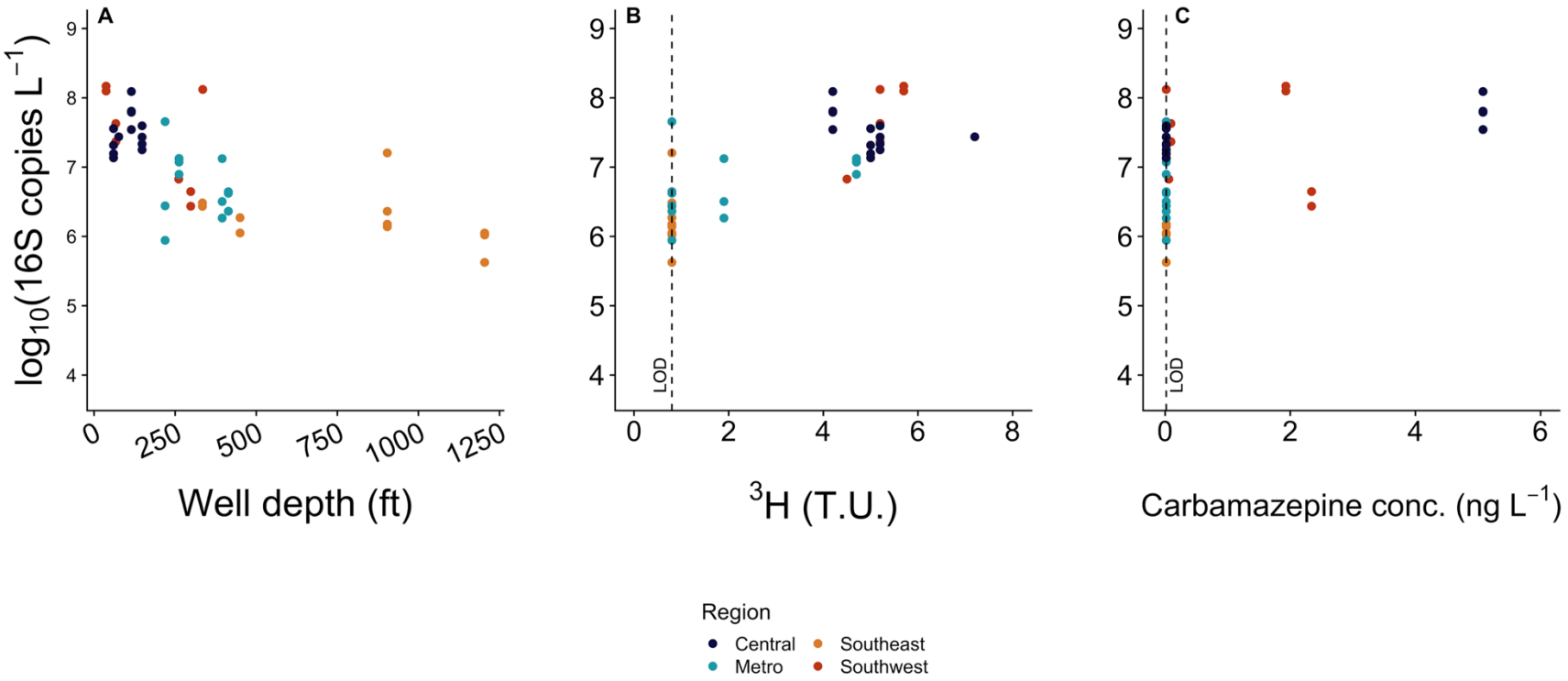

Figure 3 - Relationship between the quantities of 16S rRNA genes at each raw water sampling location to (**A**) well depth, (**B**) tritium content, and (**C**) carbamazepine concentration. Vertical dashed lines denote the limit of detection (LOD) for tritium content and carbamazepine concentration (0.8 TU and 0.018 ng L^{-1}, respectively).

3.4 Quantification of selected gene targets

Nine of 14 selected gene targets were below detection in all samples (Table AP10), including *cadF* and *ciaB* (*Campylobacter jejuni*), ITS (*Mycobacterium avium* complex), *eaeA*, *stx1*, and *stx2* (Enterohemmorrhagic *E. coli*), *virA* (*Shigella* spp. and Enteroinvasive *E. coli*), *mip* (*Legionella pneumophila*), and *hex* (Adenovirus). Of the six gene targets corresponding to *E. coli*, only two were positively detected (*ftsZ* and *uidA*) among all samples. Both genes were detected only once and were observed in the same non-potable well sample that was previously known to contain *E. coli*.

The 23S rRNA gene specific for *Enterococcus* spp. was detected at least once in 28/44 sampling systems (15 of the 16 systems), and the *atpE* and *ssrA* genes (targeting *Mycobacteria* spp. and *Legionella* spp., respectively) were detected at least once in four of the 16 systems (Figure 4) (medians and ranges of gene concentrations at each sampling location within each system are summarized in Table AP11). Positive detection of the 23S rRNA gene specific for *Enterococcus* spp. appeared random and was independent of sampling location, system, and region. All systems that were positive for the *atpE* gene contained no detectable levels in the raw water except for the non-potable well in the southwest region. The *atpE* gene was detected in the highest concentrations in two chloraminated systems and increased from source to tap in all cases. Three systems contained the *ssrA* gene in the raw water; two of these systems do not perform disinfection and the third disinfects with free chlorine. In the two non-disinfecting systems, *ssrA* concentrations at the taps remained at or above the raw water

concentrations. In contrast, the *ssrA* concentration in the chlorinated system fell below detection after disinfection.

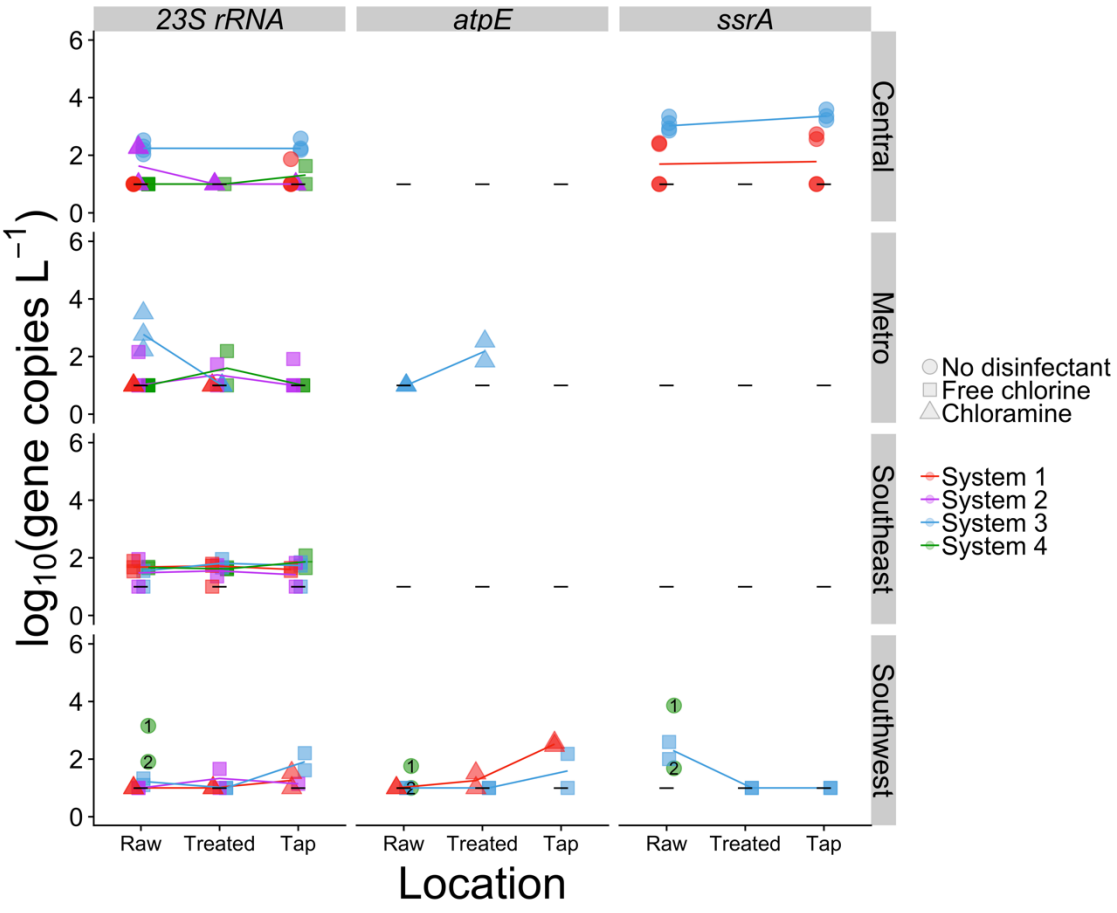

Figure 4 - Quantification of 23S rRNA specific for *Enterococcus* spp., *atpE* (*Mycobacteria* spp.), and *ssrA* (*Legionella* spp.) genes in all water systems separated by region (C = central, M = Metro, SE = Southeast, SW = Southwest). Green circles in the Southwest plot labeled '1' and '2' represent (1) the non-potable well and (2) the potable well within the SW4 system. Dashed lines represent the quantification limit (10 gene copies L^{-1}).

3.5 Community structure and composition

Sequencing of PCR-amplified 16S rRNA gene fragments resulted in a total of 5,997,930 high-quality sequences (range 4,608 – 263,669, median 75,315 sequences per

sample) in 73 samples and contained 25,056 amplicon sequence variants (ASVs). A total of 116 samples were collected, of which 43 failed a binary quality control screening likely due to insufficient template. Samples that failed quality control screening were generally collected from either treated water samples or from deep wells in the metro and southeast regions where total bacteria quantities were low. Random subsampling for diversity analyses was performed at a depth of 36,280 sequences to avoid biases arising from differing sequence depths among different samples and resulted in the omission of a single sample (one method blank). Two of the seven method blank samples passed quality control screening and are included in subsequent taxonomic composition and beta diversity figures.

Sequences belonging to the genera *Enterococcus, Mycobacteria,* and *Legionella* were observed in the samples that tested positive via qPCR. Sequences belonging to the genus *Enterococcus* were observed in 17/72 samples, including blank samples which contained relative abundances of these sequences up to 0.04%. qPCR results targeting *Mycobacteria* spp. are reflected in the sequencing results; no raw water samples contained sequences belonging to the genus *Mycobacteria* and the highest relative abundances of these sequences were seen in the treated and tap samples. Sequences belonging to the genus *Legionella* were observed in 45/72 samples, with the greatest relative abundances observed in the samples that tested positive for the *ssrA* gene via qPCR.

Bacteria community composition in water samples was primarily comprised of *Betaproteobacteria, Gammaproteobacteria,* and organisms unclassified at the class and

phylum taxonomic levels (Figure 5). In raw water samples, community composition was stable throughout the 1.5 - year long sampling period except for two systems that infrequently displayed shifts towards *Gammaproteobacteria*-dominated composition (systems C1 and M1). Treated and tap water samples that passed quality control were sparse and prevented observation of overall temporal trends in community composition. The majority of samples (62/72) contained sequences that are currently not assigned to any phyla in relative abundances up to 30% of total reads. Organisms belonging to the candidate phyla *Omnitrophica* and *Parcubacteria* were observed in nearly all raw water samples and were independent of the region from where the samples were collected. Community composition in the two systems that do not disinfect (C1 and C3) were unchanged between the source and tap water samples. In general, chlorine or chloramine disinfection tended to shift composition toward *Betaproteobacteria* and *Gammeproteobacteria*, although some systems (M1 and SW1) retained more evenness than others after disinfection and at the tap. The two method blank samples that passed quality control screening displayed similar composition to treated and tap samples (i.e., dominated by *Gammaproteobacteria*) (Figure AP1).

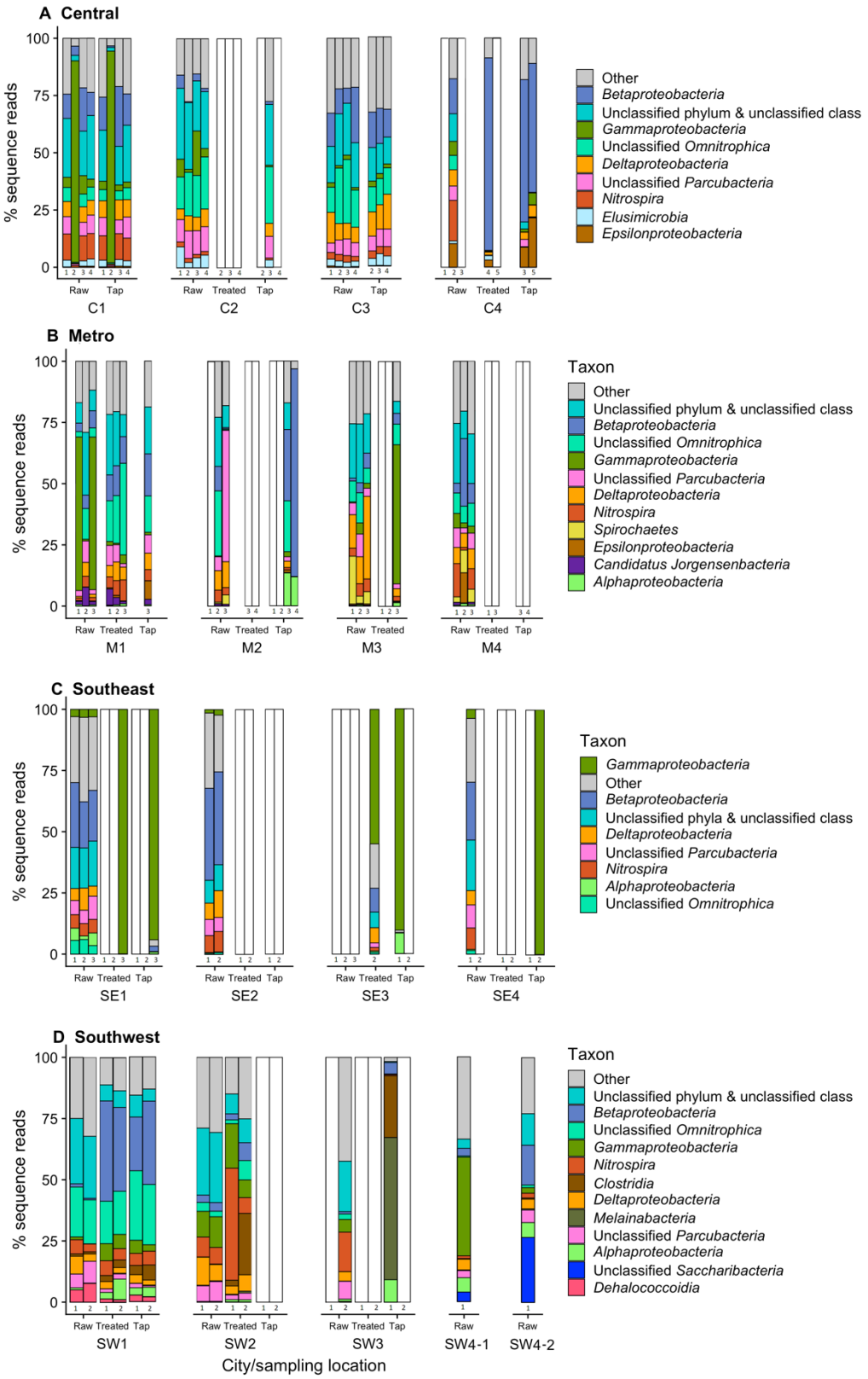

Figure 5 – Class-level bacterial community profiles of all samples by sampling location based on sequencing of 16S rRNA gene fragments. Empty bars are samples that failed quality control. Replicate samples are numbered under each bar (see Table AP12 – AP15) for dates each sample was collected).

Between sample diversity (beta diversity) of all samples by region was visualized in a principal coordinates analysis (PCoA) using the Bray-Curtis dissimilarity metric and revealed dissimilarities between most samples and tight clustering of three systems (C1, C3, and M1) (

Figure 6A) (within-sample diversity (alpha diversity) as the Shannon and Simpson indices is displayed in Figure AP2) . Weighted UniFrac and unweighted UniFrac visualizations are provided in Figures AP3 and AP4, respectively. Plotting samples by system and sample type (raw vs. treated vs. tap) within each region revealed a split in what may primarily drive community composition within different systems (

Figure 6B). Roughly half of the systems clustered distinctly and remained clustered regardless of disinfection or sampling location (systems C2, M1, SW1). Conversely, many systems within the metro, southeast, and southwest regions share similar structure in the raw water, but diverge after treatment (systems SE1, SE2, SE4 in the southeast region, or systems M3, M4 in the metro region).

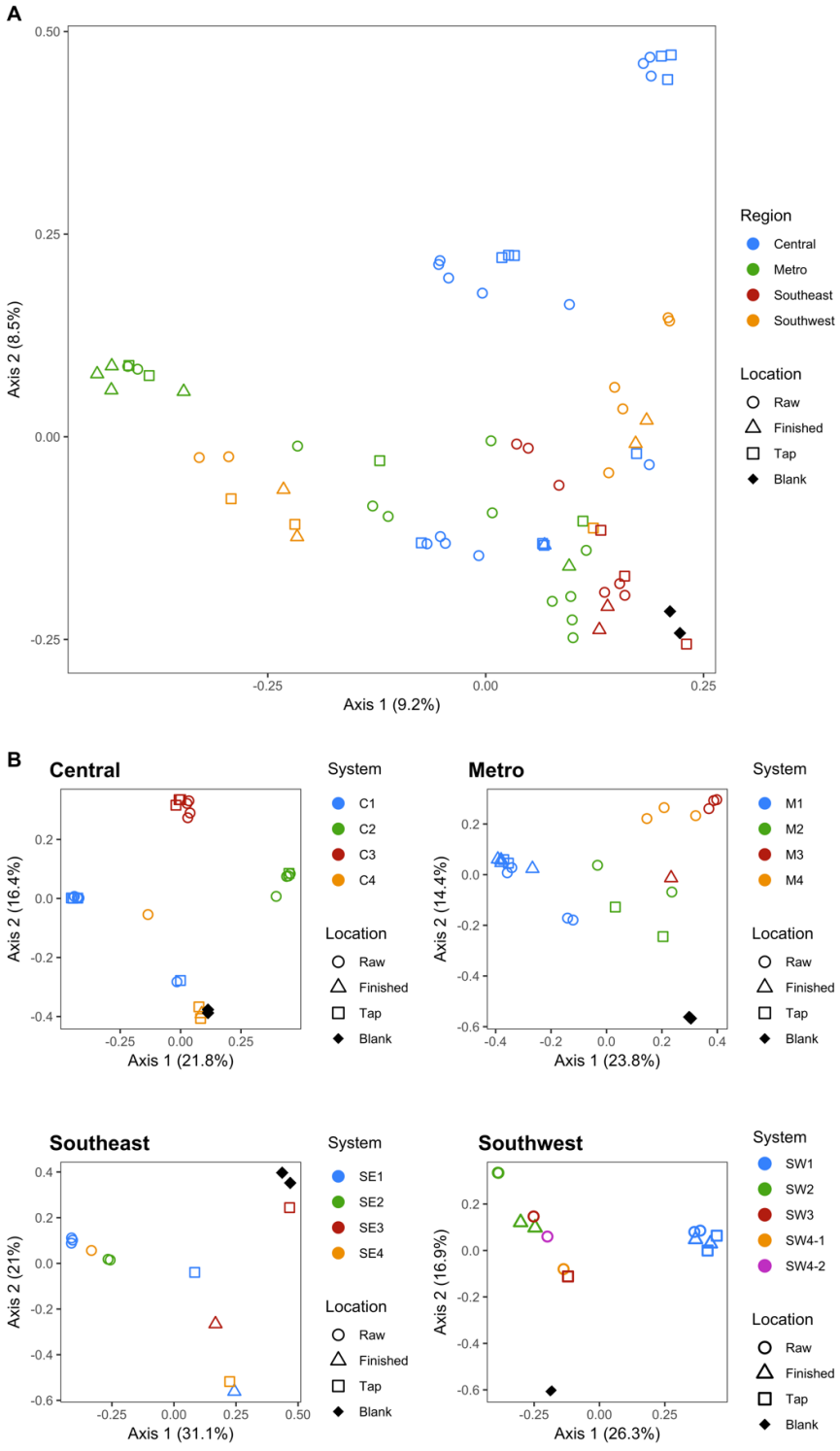

Figure 6 - Principal coordinate analysis of the Bray-Curtis dissimilarity metric for (**A**) All samples and (**B**) each sample separated by region. Color denotes system and shape denotes the sampling location type. Two method blank samples that cleared quality control screening are plotted as closed diamonds in all plots.

PERMANOVA tests revealed that well depth (R^2_{adonis} = 0.15, p_{adonis} = 0.001, $p_{betadisper}$ = 0.001), region (R^2_{adonis} = 0.15, p_{adonis} = 0.001, $p_{betadisper}$ = 0.87), and system within each region had significant effects on raw water community composition and disinfection type had a significant effect on community composition in all sample location types (p_{adonis} = 0.001, R^2_{adonis} = 0.09, $p_{betadisper}$ = 0.001) (Table AP16). Beta dispersion between groups in the well depth and disinfection type tests were significant, however, which may have confounded the PERMANOVA results[51] ($p_{betadisper}$ = 0.001 for both tests). Pairwise PERMANOVA results on the effect of well depth on raw water community composition revealed significant differences between all unique well depth pairs (shallow vs medium, shallow vs deep, medium vs deep), although dispersion was significant between the medium vs deep sample group ($p_{Wilcoxon}$ = 0.02, adj. $p_{Wilcoxon}$ = 0.05) (Table AP17). Significant effects were observed between 3/6 and 2/6 unique system-pairs in the central and metro regions, respectively. These system pairs generally included systems that displayed distinct separation in the Bray Curtis PCoA plot (C1, C2, C3, and M1). Pairwise comparisons on the effect of disinfection type revealed significant difference between all unique pairs, but beta dispersion was significant between the no disinfectant vs chloramine and free chlorine vs chloramine comparisons ($p_{Wilcoxon}$ = 6.0×10^{-4} and 7.0×10^{-3} and adj. $p_{Wilcoxon}$ = 6.0×10^{-4} and 6.0×10^{-3} for the no disinfectant vs chloramine and free chlorine vs chloramine pairs, respectively).

Well depth, tritium content, and carbamazepine concentration altogether significantly explained 76.5% ($p = 3.0 \times 10^{-3}$) of the variance observed in raw water sample ASVs via a constrained correspondence analysis (CCA) (Figure 7). Individual

variances of ASVs observed in raw water samples by these three environmental parameters were computed to determine which contributed most to the observed variance. Variance explained by well depth, tritium content, and carbamazepine concentration were 75.4%, 49.4%, and 24.6%, respectively. Well depth and tritium content significantly explained the observed variance ($p = 2.0 \times 10^{-3}$ for well depth, $p = 2.0 \times 10^{-3}$ for tritium content), but carbamazepine concentration did not ($p = 0.274$). Well depth was the dominant influence on ASVs observed in the metro and southeast region where well depths are greatest, and tritium content was the dominant influence on ASVs observed in the central and southwest regions where tritium content was largest.

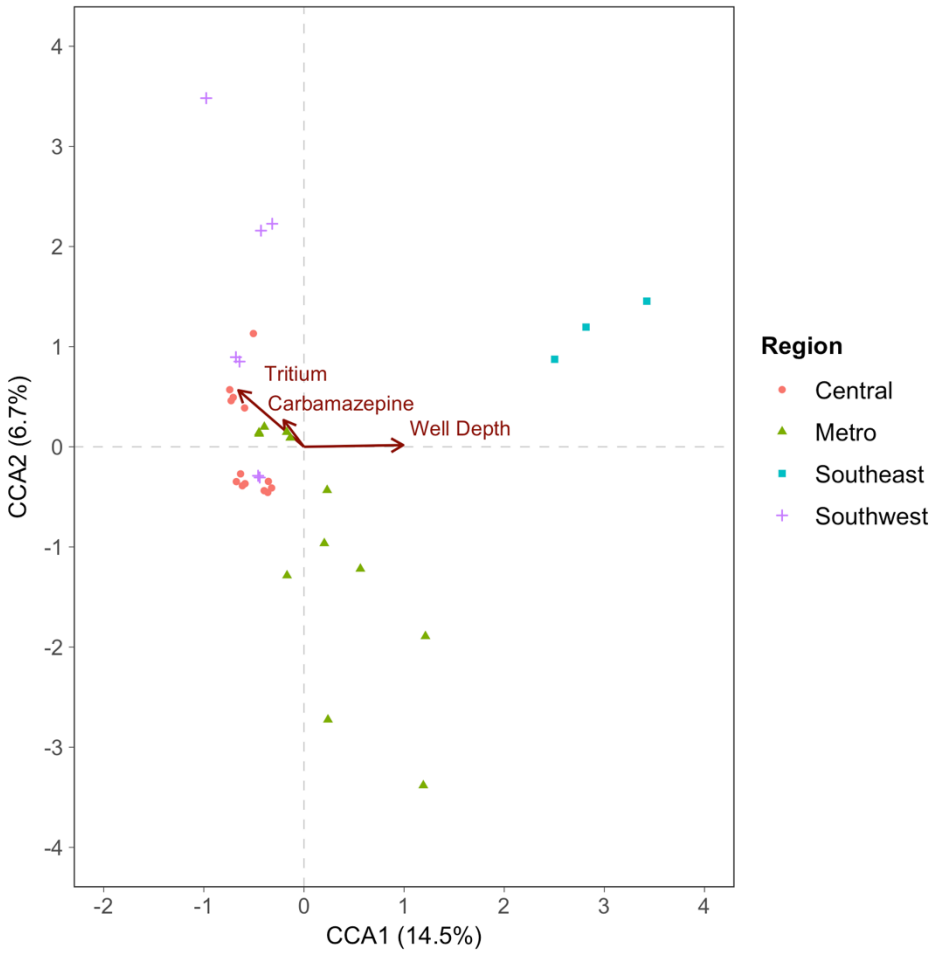

Figure 7 - Constrained correspondents analysis (CCA) on the effect of well depth, tritium content, and carbamazepine concentration on community composition in raw water samples.

4 Discussion

The principal finding from the current study is that groundwater is generally free of enteric pathogens. All samples were negative for the targeted genes from *C. jejuni*, *Shigella* spp., and Adenovirus. *E. coli*-specific genes were only detected in a well previously known to be contaminated. Furthermore, sequencing of 16S rRNA gene fragments confirmed the results from the total coliform and qPCR assays and corroborated the findings found therein. Large sample volumes (~10^3 L) allowed for direct high-resolution quantification of the selected gene targets to a detection limit of ~10 gene copies L^{-1} (except the 16S rRNA gene). The combination of high-resolution qPCR data and 16S rRNA gene sequences strongly suggests that enteric pathogens are generally absent from the public water systems that participated in this study.

Despite the absence of enteric pathogens in our samples, our results suggest that opportunistic pathogens belonging to the genera *Legionella* and *Mycobacteria* can be present in groundwater. We directly quantified concentrations of the *ssrA* and *atpE* genes (taxonomic targets are *Legionella* spp. and *Mycobacteria* spp., respectively) in 6/16 systems from source to tap via qPCR. The *ssrA* gene concentrations in raw groundwater are consistent with *Legionella*-specific gene concentrations previously reported in multiple groundwater systems.[52] An inspection of the 16S rRNA gene sequences from the same samples revealed *Legionella*-like and *Mycobacteria*-like ASVs, consistent with the qPCR results.

Although all samples were non-detect for the *mip* and ITS genes (taxonomic targets are *Legionella pneumophila* and *Mycobacterium avium* complex (MAC),

respectively), the presence of organisms belonging to the genera *Legionella* and *Mycobacteria* could still pose a public health risk for those with compromised immune systems. While *L. pneumophila* is responsible for 90-95% of reported legionellosis infections, the World Health Organization's report on the prevention of legionellosis maintains that all *Legionella* species are infectious.[53] The majority of bacteria that comprise the genus *Mycobacteria* are benign, but several species are pathogenic and can result in infections in people with compromised immune systems, such as AIDS patients.[54,55] The *Mycobacterium avium* complex (MAC), primarily composed of the species *M. avium* and *M. intracellulare,* is of particular concern in drinking water distribution systems and has been the subject of several investigations.[56–59]

In the disinfecting systems under investigation, *atpE* gene concentrations were enriched but *ssrA* gene concentrations were reduced from source to tap in all systems where these genes were detected. In the three *atpE*-positive sampling systems that maintain residual disinfectants (M3, SW1, and SW3), the *atpE* gene concentration was below the method detection limit (MDL) in the source waters and increased to about 3.0×10^2 gene copies L^{-1} at the tap. These results are consistent with previous knowledge that *Mycobacteria* are disinfectant-resistant organisms.[60,61] Conversely, the *ssrA*-positive sampling system that disinfects with free-chlorine (SW3) contained 2.0×10^2 gene copies L^{-1} in the source water and after disinfection the concentration was below the MDL (< 10 gene copies L^{-1}). There was no appreciable reduction in the *ssrA* gene concentration from source to tap in the two systems that do not disinfect (C1 and C3).

These results suggest that *Legionella* bacteria can be present in groundwater and may persist to the tap if not inactivated via disinfection.

Quantification of total biomass using the 16S rRNA gene marker revealed that disinfection temporarily reduces total biomass, but regrowth occurs in the DWDS. In 12/13 systems that disinfect before distribution, there was a substantial decrease in 16S rRNA genes between the source and treated water locations, some as high as 3.5-log removal of the gene (systems C2 and SW3). In most systems, microbial inactivation at the treated and tap sampling locations was so drastic that samples lacked sufficient template for PCR amplification and subsequent sequencing. Systems that exhibited less bacterial regrowth contained free chlorine or chloramine concentrations that were similar to the treated sampling locations (three systems in the southwest region). This research demonstrates the effectiveness of maintaining a stable residual disinfectant in the DWDS for suppressing microbial growth.

Temporal variation in total bacteria concentrations and community composition were consistent throughout the 1.5 – year sampling period. In the metro region, samples were collected over the course of six unique seasons (summer ’16, fall ’16, winter ‘16/’17, spring ’17, summer ’17, and fall ’17) and exhibited no discernable trend in total bacteria concentrations over this time period (Figure 2). The central, southeast, and southwest regions displayed similar trends, although sampling events were fewer and spanned three unique seasons at most (three seasons for the central region, and two seasons for the southeast and southwest regions). Detection frequency of genes corresponding to pathogenic organisms was too sparse to make conclusions about

temporal trends. Community composition of raw water samples across all systems contained dominant taxa that were observed repeatedly in roughly the same proportions of total reads among replicate samples (Figure 5). Raw water samples within-system tended to cluster consistently in the principal coordinates analysis (

Figure 6). Treated and tap water samples that passed quality control were insufficient to make observations regarding the seasonal effects on community composition. The temporal stability of total bacteria concentrations and raw water community composition suggests that one sampling event for each system is sufficient for understanding the dominant taxa and the evenness of the composition. Most of the sampling events in this study were conducted in the summer months (particularly in the central, southeast, and southwest regions), however, which may not have captured all seasonal variability in total bacteria concentrations or community composition.

This research suggests that total bacteria (as 16S rRNA genes) concentrations are lower in deeper wells that draw older water (tritium content < 0.8 TU), and greater in shallower wells that draw younger water (tritium content > 0.8 TU). In general, tritium content and well depth were good predictors of total biomass in raw water samples. Raw water samples collected from the metro and southeast regions (regions with the deepest wells and lowest tritium content) contained 16S rRNA gene concentrations between 10^6 – 10^7 gene copies L^{-1}, while raw water samples collected from the central and southwest regions (regions with shallower wells and higher tritium content) contained gene concentrations above 10^7 gene copies L^{-1} with several samples exceeding 10^8 gene copies L^{-1} (systems C1 and SW3). Previous investigations that have quantified total bacteria as

16S rRNA genes in groundwater have been limited to bioremediation applications where the groundwater is in an altered or contaminated state and is not designated for consumption.[9] Miller et al. quantified total bacteria concentrations in groundwater contaminated with tricholorethene and dichloroethene and reported concentrations between $10^4 – 10^7$ gene copies L^{-1}.[62] Other bioremediation investigations have reported concentrations of 16S rRNA genes in bioaugmented groundwaters or 16S rRNA genes normalized to soil mass and therefore direct comparisons with results from the current study cannot be made.[64,65] The strong correlations between well depth and tritium content to 16S rRNA gene concentrations ($\rho = 0.77$ and $\rho = -0.81$, respectively) presented herein provides practical knowledge in the design and management of groundwater-sourced drinking water supplies that may desire low bacteria concentrations.

Taxonomic assignment to 16S rRNA gene sequences revealed high relative abundances of uncultured organisms from several candidate phyla as well as organisms not currently assigned to any phyla. The sum of relative abundances of sequences belonging to *Candidatus Omnitrophica*, *Candidatus Parcubacteria*, and sequences without any phyla-level classification were nearly 50% of total sequences in the majority of samples. Bacterial composition, however, tended to shift in favor of *Proteobacteria* after disinfection, consistent with previous investigations.[66,67] These results suggest that the bacterial composition of groundwater systems is unique. Organisms belonging to the candidate phyla *Parcubacteria* have previously been observed in extreme marine environments[68] as well as shallow groundwater systems.[69,70] They have also been shown to possess extremely small cell volumes and genome sizes (< 1Mb)[71,72] that contain

highly reduced metabolic genes.[73] Metagenomic studies of these organisms may provide novel insight into the metabolic functions of organisms native to extreme (low temperature and low nutrient) aquatic environments.

Beta diversity analyses using 16S rRNA gene sequences suggests that community composition within systems are primarily driven by source water and/or disinfection. PERMANOVA tests on raw water samples indicated significant differences in composition depending on region and well depth, and the majority of variation in community composition observed was significantly explained by well depth and tritium content (as indicated by the CCA). Raw water samples compared between systems within each region displayed significant differences when compared pairwise in PERMANOVA tests and confirm the clustering seen in the PCoA (systems C1, C2, C3 in the central region, and system M1 in the metro region) (

Figure 6B). Samples from systems C2, M1, and SW1 were tightly clustered and independent of sampling location (raw vs. treated vs. tap) (

Figure 6B), suggesting that within some systems, source water is a greater driver of composition than disinfection. Some systems (SE1, SE2, and SW3), however, exhibited shifts in structure towards *Proteobacteria* (Figure 5) after disinfection, suggesting that community composition in some systems may be more susceptible to disinfection (no distinct clustering based on disinfection type was observed in a PCoA). These results suggest that source water is a primary driver of community composition and is consistent with those seen in the literature,[66,74] as well as the effects of free chlorine and chloramine disinfectants.[75] Further, these data suggest that there is

inconsistency in what drives community structure within groundwater-sourced drinking water systems. It has been reported previously that disinfectant choice may influence the abundances of corrosion-inducing microbes,[75] and that the majority of aging DWDS problems contain complications with microbiologically-induced corrosion.[76] A clearer understanding of why certain community compositions appear more resilient to the effects of residual disinfectants may provide important insight into DWDS management.

Quantitative PCR detection of genes associated with bacterial pathogens agreed well with total coliform bacteria detection via a traditional culture-based assay. In general, environmental samples that were negative for total coliforms were also negative for most of the selected gene targets. The sample volume used for the traditional indicator bacteria assay (100 mL) was approximately 4 orders of magnitude smaller than the sample volumes processed for the qPCR methods performed in this study (~1000 L). Despite the major difference in sample volumes between these two methods, the culture-based assay was a good predictor of the co-occurrence of enteric pathogens. These results strengthen the efficacy of monitoring for fecal indicator organisms as a proxy for the presence of other pathogenic microorganisms. The additional importance afforded to conventional culture-based detection methods has positive implications for water utilities who routinely perform indicator bacteria tests. From a practical perspective, the ability of the method to predict the co-occurrence of enteric pathogens with high probability combined with the simplicity of the method is a major strength.

This research was limited by the number of water supplies from which samples were collected. Further confidence could have been afforded to our findings had more

water utilities been included in the study including systems with higher risk shallow wells like those sampled in a recent investigation in Door County, Wisconsin.[6] One high-risk well was sampled from the southwest region (SW4-1) but was only sampled on one occasion due to accessibility difficulties. Replicate samples from positive control wells located in each sampling region would have provided further insight into what other pathogenic microorganisms may be present in groundwater samples that are positive for indicator bacteria.

Another potential limitation of this study is the highly-specific probe-based qPCR assays, which may have potentially yielded false negative detections. In contrast to universal dye-based qPCR assays, probe-based assays require identical reverse complementary sequences in the template target to produce a fluorescent signal. Nucleotide polymorphisms that may be present in the template targets could prevent the probe from annealing to the target DNA, thereby producing no signal. The comparison of *ssrA* gene quantities to *Legionella*-like ASVs observed from 16S rRNA sequences is potential evidence of false negative detection. Frequencies of *Legionella*-like 16S rRNA gene sequences were much higher than frequencies of *ssrA* gene detection via qPCR. Conversely, the high detection frequency of *Legionella*-like 16S rRNA gene sequences may be due to the presence of multiple copies of 16S rRNA genes in the *Legionella* genome compared to the singular occurrence of the *ssrA* gene.[77]

In contrast to false-negative detections, this study was potentially limited by false-positive detections. False-positive detection of the 23S rRNA gene specific to *Enterococcus* spp. was suspected because most samples were negative for all other gene

targets but positive with high frequency for this gene target. Use of the Basic Local Alignment Search Tool (BLAST)[78] on the GenBank sequencing database revealed a region of high sequence similarity between the genera *Enterococcus* and *Streptococcus*. Results from sequencing of 16S rRNA gene fragments confirmed the presence of *Streptococcus*-like sequences in 6/72 samples, as well as *Enterococcus*-like sequences in the method blank samples (up to relative abundances of 0.04%). The presence of *Enterococcus*-like sequences observed in the method blank samples suggests that organisms belonging to the genus *Enterococcus* were introduced to our samples through reagents used in the ultrafiltration concentrating process.

The major conclusion of this study is that groundwater supplies are generally free of enteric pathogens but sometimes contain opportunistic pathogens from the genera *Mycobacteria* and *Legionella*. This conclusion was reached based on high volume sampling techniques, direct quantification of gene targets associated with bacterial pathogens via qPCR, sequencing of PCR-amplified 16S rRNA genes, and traditional culture-based methods to enumerate indicator bacteria. These results were observed among replicate samples collected across geographically broad sampling systems. There was general agreement between qPCR methods targeting genes specific to pathogenic microorganisms and traditional indicator bacteria assays (despite major differences in samples volumes used in the two methods) and confirms the practical value of routine monitoring for indicator bacteria. The ubiquitous reliance on groundwater worldwide as a drinking water source necessitates further research of the pathogenic microorganisms that may reside in these environments.

5 Summary and conclusions

This research was performed to better characterize bacteria in groundwater supplies from source to tap throughout the State of Minnesota, particularly bacterial pathogens. A high-volume dead-end ultrafiltration approach was used to capture and recover microbes, allowing greater resolution than is afforded with conventional 1 L sample volumes. The major conclusion of this study is that groundwater supplies in Minnesota are generally free of enteric pathogens. All samples were negative for the target genes corresponding to *Campylobacter jejuni, Shigella* spp., and Adenovirus. *E. coli* – specific genes were only detected in a well with a documented history of contamination. Gene targets for the genera *Legionella* and *Mycobacteria* were detected in four of the 16 groundwater systems. If not inactivated through disinfection, this research suggests that pathogenic bacteria from the genus *Legionella* may persist to the tap. Mycobacteria were enriched from source to tap in disinfecting systems, consistent with prior knowledge that organisms from this genus are disinfectant-resistant. Pathogenic bacteria from the genera *Legionella* and *Mycobacteria* may pose a health risk to those with compromised immune systems.

Total bacteria concentrations (as 16S rRNA gene copies L^{-1}) were measured from source to tap in all systems. In raw water samples, total bacteria concentrations ranged from ~10^5 – 10^8 gene copies L^{-1}, with the greatest concentrations occurring in systems with shallow wells and young water (central region and southwest regions), and the lowest concentrations occurring in systems with the deepest wells and old water (metro and southeast regions). Total bacteria concentrations were below detection in many

treated samples indicating thorough bacterial inactivation through disinfection. At the tap, most samples displayed an increase in total bacteria, suggesting significant regrowth. There was no significant change in total bacteria concentrations from source to tap in the two non-disinfecting systems.

Sequencing of 16S rRNA gene fragments revealed highly diverse and previously uncharacterized organisms in raw water samples. 16S rRNA gene sequences associated with organisms from candidate phyla comprised nearly 50% of total sequence reads in many raw water samples. Between sample diversity (beta diversity) analyses suggests a split in what drives community composition in groundwater systems. Source water appeared to drive community composition in some systems regardless of sample location (raw vs. treated vs. tap) or disinfection, while community composition in other systems displayed shifts towards *Proteobacteria* after subjected to chlorine disinfectants.

5.1 Recommendations

This research suggests that pathogen-free groundwater is generally found with wells that are deeper than 150 ft, contain tritium content at < 4.0 TU, and no detectable carbamazepine concentration. Disinfection should be considered for systems with shallow wells (< 150 ft) due to the high total bacteria concentrations they may contain. Furthermore, shallower groundwater systems are more susceptible to bacterial contamination from a variety of sources, including poorly maintained septic systems, aging sewer infrastructure, and infiltration from surface runoff. For systems with shallow

wells or those that choose not to disinfect, increased monitoring for indicator organisms may be necessary to ensure consistent high-quality groundwater.

References

1. World Water Assessment Programme (WWAP). The United Nations World Water Development Report 2015: Water for a Sustainable World, Facts and Figures. *UN Water Rep.* 138 (2015). doi:10.1016/S1366-7017(02)00004-1
2. Macler, B. A. *et al.* Current knowledge on groundwater microbial pathogens and their control. *Hydrogeol. J.* **8,** 29–40 (2000).
3. Smeets, P. W. M. H. *et al.* The Dutch secret: How to provide safe drinking water without chlorine in the Netherlands. *Drink. Water Eng. Sci.* **2,** 1–14 (2009).
4. Wallender, E. K. *et al.* Contributing Factors to Disease Outbreaks Associated with Untreated Groundwater. *Groundwater* **52,** 886–897 (2014).
5. Murphy, H. M. *et al.* Review: Epidemiological evidence of groundwater contribution to global enteric disease, 1948–2015. *Hydrogeol. J.* **25,** 981–1001 (2017).
6. Borchardt, M. A. *et al.* Norovirus Outbreak Caused by a New Septic System in a Dolomite Aquifer. *Ground Water* **49,** 85–97 (2011).
7. Borchardt, M. A. *et al.* Vulnerability of drinking water wells in La Crosse, Wisconsin, to enteric virus contamination from surface water contributions. *Appl. Environ. Microbiol.* **70,** 5937–5946 (2004).
8. Borchardt, M. A. *et al.* Avian influenza virus RNA in groundwater wells supplying poultry farms affected by the 2015 influenza outbreak. *Environ. Sci. Technol. Lett.* **4,** 268–272 (2017).
9. Lambertini, E. *et al.* Risk of viral acute gastrointestinal illness from nondisinfected drinking water distribution systems. *Environ. Sci. Technol.* **46,** 9299–9307 (2012).
10. Lambertini, E. *et al.* Virus contamination from operation and maintenance events in small drinking water distribution systems. *J. Water Health* **9,** 799–812 (2011).
11. Gotkowitz, M. B. *et al.* Effects of Climate and Sewer Condition on Virus Transport to Groundwater. *Environ. Sci. Technol.* **50,** 8497–8504 (2016).
12. Murphy, H. M. *et al.* Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada. *Epidemiol. Infect.* **144,** 1355–1370 (2016).
13. Benedict, K. M. *et al.* Surveillance for Waterborne Disease Outbreaks Associated with Drinking Water — United States, 2013–2014. *Morb. Mortal. Wkly. Rep.* **66,** 1216–1221 (2017).
14. U.S. Environmental Protection Agency. National Primary Drinking Water Regulations: Ground Water Rule; Final Rule. (2006). Available at: https://www.gpo.gov/fdsys/pkg/FR-2006-11-08/html/06-8763.htm.

15. European Union Drinking Water Directive. Council Directive 98/83. *EC* **3,** (1998).

16. Ii, B. & Betragsanpassung, N. Langtitel Änderung Präambel / Promulgationsklausel. **2006,** 1–21 (2017).

17. Borchardt, M. A. *et al.* Incidence of enteric viruses in groundwater from household wells in Wisconsin. *Appl. Environ. Microbiol.* **69,** 1172–1180 (2003).

18. Borchardt, M. A. *et al* Viruses in Non-Disinfected Drinking Water from Municipal Wells and Community Incdience of Acute Gastrointestinal Illness. *Env. Health Perspectives.* **1272,** 1–40 (2012).

19. Leal, J. L. The Sterilization Plant of the Jersey City Water Supply Company at Boonton, N.J. *Proc. Am. Water Work. Assocation* 100–109 (1909).

20. Richardson, S. D. *et al.* Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Rev. Mutat. Res.* **636,** 178–242 (2007).

21. Stalter, D. *et al.* Fingerprinting the reactive toxicity pathways of 50 drinking water disinfection by-products. *Water Res.* **91,** 19–30 (2016).

22. Krasner, S. W. *et al.* Formation, precursors, control, and occurrence of nitrosamines in drinking water: A review. *Water Res.* **47,** 4433–4450 (2013).

23. Zhang, Y. *et al.* Accelerated chloramine decay and microbial growth by nitrification in premise plumbing. *J. Am. Water Work. Assoc.* **101,** 51–62 (2009).

24. Doria, M. de F. *et al.* Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Sci. Total Environ.* **407,** 5455–5464 (2009).

25. Rosario-Ortiz, F. *et al.* How do you like your tap water? *Science.* **351,** 912–914 (2016).

26. Doerfliger, N. *et al.* Water vulnerability assessment in karst environments: A new method of defining protection areas using a multi-attribute approach and GIS tools (EPIK method). *Environ. Geol.* **39,** 165–176 (1999).

27. Smith, C. M. *& Hill, V. R.* Dead-end hollow-fiber ultrafiltration for recovery of diverse microbes from water. *Appl. Environ. Microbiol.* **75,** 5284–5289 (2009).

28. Ishii, S. *et al.* Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. *Appl. Environ. Microbiol.* **79,** 2891–2898 (2013).

29. Rocchetti, T. T. *et al.* Validation of a multiplex real-time PCR assay for detection of *Mycobacteria* spp., *Mycobacterium tuberculosis* complex, and *Mycobacterium avium* complex directly from clinical samples by use of the BD max open system. *J. Clin. Microbiol.* **54,** 1644–1647 (2016).

30. Behr, T. *et al.* A Nested Array of rRNA Targeted Probes for the Detection and Identification of *Enterococci* by Reverse Hybridization. *Syst. Appl. Microbiol.* **23,**

563–572 (2000).

31. Benitez, A. J. & Winchell, J. M. Clinical application of a multiplex real-time PCR assay for simultaneous detection of *Legionella* species, *Legionella pneumophila*, and *Legionella pneumophila* serogroup 1. *J. Clin. Microbiol.* **51,** 348–351 (2013).

32. Radomski, N. *et al. atpE* gene as a new useful specific molecular target to quantify *Mycobacteriumin* environmental samples. *BMC Microbiol.* **13,** 277 (2013).

33. Muyzer, G. *et al.* Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59,** 695–700 (1993).

34. Lambertini, E. *et al.* Concentration of enteroviruses, adenoviruses, and noroviruses from drinking water by use of glass wool filters. *Appl. Environ. Microbiol.* **74,** 2990–2996 (2008).

35. Bartram, A. K. *et al.* Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl. Environ. Microbiol.* **77,** 3846–3852 (2011).

36. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335 (2010).

37. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

38. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13,** 581–583 (2016).

39. Edgar, R. C. *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).

40. Rognes, T. *et al.* VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4,** e2584 (2016).

41. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41,** 590–596 (2013).

42. Yilmaz, P. *et al.* The SILVA and 'all-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42,** 643–648 (2014).

43. Werner, J. J. *et al.* Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J.* **6,** 94–103 (2012).

44. Vanderford, B. J. & Snyder, S. A. Analysis of pharmaceuticals in water by isotope dilution liquid chromatography/tandem mass spectrometry. *Environ. Sci. Technol.* **40,** 7312–7320 (2006).

45. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. (2014).

46. Huston, C. & Juarez-Colunga, E. Guidelines for computing summary statistics for

data-sets containing non-detects. *Bulkley Val. Res. Cent.* (2009).

47. Benjamini, Y, *et al.* Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*. Vol . 57, **57,** 289–300 (2009).

48. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 32–46 (2006).

49. Oksanen, J. *et al.*: R functions for vegetation ecologists. *Date access* **15,** 2014 (2005).

50. Arbizu, M. Pairwiseadonis: Pairwise multilevel comparison using adonis. *R Packag. Version 0.0* **1,** (2017).

51. Warton, D. I. *et al.* Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* **3,** 89–101 (2012).

52. Wullings, B. A. & Kooij, D. Van Der. Occurrence and Genetic Diversity of Uncultured Legionella spp. in Drinking Water Treated at Temperatures below 15 ° C. **72,** 157–166 (2006).

53. McDade, J. E. *Legionella* and the Prevention of Legionellosis. *Emerg. Infect. Dis.* **14,** 1006a–1006 (2008).

54. Cosma, C. L. *et al.* The Secret Lives of the Pathogenic Mycobacteria. *Annu. Rev. Microbiol.* **57,** 641–676 (2003).

55. Falkinham, J. O. Nontuberculous mycobacteria in the environment. *Clin. Chest Med.* **23,** 529–551 (2002).

56. Berry, D. *et al.* Mycobacterium avium infections of acanthamoeba strains: Host strain variability, grazing-acquired infections, and altered dynamics of inactivation with monochloramine. *Appl. Environ. Microbiol.* **76,** 6685–6688 (2010).

57. Falkinham, J. O. *et al.* Factors Influencing Numbers of *Mycobacterium avium*, *Mycobacterium* intracellulare, and other *Mycobacteria* in drinking water distribution systems *Appl. Environ. Microbiol.* **67,** 1225–1231 (2001).

58. Glover, N. *et al.* The isolation and identification of *Mycobacterium avium* complex (MAC) recovered from Los Angeles potable water, a possible source of infection in AIDS patients. *Int. J. Environ. Health Res.* **4,** 63–72 (1994).

59. Feazel, L. M. *et al.* Opportunistic pathogens enriched in showerhead biofilms. *Proc. Natl. Acad. Sci.* **106,** 16393–16399 (2009).

60. Briones, A. & Raskin, L. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *Curr. Opin. Biotechnol.* **14,** 270–276 (2003).

61. Gomez-Alvarez, V. *et al.* Metagenomic Analyses of Drinking Water Receiving Different Disinfection Treatments. *Appl. Environ. Microbiol.* (2012).

doi:10.1128/AEM.01018-12

62. Miller, T. R. *et al.* Bacterial community analysis of shallow groundwater undergoing sequential anaerobic and aerobic chloroethene biotransformation. *FEMS Microbiol. Ecol.* **60,** 299–311 (2007).

63. Winded, C. *et al.* Depth-resolved quantification of anaerobic toluene degraders and aquifer microbial community patterns in distinct redox zones of a tar oil contaminant plume. *Appl. Environ. Microbiol.* **74,** 792–801 (2008).

64. Lee, P. K. H. *et al.* Quantifying genes and transcripts to assess the in situ physiology of 'Dehalococcoides' spp. in a trichloroethene-contaminated groundwater site. *Appl. Environ. Microbiol.* **74,** 2728–2739 (2008).

65. Ritalahti, K. M. *et al.* Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains. *Appl. Environ. Microbiol.* **72,** 2765–2774 (2006).

66. Bautista-de los Santos, Q. M. *et al.* Emerging investigators series: microbial communities in full-scale drinking water distribution systems – a meta-analysis. *Environ. Sci. Water Res. Technol.* **2,** 631–644 (2016).

67. Tokajian, S. T. *et al.* Phylogenetic assessment of heterotrophic bacteria from a water distribution system using 16S rDNA sequencing. *Can. J. Microbiol.* **51,** 325–335 (2005).

68. León-Zayas, R. *et al.* The metabolic potential of the single cell genomes obtained from the Challenger Deep, Mariana Trench within the candidate superphylum Parcubacteria (OD1). *Environ. Microbiol.* **19,** 2769–2784 (2017).

69. Lin, X. *et al.* Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *ISME J.* **6,** 1665–1676 (2012).

70. Lin, X. *et al.* Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ. Microbiol.* **14,** 414–425 (2012).

71. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499,** 431–437 (2013).

72. Luef, B. *et al.* Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* **6,** 1–8 (2015).

73. Nelson, W. C. & Stegen, J. C. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* **6,** 1–14 (2015).

74. Eichler, S. *et al.* Composition and Dynamics of Bacterial Communities of a Drinking Water Supply System as Assessed by RNA- and DNA-Based 16S rRNA Gene Fingerprinting. *Appl. Environ. Microbiol.* **72,** 1858–1872 (2006).

75. Stanish, L. F. *et al.* Factors Influencing Bacterial Diversity and Community Composition in Municipal Drinking Waters in the Ohio River Basin, USA. *PLoS*

One **11,** e0157966 (2016).

76. U.S. Environmental Protection Agency. 2007 Drinking Water Infrastructure Needs Survey and Assessment. (2009).

77. Waak, M. *et al.* Occurrence of Legionella spp. in Water-Main Biofilms from Two Drinking Water Distribution Systems. (2018). doi:10.1021/acs.est.8b01170

78. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134 (2012).

Appendix

Table AP1 - Detailed qPCR assay information for each gene target

Taxonomic Target	Target gene name	Gene product	Primer[a] & Probe[b] (5'-3' sequence)	Protocol	Primer conc (nM)	Probe con (nM)
Campylobacter jenjuni	*cadF*	Fibronectin-binding protein	F: TGC TAT TAA AGG TAT TGA TGT RGG TGA R: GCA GCA TTT GAA AAA TCY TCA T UPL 039	95°C for 5 min, followed by 45 cycles of 15 s at 95°C and 1 min at 60°C	900	400
Campylobacter jenjuni	*ciaB*	Invasion antigen B	F: GCG TTT TGT GAA AAA GAT GAA GAT AG R: GGT GAT TTT ACT TTC ATC CAA GC UPL 137		900	400
Mycobacterium avium complex (MAC)	ITS	Invasion antigen B	F: TTG GGC CCT GAG ACA ACA CT R: GCA ACC ACT ATC CAA TAC TCA AAC AC P: CCG TGT GGA GTC CCT CCA TCT TGG		1800	400
General *E. coli*	*ftsZ*	Cell division protein	F: CTG GTG ACC AAT AAG CAG GTT R: CAT CCC ATG CTG CTG GTA G UPL 071		900	400
General *E. coli*	*uidA*	Beta-D-glucuronidase	F: CCC TTA CGC TGA AGA GAT GC R: TTC ATC AAT CAC CAC GAT GC UPL 113		900	400
Enterohemorrhagic *E. coli* (EHEC)	*eaeA*	Intimin	F: GGC GAA TAC TGG CGA GAC TA R: GGC GCT CAT CAT AGT CTT TCT T UPL 028		900	400
Enterohemorrhagic *E. coli* (EHEC)	*stx1*	Shiga toxin 1 subunit A	F: TGT AAT GAC TGC TGA AGA TGT TGA T R: TCC ATG ATA RTC AGG CAG GA UPL 060		900	400
Enterohemorrhagic *E. coli* (EHEC)	*stx2*	Shiga toxin 2 subunit A	F: TCT GGC GTT AAT GGA GTT YAG R: GTG ACA GTG ACA AAA CGC AGA UPL 126		900	400
Shigella spp. and enteroinvasive *E. coli*	*virA*	Secreted VirG-processing protein	F: GGC AAT CTC TTC ACA TCA CG R: TTC GGA CAT AAT TTG GGC ATA UPL 006		900	400

Enterococcus spp.	23S rRNA	Large subunit, ribosomal RNA	F: GAG AAA TTC CAA ACG AAC TTG R: CAG TGC TCT ACC TCC ATC ATT P: TGG TTC TCT CCG AAA TAG CTT TAG GGC TA		500	100
Legionella spp.	*ssrA*	Transfer-messenger RNA	F: GGG CGA CCT GGC TTC R: GGT CAT CGT TTG CAT TTA TAT TTA P: ACG TGG GTT GCA A		500	100
Legionella pneumophila	*mip*	Macrophage infectivity potentiator	F: TTG TCT TAT AGC ATT GGT GCC G R: CCA ATT GAG CGC CAC TCA TAG P: CGG AAG CAA TGG CTA AAG GCA TGC A		500	100
Legionella pneumophila serogroup 1	*wzm*	ABC transporter of LPS O-antigen	F: TGC CTC TGG CTT TGC AGT TA R: CAC ACA GGC ACA GCA GAA ACA P: TTT ATT ACT CCA CTC CAG CGA T		500	100
Mycobacteria spp.	*atpE*	ATP synthase C chain AtpE	F: CGG YGC CGG TAT CGG YGA R: CGA AGA CGA ACA RSG CCA T P: ACS GTG ATG AAG AAC GGB GTR AA	95°C for 45 s, followed by 40 cycles of 3 s at 95°C and 30 s at 60°C	500	50
All Bacteria	16S rRNA	Small subunit, ribosomal RNA	F: ACT CCT ACG GGA GGC AGC AG R: ATT ACC GCG GCT GCT GG -	95°C for 1 min, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C	500	100
Adenovirus	*hex*	Hexon protein for capsid coat	F: GGA CGC CTC GGA GTA CCT GA R: CGC TGI GAC CIG TCT GTG G P: CAC CGA TAC GTA CTT CAG CCT GGG T	95°C for 10 min, followed by 45 cycles of 15 s at 95°C and 1 min at 6°C	500	100

[a]Forward and reverse primer sequences are preceded by the letters ‘F’ and ‘R’, respectively.
[b]Probe sequences are preceded by the letter ‘P’. Items containing "UPL" followed by a number represent proprietary probe sequences from the Universal ProbeLibrary® (Roche Molecular Systems. Inc, Pleasanton, CA)

Solid phase extraction procedure for carbamazepine quantification

Prior to the SPE, tubing and stoppers were pre-washed with 5mL methanol and acetonitrile (Thermo Fisher Scientific, Waltham, MA) per tube, then rinsed with pH 2.0 ultrapure water. SPE cartridges were loaded into the SPE manifold and preconditioned with 5mL methyl tert-butyl ether (MTBE) and methanol, then rinsed with ultrapure water. Tubing and stoppers were inserted into the SPE cartridges drawing water samples at a flow rate of 10 mL min^{-1}. After drawing all water sample volumes through the system, SPE cartridges were rinsed with 5 mL Nanopure water, then dried for 150 seconds at a maximum vacuum pressure of 20 mmHg. SPE cartridges were eluted with 5 mL of methanol followed by 5 mL of a 1:9 (v/v) methanol/MTBE solution into 550^{o}F furnace-combusted glass centrifuge tubes. Afterwards, the eluate was concentrated with a gentle stream of nitrogen gas and Nanopure water was added to achieve a final volume of 400 μL. Concentrated samples were stored at 4^{o}C until analysis by liquid chromatography and mass spectrometry.

Measurement of carbamazepine and carbamazepine isotope concentrations in environmental samples began with separation using a NanoAcquity Ultra Performance Liquid Chromatography (UPLC) instrument (Waters Corporation, Milford, MA) with a 0.5 mm x 150 mm ZORBAX SB-C18 column containing 5 μm particle size (Agilent Technologies, Santa Clara, CA). Mobile-phase solutions consisted of a binary gradient of 5 mM ammonium acetate and 100% methanol operating at 20 μL min^{-1}. The mobile-phase gradient began with 10% methanol (90% 5 mM ammonium acetate) for 5 minutes, then linearly increased to 98% methanol (2% 5 mM ammonium acetate) over an

additional 5 minutes where it was held for 10 minutes (total of 20 minutes per injection). After separation, carbamazepine and carbamazepine isotope species entered a TSQ Quantum Discovery MAX tandem mass spectrometer (Thermo Fisher Scientific, Waltham, MA) based on precursor ion molecular weights summarized in Table AP2.

Table AP2 - Carbamazepine and carbamazepine isotope parameters for LC-MS/MS

Chemical species	Retention time (min)	Precursor ion (m/z)	Product Ion (m/z)	Collision Energy (eV)	LOD (ng/L)	LOQ (ng/L)
Carbamazepine	7.25	237.4	194	20	0.018	0.059
Carbamazepine-d_{10}	7.25	247.4	204	20	-	-

Table AP3 - Total coliforms and E. coli measurements for all regions. Each observation is reported, separated by commas.

Region/system	Location	Coliforms (MPN 100mL^{-1})	*E. coli* (MPN 100mL^{-1})	# Obs.
CENTRAL				
C1	Raw	0, 0, 0, 0	0, 0, 0, 0	4
	Tap	0, 12.2, 0	0, 0, 0	3
C2	Raw	0, 0, 0, 0	0, 0, 0, 0	4
	Treated	0, 0, 0	0, 0, 0	3
	Tap	0, 0, 0	0, 0, 0	3
C3	Raw	0, 0, 0, 0	0, 0, 0, 0	4
	Tap	0, 0, 0	0, 0, 0	3
C4	Raw	0, 0, 0	0, 0, 0	3
	Treated	0, 0	0, 0	2
	Tap	0, 0	0, 0	2
METRO				
M1	Raw	0, 0, 0	0, 0, 0	3
	Treated	0, 0, 0	0, 0, 0	3
	Tap	0	0	1
M2	Raw	0, 0, 0, 0	0, 0, 0, 0	4
	Treated	0, 0	0, 0	2
	Tap	0, 0, 0, 0	0, 0, 0, 0	4
M3	Raw	0, 0, 0	0, 0, 0	3
	Treated	0, 0, 0	0, 0, 0	3
M4	Raw	0, 0, 0	0, 0, 0	3
	Treated	0, 0	0, 0	2
	Tap	0, 5.2	0, 0	2
SOUTHEAST				
SE1	Raw	0, 1, 0	0, 0, 0	3
	Treated	0, 1, 0	0, 0, 0	3
	Tap	0, 0, 0	0, 0, 0	3
SE2	Raw	0, 1	0, 0	2
	Treated	0, 3.1	0, 0	2
	Tap	0, 1	0, 0	2
SE3	Raw	0, 0	0, 0	2
	Treated	0, 0	0, 0	2
	Tap	0, 0	0, 0	2
SE4	Raw	0, 0	0, 0	2
	Treated	0, 1	0, 0	2
	Tap	1, 0	0, 0	2
SOUTHWEST				
SW1	Raw	0, 0	0, 0	2
	Treated	0, 0	0, 0	2
	Tap	0, 0	0, 0	2
SW2	Raw	0, 0	0, 0	2
	Treated	0, 0	0, 0	2
	Tap	0, 0	0, 0	2
SW3	Raw	0, 0	0, 0	2
	Treated	0, 0	0, 0	2
	Tap	0, 0	0, 0	2
SW4-1	Raw	727	249.5	1
SW4-2	Raw	0	0	1

Table AP4 – Mean total, free, and combined chlorine concentrations at treated and tap water locations for each system

Region/system	Location	Total chlorine (mg Cl_2 L^{-1})		Free chlorine (mg Cl_2 L^{-1})		Combined chlorine (mg Cl_2 L^{-1})	
		Mean	Std dev[a]	Mean	Std dev	Mean	Std dev
CENTRAL							
C1	Treated	No disinfection		No disinfection		No disinfection	
	Tap	No disinfection		No disinfection		No disinfection	
C2	Treated	3.84	0.16	1.89	0.02	1.86	0.03
	Tap	2.49	1.05	0.17	0.1	2.32	1.07
C3	Treated	No disinfection		No disinfection		No disinfection	
	Tap	No disinfection		No disinfection		No disinfection	
C4	Treated	0.37	n.a.	0.31	n.a.	0.06	n.a.
	Tap	0.05	n.a.	0	n.a.	0.52	n.a.
METRO							
M1	Treated	1.28	0.23	0.67	0.84	0.62	0.62
	Tap	1.51	n.a.	0.43	n.a.	1.08	n.a.
M2	Treated	1.36	n.a.	0.76	n.a.	0.61	n.a.
	Tap	1.32	n.a.	1.67	n.a.	0.84	n.a.
M3	Treated	2.04	0.5	0.1	0.04	1.58	0.04
	Tap	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
M4	Treated	1.02	n.a.	0.8	n.a.	0.22	n.a.
	Tap	0.73	n.a.	0.58	n.a.	0.16	n.a.
SOUTHEAST							
SE1	Treated	1.03	0.2	0.75	0.2	0.28	0.11
	Tap	0.75	0.18	0.65	0.14	0.1	0.04
SE2	Treated	1.31	n.a.	0.26	n.a.	0.53	n.a.
	Tap	0.48	n.a.	0.42	n.a.	0.07	n.a.
SE3	Treated	0.45	n.a.	0.3	n.a.	0.15	n.a.
	Tap	0.74	n.a.	0.55	n.a.	0.19	n.a.
SE4	Treated	1.89	n.a.	1.15	n.a.	0.74	n.a.
	Tap	0.65	n.a.	0.57	n.a.	0.08	n.a.
SOUTHWEST							
SW1	Treated	2.23	n.a.	0.08	n.a.	2.15	n.a.
	Tap	2.2	n.a.	0.11	n.a.	2.09	n.a.
SW2	Treated	0.97	n.a.	0.81	n.a.	0.16	n.a.
	Tap	0.94	n.a.	0.7	n.a.	0.24	n.a.
SW3	Treated	1.35	n.a.	0.95	n.a.	0.41	n.a.
	Tap	1.31	n.a.	1.03	n.a.	0.28	n.a.

[a]Standard deviations labeled 'n.a.' were not computed because sample observations were less than three.

Table AP5 - Water quality parameters for the central region

City		C1		C2			C3		C4		
		Raw	Tap	Raw	Treated	Tap	Raw	Tap	Raw	Treated	Tap
Nitrate (mg-N L^{-1})	Mean	0.17	0.71	0	0	0.01	0	0.09	0.13	-	0
	Std dev[a]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1
Nitrite (mg-N L^{-1})	Mean	0	0	0	0	0.001	0	0	0.005	-	0.004
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1
Chloride (mg L^{-1})	Mean	16.4	19.2	18.8	29.56	27.4	31.7	19.3	6.4	-	8.2
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1
Sulfate (mg L^{-1})	Mean	50	53.4	1.13	1.17	9.62	39.9	18.8	31.4	-	28.2
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1
pH	Mean	7.26	7.04	7.86	7.58	7.73	7.32	7.39	7.75	-	7.6
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1
Temp (ºF)	Mean	62.6	65.7	56.5	58.3	65.1	59.5	67.1	63.3	-	65.3
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
	# Obs.	1	1	1	1	1	1	1	1	-	1

[a]Standard deviations labeled 'n.a.' were not computed because sample observations were less than three.

Table AP6 - Water quality parameters for the metro region

City		M1			M2			M3		M4		
Sampling location		Raw	Trated	Tap	Raw	Treated	Tap	Raw	Treated	Raw	Treated	Tap
Nitrate (mg-N L^{-1})	Mean	-	-	-	0	0	0	0	-	-	-	-
	Std dev[a]	-	-	-	n.a.	n.a.	n.a.	n.a.	-	-	-	-
	# Obs.	-	-	-	1	1	1	1	-	-	-	-
Nitrite (mg-N L^{-1})	Mean	-	-	-	0	0	0	0.003	-	-	-	-
	Std dev	-	-	-	n.a.	n.a.	n.a.	n.a.	-	-	-	-
	# Obs.	-	-	-	1	1	1	1	-	-	-	-
Chloride (mg L^{-1})	Mean	-	-	-	5.88	13.1	1.55	-	-	-	-	8.37
	Std dev	-	-	-	8.7	n.a.	n.a.	-	-	-	-	n.a.
	# Obs.	-	-	-	3	2	2	-	-	-	-	2
Sulfate (mg L^{-1})	Mean	-	-	-	7.58	10.2	4.64	-	-	-	-	20.5
	Std dev	-	-	-	4.9	n.a.	n.a.	-	-	-	-	n.a.
	# Obs.	-	-	-	3	2	2	-	-	-	-	2
pH	Mean	-	-	-	7.99	7.4	7.32	-	-	-	-	-
	Std dev	-	-	-	n.a.	n.a.	n.a.	-	-	-	-	-
	# Obs.	-	-	-	2	2	1	-	-	-	-	-
Temp (ºF)	Mean	-	-	-	55.3	54.3	62.4	-	-	-	-	70.5
	Std dev	-	-	-	n.a.	n.a.	n.a.	-	-	-	-	n.a.
	# Obs.	-	-	-	2	2	1	-	-	-	-	1

[a]Standard deviations labeled 'n.a.' were not computed because sample observations were less than three.

Table AP7 - Water quality parameters for the southeast region

City		SE1			SE2			SE3			SE4		
Sampling location		Raw	Treated	Tap	Raw	Treated	Tap	Raw	Treated	Tap	Raw	Treated	Tap
Nitrate (mg-N L^{-1})	Mean	0.015	0	0	0.13	0.13	0	0	0	0	0	0	0
	Std dev[a]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	1	1	1	1	1	1	1
Nitrite (mg-N L^{-1})	Mean	0.002	0	0	0.002	0.002	0.003	0	0.001	0	0.001	0	0
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	2	1	1	1	1	1	1
Chloride (mg L^{-1})	Mean	0.84	3.20	4.55	0.07	1.5	1.35	1.59	4.45	3.2	2.49	3.1	1.42
	Std dev	0.08	1.5	3.0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	3	3	3	1	1	1	2	1	2	1	1	1
Sulfate (mg L^{-1})	Mean	26.0	24.8	22.5	20.9	20.9	21.8	40.8	39.2	38.9	23.0	22.7	22.9
	Std dev	5.8	5.7	12	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	3	3	3	1	1	1	2	1	2	1	1	1
pH	Mean	7.4	7.53	7.49	7.46	7.36	7.47	7.36	7.27	7.42	7.31	7.33	7.22
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	1	1	1	1	1	1	1	1	1
Temp (°F)	Mean	61.3	66.7	66	63.9	60.8	67.6	63.3	68.4	72.8	65.8	64	70.8
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	1	1	1	1	1	2	2	2	2

[a]Standard deviations labeled 'n.a.' were not computed because sample observations were less than three.

Table AP8 - Water quality parameters for the southwest region

City		SW1			SW2			SW3			SW4-1	SW4-2
Sampling location		Raw	Treated	Tap	Raw	Treated	Tap	Raw	Treated	Tap	Raw	Raw
Nitrate (mg-N L^{-1})	Mean	0	0.15	0.15	1.23	1.44	0.85	1.21	1.26	0.98	4.95	6.6
	Std dev[a]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	2	2	2	2	1	1
Nitrite (mg-N L^{-1})	Mean	0.004	0.004	0.003	0.001	0	0	0.019	0.002	0.003	0	0.004
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	2	2	2	2	1	1
Chloride (mg L^{-1})	Mean	1.53	4.79	5.18	11.9	17.2	17.5	17.5	13.1	12.8	15.3	16.8
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	2	2	2	2	1	1
Sulfate (mg L^{-1})	Mean	471	398	402	33.5	53.5	51.1	124	91.8	90.3	21.9	24.5
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	2	2	2	2	2	2	2	2	2	1	1
pH	Mean	7.32	8.57	8.82	7.40	7.39	7.38	7.25	7.19	7.18	6.97	7.05
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	1	2	1	1	1	1	1	1	1	1	1
Temp (^{o}F)	Mean	55.6	53.7	60.3	55.8	56.3	67.1	55.6	61.6	65.6	60.8	64.4
	Std dev	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	# Obs.	1	2	1	1	1	2	1	2	2	1	1

[a]Standard deviations labeled 'n.a.' were not computed because sample observations were less than three.

Table AP9 - Sampling bins for Kruskal-Wallis test on temporal effect on 16S rRNA gene concentrations.

Group	Sampling period	# of samples[a]
Summer 2016	June 2016 - August 2016	23
Fall 2016	September 2016 - December 2016	7
Summer 2017	May 2017 - August 2017	62
Fall 2017	September 2017 - December 2017	17

[a]All samples collected during the corresponding sampling period were grouped together, regardless of sampling location type (raw vs. treated. vs. tap)

Table AP10 - Detection frequency table of qPCR gene targets corresponding to pathogenic microorganisms

System	Location type	Number of samples	No. of positive detections													
			cadF	*ciaB*	*atpE*	ITS	*ftsZ*	*uidA*	*eaeA*	*stx1*	*stx2*	*virA*	23S rRNA	*ssrA*	*mip*	*hex*
C1	Raw	4	-	-	-	-	-	-	-	-	-	-	-	2	-	-
	Tap	4	-	-	-	-	-	-	-	-	-	-	1	2	-	-
C2	Raw	4	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Treated	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Tap	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C3	Raw	4	-	-	-	-	-	-	-	-	-	-	4	4	-	-
	Tap	4	-	-	-	-	-	-	-	-	-	-	3	3	-	-
C4	Raw	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
M1	Raw	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Treated	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Tap	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M2	Raw	3	-	-	-	-	-	-	-	-	-	-	1	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
	Tap	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M3	Raw	3	-	-	-	-	-	-	-	-	-	-	3	-	-	-
	Treated	3	-	-	2	-	-	-	-	-	-	-	-	-	-	-
M4	Raw	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SE1	Raw	3	-	-	-	-	-	-	-	-	-	-	3	-	-	-
	Treated	3	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Tap	3	-	-	-	-	-	-	-	-	-	-	3	-	-	-
SE2	Raw	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
SE3	Raw	3	-	-	-	-	-	-	-	-	-	-	3	-	-	-
	Treated	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-

System	Location type	Number of samples	No. of positive detections													
			cadF	*ciaB*	*atpE*	ITS	*ftsZ*	*uidA*	*eaeA*	*stx1*	*stx2*	*virA*	23S rRNA	*ssrA*	*mip*	*hex*
SE4	Raw	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-
SW1	Raw	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Treated	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-
	Tap	2	-	-	2	-	-	-	-	-	-	-	1	-	-	-
SW2	Raw	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-
	Tap	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-
SW3	Raw	2	-	-	-	-	-	-	-	-	-	-	2	2	-	-
	Treated	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Tap	2	-	-	1	-	-	-	-	-	-	-	2	-	-	-
SW4-1	Raw	1	-	-	1	-	1	1	-	-	-	-	1	1	-	-
SW4-2	Raw	1	-	-	-	-	-	-	-	-	-	-	1	1	-	-

Table AP11 - Medians and ranges of gene target concentrations at each sampling location within each system.

System/ sampling loc.		Log$_{10}$(gene copies L^{-1})					
		23S rRNA		*atpE*		*ssrA*	
		Median	Range[a]	Median	Range	Median	Range
C1	Raw	< 1.0	n.a.	< 1.0	n.a.	1.7	< 1.0 - 2.4
	Tap	< 1.0	< 1.0 - 1.9	< 1.0	n.a.	1.8	< 1.0 - 2.7
C2	Raw	1.60	< 1.0 - 2.3	< 1.0	n.a.	< 1.0	n.a.
	Treated	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Tap	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
C3	Raw	2.2	2.0 - 2.5	< 1.0	n.a.	3.0	2.9 - 3.3
	Tap	2.2	2.2 - 2.6	< 1.0	n.a.	3.4	3.2 - 3.6
C4	Raw	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Treated	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.3	< 1.0 - 1.6	< 1.0	n.a.	< 1.0	n.a.
M1	Raw	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Treated	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Tap	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
M2	Raw	< 1.0	< 1.0 - 2.2	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.4	< 1.0 - 1.7	< 1.0	n.a.	< 1.0	n.a.
	Tap	< 1.0	< 1.0 - 1.9	< 1.0	n.a.	< 1.0	n.a.
M3	Raw	2.8	2.2 - 3.5	< 1.0	n.a.	< 1.0	n.a.
	Treated	< 1.0	n.a.	2.2	1.8 - 2.5	< 1.0	n.a.
M4	Raw	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.6	< 1.0 - 2.2	< 1.0	n.a.	< 1.0	n.a.
	Tap	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
SE1	Raw	1.7	1.5 - 1.9	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.7	< 1.0 - 1.8	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.6	1.5 - 1.6	< 1.0	n.a.	< 1.0	n.a.
SE2	Raw	1.5	< 1.0 - 2.0	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.5	1.4 - 1.7	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.4	< 1.0 - 1.8	< 1.0	n.a.	< 1.0	n.a.
SE3	Raw	1.5	< 1.0 - 1.5	< 1.0	n.a.	< 1.0	n.a.
	Treated	2	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.7	< 1.0 - 1.8	< 1.0	n.a.	< 1.0	n.a.
SE4	Raw	1.7	1.6 - 1.7	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.6	1.6 - 1.7	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.9	1.6 - 2.1	< 1.0	n.a.	< 1.0	n.a.
SW1	Raw	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Treated	< 1.0	n.a.	1.26	< 1.0 - 1.5	< 1.0	n.a.
	Tap	1.3	< 1.0 - 1.5	2.52	2.5 - 2.6	< 1.0	n.a.
SW2	Raw	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Treated	1.3	< 1.0 - 1.7	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.1	< 1.0 - 1.1	< 1.0	n.a.	< 1.0	n.a.
SW3	Raw	1.2	1.1 - 1.3	< 1.0	n.a.	2.3	2.0 - 2.6
	Treated	< 1.0	n.a.	< 1.0	n.a.	< 1.0	n.a.
	Tap	1.9	1.6 - 2.2	1.6	< 1.0 - 2.2	< 1.0	n.a.
SW4-1[b]	Raw	3.2	n.a.	1.8	n.a.	3.9	n.a.
SW4-2	Raw	1.9	n.a.	< 1.0	n.a.	1.7	n.a.

[a]Ranges not applicable (n.a.) either due to one sampling event or all sampling events were below quantification limit
[b]Non-potable well

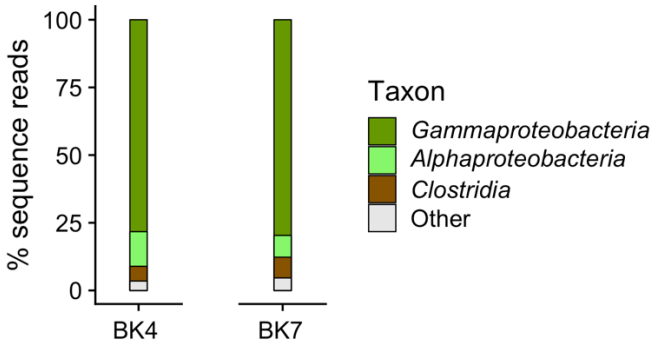

Figure AP1 - Taxonomic composition (class-level) of method blank samples that passed quality control screening

Table AP12 - Replicate sampling dates for the central region

System	Sampling location	Replicate #	Date collected
C1	Raw	1	7/25/16
	Raw	2	8/29/16
	Raw	3	5/24/17
	Raw	4	7/18/17
	Tap	1	7/25/16
	Tap	2	8/29/16
	Tap	3	5/24/17
	Tap	4	7/18/17
C2	Raw	1	7/25/16
	Raw	2	8/29/16
	Raw	3	5/24/17
	Raw	4	7/19/17
	Treated	2	8/29/16
	Treated	3	5/24/17
	Treated	4	7/19/17
	Tap	2	8/29/16
	Tap	3	5/24/17
	Tap	4	7/19/17
C3	Raw	1	7/25/16
	Raw	2	8/29/16
	Raw	3	5/24/17
	Raw	4	7/19/17
	Tap	2	8/29/16
	Tap	3	5/24/17
	Tap	4	7/19/17
C4	Raw	1	7/25/16
	Raw	2	7/25/16
	Raw	3	8/29/16
	Treated	4	5/24/17
	Treated	5	7/19/17
	Tap	3	8/29/16
	Tap	5	7/19/17

Table AP13 - Replicate sampling dates for the metro region

System	Sampling location	Replicate #	Date collected
M1	Raw	1	7/6/16
	Raw	2	12/14/16
	Raw	3	6/6/17
	Treated	1	7/6/16
	Treated	2	12/14/16
	Treated	3	6/6/17
	Tap	3	6/6/17
M2	Raw	1	6/13/16
	Raw	2	8/31/17
	Raw	3	10/18/17
	Treated	3	8/31/17
	Treated	4	10/18/17
	Tap	1	6/13/16
	Tap	2	9/24/16
	Tap	3	5/18/17
	Tap	4	10/18/17
M3	Raw	1	6/22/16
	Raw	2	10/12/16
	Raw	3	6/22/17
	Treated	1	6/22/16
	Treated	2	10/12/16
	Treated	3	6/22/17
M4	Raw	1	6/7/16
	Raw	2	11/2/16
	Raw	3	6/1/17
	Treated	1	6/7/16
	Treated	3	6/1/17
	Tap	3	6/1/17
	Tap	4	12/7/16

Table AP14 - Replicate sampling dates for the southeast region

System	Sampling location	Replicate #	Date collected
SE1	Raw	1	6/13/17
	Raw	2	8/16/17
	Raw	3	10/23/17
	Treated	1	6/13/17
	Treated	2	8/16/17
	Treated	3	10/23/17
	Tap	1	6/13/17
	Tap	2	8/16/17
	Tap	3	10/23/17
SE2	Raw	1	6/27/17
	Raw	2	8/2/17
	Treated	1	6/27/17
	Treated	2	8/2/17
	Tap	1	6/27/17
	Tap	2	8/2/17
SE3	Raw	1	6/13/17
	Raw	2	6/13/17
	Raw	3	8/16/17
	Treated	2	8/16/17
	Tap	1	6/13/17
	Tap	2	8/16/17
SE4	Raw	1	6/14/17
	Raw	2	8/2/17
	Raw	3	6/14/17
	Treated	2	8/2/17
	Tap	1	6/14/17
	Tap	2	8/2/17

Table AP15 - Replicate sampling dates for the southwest region

System	Sampling location	Replicate #	Date collected
SW1	Raw	1	7/5/17
	Raw	2	10/2/17
	Treated	1	7/5/17
	Treated	2	10/2/17
	Tap	1	7/5/17
	Tap	2	10/2/17
SW2	Raw	1	7/5/17
	Raw	2	10/2/17
	Treated	1	7/5/17
	Treated	2	10/2/17
	Tap	1	7/6/17
	Tap	2	10/3/17
SW3	Raw	1	7/6/17
	Raw	2	10/3/17
	Treated	1	7/6/17
	Treated	2	10/3/17
	Tap	1	7/6/17
	Tap	2	10/3/17
SW4-1	Raw	1	10/3/17
SW4-2	Raw	1	10/3/17

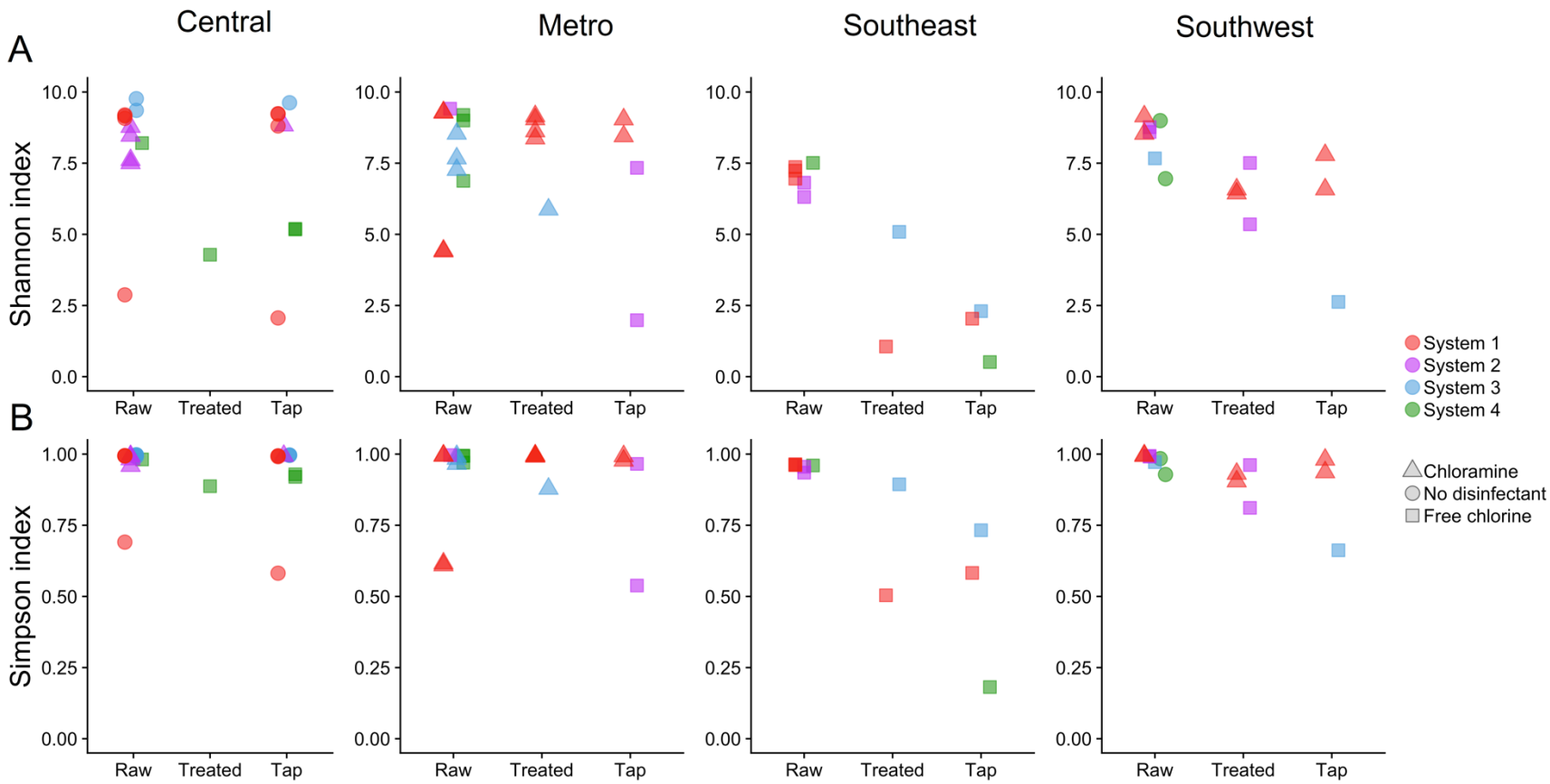

Figure AP2 - Within-sample diversity (alpha diversity) by the (**A**) Shannon index and (**B**) Simpson index of samples that passed quality control screening for sequencing analysis in the central, metro, southeast, and southwest regions.

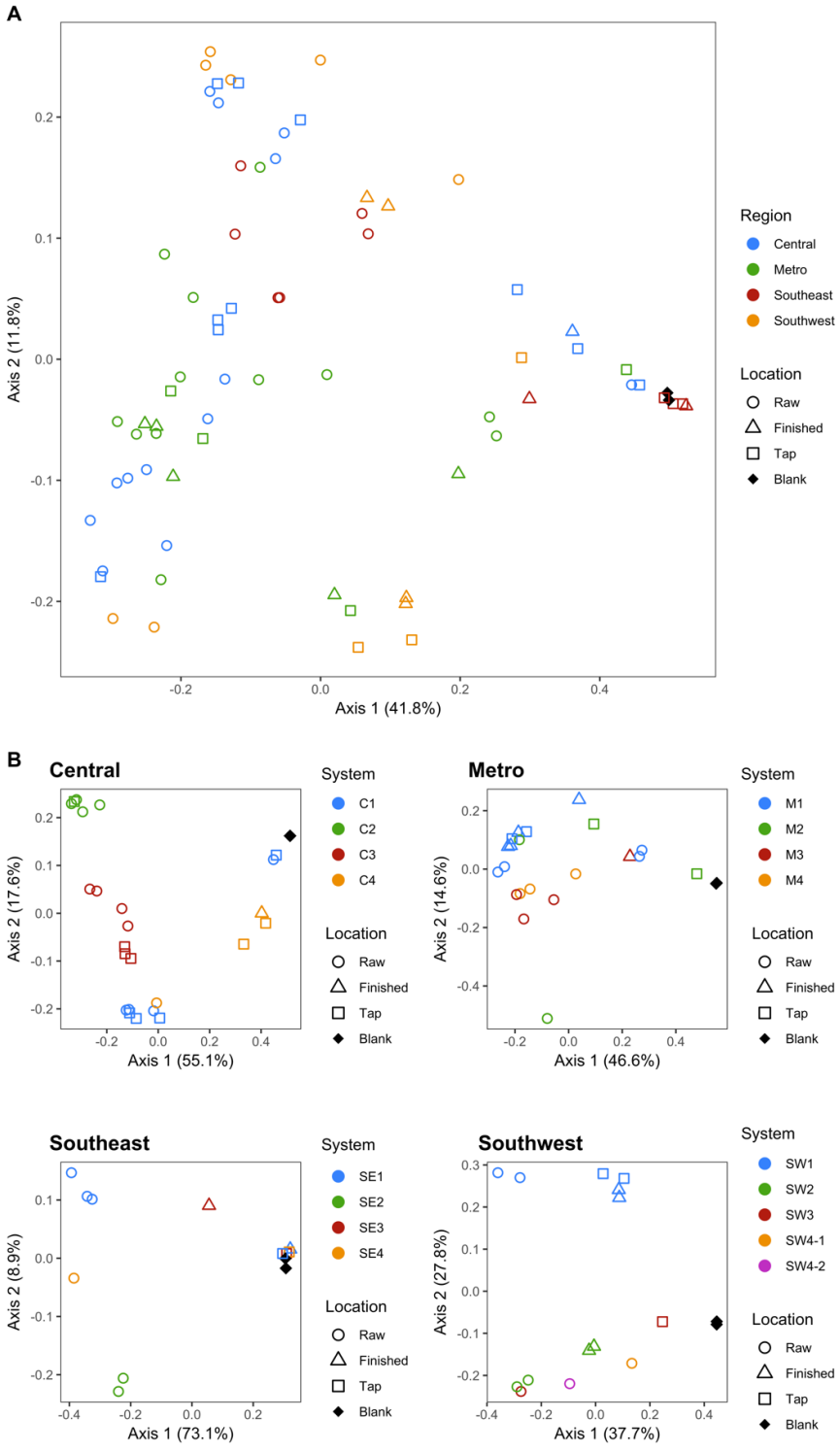

Figure AP3 - Principal coordinate plots of the weighted unifrac dissimilarity metric for (**A**) All samples and (**B**) each sample separated by region. Color denotes system and shape denotes the sampling location type. Two method blank samples that cleared quality control screening are plotted as closed diamonds in all plots.

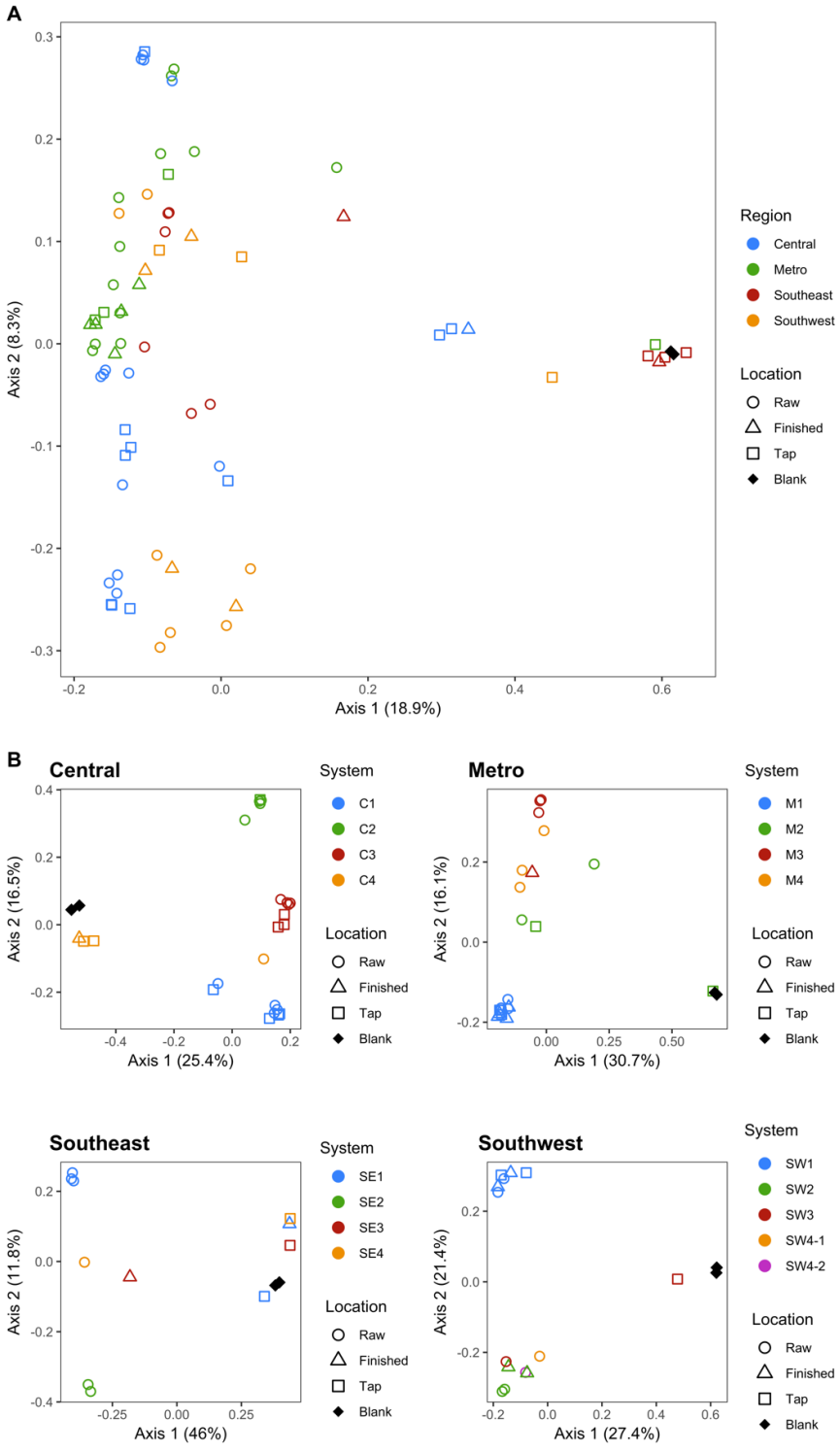

Figure AP4 - Principal coordinate analysis plots of the unweighted unifrac dissimilarity metric for (**A**) All samples and (**B**) each sample separated by region. Color denotes system and shape denotes the sampling location type. Two method blank samples that cleared quality control screening are plotted as closed diamonds in all plots.

Table AP16 - Overall PERMANOVA results and differences in beta dispersion results. Significant p-values are in bold

	R^2_{adonis}	p_{adonis}	$p_{betadisper}$
Well depth* (raw water only)	0.15	**0.001**	**0.001**
Region (raw water only)	0.15	**0.001**	0.87
Disinfection type	0.09	**0.001**	**0.001**
System within each region (raw water only)			
Central	0.64	**0.001**	0.5
Metro	0.56	**0.001**	0.69
Southeast	0.88	**0.017**	**0.0014**
Southwest	0.95	**0.011**	**0.001**

*Shallow = 0 - 200 ft, medium = 200 - 400 ft, deep = > 400 ft

Table AP17 - Pairwise PERMANOVA and beta dispersion comparisons for all tested variables. Adjusted p-values are adjusted with the false discovery rate (FDR) adjustment. Significant p-values are in bold

	Pair-wise PERMANOVA			Pair-wise beta dispersion	
	R^2_{adonis}	p_{adonis}	adj. p_{adonis}	$p_{Wilcoxon}$	adj. $p_{Wilcoxon}$
Effect of well depth					
Medium vs. shallow	0.11	**0.001**	**0.001**	0.6	0.6
Medium vs. deep	0.20	**0.001**	**0.001**	**0.017**	0.05
Shallow vs. deep	0.28	**0.001**	**0.001**	0.07	0.1
Effect of region					
Metro vs southwest	0.15	**0.02**	**0.02**	0.59	0.71
Metro vs southeast	0.21	**0.003**	**0.005**	0.21	0.43
Metro vs central	0.16	**0.001**	**0.002**	0.77	0.77
Southwest vs southeast	0.35	**0.001**	**0.002**	0.18	0.43
Southwest vs central	0.13	**0.013**	**0.016**	0.54	0.71
Effect of system (central)					
C3 vs C1	0.65	**0.03**	0.06	0.69	0.82
C3 vs C2	0.85	**0.03**	0.06	0.49	0.73
C3 vs C4	0.64	0.20	0.24	0.40	0.73
C1 vs C2	0.75	**0.03**	0.06	0.89	0.89
C1 vs C4	0.26	0.40	0.40	0.40	0.73
C2 vs C4	0.80	0.20	0.24	0.40	0.73
Effect of system (metro)					
M2 vs M1	0.36	0.20	0.20	1.00	1.00
M2 vs M4	0.50	0.10	0.12	0.80	1.00
M2 vs M3	0.65	0.10	0.12	0.20	1.00
M1 vs M4	0.63	**0.02**	0.11	1.00	1.00
M1 vs M3	0.73	**0.04**	0.11	1.00	1.00
M4 vs M3	0.57	0.10	0.12	1.00	1.00
Effect of system (southeast)					
SE2 vs SE1	0.94	0.10	0.30	0.20	0.60
SE2 vs SE4	0.98	0.33	0.33	0.67	0.67
SE1 vs SE4	0.79	0.25	0.33	0.50	0.67
Effect of system (southwest)					
SW4-1 vs SW4-2	1.00	-	-	-	-
SW4-1 vs SW1	0.93	0.33	0.33	0.48	0.67

SW4-1 vs SW3	1.00	-	-	-	-
SW4-1 vs SW2	0.96	0.33	0.33	0.67	0.67
SW4-2 vs SW1	0.94	0.33	0.33	0.48	0.67
SW4-2 vs SW3	1.00	-	-	-	-
SW4-2 vs SW2	0.95	0.33	0.33	0.67	0.67
SW1 vs SW3	0.94	0.33	0.33	0.48	0.67
SW1 vs SW2	0.96	0.33	0.33	0.22	0.67
SW3 vs SW2	0.94	0.33	0.33	0.67	0.67
Effect of disinfectant type					
No disinfectant vs free chlorine	0.12	**0.001**	**0.001**	0.24	0.24
No disinfectant vs chloramine	0.15	**0.001**	**0.001**	**0.0006**	**0.0006**
Free chlorine vs chloramine	0.46	**0.001**	**0.001**	**0.007**	**0.006**