Electronic Supplementary Information

Cast iron drinking water pipe biofilms support diverse microbial communities containing antibiotic resistance genes, metal resistance genes, and class 1 integrons

Lee K. Kimbell[1], Emily Lou LaMartina[2], Anthony D. Kappell[1], Jingwan Huo[3], Yin Wang[3], Ryan J. Newton[2], Patrick J. McNamara[1*]

1. Marquette University, Department of Civil, Construction & Environmental Engineering
2. University of Wisconsin-Milwaukee, School of Freshwater Sciences
3. University of Wisconsin-Milwaukee, Department of Civil and Environmental Engineering
*corresponding author; e-mail: Patrick.McNamara@marquette.edu
1637 W Wisconsin Ave
Milwaukee, WI 53233, USA

Content

Materials and Methods: Description of qPCR Methodology and 16S rRNA amplicon sequencing S2

Supplementary Table 1: Description of qPCR Methods and Results S3

Supplementary Table 2: MIQE Guidelines for qPCR and ddPCR experiments S5

Supplementary Figure 1: Diagram of biofilm samples collected from cast iron drinking water main S9

Supplementary Figure 2: Absolute abundance of 16S rRNA genes in different biofilm microenvironments from a cast-iron drinking water main as measured with ddPCR and qPCR S10

Supplementary Figure 3: Comparison of absolute gene abundance of antibiotic resistance genes, metal resistance genes, and *intI*1 as measured with ddPCR and qPCR S11

Supplementary Figure 4: Relative abundance of antibiotic resistance genes, metal resistance genes, and *intI*1 in different biofilm microenvironments from a cast-iron drinking water main as measured with ddPCR S12

Supplementary Figure 5: Shannon alpha diversity measurements of biofilm microbial communities characterized by 16S rRNA gene sequencing S13

Supplementary Figure 6: Spearman's rank sum correlation between relative abundance of most abundant bacterial genera observed in biofilm samples from the cast iron water main S14

Pages: 15, Figures: 6, Tables: 2

Materials and Methods

Quantitative PCR (qPCR) Assays

A subset of biofilm samples was initially analyzed at 5, 10, 50, and 100-fold dilutions to determine the optimum dilution for gene quantification. A 10-fold dilution was found to yield optimum quantitation for DNA extracts and was utilized for all samples. Each qPCR assay consisted of a total reaction volume (20 µL) with 10 µL PowerUp™ SYBR® Green Master Mix, 2 µL each of forward and reverse primers (10 µM), 5 µL of diluted DNA extract, and 1 µL molecular-grade water.

Following each qPCR assay, melt curves were generated and analyzed to verify specific target amplification based on positive controls. Gene concentrations for each sample were quantified in triplicate, and the mean value was used for subsequent statistical analysis. If only two of the three replicates yielded positive detections in the qPCR assay, then the mean value of the two positive replicates was used in subsequent analyses (Kimbell et al., 2018; Pruden et al., 2012). Standard curves for qPCR assays were produced by ten-fold serial dilution of plasmid DNA yielding 10^6 to 10^0 copies per reaction. Standard curves and negative controls were conducted in triplicate and were included with each 96-well plate. All qPCR negative controls failed to yield amplification above the limit of quantification for each assay. Copy numbers of target genes were $\log_{10}$ transformed to meet the assumptions of normality for statistical analysis (Burch et al., 2013). Relative abundance of each target gene was normalized to 16S rRNA gene copies for statistical analysis. All qPCR assays were conducted on a Roche LightCycler® 96 (Roche Molecular Diagnostics, Pleasanton, CA) at the Marquette University Water Quality Center in Milwaukee, WI.

Droplet Digital PCR (ddPCR) Assays

All details for ddPCR assays are provided in the manuscript text. **Supplementary Table 2** shows the MIQE guidelines that were followed for qPCR and ddPCR experiments.

Supplementary Table 1. Primers, annealing temperatures, amplification efficiencies, and R^2 values for qPCR analysis of target genes.

Gene	Annealing Temperature (°C)	Forward Primer & Reverse Primer	Amplicon Size (bp)	Efficiency Average (%)	R^2	Reference
16S rRNA	60	F- (5'-CCTACGGGAGGCAGCAG-3') R- (5'-ATTACCGCGGCTGCTGG-3')	202	98.5%	0.98	(Muyzer et al., 1993)
bla$_{SHV}$	64	F- (5'-CGCTTTCCCATGATGAGCACCTTT-3') R- (5'-TCCTGCTGGCGATAGTGGATCTTT-3')	94	95%	1.0	(Xi et al., 2009)
bla$_{TEM}$	60	F- (5'-GCKGCCAACTTACTTCTGACAACG-3') R- (5'-CTTTATCCGCCTCCATCCAGTCTA-3')	257	107%	0.98	(Marti et al., 2013)
*cop*A	63	F- (5'-ATGTGGAACSARATGCGKATGA-3') R- (5'-AGYTTCAGGCCSGGAATACG-3')	193	101%	0.99	(Roosa et al., 2014)
*czc*D	55	F- (5'-TCATCGCCGGTGCGATCATCAT-3') R- (5'-TGTCATTCACGACATGAACC-3')	272	99%	0.98	(Roosa et al., 2014)
*intI*1	60	F- (5'-CCTCCCGCACGATGATC-3') R- (5'-TCCACGCATCGTCAGGC-3')	280	96.5%	1.0	(Goldstein et al., 2001)
*sul*1	60	F- (5'-CCGTTGGCCTTCCTGTAAAG-3') R- (5'-TTGCCGATCGCGTGAAGT-3')	67	97%	1.0	(Burch et al., 2013)
tet(L)	60	F- (5'-TCGTTAGCGTGCTGTCATTC-3') R- (5'-GTATCCCCACCAATGTAGCCG-3')	276	99%	0.99	(Ng et al., 2001)

Supplementary Table 2: MIQE Guidelines/Checklist

Item to check	Importance*	Comments
Experimental Design		
Definition of experimental and control groups	E	Yes
Number within each group	E	Yes
Assay carried out by core lab or investigator's lab?	D	Yes
Power analysis	D	NA
Sample		
Description	E	Yes
Volume or mass of sample processed	E	Yes
Microdissection or microdissection	E	NA
Processing Procedure	E	Yes
If frozen – how and how quickly?	E	Yes
If fixed – with what, how quickly?	E	NA
Sample storage conditions and duration	E	Yes
Nucleic acid extraction		
Quantification – instrument/method	E	Yes
Storage conditions: temperature, concentrations, duration, buffer	E	Yes
DNA or RNA quantification	E	Yes
Quality/integrity, instrument/method, e.g. RNA integrity/R quality index and trace or 3':5'	E	NA
Template structural information	E	Yes
Template modification (digestion, sonication, preamplification., etc.)	E	NA
Template treatment (initial heating or chemical denaturation)	E	NA
Inhibition dilution or spike	E	NA
DNA contamination assessment of RNA sample	E	NA
Details of DNase treatment where performed	E	NA
Manufacturer of reagents used and catalogue number	D	Yes
Storage of nucleic acid: temperature, concentration, duration, buffer	E	Yes
RT (if necessary)		
cDNA priming method + concentration	E	NA
One- or 2-step protocol	E	NA
Amount of RNA used per reaction	E	NA
Detailed reaction components and conditions	E	NA
RT efficiency	D	NA
Estimated copies measures with and without addition of RT	D	NA
Manufacturer of reagents used and catalogue number	D	NA
Reaction volume (for 2-step RT reaction)	D	NA
Storage of cDNA: temperature, concentration, duration, buffer	D	NA
ddPCR target information		
Sequence accession number	E	Yes

Amplicon location	D	No
Amplicon length	E	Yes
In silico specificity screen (BLAST, etc.)	E	No
Pseudogenes, retro-pseudogenes or other homologs	D	NA
Sequence alignment	D	NA
Secondary structure analysis of amplicon and GC content	D	NA
Location of each primer by exon and intron	E	Yes
Where appropriate, which splice variants are targeted?	E	No
ddPCR oligonucleotides		
Primer sequences and/or amplicon context sequence	E	Yes
RTPrimerDB identification number	D	NA
Probe sequences	D	NA
Location and identity of any modifications	E	NA
Manufacturer or oligonucleotides	D	Yes
Purification method	D	NA
ddPCR protocol		
Complete reaction conditions	E	Yes
Reaction volume and amount of RNA/cDNA/DNA	E	Yes
Primer, (probe), Mg^{++} and dNTP concentrations	E	Yes
Polymerase identity and concentration	E	Yes
Buffer/kit catalogue no. and manufacturer	E	Yes
Exact chemical constitution of the buffer	D	NA
Additives (SYBR green I, DMSO, etc.)	E	Yes
Plate/tubes catalogue No and manufacturer	D	Yes
Complete thermocycling parameters	E	Yes
Reaction setup	D	Yes
Gravimetric or volumetric dilutions	D	Yes
Total PCR reaction volume prepared	D	Yes
Partition number	E	NA
Individual partition	E	NA
Total volume of the partitions measured (effective reaction size)	E	NA
Partition volume variance/SD	D	Yes
Comprehensive details and appropriate use of controls	E	Yes
Manufacturer of ddPCR instrument	E	Yes
ddPCR validation		
Optimization data for the assay	D	Yes
Specificity (when measuring rare mutations, pathogen sequences etc.)	E	NA
Limit of detection of calibration control	D	Yes
If multiplexing, comparison with singleplex assays	E	NA
Data analysis		
Mean copies per partition	E	Yes
ddPCR analysis program	E	Yes
Outlier identification and disposition	E	NA
Results of no-template controls	E	Yes
Examples of positive(s) and negative experimental results as supplemental data	E	Yes
Where appropriate, justification of number and choice of reference genes	E	NA
Where appropriate, description of normalization method	E	NA
Number and concordance of biological replicates	D	Yes
Number and stage (RT of ddPCR) of technical replicates	E	NA

Repeatability (intraassay variation)	E	NA
Reproducibility (interassay/user/lab etc. variation)	D	Yes
Experimental variance or CI	E	Yes
Statistical methods used for analysis	E	Yes
Data submission using RDML (real-time PCR data markup language)	D	No

* Essential information (E) submitted with the manuscript if applicable to the study. Desirable information (D) submitted if possible. NA = not applicable, Yes = provided in manuscript, No = not performed or provided in manuscript.
*Table adapted from (Bustin et al., 2009).

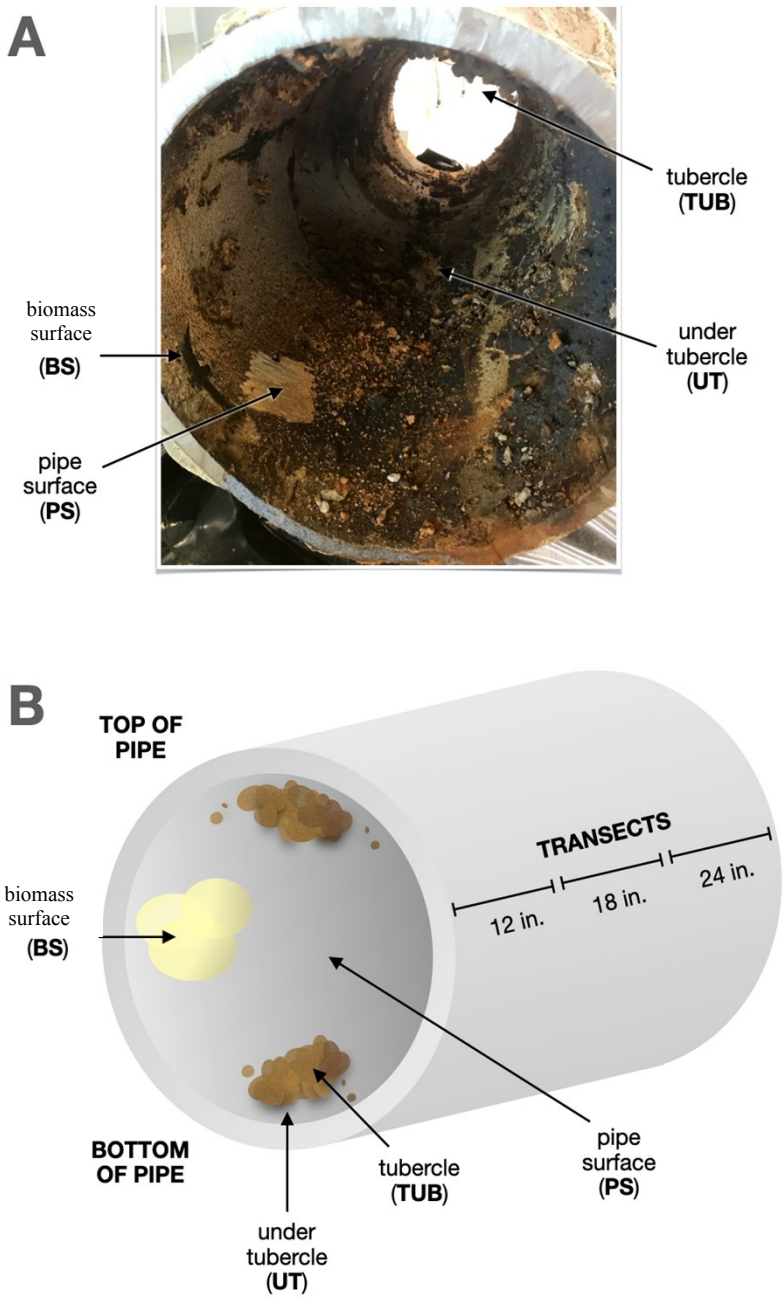
Illumina Sequencing of Biofilm Samples

To prepare biofilm microbial communities for DNA sequence analysis, the V4 hypervariable region of 16S rRNA genes was PCR-amplified from purified DNA using primers 515F/806R with Nextera adapters (Illumina, Inc., San Diego, CA) (Caporaso et al., 2012; Garner et al., 2019). One extraction blank and mock community (#HM-782D, BEI Resources) were included in the sample set. Reactions (25 µL total volume) were set up as follows: 12.5 µL KAPA HiFi HotStart ReadyMix (Roche Sequencing, Pleasanton, CA), 5 µM 515F primer, 5 µM 806R primer, 7.5 µL HyClone DNA-free water, and 2 µL diluted DNA extract. PCR thermal cycling conditions were as follows: 5 min at 95°C, 40 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 90 s, followed by 5 min at 72°C. PCR products were initially screened by electrophoresis on 1.5% agarose gels. Triplicate PCR products were pooled in ~75 µL samples, with amplicons less than 100 bp removed using 56 µL Agencourt AMPure XP beads in each (Beckman Coulter, Brea, CA). Final products were resuspended in 40 µL 10 mM Tris-EDTA. Sample library preparation was prepared according to the Illumina MiSeq protocol in the Nextera XT Index kit (Illumina, Inc., San Diego, CA) using 2X KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland). Indexed [LL1] [LL1]PCR amplicons were cleaned with AMPure XP beads and the SequalPrep Normalization kit (Thermo Fisher Scientific, Waltham, MA). Sequencing was conducted on the Illumina MiSeq platform using a 2x250-cycle paired-end protocol at the University of Wisconsin-Milwaukee Great Lakes Genomics Center in Milwaukee, WI (http://greatlakesgenomics.uwm.edu).

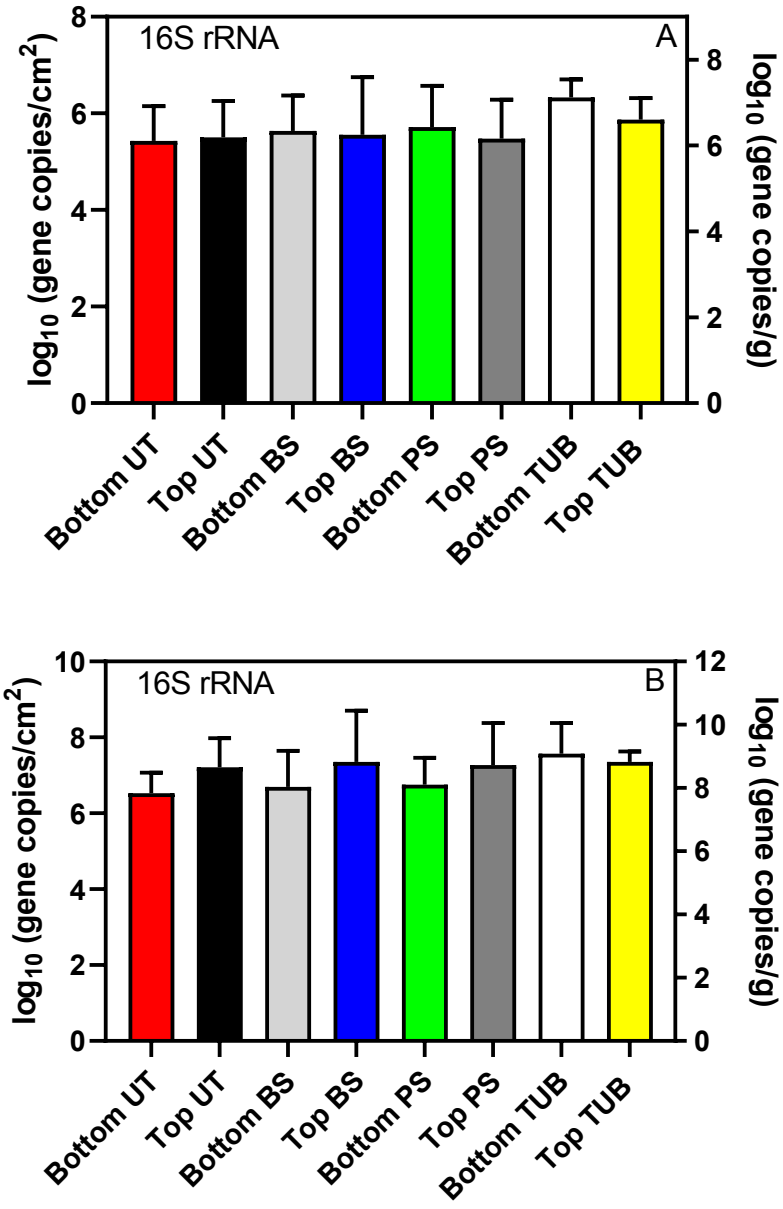
Primer and barcode sequences were removed from reads using cutadapt (Martin, 2011). Processing of sequence reads was conducted using the open statistical program 'R' utilizing the 'DADA2' package, which is a model-based approach for correcting amplicon errors without sequence similarity clustering to construct amplicon sequence variants (ASVs).(Rosen et al., 2012) Based on quality profiles, the last 20 base pairs of forward and reverse reads were removed. Two maximum expected errors were permitted. Any reads with ambiguous bases (N) were removed and reads with quality scores lower than 10 were removed. Chimeric reads were checked using the "consensus" method and removed. Taxonomic classification for the resulting ASVs was assigned using the SILVA v132 reference database.(Quast et al., 2013)

ASVs that were classified as mitochondria, chloroplast, or eukaryota were removed. Additional thresholds were set to identify and remove ASVs potentially derived from the mock community, extraction/PCR blank, and non-target samples that were included in the sequencing run: 1) any ASVs that were exact matches to those in the mock community were removed; 2) any ASVs that were three times more abundant in the blank sample (negative control) than across the dataset were removed; 3) any ASVs that were ten times more abundant in the non-target dataset than the drinking water pipe biofilm samples were removed.

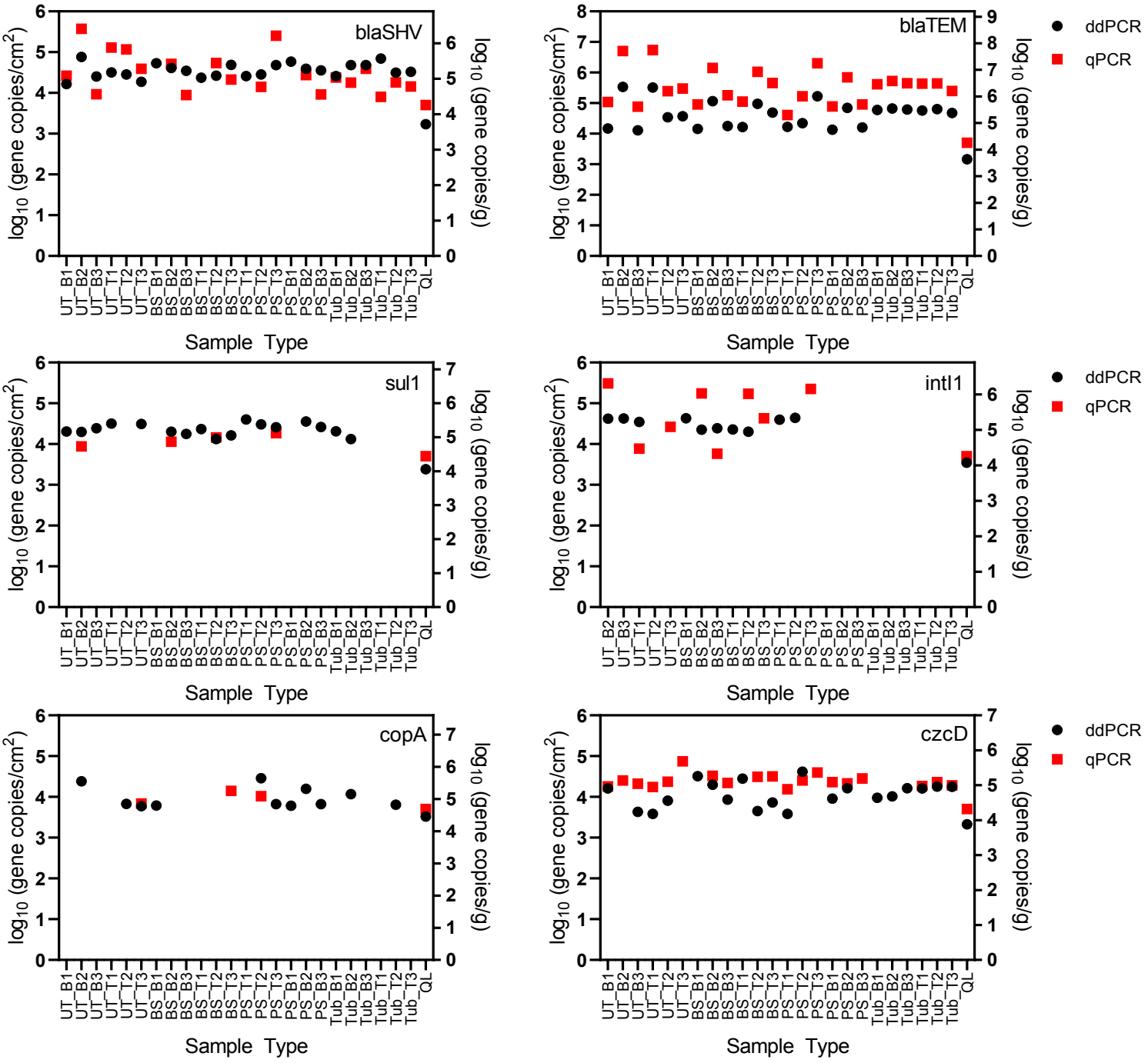
Code for these analyses can be found on GitHub: https://github.com/NewtonLabUWM/DrinkingWaterPipe_Ecology.



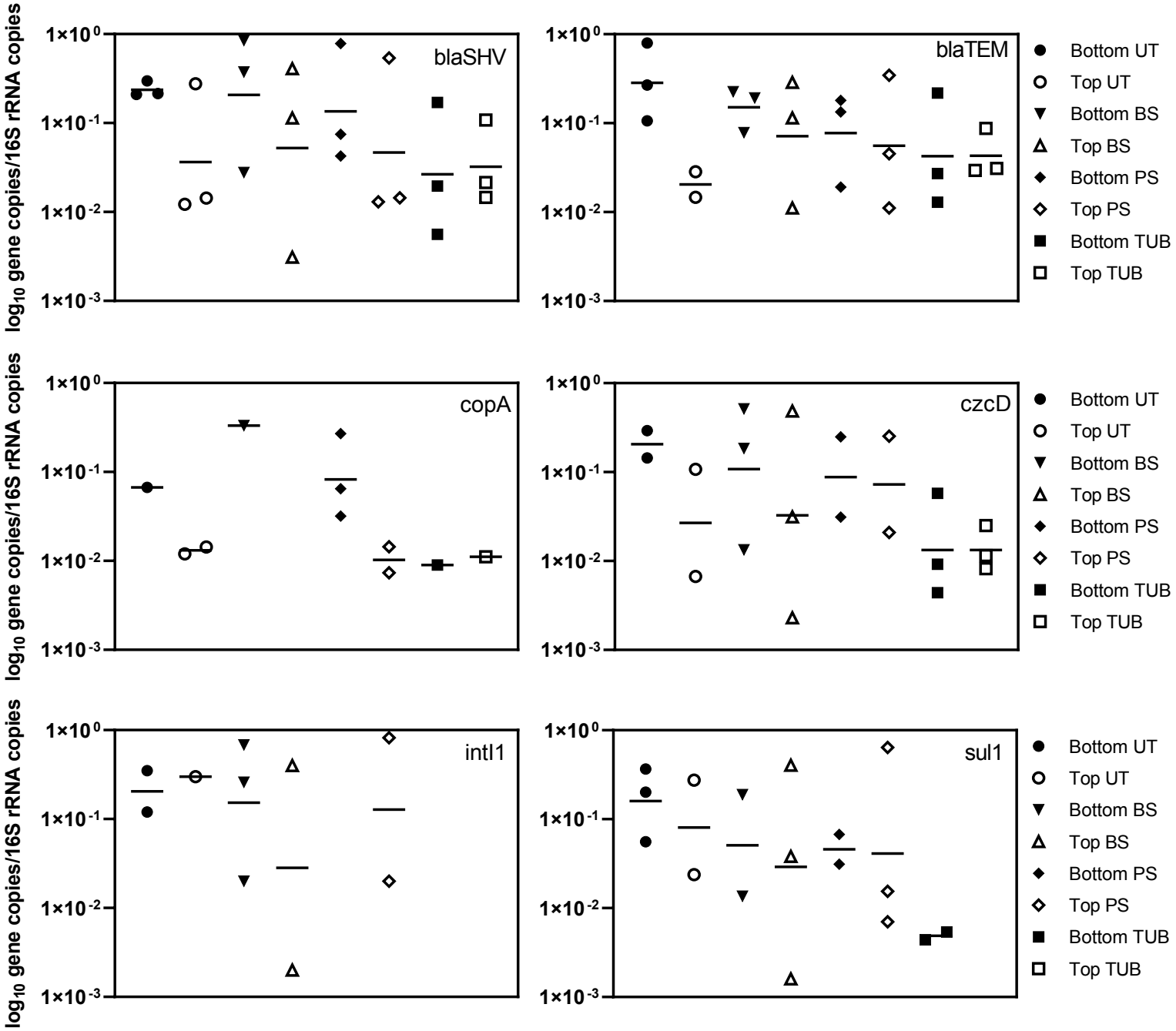
Supplementary Figure 1. Diagram of different biofilm sample locations from chloraminated cast iron drinking water main. Sample locations included biomass surface (BS), pipe surface scrape (PS), tubercles (TUB), and under tubercles (UT). Biomass surface samples were collected in areas with increased corrosion deposits and/or biofilm development. Pipe surface samples were collected by swabbing areas of the pipe with the least amounts of biofilm development.



Supplementary Figure 2. Absolute abundance of 16S rRNA genes in different biofilm microenvironments from a cast-iron drinking water main as measured with droplet digital PCR (ddPCR) (Figure A) and quantitative PCR (qPCR) (Figure B). The biofilm microenvironments include biomass surface (BS), pipe surface (PS), tubercle (TUB), and under tubercle (UT). Each bar represents a different sample category which includes triplicate biofilm samples (n=3). The mean absolute gene abundance (left y-axis: $\log_{10}$ gene copies/cm^2, right y-axis: $\log_{10}$ gene copies/g) for each of the biofilm samples is plotted on the y-axis with one standard deviation. The right y-axis only applies to corrosion tubercle (TUB) samples, which were analyzed by weight.

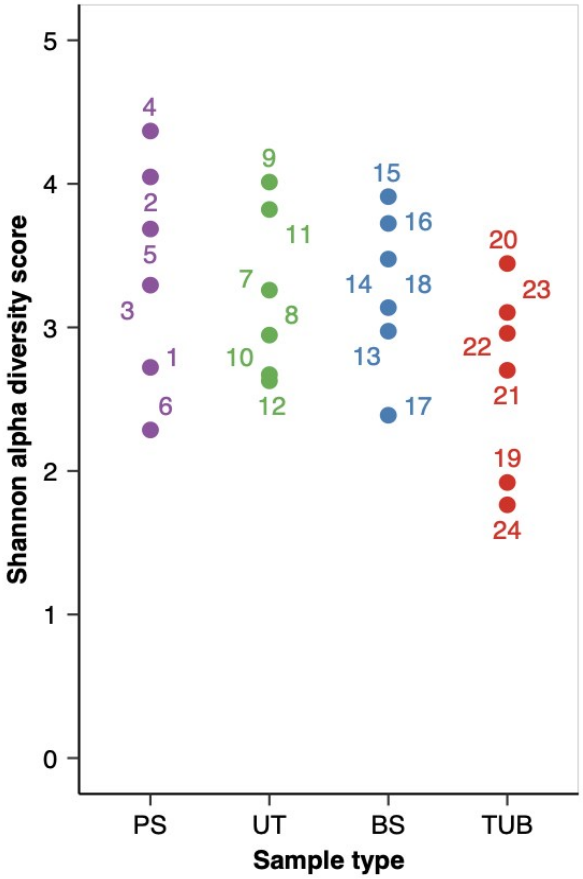


Supplementary Figure 3: Comparison of absolute gene abundance of antibiotic resistance genes, metal resistance genes, and *intI*1 as measured with ddPCR and qPCR assays. The biofilm microenvironments include biomass surface (BS), pipe surface (PS), tubercle (TUB), and under tubercle (UT). Each biofilm sample is also categorized by top (T) or bottom (B) pipe sample location. Each point represents the average of technical replicates for each biofilm sample (n=3). The mean absolute gene abundance (left y-axis: $\log_{10}$ gene copies/cm^2, right y-axis: $\log_{10}$ gene copies/g) for each of the biofilm samples is plotted. The right y-axis only applies to corrosion tubercle (TUB) samples, which were analyzed by weight. ddPCR concentrations are denoted with solid black circles and qPCR concentrations are denoted with red squares. The quantification limit (QL) is also plotted for each gene.

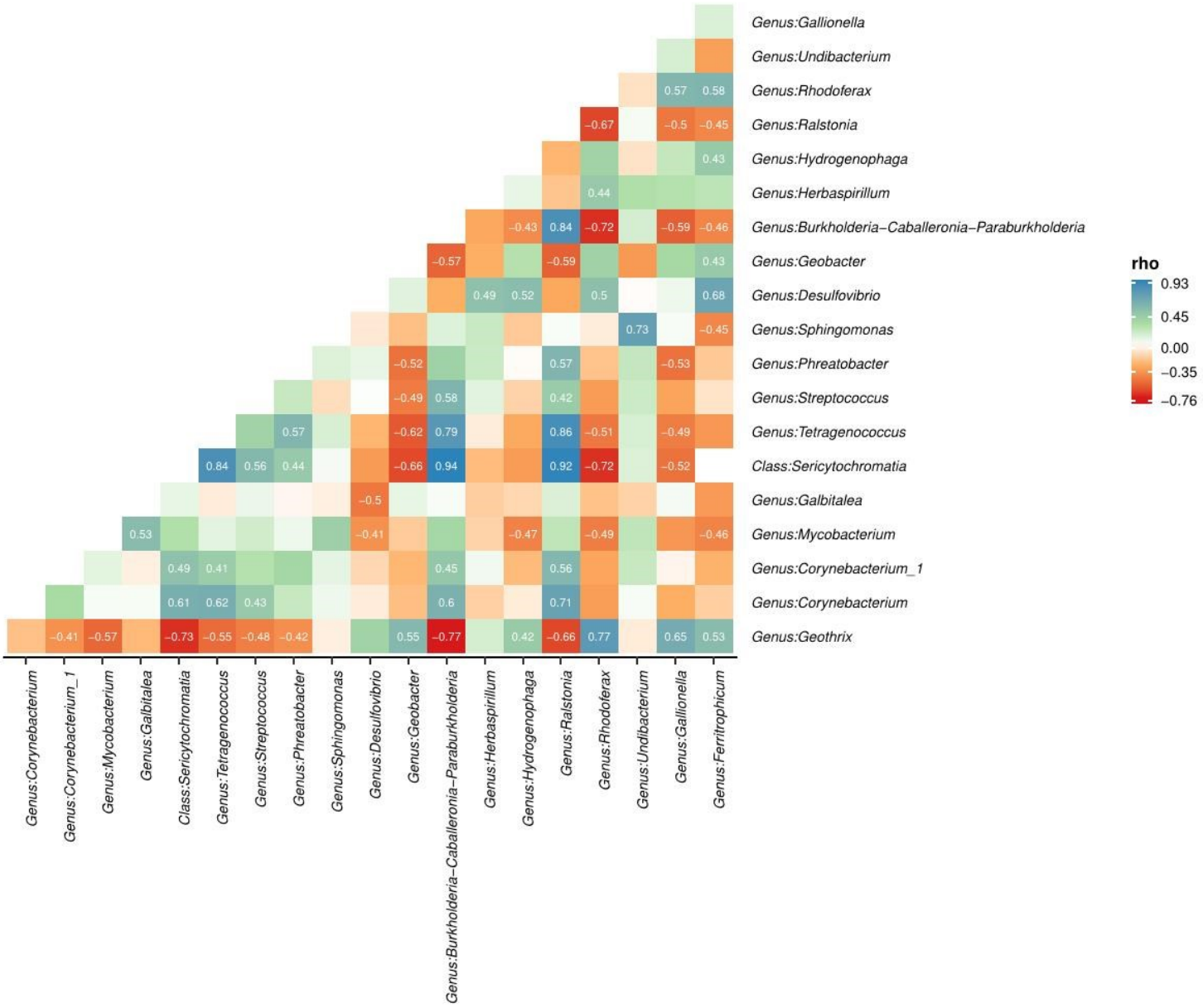


Supplementary Figure 4. Relative abundance of antibiotic resistance genes, metal resistance genes, and *intI*1 in different biofilm microenvironments from a cast-iron drinking water main as measured with droplet digital PCR (ddPCR). The biofilm microenvironments include biomass surface (BS), pipe surface (PS), tubercle (TUB), and under tubercle (UT). Each biofilm sample is also categorized by top or bottom pipe sample location. Each point represents the average of technical replicates for each biofilm sample (n=3).

	Sample type	Pipe location	Distance into pipe
1	Pipe surface (PS)	Bottom	18 in
2	PS	Bottom	24 in
3	PS	Bottom	24 in
4	PS	Top	12 in
5	PS	Top	18 in
6	PS	Top	24 in
7	Under tubercle (UT)	Bottom	12 in
8	UT	Bottom	18 in
9	UT	Bottom	24 in
10	UT	Top	12 in
11	UT	Top	18 in
12	UT	Top	24 in
13	Biomass surface (BS)	Bottom	12 in
14	BS	Bottom	18 in
15	BS	Bottom	24 in
16	BS	Top	12 in
17	BS	Top	12 in
18	BS	Top	18 in
19	Tubercle (TUB)	Bottom	18 in
20	TUB	Bottom	18 in
21	TUB	Bottom	24 in
22	TUB	Top	12 in
23	TUB	Top	18 in
24	TUB	Top	24 in



Supplementary Figure 5. Sample information for drinking water pipe samples (left). Shannon alpha diversity measurements of microbial communities characterized by 16S rRNA gene sequencing (right). Color of points denote biofilm microenvironments (sample type). Point labels indicate sample number as in table on left.



Supplementary Figure 6. Relationships between relative abundance of most abundant bacterial genera observed in biofilm samples from cast iron water main. Color denotes the result from correlation analysis using Spearman's rank sum correlation with Spearman's rho value plotted for each comparison. The Rho value for statistically significant relationships are also included (p values < 0.05). The lowest available taxonomic classification for each observed ASV is provided.

References

Burch, T.R., Sadowsky, M.J., LaPara, T.M., 2013. Air-drying beds eliminate antibiotic resistance genes and class 1 integrons in residual municipal wastewater solids. Environ. Sci. Technol. 9965–9971.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622. https://doi.org/10.1373/clinchem.2008.112797

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Garner, E., Inyang, M., Garvey, E., Parks, J., Glover, C., Dickenson, E., Sutherland;, J., Salveson, A., Edwards, M.A., Pruden, A., 2019. Impact of blending for direct potable reuse on premise plumbing microbial ecology and regrowth of opportunistic pathogens and antibiotic resistant bacteria. Water Res. 151, 75–86. https://doi.org/10.1016/j.watres.2018.12.003

Goldstein, C., Lee, M.D., Sanchez, S., Phillips, B., Register, B., Grady, M., Liebert, C., Summers, A.O., White, D.G., Maurer, J.J., Hudson, C., 2001. Incidence of Class 1 and 2 Integrases in Clinical and Commensal Bacteria from Livestock , Companion Animals , and Exotics. Antimicrob. Agents Chemother. 45, 723–726. https://doi.org/10.1128/AAC.45.3.723

Kimbell, L.K., Kappell, A.D., McNamara, P.J., 2018. Effect of pyrolysis on the removal of antibiotic resistance genes and class I integrons from municipal wastewater biosolids. Environ. Sci. Water Res. Technol. 4, 1807–1818. https://doi.org/10.1039/C8EW00141C

Marti, E., Jofre, J., Balcazar, J.L., 2013. Prevalence of Antibiotic Resistance Genes and Bacterial Community Composition in a River Influenced by a Wastewater Treatment Plant. PLoS One 8, e78906. https://doi.org/10.1371/journal.pone.0078906

Muyzer, G., Waal, E. de, Uitierlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environmantal Microbiol. 59, 695–700.

Ng, L.K., Martin, I., Alfa, M., Mulvey, M., 2001. Multiplex PCR for the detection of tetracycline resistant genes. Mol. Cell. Probes 15, 209–215. https://doi.org/10.1006/mcpr.2001.0363

Pruden, A., Arabi, M., Storteboom, H.N., 2012. Correlation between upstream human activities and riverine antibiotic resistance genes. Environ. Sci. Technol. 46, 11541–11549. https://doi.org/10.1021/es302657r

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596. https://doi.org/10.1093/nar/gks1219

Roosa, S., Wattiez, R., Prygiel, E., Lesven, L., Billon, G., Gillan, D.C., 2014. Bacterial metal resistance genes and metal bioavailability in contaminated sediments. Environ. Pollut. 189,

143–151. https://doi.org/10.1016/j.envpol.2014.02.031

Rosen, M.J., Callahan, B.J., Fisher, D.S., Holmes, S.P., 2012. Denoising PCR-amplified metagenome data. BMC Bioinformatics 13. https://doi.org/10.1186/1471-2105-13-283

Xi, C., Zhang, Y., Marrs, C.F., Ye, W., Simon, C., Foxman, B., Nriagu, J., 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Appl. Environ. Microbiol. 75, 5714–5718. https://doi.org/10.1128/AEM.00382-09