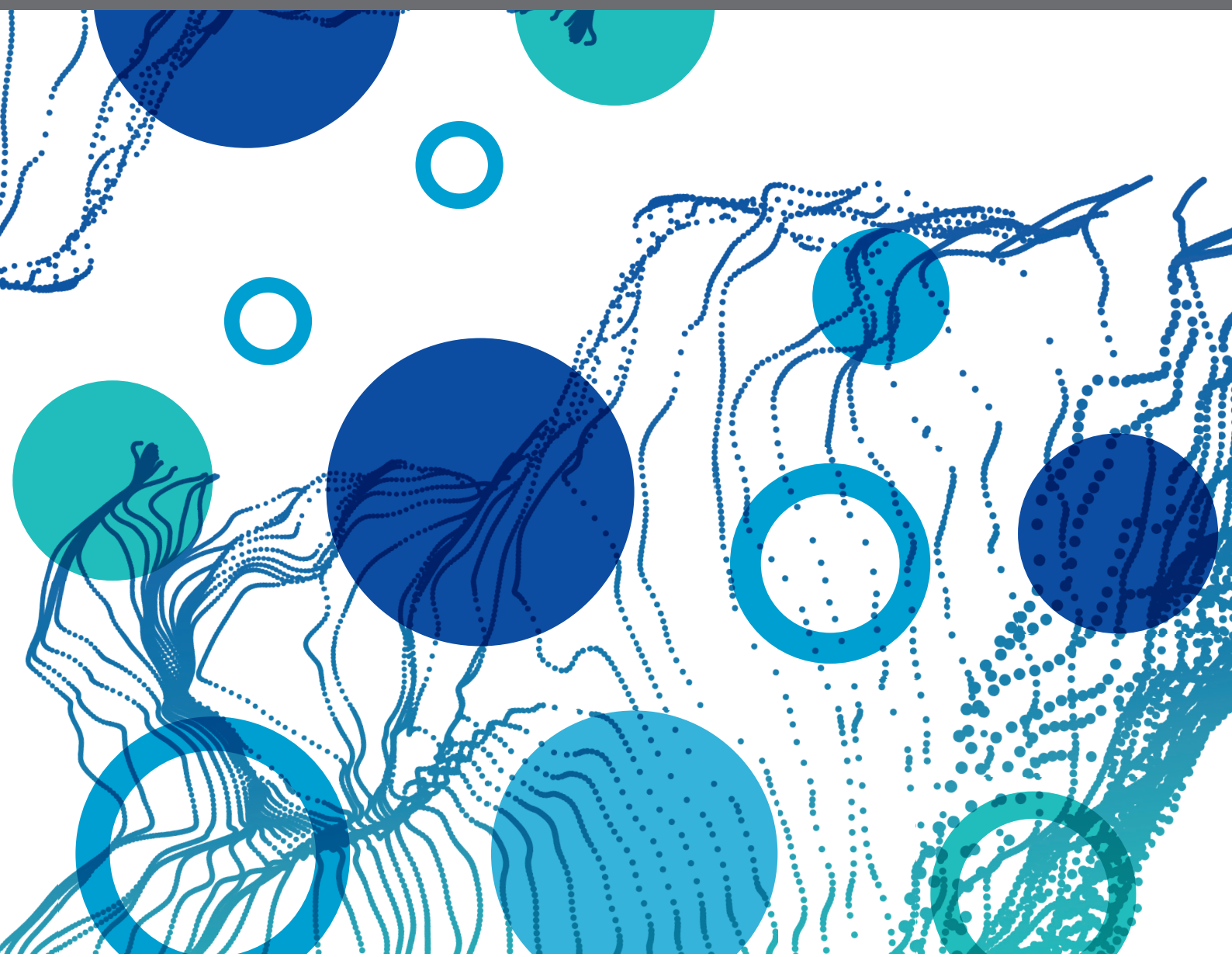
FUNCTIONAL DIVERSITY OF AQUATIC MICROORGANISMS AND THEIR ROLES IN WATER QUALITY

EDITED BY: Jingqiu Liao, Yi Huang, Jasna Kovac and Daniel Lowell Weller

PUBLISHED IN: Frontiers in Water



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-370-2
DOI 10.3389/978-2-88976-370-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FUNCTIONAL DIVERSITY OF AQUATIC MICROORGANISMS AND THEIR ROLES IN WATER QUALITY

Topic Editors:
Jingqiu Liao, Columbia University, United States
Yi Huang, Peking University, China
Jasna Kovac, The Pennsylvania State University (PSU), United States
Daniel Lowell Weller, University of Rochester, United States

Citation: Liao, J., Huang, Y., Kovac, J., Weller, D. L., eds. (2022). Functional Diversity of Aquatic Microorganisms and Their Roles in Water Quality. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-370-2

Table of Contents

frontiers in Water

EDITORIAL
published: 25 March 2022
doi: 10.3389/frwa.2022.879155

Editorial: Functional Diversity of Aquatic Microorganisms and Their Roles in Water Quality

Jingqiu Liao[1]*, *Daniel L. Weller*[2,3], *Jasna Kovac*[4] *and Yi Huang*[5]

[1] Department of Systems Biology, Columbia University, New York, NY, United States, [2] Department of Environmental and Forest Biology, State University of New York College of Environmental Science and Forestry, Syracuse, NY, United States, [3] Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, United States, [4] Department of Food Science, The Pennsylvania State University, University Park, PA, United States, [5] College of Environmental Sciences and Engineering, Peking University, Beijing, China

Keywords: water quality, pathogen, environmental microbial community, agricultural water, wastewater, recreational water, human health

Editorial on the Research Topic

Functional Diversity of Aquatic Microorganisms and Their Roles in Water Quality

As an important component of aquatic environments, microorganisms are functionally diverse and play a crucial role in water quality and human health. Through a wide spectrum of biogeochemical processes, they can improve the sustainability and stability of aquatic ecosystems and positively influence water quality (Zinger et al., 2012). For example, many aquatic microorganisms are capable of removing chemical hazards such as heavy metals (Dixit et al., 2015) and petroleum (Zaki et al., 2015), and of catalyzing wastewater treatment processes (Daims et al., 2006). Meanwhile, contamination of agricultural, recreational, and drinking water by human pathogens is of public health concern due to the potential for transmission to humans directly (e.g., through drinking, dermal contact) and indirectly (e.g., through the use of contaminated water to irrigate crops) (Egli et al., 2002). Thus, characterizing the functional diversity of aquatic microorganisms and understanding their impact on water environments represent a key research need. However, despite the importance, aquatic microorganisms in many regions in the world are understudied. As a result, affordable, targeted tools for assessing and managing water quality are limited. This Research Topic aimed to improve the understanding of the distribution, structure, and function of aquatic microbial populations with a specific focus on potential approaches for improving the assessment and management of water quality. We received contributions across a wide range of topics, including microbial dynamics in wastewater and surface water, environmental factors driving the distribution of foodborne pathogens in water, and novel approaches to water quality monitoring and microbial risk assessment. Here, we briefly summarize these contributions by their relevance to three water types—wastewater, agricultural water, and recreational water.

WASTEWATER

Wastewater treatment and wastewater reuse is essential to the global demand for water in the era of climate change (Yadav et al., 2020). Microorganisms are responsible for biological wastewater treatment processes such as removal of organic matter, toxic metals, odors, nutrients including nitrogen and phosphorus (Wagner et al., 2002; Ferrera and Sanchez, 2016). Therefore, knowledge on the structure and functions of microbial communities and populations in wastewater are crucial for the development and optimization of wastewater treatment systems. Huang et al.

OPEN ACCESS

Edited and reviewed by:
Boris M. Van Breukelen, Delft University of Technology, Netherlands

***Correspondence:**
Jingqiu Liao
jl5897@cumc.columbia.edu

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *18 February 2022*
Accepted: *03 March 2022*
Published: *25 March 2022*

Citation:
Liao J, Weller DL, Kovac J and Huang Y (2022) Editorial: Functional Diversity of Aquatic Microorganisms and Their Roles in Water Quality. Front. Water 4:879155. doi: 10.3389/frwa.2022.879155

investigated temporal and spatial changes in composition and putative function of bacterial communities in wastewater stabilization ponds (WSP), a commonly used system to treat municipal wastewater in the Canadian Arctic. They identified temperature as a major factor influencing the bacterial community structure and potential removal of nutrients in the WSPs. In a preliminary study, Cruz et al. examined differences in the composition and richness of bacterial communities in treated and untreated wastewater from different sources (e.g., hospitals, fish farming sites, sewage effluents, and surface waters) in Sri Lanka and Philippines.

AGRICULTURAL WATER

Given the central role of water in food production, ensuring and monitoring agricultural water is essential to the maintenance of a safe and secure food supply (Molden et al., 2010). However, contamination of agricultural water by fecal contaminants and human pathogens in agricultural water presents a key human health risk (Rodrigues et al., 2019). To effectively manage health risks associated with contamination of agricultural water, it is important to understand factors that are associated with fecal and pathogen contamination. Murphy et al. showed that weather, especially temperature, was strongly associated with *Salmonella* contamination of north and South Florida agricultural water sources. Bihn et al. found that water specific conductance, pH, and turbidity were slightly associated with fecal indicator bacteria levels in NY and TN, USA and highlighted the importance of using multiple data types (e.g., bacterial, physicochemical) when assessing irrigation water quality. Belias et al. developed machine learning models using water quality parameters, surrounding land-use, weather conditions, and sampling site characteristics as predictors to predict pathogen presence in canals used to source water for produce production in the Southwestern US agricultural water. Belias et al. findings suggest that predictive models could supplement or provide an alternative to current approaches for monitoring agricultural water quality. Smith et al. detected stable spatial and temporal changes in phytoplankton functional groups in two agricultural irrigation ponds in Maryland and found that water quality parameters were associated with certain phytoplankton functional groups.

RECREATIONAL WATER

Exposure to contaminated recreational water represents another pathway for human pathogen transmission (Fewtrell and Kay, 2015). While understanding these infectious agents of concern, including fecal indicator bacteria in recreational water is of great importance, additional information on factors associated with pathogen contamination of recreational waterways and tools for assessing contamination is needed. By analyzing water samples collected along an urban-rural gradient, Weller et al. found that temperature, nutrient, sediment levels, and land use are key factors associated with the levels of fecal indicator bacteria in Upstate New York streams used for both agricultural and recreation. Fernández-Baca et al. developed a rapid, in-field workflow for detecting fecal indicator bacteria in recreation water that takes substantially less time compared to culture-based assays. Tools like that developed by Fernández-Baca et al. can facilitate risk management (e.g., opening and closing of beaches), by providing decision-makers with key data faster. Brooks and Rose compared models for evaluating the long-term persistence of enteric markers in sewage spiked river water and found that log-linear models overestimated the persistence of enteric markers and suggested that best fit models with dynamic decay rates in water quality models are more accurate in evaluating the decay of enteric markers.

CONCLUDING REMARKS

Over the last several decades, efforts in understanding water quality have moved beyond its physical and chemical attributes to biological attributes (Fewtrell et al., 2001). The growing interests in the role of microorganisms in water quality is well reflected by the large number of contributions received in this Research Topic. It is exciting to see active research characterizing the composition, function, diversity of microbial communities, populations, and pathogens in water across different types including wastewater, agricultural water, and recreational water, as well as developing useful tools for water quality monitoring. This Research Topic has provided insights and solutions for both favorable and unfavorable consequences of water quality influenced by microorganisms. With the emerging technologies (e.g., high-throughput sequencing) in profiling microorganisms and broad application of deep learning in the field of environmental microbiology, we are optimistic to a new level of water management and public health improvement.

AUTHOR CONTRIBUTIONS

All authors have made a substantial contribution in writing and revising the editorial. All authors contributed to the article and approved the submitted version.

REFERENCES

Daims, H., Taylor, M. W., and Wagner, M. (2006). Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol.* 24, 483–89. doi: 10.1016/j.tibtech.2006.09.002

Dixit, R., Malaviya, D., Pandiyan, K., Singh, U. B., Sahu, A., Shukla, R., et al. (2015). Bioremediation of heavy metals from soil and aquatic environment: an overview of principles and criteria of fundamental processes. *Sustainability* 7, 2189–2212. doi: 10.3390/su7022189

Egli, T., Köster, W., and Meile, L. (2002). Pathogenic microbes in water and food: changes and challenges. *FEMS Microbiol. Rev.* 26, 111–12. doi: 10.1111/j.1574-6976.2002.tb00603.x

Ferrera, I., and Sanchez, O. (2016). Insights into microbial diversity in wastewater treatment systems: how far have we come? *Biotechnol. Adv.* 34, 790–802. doi: 10.1016/j.biotechadv.2016.04.003

Fewtrell, L., and Bartram, J. (2001). *Water Quality: Guidelines, Standards & Health*. London: IWA Publishing.

Fewtrell, L., and Kay, D. (2015). Recreational water and infection: a review of recent findings. *Curr. Environ. Health Rep.* 2, 85–94. doi: 10.1007/s40572-014-0036-6

Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M. A., and Kijne, J. (2010). Improving agricultural water productivity: between optimism and caution. *Agric. Water Manage.* 97, 528–35. doi: 10.1016/j.agwat.2009.03.023

Rodrigues, C., da Silva, A. L. B. R., and Dunn, L. L. (2019). Factors impacting the prevalence of foodborne pathogens in agricultural water sources in the Southeastern United States. *Water* 12, 51. doi: 10.3390/w12010051

Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., and Daims, H. (2002). Microbial community composition and function in wastewater treatment plants. *Antonie van Leeuwenhoek* 81, 665–80. doi: 10.1023/A:1020586312170

Yadav, B., Pandey, A. K., Kumar, L. R., Kaur, R., Yellapu, S. K., Sellamuthu, B., et al. (2020). "1 - Introduction to wastewater microbiology: special emphasis on hospital wastewater," in *Current Developments in Biotechnology and Bioengineering*, eds R. D. Tyagi, B. Sellamuthu, B. Tiwari, S. Yan, P. Drogui, X. Zhang, et al. (Amsterdam; Oxford; Cambridge: Elsevier), 1–41.

Zaki, M. S., Authman, M. M., and Abbas, H. H. (2015). Bioremediation of petroleum contaminants in aquatic environments. *Life Sci. J.* 12, 127–39.

Zinger, L., Gobet, A., and Pommier, T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Mol. Ecol.* 21, 1878–96. doi: 10.1111/j.1365-294X.2011.05362.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

BRIEF RESEARCH REPORT
published: 27 July 2021
doi: 10.3389/frwa.2021.695207

Comparing Log-Linear and Best-Fit Models to Evaluate the Long-Term Persistence of Enteric Markers in Sewage Spiked River Water

Yolanda M. Brooks[1]* *and Joan B. Rose*[1,2]

[1] Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, United States,
[2] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, United States

Water quality models use log-linear decay to estimate the inactivation of fecal indicator bacteria (FIB). The decay of molecular measurements of FIB does not follow a log-linear pattern. This study examined the factors associated with the persistence of *Escherichia coli uidA*, enterococci *23S rDNA*, and *Bacteroides thetataiotaomicron 1,6 alpha mannanase* in microcosms containing 10% (vol/vol) sewage spiked river water stored at 4°C for up to 337 days. The study estimated the markers' persistence with log-linear models (LLMs) to the best-fit models, biphasic exponential decay (BI3) and log-logistic (JM2) and compared the estimates from the models. Concentrations of *B. thetataiotaomicon* decreased to levels below detection after 31 days in storage and were not fit to models. BI3 and JM2 were fit to *E. coli* and enterococci, respectively. LLMs had larger Bayesian information criterion values than best-fit models, indicating poor fit. LLMs over-estimated the time required for 90% reduction of the indicators (T_{90}) and did not consider dynamic rates of decay. Time in storage and indicator species were associated with the persistence of the markers ($p < 0.001$). Using the T_{90} values of the best-fit models, enterococci was the most persistent indicator. Our data supports the use of best fit models with dynamic decay rates in water quality models to evaluate the decay of enteric markers.

Keywords: fecal pollution, persistence time, water quality, best fit analysis, *Bacteroides* as indicator of fecal matter, enterococci, *Escherichia coli*, molecular markers

OPEN ACCESS

Edited by:
Jasna Kovac, Pennsylvania State University (PSU), United States

Reviewed by:
Meredith B. Nevers, United States Geological Survey (USGS), United States
Hyatt Green, SUNY College of Environmental Science and Forestry, United States
Autumn Kraft, United States Department of Agriculture (USDA), United States

***Correspondence:**
Yolanda M. Brooks
ybrooks@sjcme.edu

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *14 April 2021*
Accepted: *15 June 2021*
Published: *27 July 2021*

Citation:
Brooks YM and Rose JB (2021) Comparing Log-Linear and Best-Fit Models to Evaluate the Long-Term Persistence of Enteric Markers in Sewage Spiked River Water. Front. Water 3:695207. doi: 10.3389/frwa.2021.695207

INTRODUCTION

In the United States, environmental waters that are impaired or in danger of being impaired are assigned remediation plans aimed at restoring designated uses (United States, 2016). Forecasting the restoration of an impaired watershed is partially dependent on the development of water quality models that estimate the concentrations of fecal indicator bacteria (FIB), while simultaneously calculating the transport and fate (particularly the survival) in the aquatic ecosystem. Water quality models have been slow to adapt data generated from molecular measurements of FIB even as these measurements are finding use in rapid "swimmable" assessments. Compared to concentrations of culturable enterococci, measurements of enterococci *23S rDNA* (Entero1) had a stronger correlation to gastrointestinal illness incidence of children visiting beaches near point sources of treated sewage at Lake Michigan, Michigan and Indiana; and Lake Erie, Ohio (Wade et al., 2008). Therefore, water quality models that incorporate molecular measurements of FIB could provide better resolution of the risk of waterborne illness.

Microbial source tracking (MST) with molecular methods can identify and quantify pollution associated with specific hosts in watersheds of impaired rivers (Ballesté et al., 2020; Brooks et al., 2020). Identification of host-associated pollution is important because different origins of pollution are suggested to pose varying risks of adverse health effects (Soller et al., 2010). Multiple studies have demonstrated that human-associated markers can be detected in various types of environmental waters (Reischer et al., 2007; Rusiñol et al., 2014; Ahmed et al., 2019). The concentrations of the markers can also be incorporated into water quality models to improve forecasting of an impaired watershed (Jeong et al., 2019). Such analysis could offer better assessment of pollution associated with higher risk to public health.

Two commonly used water quality models, HSPF, and SWAT, use log-linear models (LLMs) to calculate the survival of culturable FIB in aquatic environments (Johnson et al., 1984; Gassman et al., 2007). A meta-analysis determined that there are three general patterns of survival curves of culturable *E. coli* in water microcosms: biphasic log-linear, shouldering followed by log-linear decay, and log-linear (Blaustein et al., 2013). These studies suggest that persistence models could accommodate the dynamic decay of enteric markers.

There are few studies that have evaluated the persistence of FIB in water over longer timescales. A previous study determined that the time needed for the 90% reduction (T_{90}) of culturable *E. coli* (EC) and enterococci (ENT) was >60 days in microcosms containing sediment covered with freshwater spiked with sewage effluent and stored at 10°C (Pote et al., 2009). Further research of the long-term persistence of FIB would provide better forecasting power for water quality models.

In order to increase the accuracy and precision of water quality models, we describe the long term persistence (up to 337 days) of three enteric molecular markers, *B. thetaiotaomicron 1,6 alpha-mannanase*, enterococci *23S rDNA*, and *E. coli uidA*, stored in microcosms of 10% (vol/vol) sewage spiked river water stored in the dark and at 4°C. Enterococci *23S rRNA* was chosen because it is a standard method to evaluate fecal pollution in recreational waters (USEPA, 2015). *E. coli uidA* was chosen because the marker is species specific and the United States Environmental Protection Agency established criteria regarding the concentration of culturable *E.coli* in recreational waters (USEPA, 2012). *B. thetaiotaomicron 1,6 alpha-mannanase* was chosen to evaluate the persistence of human associated pollution. The temperature was chosen because it represented a conservative estimate of the persistence of enteric markers in many temparate freshwater lakes. Microcosms were stored in the dark during the duration of this study. Sunlight inactivation was not evaluated in this study because previous studies have demonstrated that freshwater microcosms containing either human or cattle sewage demonstrated no significant difference in the decay of molecular targets in the presence of sunlight compared to microcosms stored in the dark (Dick et al., 2010; Korajkic et al., 2019).

The specific objectives of this study were to:

1. Evaluate the mathematical relationships of experimental data of *Escherichia coli uidA*, enterococci *23S rRNA*, and *Bacteroides thetataiotaomicron alpha-mannanase* measured from microcosms of 10% sewage spiked river water stored for up to 337 days at 4°C using log-linear and best-fit methods. Compare the mathematical relationships of the best-fit models to log-linear models. Compare the times required for 90% and 99% reductions of the initial concentrations of the markers using best-fit models and a corresponding log-linear model.
2. Evaluate the associations of time in storage, and indicator species to the observed persistence of the three markers.

METHODS

Microcosm Set-Up

Water from the Red Cedar River (6 L) was autoclaved for 1 h in multiple 2 L plastic containers. After cooling to room temperature, the water was spiked with 10% (vol/vol) raw sewage from the East Lansing Wastewater Treatment and equally divided into 12 autoclaved 1 L Nalgene polypropylene bottles (ThermoFisher Scientific, Waltham, MA). Neither the river water nor the raw sewage were evaluated for the three genetic markers prior to the set-up of the microcosms. Each microcosm contained ∼500 mL of liquid and no sediment. All microcosms were stored in the dark at 4°C.

Sample Processing and DNA Extraction

On each of the following days after seeding the raw sewage into the river water: 0, 31, 61, 93, 128, 155, 187, 219, 248, 279, 306, and 337, a microcosm was shaken 25 times (EPA-DNA, USEPA, 2015) and triplicate aliquots of 100 ml from one microcosm were membrane filtered onto three separate Nucleopore Track Etch polycarbonate membrane filters (0.45 μm pore size, 47 mm diameter, Whatman Inc., Piscataway, NJ). The three filters per sampling day represented technical replicates. DNA extraction from all samples was performed using USEPA Method 1611.1 (EPA-DNA, USEPA, 2015). Specifically, three filters were used to concentrate cells within each time point. At each time point, a filtration negative of autoclaved phosphate buffer saline accompanied the filtration of the samples. The extraction method in USEPA Method 1611.1 mechanically shears the cells. The first step is 1 min of bead beating of the filters at max speed in 600 μL DNA extraction buffer with 0.3 mm glass beads. Two centrifugation steps were completed to separate glass beads and other cell debris from the 350 μL-supernatant containing the DNA. The supernatants were diluted 1:5 in AE buffer and were stored at −80°C.

qPCR Quantification of Three Indicators of Fecal Pollution

The primers, probes, amplified sequence length, and qPCR protocols for *Bacteroides thetataiotaomicron alpha mannanase* (BT-am), *Escherichia coli uidA* (EC-uidA), and *Enterococcus* spp. *23S rDNA* (ENT-23) are described in **Table 1**. The qPCR reagents of the ENT-23, EC-uidA and BT-am were: 10 μl of LightCycler 480 Probes 25 MasterMix (Roche, Indianapolis, IN), 0.5 μM of forward and reverse primers (0.3 μM of BT-am forward and reverse primer concentration), 0.2 μM of the

TABLE 1 | Details of the qPCR assays (primers, probes, amplified sequence length, and qPCR protocols) of the genetic markers, enterococci *23S rDNA* (ENT-23), *Escherichia coli uidA* (EC-uidA), and *Bacteroides thetataiotaomicron alpha mannanase* (BT-am).

Source species and gene (marker name)	Primer and probe sequences (5′-3′)	Sequence length (bp)	qPCR protocol	References
Bacteroides thetataiotaomicron alpha mannanase (BT-am)	CATCGTTCGTCAGCAGTAACA CCAAGAAAAAGGGACAGTGG FAM-CAGCAGGT-NFQ[†]	63	15 min at 95°C **45 cycles:** 15 s at 94°C 60 s at 60°C 5 s for 72°C	Yampara-Iquise et al., 2008
Escherichia coli uidA (EC-uidA)	CAATGGTGATGTCAGCGTT ACACTCTGTCCGGCTTTTG 6FAM-TTGCAACTGGACAAGGCACCAGC-BBQ	163	10 min at 95°C **40 cycles:** 6 s at 95°C 8 s at 58°C 8 s at 72°C	Srinivasan et al., 2011
Enterococcus spp. *23S rDNA* (ENT-23)	AGAAATTCCAAACGAACTTG CAGTGCTCTACCTCCATCATT 6FAM-GGTTCTCTCCGAAATAGCTTTAGGGCTA-TAMRA	91	10 min at 95°C **40 cycles:** 15 s at 95°C 30 s at 60°C 15 s at 72°C	Frahm and Obst, 2003

[†] *Primer represents the Roche Universal Probe Libraries Primer 62.*

probe (0.1 μM BT-am probe), 2 mg/ml bovine serum album (BSA), 1 mM of MgCl2 (EC-uidA only), 5 μl of DNA sample, and enough nuclease free water added to equal a 20 μl reaction volume. The following components were analyzed in each qPCR assay: three wells of a dilution step from the standard curve such as the 10^6 or 10^7 copies/rxn for ENT, 10^5 or 10^7 copies/rxn for EC, and or 10^3 or 10^4 copies/rxn for BT, a well containing nuclease free water, and duplicates of method blanks for each storage timepoint (DNA extracts of sterile phosphate buffer saline membrane filtered at each time point). Each DNA extract from the experiment was run in duplicate to represent analytical replicates. The analytical replicates were averaged to represent each technical replicate. Information regarding the performance of the standards curves, including the theoretical limit of detection, limit of quantification, and efficiency of the standard curve are included in **Supplementary Table 1**.

Quantification was established with a standard curve. Overnight cultures of *Enterococcus faecalis* ATCC strain 19433 and *E. coli* ATCC strain 15597 were made, and genomic DNA was extracted using Qiamp DNA Mini Kit (Qiagen Inc, Valencia, CA). The qPCR measurements were reported as cellular equivalents (CE) of BT, ENT, or EC per 100 ml-water sample to compare the data to survival curves of culturable FIB in previous studies. All DNA extracts were diluted 1:5 before addition to the qPCR reagents to decrease the chances of inhibition, which also decreased the lowest detected concentration of BT, EC, and ENT to 2.68 * 10^2, 1.30 * 10^5, and 1.38 * 10^3 CE/100 ml-water sample, respectively. Genomic DNA was quantified of *B. thetataiotaomicron* (ATCC, 29148) with NanoDrop 1000 (ThermoFisher Scientific, Waltam, MA) to obtain copy number per rxn. Subsequently, 1:10 serial dilutions with dilution steps of the genomic DNA of the three indicators were made. The dilution steps spanned from 10^0 to 10^7 copies/rxn. One standard curve was used for both qPCR runs and contained triplicates of each dilution step. For all standard curves, the R^2 was > 0.96. qPCR inhibition (defined as technical duplicates differing by > 3.32 Ct) was not noted in the experiment.

Persistence Modeling and Statistical Analysis

The statistical and modeling analyses considered all data points, including technical replicates that measured ENT, EC, and BT (statistical analysis only). The persistence of BT was not evaluated using a best fit model, or log-linear model because there were only three timepoints that were above the detection limit (**Figure 1A**). In the EC dataset, non-detects were evaluated at the detection limit in the persistence modeling and statistical analyses, as previously advised (USEPA, 1991). Model selection and shape of best-fit models of EC were affected by non-detect samples that occurred at $t = 219$ and 248 days (**Figure 1C**).

The average CE concentrations of each indicator were transformed to the relative persistence of the original concentration, N/N_0, where N was the average concentration of the indicator from three technical replicates (CE/100 ml-sample) after t days in storage, and N_0 was the average concentration of the indicator from three technical replicates at initial sampling ($t = 0$). A model fitting tool in R (R Development Core Team, 2013) provided by Drs. Kyle Enger and Jade Mitchell used maximum likelihood estimations to evaluate 17 linear and non-linear models (qmra.canr.msu.edu) to the data. The best-fit model with the lowest value of the Bayesian information criterion (BIC) was chosen to represent the persistence of the indicators. If the BIC value of biphasic log-linear model (BI3) was ≤ 2 units from the smallest BIC value, then BI3 was chosen; as ≤ 2 units difference are considered equivalent (Bolker, 2008). These two best-fit models evaluated the indicators: biphasic log linear model, BI3 (Carret et al., 1991); and log-logistic model, JM2 (Juneja et al., 2003). The model fitting tool also evaluated the persistence of the indicators using the log-linear model, LLM (Chick, 1908). Descriptions of the equations are in **Table 2**. The T_{90}, and T_{99} values (the time required for 1 and 2 log reductions of the initial concentrations of the indicators, respectively) were calculated by substituting -1.0 (or -2.0 for T_{99}) $= \text{Log}_{10}(N/N_0)$, and solving for t.

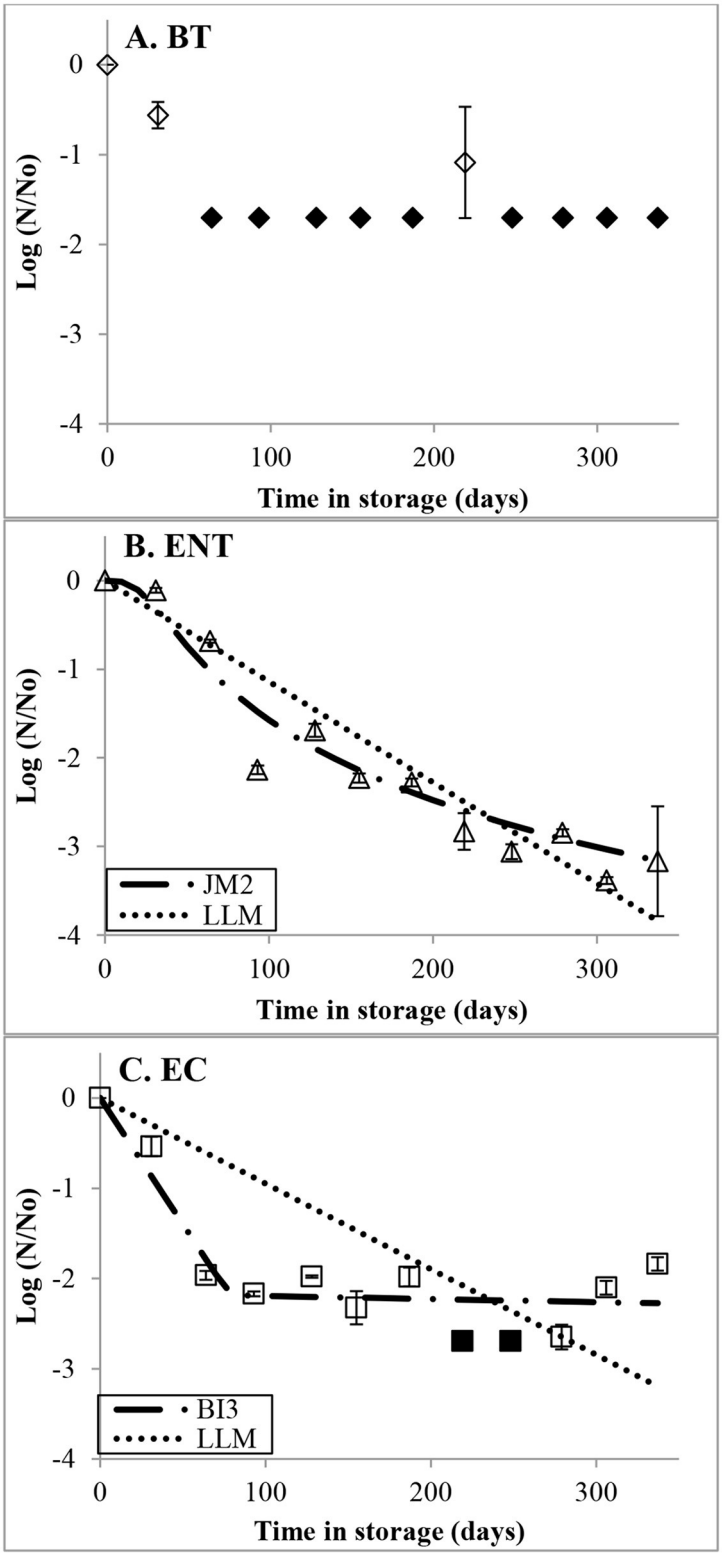


FIGURE 1 | The relative persistence of **(A)** *Bacteroides thetaotiaomicron,* BT; **(B)** enterococci, ENT; and **(C)** *Escherichia coli*, EC measured from microcosms containing sewage spike river water stored at 4°C for up to 337 days. Filled in data points illustrate the data that were below the detection limit and represent the indicator specific detection limit. Each data point represents the average relative persistence of three technical replicates. The error bars represent one standard error. The lines represent the following models: log-linear (LLM), biphasic log-linear (BI3), or log-logistic models (JM2).

TABLE 2 | Descriptions and properties of the log-linear model and best-fit equations (biphasic log-linear, and log-logistic) that mathematically evaluated the relationship of enterococci *23S rDNA*, and *Escherichia coli uidA* from microcosms containing sewage spike river water stored at 4°C for up to 337 days.

Models	Equation†	References
Log-linear (LLM)	$Log_{10}(N/N_0) = Log_{10}(e^{-k_1 t})$	Chick, 1908
Biphasic log-linear (BI3)	for 0 ≤t <x: $Log_{10}(N/N_0) = Log_{10}(e^{-k_1 t})$ for t ≥ x: $Log_{10}(N/N_0) = Log_{10}(e^{-k_1 t + k_2 * (t-x)})$	Carret et al., 1991
Log-logistic model (JM2)	$Log_{10}(N/N_0) = -Log_{10}(1 + e^{K_1 + k_2 * ln(t)})$	Juneja et al., 2003

† *K_n = decay constant; t = time in days; $Log_{10}(N/N_0)$ = relative persistence of the indicators, where N is the indicator concentration after t days, and N_0 is the initial concentration normalized to cell equivalents (CE) per 100 ml-water sample.*

TABLE 3 | Descriptions of the equation parameters, BIC values, predicted T_{90}, and T_{99} values (days) calculated from the best-fit and log-linear models that evaluated the persistence of the cell equivalents of *E. coli* (EC) and enterococci (ENT) measured from microcosms containing sewage spiked river water stored at 4°C for up to 337 days.

Indicator	Model†	BIC value	Predicted T_{90} (days)	Predicted T_{99} (days)	Model parameters
EC	BI3	34.37	35.80	71.59	$k_1 = 0.06$; $k_2 = 0.06$; $bpt\ x = 78.25$
	LLM	53.59	107.01	214.02	$k_1 = 0.02$
ENT	JM2	34.90	62.93	138.94	$k_1 = -10.34$; $k_2 = 3.03$
	LLM	43.80	92.10	184.19	$k_1 = 0.03$

† *The models were: log-linear (LLM); biphasic log-linear (BI3); and log-logistic decay (JM2).*

Multiple linear regression analysis evaluated the associations of the observed relative persistence of the indicators to time in storage and indicator species. The analysis was performed with SPSS 22.0 (SPSS, Inc., Chicago, IL) using the following equation: $Log_{10}(N/N_0) = b_S \times S + b_I \times I + b_o$, where N/N_0 = relative persistence as described previously; $b_0 = \mathrm{Log}_{10}(N/N_0)$ intercept; and b with a letter subscript represented the coefficients of the following independent variables: S = time in storage (0 – 337 days) and I = indicator (BT, EC, and ENT).

RESULTS

Multiple Linear Regression Analysis

The regression equation for the dataset was: $Log_{10}(N/N_0) = 0.006 \times S + 0.307 \times I - 0.146$, with $R^2 = 0.549$ ($p < 0.001$; $n = 108$). Time in storage and indicator species ($p < 0.001$) were significantly associated with the relative persistence of the three indicators. This indicates that the longer the time in storage at 4°C, the greater the decrease in the concentration of the genetic marker compared to initial time. The persistence of each marker was significantly different and BT was the least persistent and ENT was the most persistent.

Description of the Observed Persistence of the Three Molecular Markers

At initial sampling, the average concentrations of BT, EC, and ENT (and standard error) were: 6.81 * 10^4 (5.08 * 10^3), 3.18 * 10^8 (3.99 * 10^7), and 1.01 * 10^7 CE (2.58 * 10^6) per 100 ml- water sample, respectively. The average concentrations of BT and EC were below detection in 9 and 2 time-points, respectively, for the complete data set. The most and least persistent markers were EC and BT, respectively (**Figures 1A–C**). Concentrations of BT were above detection at $t = 0$, 31, and 219 days (**Figure 1A**). The average concentration of BT in one technical replicate at $t = 219$ days was above the detection limit (**Figure 1A**). At $t = 337$ days, the observed relative persistence of BT was below detection representing ≤ -1.17 log decay. Overall, the observed relative persistence of EC and ENT decreased over the duration of the experiment by −1.84 log and −3.17 log, respectively, at $t =$ 337 days (**Figures 1B,C**). The observed relative persistence of EC fluctuated around −2 log during $t = 61$ – 337 days (**Figure 1C**).

Comparison of the Persistence of Three Enteric Markers Predicted From the Best-Fit and Log-Linear Models

Table 3 outlines the parameters of the selected best-fit models, log-logistic (JM2), and biphasic log-linear (BI3), and LLM along with their corresponding BIC values, and T_{90} and T_{99} values. JM2 and BI3 were selected for ENT and EC, respectively.

Non-detects were observed in the EC dataset and affected the decay rates calculated from the best-fit models and LLMs (**Figure 1C**). After 78.25 days in storage, LLM calculated a faster rate of decay for EC than BI3 (**Table 3**). Also, LLM calculated a faster decay rate for ENT after ∼250 days in storage than JM2. For both datasets, LLMs had larger BIC values than the corresponding best-fit models (**Table 3**).

The best-fit models demonstrated that the decay of ENT and EC during the experiment's duration was dynamic (**Figures 1B,C**). During $t = 0$ – 78.25 days, BI3 calculated a rapid decay of EC, afterwards during $t = 78.25$ – 337 days the decay waned (**Figure 1C**, **Table 3**). JM2 estimated that the relative persistence of ENT experienced shouldering followed by rapid decay, and slower decay toward the end of the experiment (**Figure 1B**). The best fit models calculated that the concentrations of EC and ENT at $t = 337$ days relative to the initial concentrations were: −2.27 log and −3.12 logs, respectively, and were larger than the concentrations derived from the corresponding LLMs (−3.15 log and −3.66 log, respectively).

Comparison of the Predicted T_{90} and T_{99} Values From the Best-Fit and Log-Linear Models

The T_{90} and T_{99} values, time required to reduce the initial concentrations of the indicators by 90%, and 99%, respectively, were calculated from the best-fit models and LLMs of the ENT and EC datasets (**Table 3**). The T_{90} and T_{99} values of

ENT calculated from LLM (92.1 and 184.19 days, respectively) were larger than the values from JM2 (62.93 and 138.94 days, respectively). The LLM model of EC predicted that T_{90} and T_{99} = 107.01 and 214.02 days, respectively, and were larger than the values predicted from BI3 (35.80, and 71.59 days, respectively). Based on the T_{90} values calculated from the best-fit models, the relative order of persistence of the indicators was EC < ENT, while the order of persistence based on the T_{90} values from the LLMs was ENT < EC.

DISCUSSION

This is one of the first studies to use experimental data to describe the long-term persistence of naturally occurring *Bacteroides thetataiotaomicron alpha-mannanase,* enterococci *23S rRNA*, and *Escherichia coli uidA* measured from microcosms containing sewage spiked river water stored at 4°C. Using linear regression analysis, the time in storage and indicator species ($p < 0.001$) were significantly related to the persistence of the markers. The order of persistence of the markers was BT < EC < ENT. These results indicate that the persistence of the indicators is specific to the species in a long-term evaluation of persistence. In a 28-day evaluation of persistence of fecal indicator markers in liquid microcosms and storage on solid surface, time in storage was also significantly associated to persistence of the molecular markers of water stored at 4°C while indicator species was not significant (Brooks et al., 2015). The finding in this study and in Brooks et al. (2015) indicate that indicator species significantly affect marker persistence over longer time frames that may not be apparent in shorter time frames such as up to 28 days.

We compared the mathematical relationships between the concentrations of indicators and time in storage (up to 337 days) from log-linear decay models with a constant decay rate to best fit models, biphasic log-linear decay model (BI3) and log-logistic model (JM2). Overall, the best-fit models responded well to changes in decay rates over the course of the experiment including the modeling of effects such as shouldering and slower decay as the experiment progressed (**Figures 1B,C**). Similarly, a meta-analysis of the inactivation of culturable *E. coli* in water determined that the most popular inactivation model was biphasic log-linear decay, with an initial rapid decay followed by slower decay (Blaustein et al., 2013). Another meta-analysis of the inactivation of culturable *E. coli* in water observed that a large portion of the dataset was better represented by "a piecewise log-linear model" (Stocker et al., 2014). Likewise in ENT, inactivation of culturable enterococci and ENT-23 in marine water during winter for up to 8 days demonstrated an initial shouldering of the concentrations of the indicator organisms and markers (Mattioli et al., 2017).

There are a few explanations for the variation of the decay over time. The presence of BT in one technical replicate at $t = 219$ days could be due to a sub-population more resistant to degradation. Another possible scenario is that cells may have attached to a solid surface such as particulate matter in the water, which could enhance persistence. Previously, a simulation of water columns of brackish and freshwater estimated that the attachment of cells of enterococci to a particle can increase its persistence in the water column (Myers and Juhl, 2020). Additionally, the rapid decay of ENT and EC at the beginning of the experiment may be due to the presence of sub-populations with distinct decay rates (Rogers et al., 2011). Specifically, the survival of cultivable *E. coli* in microcosms of filtered estuarine water microcosms was increased in cells from the B1 phylotype (Berthe et al., 2013). DNA fingerprinting of cultivable enterococci isolated from the Lake Superior watershed indicated that sub-populations of *Enterococcus* sp. have differing decay rates and some populations can grow in the watershed (Ran et al., 2013). The rapid decay of the indicators could also be a result of ENT and EC exceeding the carrying capacity of the microcosm, causing rapid decay until the carrying capacity is achieved (Easton et al., 2005). This theory may explain the > 1 log variation of the concentration of EC during $t = 61 – 306$ days. The increase in the concentration of EC during $t = 248 – 337$ days could indicate a subpopulation of *E. coli* that are able to utilize resources to grow in the microcosm to achieve a higher cell density. The above studies suggest that the evaluation of the survival or persistence of FIB in environmental waters may be better represented with water quality models that have dynamic decay rates and scenarios for possible regrowth.

In our study, we compared the T_{90} and T_{99} values calculated from the best-fit models and LLMs. Overall, the T_{90} and T_{99} values from LLMs were larger than from the corresponding best-fit models, indicating that LLMs overestimated the persistence of ENT and EC. This overestimation could delay progress toward restoring impaired environmental waters. Additionally, the T_{90} values calculated from LLMs and best-fit models predicted different relative orders of persistence for the enteric markers. A previous study used LLMs to estimate that genetic markers from sewage sourced ENT and EC seeded into marine and freshwater *in-situ* microcosms had T_{90} values that ranged 51 – 335 days and 21 – 54 days, respectively (Sagarduy et al., 2019). Similarly, in our study, LLMs calculated a similar fining for EC (T_{90} = 107.01 days), and the best-fit models for ENT and EC (T_{90} = 62.93 and 35.8 days, respectively) were within the T_{90} values of the above-mentioned study.

Currently, LLMs are often used to estimate the persistence and loads of fecal indicators in water quality models that forecast restoration of impaired environmental waters. Our findings indicate that subpopulations of ENT and EC are persistent in microcosm at 4°C for up to 337 days (**Figures 1B,C**), which provide further evidence that FIB can persist in water at temperatures ≤ 4°C (Davenport et al., 1976). The data from our study indicate that water quality models should make provisions to estimate the role of sub-populations of persistent FIB in maximum daily load assessments and utilize best-fit models with variable decay rates to better estimate the decay of enteric markers in impaired environmental waters. Such information can guide water quality models into a better understanding of more realistic maximum loads of fecal indicators as well as improve forecasting of the time required to remediate impaired environmental waters. Such improvements in the models can allow for the better protection of public health by

increasing the accuracy of the estimation of pollution present in impaired watersheds.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YMB designed and conducted experiment, analyzed the data, and wrote the manuscript. JBR designed the experiment, provided financial support, guided data analysis, and provided critical feedback in writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was partially funded by the Environmental Protection Agency Great Lakes Restoration Initiative awards #115355 and #3002639641, and the National Science Foundation Water, Sustainability and Climate award (EAR-1039122).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.695207/full#supplementary-material

REFERENCES

Ahmed, W., O'Dea, C., Masters, N., Kuballa, A., Marinoni, O., and Katouli, M. (2019). Marker genes of fecal indicator bacteria and potential pathogens in animal feces in subtropical catchments. *Sci. Total Environ.* 656, 1427–1435. doi: 10.1016/j.scitotenv.2018.11.439

Ballesté, E., Demeter, K., Masterson, B., Timoneda, N., Sala-Comorera, L., and Meijer, W. G. (2020). Implementation and integration of microbial source tracking in a river watershed monitoring plan. *Sci. Total Environ.* 736:139573. doi: 10.1016/j.scitotenv.2020.139573

Berthe, T., Ratajczak, M., Clermont, O., Denamur, E., and Petit, F. (2013). Evidence for coexistence of distinct *Escherichia coli* populations in various aquatic environments and their survival in estuary water. *Appl. Environ. Microbiol.* 79, 4684–4693. doi: 10.1128/AEM.00698-13

Blaustein, R. A., Pachepsky, Y., Hill, R. L., Shelton, D. R., and Whelan, G. (2013). Escherichia coli survival in waters: temperature dependence. *Water Res.* 47, 569–578. doi: 10.1016/j.watres.2012.10.027

Bolker, B. M. (2008). *Ecological Methods and Data in R*. Princeton, NJ: Princeton University Press.

Brooks, Y., Aslan, A., Tamrakar, S., Murali, B., Mitchell, J., and Rose, J. B. (2015). Analysis of the persistence of enteric markers in sewage polluted water on a solid matrix and in liquid suspension. *Water Res.* 76, 201–212. doi: 10.1016/j.watres.2015.02.039

Brooks, Y. M., Spirito, C. M., Bae, J. S., Hong, A., Mosier, E. M., Sausele, D. J., et al. (2020). Fecal indicator bacteria, fecal source tracking markers, and pathogens detected in two Hudson River tributaries. *Water Res.* 171:115342. doi: 10.1016/j.watres.2019.115342

Carret, G., Flandrois, J. P., and Lobry, J. R. (1991). Biphasic kinetics of bacterial killing by quinolones. *J. Antimicrob. Chemother.* 27, 319–327. doi: 10.1093/jac/27.3.319

Chick, H. (1908). An Investigation of the Laws of Disinfection. *J. Hyg. (London)* 8, 92–158. doi: 10.1017/S0022172400006987

Davenport, C. V., Sparrow, E. B., and Gordon, R. C. (1976). Fecal indicator bacteria persistence under natural conditions in an ice-covered river. *Appl. Envir. Microbiol.* 32, 527–536. doi: 10.1128/aem.32.4.527-536.1976

Dick, L. K., Stelzer, E., A., Bertke, E. E., Fong, D. L., and Stoeckel, D. M. (2010). Relative decay of Bacteroidales microbial source tracking markers and cultivated *Escherichia coli* in freshwater microcosms. *Appl. Environ. Microbiol.* 76, 3255–3262. doi: 10.1128/AEM.02636-09

Easton, J. H., Gauthier, J. J., Lalor, M. M., and Pitt, R. E. (2005). Die-off of pathogenic *E. coli* o157:h7 in sewage contaminated waters. *J. Am. Water Resour. Assoc.* 41, 1187–1193. doi: 10.1111/j.1752-1688.2005.tb03793.x

Frahm, E., and Obst, U. (2003). Application of the fluorogenic probe technique (TaqMan PCR) to the detection of *Enterococcus* spp. and *Escherichia coli* in water samples. *J. Microbiol. Methods* 52, 123–131. doi: 10.1016/S0167-7012(02)00150-1

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). The soil and water assessment tool: historical development, applications, and future research directions. *Invit. Rev. Ser.* 50, 1211–1250. doi: 10.13031/2013.23637

Jeong, J., Wagner, K., Flores, J. J., Cawthon, T., Her, Y., Osorio, J., et al. (2019). Linking watershed modeling and bacterial source tracking to better assess *E. coli* sources. *Sci. Total Environ.* 648, 164–175. doi: 10.1016/j.scitotenv.2018.08.097

Johnson, R. C., Imhoff, J. L., Kittle, J. L., and Donigan, A. S. (1984). *Hydrological Simulation Program–Fortran (HSPF), User Manual for Release 8.0, EPA/600/R-97/080*, 755p. Athens, GA: USEPA.

Juneja, V. K., Marks, H. M., Mohr, T., Heiri, O., Lotter, A. F., Lemcke, G., et al. (2003). Predictive thermal inactivation model for effects of temperature, sodium lactate, NaCl, and sodium pyrophosphate on salmonella serotypes in ground beef. *Appl. Environ. Microbiol.* 69, 5138–5156. doi: 10.1128/AEM.69.9.5138-5156.2003

Korajkic, A., McMinn, B. R., Ashbolt, N. J., Sivaganesan, M., Harwood, V. J., and Shanks, O. C. (2019). Extended persistence of general and cattle-associated fecal indicators in marine and freshwater environment. *Sci. Total Environ.* 650, 1292–1302. doi: 10.1016/j.scitotenv.2018.09.108

Mattioli, M. C., Sassoubre, L. M., Russell, T. L., and Boehm, A. B. (2017). Decay of sewage-sourced microbial source tracking markers and fecal indicator bacteria in marine waters. *Water Res.* 108, 106–114. doi: 10.1016/j.watres.2016.10.066

Myers, E. M., and Juhl, A. R. (2020). Particle association of Enterococcus sp. increases growth rates and simulated persistence in water columns of varying light attenuation and turbulent diffusivity. *Water Res.* 186:116140. doi: 10.1016/j.watres.2020.116140

Pote, J., Haller, L., Kottelat, R., Sastre, V., Arpagaus, P., and Wildi, W. (2009). Persistence and growth of faecal culturable bacterial indicators in water column and sediments of Vidy Bay, Lake Geneva, Switzerland. *J. Environ. Sci. (China)* 21, 62–69. doi: 10.1016/S.1001-0742(09)60012-7

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ran, Q., Badgley, B. D., Dillon, N., Dunny, G. M., and Sadowsky, M. J. (2013). Occurrence, genetic diversity, and persistence of enterococci in a Lake Superior watershed. *Appl. Environ. Microbiol.* 79, 3067–3075. doi: 10.1128/AEM.03908-12

Reischer, G. H., Kasper, D. C., Steinborn, R., Farnleitner, A. H., and Mach, R. L. (2007). A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Lett. Appl. Microbiol.* 44, 351–356. doi: 10.1111/j.1472-765X.2006.02094.x

Rogers, S. W., Donnelly, M., Peed, L., Kelty, C., A., Mondal, S., Zhong, Z., et al. (2011). Decay of bacterial pathogens, fecal indicators, and real-time quantitative PCR genetic markers in manure-amended soils. *Appl. Environ. Microbiol.* 77, 4839–4848. doi: 10.1128/AEM.02427-10

Rusiñol, M., Fernandez-Cassi, X., Hundesa, A., Vieira, C., Kern, A., Eriksson, I., et al. (2014). Application of human and animal viral microbial source tracking

tools in fresh and marine waters from five different geographical areas. *Water Res*. 59, 119–129. doi: 10.1016/j.watres.2014.04.013

Sagarduy, M., Courtois, S., Del Campo, A., Garmendia, J. M., and Petrau, A. (2019). Differential decay and prediction of persistence of *Enterococcus* spp. and *Escherichia coli* culturable cells and molecular markers in freshwater and seawater environments. *Int. J. Hyg. Environ. Health* 222, 695–704. doi: 10.1016/j.ijheh.2019.04.011

Soller, J. A., Schoen, M. E., Bartrand, T., Ravenscroft, J. E., and Ashbolt, N. J. (2010). Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Res*. 44, 4674–4691. doi: 10.1016/j.watres.2010.06.049

Srinivasan, S., Aslan, A., Xagoraraki, I., Alocilja, E., and Rose, J. B. (2011). Escherichia coli, enterococci, and Bacteroides thetaiotaomicron qPCR signals through wastewater and septage treatment. *Water Res*. 45, 2561–2572. doi: 10.1016/j.watres.2011.02.010

Stocker, M. D., Pachepsky, Y. A., and Shelton, D. R. (2014). Performance of weibull and linear semi-logarithmic models in simulating inactivation in waters. *J. Environ. Qual*. 43:1559. doi: 10.2134/jeq2014.01.0023

United States (2016). *33 U.S.C. S. § 1252. United States*. Washington, DC: United States Congress.

USEPA (1991). *Technical Support Document for Water Quality-Based Toxics Control*. Washington, DC: USEPA.

USEPA (2012). *Recreational Water Quality Criteria*. Washington, DC: USEPA.

USEPA (2015). *Method 1611.1: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR)*. Washington, DC: USEPA.

Wade, T. J., Calderon, R. L., Brenner, K. P., Sams, E., Beach, M., Haugland, R., et al. (2008). High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. *Epidemiology* 19, 375–383. doi: 10.1097/EDE.0b013e318169cc87

Yampara-Iquise, H., Zheng, G., Jones, J. E., and Carson, C. A. (2008). Use of a Bacteroides thetaiotaomicron-specific alpha-1-6, mannanase quantitative PCR to detect human faecal pollution in water. *J. Appl. Microbiol*. 105, 1686–1693. doi: 10.1111/j.1365-2672.2008.03895.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/frwa.2021.693631

Cross-Validation Indicates Predictive Models May Provide an Alternative to Indicator Organism Monitoring for Evaluating Pathogen Presence in Southwestern US Agricultural Water

*Alexandra Belias[1], Natalie Brassill[2], Sherry Roof[1], Channah Rock[2], Martin Wiedmann[1] and Daniel Weller[1,3,4]**

[1] Department of Food Science, Cornell University, Ithaca, NY, United States, [2] Department of Environmental Science, University of Arizona, Maricopa, AZ, United States, [3] Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States, [4] Department of Environmental and Forestry Biology, SUNY College of Environmental Science and Forestry, Syracuse, NY, United States

OPEN ACCESS

Edited by:
Nynke Hofstra, Wageningen University and Research, Netherlands

Reviewed by:
Petros Kokkinos, Hellenic Open University, Greece
Baiba Vilne, Riga Stradiņš University, Latvia

***Correspondence:**
Daniel Weller
wellerd2@gmail.com

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *11 April 2021*
Accepted: *07 July 2021*
Published: *05 August 2021*

Citation:
Belias A, Brassill N, Roof S, Rock C, Wiedmann M and Weller D (2021) Cross-Validation Indicates Predictive Models May Provide an Alternative to Indicator Organism Monitoring for Evaluating Pathogen Presence in Southwestern US Agricultural Water. Front. Water 3:693631. doi: 10.3389/frwa.2021.693631

Pathogen contamination of agricultural water has been identified as a probable cause of recalls and outbreaks. However, variability in pathogen presence and concentration complicates the reliable identification of agricultural water at elevated risk of pathogen presence. In this study, we collected data on the presence of *Salmonella* and genetic markers for enterohemorrhagic *E. coli* (EHEC; PCR-based detection of *stx* and *eaeA*) in southwestern US canal water, which is used as agricultural water for produce. We developed and assessed the accuracy of models to predict the likelihood of pathogen contamination of southwestern US canal water. Based on 169 samples from 60 surface water canals (each sampled 1–3 times), 36% (60/169) and 21% (36/169) of samples were positive for *Salmonella* presence and EHEC markers, respectively. Water quality parameters (e.g., generic *E. coli* level, turbidity), surrounding land-use (e.g., natural cover, cropland cover), weather conditions (e.g., temperature), and sampling site characteristics (e.g., canal type) data were collected as predictor variables. Separate conditional forest models were trained for *Salmonella* isolation and EHEC marker detection, and cross-validated to assess predictive performance. For *Salmonella*, turbidity, day of year, generic *E. coli* level, and % natural cover in a 500–1,000 ft (∼150–300 m) buffer around the sampling site were the top 4 predictors identified by the conditional forest model. For EHEC markers, generic *E. coli* level, day of year, % natural cover in a 250–500 ft (∼75–150 m) buffer, and % natural cover in a 500–1,000 ft (∼150–300 m) buffer were the top 4 predictors. Predictive performance measures (e.g., area under the curve [AUC]) indicated predictive modeling shows potential as an alternative method for assessing the likelihood of pathogen presence in agricultural water. Secondary conditional forest models with generic *E. coli* level excluded as a predictor showed <0.01 difference in AUC as compared to the AUC values for the original models (i.e., with generic *E. coli* level included

as a predictor) for both *Salmonella* (AUC = 0.84) and EHEC markers (AUC = 0.92). Our data suggests models that do not require the inclusion of microbiological data (e.g., indicator organism) show promise for real-time prediction of pathogen contamination of agricultural water (e.g., in surface water canals).

Keywords: *Salmonella, E. coli*, agricultural water, Arizona, produce safety, predictive modeling

INTRODUCTION

Salmonella spp. and pathogenic *Escherichia coli* (such as enterohemorrhagic *E. coli*; EHEC) are common etiological agents of foodborne outbreaks and recalls linked to produce commodities. From 2004 to 2012, *Salmonella* has caused 71 and 40 outbreaks linked to produce in the United States (US) and European Union (EU), respectively (Callejon et al., 2015). During the same time frame, pathogenic *E. coli* have caused 46 and 7 outbreaks linked to produce in the US and EU, respectively (Callejon et al., 2015). Both livestock and wildlife have been identified as possible sources of *Salmonella* and EHEC in preharvest produce environments (Delaquis et al., 2007; Hanning et al., 2009), with fecal matter from livestock operations being a known source of foodborne pathogens in surface water (Lu et al., 2004; Delaquis et al., 2007; Hanning et al., 2009). As such, application of surface waters to in-field produce has been identified as a potential route for pathogen contamination of produce (Mootian et al., 2009; Park et al., 2012; Castro-Ibanez et al., 2015; Liu et al., 2018). In fact, several outbreaks are thought to have been caused by application of contaminated water to preharvest produce (U. S. Centers for Disease Control Prevention, 2006, 2018a,b; Greene et al., 2008), including a 2018 *E. coli* O157:H7 outbreak linked to romaine lettuce grown in Arizona, which caused 210 illnesses and 5 deaths (U. S. Centers for Disease Control Prevention, 2018b).

While enteric pathogens, such as EHEC and *Salmonella*, are known surface water contaminants, they are present sporadically and at low levels, complicating detection and limiting the value of testing surface water for pathogens (Jamieson et al., 2004; Pachepsky et al., 2011). Instead, monitoring programs often test for indicator organisms, which are used to assess the hygienic quality of water and the likelihood of fecal contamination; indicator organisms include enterococci, fecal coliforms, and generic *E. coli* (Jamieson et al., 2004; Pachepsky et al., 2011). *E. coli* is used as an indicator for monitoring fecal contamination in agricultural water by industry agreements and by government regulations (Tam and Petersen, 2014; U. S. Food Drug Administration, 2015; California Leafy Greens Marketing Agreement, 2020; Freshcare, 2020); see **Table 1** for details.

There are however several drawbacks to using generic *E. coli* as an indicator of fecal contamination for surface water (Pachepsky et al., 2016; Wall et al., 2019). While some studies have established associations between generic *E. coli* levels and pathogen presence (Holvoet et al., 2014; Lopez-Galvez et al., 2014; Stea et al., 2015; Bradshaw et al., 2016; Truchado et al., 2018; Weller et al., 2020b), several studies have not shown similar associations (Haley et al., 2009; Shelton et al., 2011; Benjamin et al., 2013; Falardeau et al., 2017). *E. coli* has also been shown to survive for extended periods of time in the preharvest environment (Lu et al., 2004; Franz et al., 2014; Allende et al., 2018); therefore, a high level of generic *E. coli* does not necessarily indicate recent fecal contamination. High levels of indicator organisms also do not necessarily indicate the presence of pathogens, and alternatively, the absence or low levels of indicator organisms do not necessarily indicate the absence of pathogens (Haley et al., 2009; Shelton et al., 2011; Benjamin et al., 2013; Stea et al., 2015; Falardeau et al., 2017). Lastly, generic *E. coli* testing takes ∼24 h to complete and generic *E. coli* levels in surface waters can vary substantially over short time periods (Lothrop et al., 2018; Weller et al., 2020b); therefore, it is impossible to know the generic *E. coli* level in irrigation water at the time of its application.

Two previous studies have proposed the use of machine learning models for predicting pathogen presence in agricultural water; Weller et al. (2020c) utilized machine learning models to predict *Salmonella* presence and EHEC marker detection in New York streams and Polat et al. (2020) utilized machine learning models to predict *Salmonella* presence in Florida ponds. While previous studies have explored the use of machine learning models for predicting pathogen contamination in specific produce growing areas, further model development is essential to verify that machine learning represents a viable approach in different locations and types of surface waters. We thus collected data on surface water quality and used several approaches, including machine learning, to model the presence enteric pathogens in southwestern US canal water. Regression was used as a preliminary assessment to determine which variables were associated with pathogen presence. Conditional forest models were used for prediction because they can utilize large numbers of predictors and better able to handle complex and messy data than regression models (Kuhn and Johnson, 2013). While modeling alone will not improve the safety of produce, these models can indicate when corrective actions (e.g., water treatments) should be applied to reduce the risk of recalls and illnesses associated with produce (Savichtcheva et al., 2007; Allende and Monaghan, 2015). The southwestern US was selected for this study, because (i) it is a major produce growing region, (ii) there has been a high-profile outbreak associated with romaine lettuce contamination linked to irrigation water, and (iii) there is limited information on microbial quality of southwestern US canals (Lothrop et al., 2018; Weller et al., 2020b). As such, the specific objectives of this study were to (i) identify land use, water quality, weather, and other sampling site specific variables associated with *Salmonella* presence and EHEC marker detection (i.e., *stx* and *eaeA* detection) in southwestern US canal water, (ii) determine the feasibility of predicting the

TABLE 1 | Microbial quality requirements or regulations established in different countries or regions for agricultural water applied to pre-harvest produce.

Region	Organization	Requirement	References
United States[a]	US FDA	20 water samples over a 2–4-year period must be collected: (i) geometric mean of <126 CFU generic *E. coli*/ 100 ml and (ii) a statistical threshold value (i.e., the 90th percentile) of <410 CFU generic *E. coli*/ 100 mL	21 C.F.R. § 112.44, 2019
Europe	European Union	<100 CFU generic *E. coli*/100 ml of water[c]	European Commission, 2017
British Columbia, Canada	British Columbia Ministry of Agriculture	<77 CFU generic *E. coli*/100 ml of water[c]	Tam and Petersen, 2014
Australia	Freshcare	<100 CFU generic *E. coli*/ 100 ml of water[c]	Freshcare, 2020
California and Arizona (leafy greens only)[b]	Leafy Greens Marketing Agreement (LGMA)	*Furrow irrigation or overhead irrigation applied >21 days prior to harvest:* 100 ml of water should be collected at least monthly and the rolling geometric mean of the generic *E. coli* levels in the 5 most recent samples must be <126 CFU/100 ml and no sample may have an *E. coli* level >576 CFU/100 ml. *Overhead irrigation applied <21 days prior to harvest:* generic *E. coli* should not be detected in the water.	California Leafy Greens Marketing Agreement, 2020

[a] *As of January 2021, this policy is not being enforced and is currently under review (U. S. Food Drug Administration, 2015). This requirement is established under the Food Safety Modernization Act (FSMA).*
[b] *This is a voluntary agreement.*
[c] *No sampling scheme for water collection is specified in this regulation.*

likelihood of *Salmonella* presence and EHEC marker detection in southwestern US canal water, and (iii) determine if only real-time variables (i.e., no microbial testing) can be used to predict the likelihood of *Salmonella* presence and EHEC marker detection.

MATERIALS AND METHODS

Experimental Design

A longitudinal study was conducted to assess agricultural water quality in the southwestern US. Water was sampled from 60 canals that provide water for irrigation from January 30th to November 19th, 2018; sampling was performed approx. twice a week every 1–2 weeks except in July when no sampling was performed (see **Supplementary Figure 1** for the exact sampling dates). Each canal was sampled 1–3 times for a total of 169 samplings; all samples from a given canal were collected at the same site. Sampling sites were randomly selected from irrigation districts where produce was grown and where permission was given using ArcGIS. Ground truthing was then performed to identify a location as close as possible to the randomly generated GPS coordinates for each site. A site survey was conducted to collect information on features present at a given site (see **Supplementary Table 1**).

Sample Collection and Processing

At each sampling, two 10 L water samples (one per pathogen) and 1 L of water (for enumeration of *E. coli* and turbidity levels) were collected. Dissolved oxygen, pH, conductivity, and temperature of the canal water were measured using a Hach HQ40d meter (Loveland, CO, United States). Water surface flow was measured using the float method as described by Gore and Banning (2017). After collection, all samples were put on ice until processing. The 10 L samples were processed <18 h after collection and the 1 L sample for *E. coli* level and turbidity was processed <6 h after collection.

Laboratory testing of all samples was performed as described in Weller et al. (2020b). Briefly, generic *E. coli* enumeration was performed on a 100 ml aliquot of the 1 L sample using the Colilert Quanti-Tray 2000 kit (IDEXX, Westbrook, ME, United States), according to the manufacturer's instructions. Water turbidity was measured using the Hach 2100Q Portable Turbidimeter. The 10 L water samples were processed using the modified Moore swab (mMS) method (Sbodio et al., 2013). Each water sample was gravity-filtered through a separate mMS, placed in a separate Whirl-Pak bag (Nasco, Fort Atkinson, WI, United States), and processed as described below for either *Salmonella* presence or EHEC marker detection.

Salmonella Isolation

Briefly, 225 ml of buffered peptone water with 20 mg/L novobiocin (BPW+N) was added to the Whirl-Pak bag with the mMS, followed by incubation at 35°C for 24 h. BPW+N inoculated with *Salmonella* Typhimurium (FSL F6-0826; http://www.foodmicrobetracker.com) and uninoculated BPW+N were included as positive and negative controls, respectively. Following incubation, a 1 ml aliquot of the enrichment was transferred to a sterile tube and was shipped overnight on ice to Cornell University for further processing. Upon arrival, all enrichments were used within 2 h for a *Salmonella* screen using the BAX real-time *Salmonella* assay (Hygiena, Wilmington, DE, United States). BAX PCR-positive samples were culture confirmed. 1 and 0.1 ml of the BPW+N enrichment were added to 9 ml of tetrathionate broth (TT; Oxoid) supplemented with 200 μl of I2-KI and 100 μl of Brilliant Green and 9.9 ml of Rappaport Vassiliadis broth (RV; Acros Organic, Geel, Belgium), respectively. The TT and RV broth were incubated in a shaking water bath at 42°C for 24 h. Following

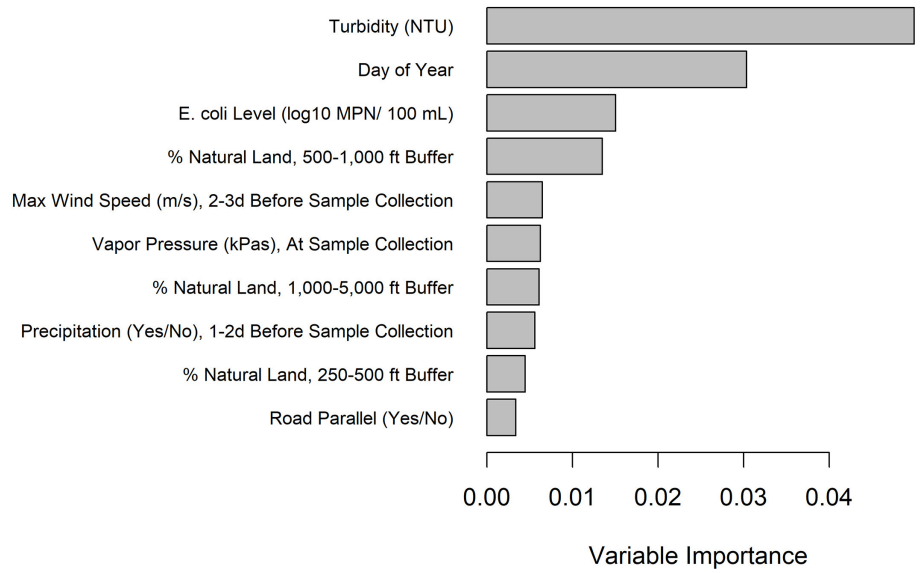


FIGURE 1 | Variable importance values for the *Salmonella* conditional forest model with generic *E. coli* included as a predictor. Only the top 10 predictors are included and are listed from most to least important.

incubation, 50 μl of each broth were streaked onto *Salmonella* CHROMagar (DRG International, Springfield, NJ, United States) and xylose lysine deoxycholate agar (XLD; Neogen, Lansing, MI, United States) plates (i.e., 1 plate for TT on CHROMagar, 1 plate for TT on XLD, 1 plate for RV of CHROMagar, and 1 plate for RV on XLD). The CHROMagar and XLD plates were then incubated at 37 and 35°C, respectively, for 24 h. After incubation, PCR of the *invA* gene was performed on presumptive *Salmonella* colonies according to the protocol described by Kim et al. (2007). If possible, 2 characteristic *Salmonella* colonies per media type (mauve colonies on CHROMagar and black colonies on XLD) were selected for PCR (4 colonies in total). If no characteristic colonies were present, up to 12 non-characteristic colonies were selected for PCR (blue colonies on CHROMagar and red colonies on XLD). All isolates were stored as 15% glycerol stocks at −80°C.

EHEC Marker Detection

PCR-based detection of *stx* and *eaeA* from the mMS was performed. 225 ml of tryptic soy broth with 10 g/L casamino acids and 8 mg/L of novobiocin (TSB+N) was added to the Whirl-Pak bag with the mMS, followed by incubation at 41°C for 24 h. TSB+N inoculated with *E. coli* O157:H7 (FSL F6-0699; http://www.foodmicrobetracker.com) and uninoculated TSB+N were included as positive and negative controls, respectively. Following incubation, a 1 ml aliquot of the enrichment was shipped overnight on ice to Cornell University for further processing. All enrichments were used within 2 h of arrival to perform a PCR screen using the BAX real-time Shiga-toxin producing *E. coli* (STEC) assay (Hygiena) according to the manufacturer's instructions to determine if the *eaeA* and/or *stx1/2* genes were present in the sample. If both *eaeA* and *stx1/2* were detected in a sample, the sample was classified as positive for "EHEC markers." However, such results could indicate either (i) both genes were present in a single organism (indicating presence of EHEC) or (ii) genes were present in separate organisms (e.g., *eaeA* indicates enteropathogenic *E. coli* presence, *stx1/2* indicates STEC presence).

Land Use Data Collection

Land use data around the sampling sites were extracted from the 2016 National Land Cover Database (NLCD; https://www.mrlc.gov/) and quantified using ESRI ArcGIS Pro 2.4.0. The percentage of land under (i) developed open space, (ii) developed (combines low-, medium-, and high-intensity developed cover), (iii) barren, (iv) natural (combines forest and wetland), (v) pasture/ hay, and (vi) crop cover at various intervals around each sampling site were calculated (Yang et al., 2018). The intervals considered were: <250 ft (< ∼75 m), 250–500 ft (∼75–150 m), 500–1,000 ft (∼150–300 m), 1,000–5,000 ft (∼300–1,525 m), and 500–10,000 ft (∼1,525–3,050 m). These buffer areas were selected, as they most closely represent the distances included in the California and Arizona Leafy Green Marketing Agreements Food Safety Practices (California Leafy Greens Marketing Agreement, 2020) metrics; while it would have been useful to characterize land use directly adjacent to the canals, an accurate map of the canal networks was not available for the study area. The number of concentrated animal feeding operations (i.e., an animal feeding operation with >1,000 animal units confined on a site for more than 45 days of the year; CAFOs) within 10,000 ft of each site was also calculated.

Weather Data Collection

Temperature, solar radiation, precipitation, wind speed, and vapor pressure data were obtained from the University of Arizona (cals.arizona.edu/AZMET/). ESRI ArcGIS Pro 2.4.0 was used

to identify the weather station closest to each of the sampling sites. Weather data were cleaned in R version 4.0.0 (R Core Team, 2021) and used to calculate weather at the time of sample collection and for the (i) 0–12 h, (ii) 12–24 h, (iii) 1–2 days, and (iv) 2–3 days prior to sample collection. Due to the small amount of precipitation during the study, all precipitation variables were converted to a binary factor to indicate if there was precipitation (>0 mm) or if there was no precipitation (=0 mm).

Regression Analysis

All data cleaning, visualization, and analyses were performed in R (R Core Team, 2021). A description of all variables used in analyses are provided in **Supplementary Table 1**. All analyses were performed separately for *Salmonella* presence and EHEC marker detection. Logistic regression was used as a preliminary assessment to characterize associations between site specific (i.e., data on features present at each site, see **Supplementary Table 1** for details), water quality, land-use, and weather variables and *Salmonella* presence and EHEC marker detection. Conditional forest analysis was used to determine if these variables could be used to predict *Salmonella* presence or EHEC marker detection.

For logistic regression, normalization and scaling of all numeric variables was performed using the "caret" package (Kuhn, 2020). Univariable logistic regression was performed, using the "lme4" package (Bates et al., 2015), to determine which of the explanatory variables listed in **Supplementary Table 1** were associated with *Salmonella* and EHEC marker presence. The day of year (number of days since Jan 1st) and irrigation district were included in each univariable model as random effects to account for temporal and spatial autocorrelation. Following univariable regression, continuous variables with $P<0.1$ were included in a principal component analysis (PCA) for variable reduction. PCA was performed using the prcomp function, such that the number of components retained must explain ≥90% of the variation in the data and each retained variable could only have major loading on one principal component. PCA was performed separately for *Salmonella* and EHEC marker presence. One representative continuous variable from each principal component, as well as all categorical variables significant at $P < 0.1$ by univariable analysis (categorical variables cannot be included in PCA) were included in the initial multivariable logistic regression models (implanted using the "lme4" package; Bates et al., 2015). Day of year and irrigation district were included each multivariable model as random effects. Backwards selection based on AIC (Akaike Information Criterion) was performed; the final selected model was the simplest model with an AIC value that was at least 2 less than the next simplest model. Model fit was assessed using the protocol described by Beauvais et al. (2018) to determine if model assumptions were met. Variance inflation factors were also calculated to test for multicollinearity.

Conditional Forest Analysis

Conditional forest analysis was used to determine if sampling site, water quality, land-use, and weather variables could be used to predict *Salmonella* presence and EHEC marker detection, as it can handle missing data, skewed data, and is robust to small sample sizes. Imputation could not be performed because >10% of observations were missing for some variables, and so imputation could introduce bias into the results. No additional machine learning algorithms were tested, as a comprehensive comparison of 23 learners for predicting enteric pathogen presence in New York streams found that conditional forest models performed well for the type of data used in the study presented here (Weller et al., 2020c). The "mlr" (Bischl et al., 2016) and "party" (Hothorn et al., 2006; Strobl et al., 2007, 2008) packages were used for model training and testing. Oversampling was performed to account for imbalanced training data. Repeated (5 iterations) 5-fold cross-validation was used to tune hyperparameters (i.e., mtry, minbucket, and mincriterion) to maximize AUC (area under the curve) and minimize overfitting. For each forest, 20,001 trees were fit. Following hyperparameter tuning, models were trained, and model testing was performed. While a separate testing data set would have been preferable to better evaluate the predictive performance of these models, one was not available. Instead, cross-validation was performed as part of model training to estimate performance measures. Variable importance scores were calculated (Strobl et al., 2007, 2008) and partial dependence plots were fit for the 4 top-ranked variables for each pathogen. While using conditional importance scores would have been preferable to account for correlation between variables, it could not be calculated due to missing data (>10% of observations were missing for some variables); as a result, the variable importance scores reported here may be biased by this correlation. Even with this limitation, we determined conditional forest was a good option in this case due to its ability to handle a large number of predictors and small sample sizes.

We also evaluated if the inclusion of generic *E. coli* levels as an input variable would substantially improve the performance of the conditional forest models. To do so, separate conditional forest models were re-run (one per pathogen) as described above but excluding generic *E. coli* level as a predictor. Performance measures were used to compare the models that included and excluded *E. coli* levels.

All models presented here, were developed as a proof of concept. As such, these models should not be used to guide on-farm decision making, and instead should be used as a starting point for the development of field-ready models (i.e., that can be used by stakeholders to predict pathogen presence in agricultural water) as part of future studies (e.g., using larger datasets, validated using an independent test dataset).

RESULTS

General Water Quality

In total, 169 samples were collected between January 30th and November 19th, 2018 from 60 canals; each canal is referred to as a "site." The sites were within an ∼28,000 km^2 area representing 9 irrigation districts. On average, the majority of land in the 10,000 ft surrounding the sites was classified as cropland or natural cover (**Supplementary Table 3**). The mean generic *E. coli* level in the water samples was 1.4 $\log_{10}$ MPN/100 ml (standard deviation = 0.7 $\log_{10}$ MPN/100 ml) and the mean turbidity in

the water samples was 32.7 NTU (standard deviation = 92.7 NTU); similar statistics for other variables can be found in **Supplementary Tables 2–4**.

Pathogen Testing Results

The overall prevalence of *Salmonella* was 36% (60/169). Of the 60 sites, 20 were *Salmonella* negative on all samplings, however, 2 of those sites were only sampled once. Conversely, 9 sites were positive on all samplings; (1, 5, and 3 of these sites were sampled 1, 2, and 3 times, respectively). The overall prevalence of EHEC markers (i.e., both *stx1/2* and *eaeA*) in the water samples was 21% (36/169). Thirty-three sites were negative for EHEC markers on all samplings, however, 3 of those sites were only sampled once. Conversely, 2 sites were positive on all samplings for EHEC markers (both sites were sampled 3 times).

Regression and Conditional Random Modeling of *Salmonella* Contamination

Multivariable regression was used to determine which variables were associated with *Salmonella* presence. The percent of developed open space (e.g., large-lot single family homes, golf courses, parks) in the 1,000–5,000 ft buffer area and precipitation 1–2 days before sample collection were the only variables retained in the final model. The percent of developed open space in the 1,000–5,000 ft buffer area was negatively associated with the log-odds of a sample being *Salmonella*-positive ($P = 0.036$, **Table 2**). Rain 1–2 days before sample collection was positively associated with the log-odds of a *Salmonella* positive sample ($P = 0.015$, **Table 2**). Results of univariable regression and PCA are detailed in **Supplementary Table 5**.

In addition to multivariable regression, we performed conditional forest analysis to predict *Salmonella* presence. The 10 top-ranked predictors included (i) three predictors related to natural cover, (ii) three weather-related predictors, (iii) two water quality related predictors, (iv) one temporal predictor, and (v) one sampling site predictor (**Figure 1**). Only "precipitation 1–2 days before sampling," was also retained in the final multivariable regression model and included in the 10 top-ranked predictors by the forest (**Figure 1**). While not retained in the multivariable regression analysis, several of the 10 top-ranked predictors in the conditional forest were significant according to univariable regression (**Supplementary Table 5**). Given that conditional forest is better able to handle complex (e.g., interactions between features) and messy (e.g., missing data) data than regression, these differences are not unexpected (see Weller et al. 2020c for more information). Partial dependence plots were fit to visualize the relationship between the 4 top-ranked predictors in the conditional forest model (**Figure 2**), which were, in order, turbidity, day of year, generic *E. coli* level, and percent natural cover in the 500–1,000 ft buffer area.

The AUC (area under the curve) and kappa score for the *Salmonella* conditional forest model were 0.84 and 0.51, respectively (**Table 3**). When the probability threshold was set to 0.5 (i.e., to label a sample as positive, the predicted probability of that sample being positive for *Salmonella* must be 0.5 or greater), the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 0.69, 0.82, 3.83, and 0.38, respectively. The sensitivity of 0.69 indicates there is a 0.31 false negative rate, or 31% of the time the model will predict a sample as being negative for *Salmonella* when it is truly positive. Since the model predicts the probability of *Salmonella* being present in a sample (i.e., a continuous outcome), a probability threshold is needed to dichotomize the predicted pathogen status as positive or negative (**Table 2**). If a binary outcome (as opposed to the continuous outcome generated by the forest algorithm) is needed when applying a predictive model, sensitivity and specificity can be adjusted by changing the threshold value (**Table 4**). For instance, if the threshold value was set at 0.4, the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio would be 0.78, 0.71, 2.70, and 0.31, respectively (**Table 4**).

Regression and Conditional Random Modeling of EHEC Marker Presence

Generic *E. coli* level, precipitation 12–24 h before sample collection, if there was a point of discharge (i.e., there was ground water well discharge into the canal) visible from the site, and if there was a road crossing visible from the site were retained in the final EHEC regression model (**Table 2**). Generic *E. coli* level ($P < 0.001$) and precipitation 12–24 h before sample collection ($P = 0.007$) were positively associated with the log-odds of EHEC marker detection ($P < 0.001$). A point of discharge and a road crossing visible from the site were negatively associated with the log-odds of EHEC marker detection.

A conditional forest model was also fit to predict EHEC marker detection. The top 10 ranked predictors included (i) six land cover predictors (five related to natural cover), (ii) one weather predictor, (iii) two water quality predictors, and (iv) one temporal predictor (**Figure 3**). While generic *E. coli* level and precipitation 12–24 h before sample collection were included in the 10 top-ranked predictors in the forest model, the presence of a point of discharge and road crossing adjacent to the sampling sites were not among the 10 top-ranked predictors in the forest (**Figure 3**). For the EHEC forest, partial dependence plots were fit for the 4 predictors, which were (i) generic *E. coli* level, (ii) day of year, (iii) percent of natural cover in the 250–500 ft buffer area, and (iv) percent of natural cover in the 500–1,000 ft buffer area (**Figure 4**).

The AUC and kappa score for the EHEC forest were 0.92 and 0.66, respectively (**Table 3**). The sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 0.78, 0.88, 6.27, and 0.25, respectively for the EHEC marker model at a threshold value of 0.5 (**Table 4**). As with the *Salmonella* forest, changing the threshold value could improve performance measures that rely on dichotomizing the predicted probability of EHEC marker detection (e.g., sensitivity; **Table 4**).

Performance of Models That Do Not Include Generic *E. coli* Levels as a Predictor

To determine if including generic *E. coli* levels in the conditional forest models substantially improved predictive performance,

TABLE 2 | Results of mixed effects regression models[a] that characterize the relationship between pathogen detection (*Salmonella* presence and EHEC marker detection) and environmental variables (e.g., land use, weather, sampling site characteristics, and water quality factors).

Outcome	Variable[b]	Log odds	95% CI[c]
Salmonella presence	Intercept	−0.9	(−3.2, 0.8)
	% Developed (open) Cover, 1,000–5,000 ft Buffer	−0.5	(−1.0, −0.1)
	Precipitation, 1–2 days (Yes)[d]	3.5	(1.0, 7.0)
EHEC marker	Intercept	0.3	(−2.3, 1.1)
	E. coli level (log_{10} MPN/ 100 ml)	1.7	(1.0, 2.6)
	Point of Discharge Present (Yes)[e]	−1.8	(−4.0, −0.1)
	Road Crossing Present (Yes)[f]	−1.8	(−3.5, −0.3)
	Precipitation, 12–24 h (Yes)[d]	3.7	(1.2, 6.7)

[a] *The day of year and irrigation district were included in the models as random effects.*
[b] *For the Salmonella presence model, the residual variance and standard deviation for the day of year are 0.7302 and 0.8545, respectively, and the residual variance and standard deviation for the irrigation district are 2.9023 and 1.7036, respectively. For the EHEC maker model, the residual variance and standard deviation for the day of year are 0.8895 and 0.9431, respectively, and the residual variance and standard deviation for the irrigation district are 0.1269 and 0.3562, respectively.*
[c] *95% CI, 95% confidence interval.*
[d] *Indicates if there was precipitation in the time frame specified before sample collection. Baseline is no precipitation.*
[e] *Indicates if a point of discharge (i.e., ground water well discharge into the canals) is present adjacent to the sampling site. Baseline is no point of discharge.*
[f] *Indicates if a road crossing is present adjacent to the sampling site. Baseline is no road crossing.*

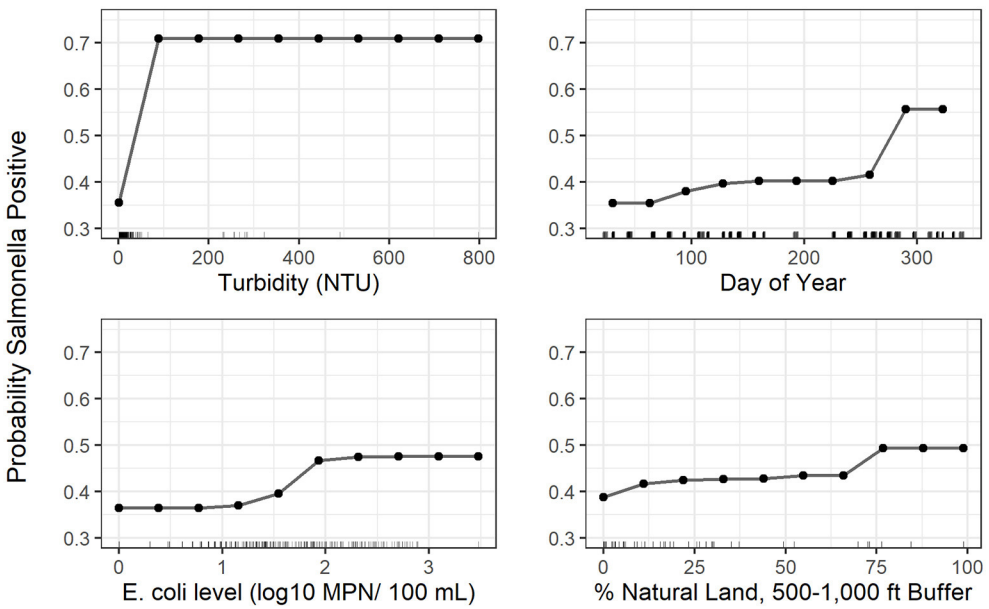


FIGURE 2 | Partial dependence plots for the 4 top-ranked predictors according to variable importance in the *Salmonella* conditional forest model with generic *E. coli* included as a predictor; The plots indicate how the predicted probability of a water sample being positive for *Salmonella* presence changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.

we re-ran the forest models without generic *E. coli* level as a predictor (see **Supplementary Figures 2–5** for variable importance and partial dependence plots). There were no substantial differences in performance between the models with and without generic *E. coli* included as a predictor (**Table 3**). For instance, the AUC values were 0.84 and 0.83 for the *Salmonella* forests that included and excluded generic *E. coli* levels, respectively. Similarly, the AUC values were 0.92 and 0.91 for the EHEC forests that included and excluded generic *E. coli* levels, respectively.

DISCUSSION

The current study assessed *Salmonella* presence and EHEC marker detection in southwestern US canals used for agricultural water. Regression was used to identify associations between environmental conditions and *Salmonella* presence and EHEC marker detection. The data were also utilized to determine if forest-based models were a feasible approach for predicting *Salmonella* presence and EHEC marker detection in canals. While these models were developed as a proof of concept,

TABLE 3 | Performance measures for the conditional random forest models displaying the relationship of the pathogen detection outcomes (*Salmonella* presence and EHEC marker detection) with the land use, weather, sampling site characteristics, and water quality predictors with generic *E. coli* level included and excluded as a predictor.

Outcome	*E. coli*[a]	AUC[b]	Kappa[c]	Sensitivity[d]	Specificity[d]	LR+[d,e]	LR-[d,f]	DOR[g] (95% CI[h])
Salmonella	Yes	0.84	0.51	0.69	0.82	3.83	0.38	10.06 (7.38, 13.70)
	No	0.83	0.49	0.69	0.80	3.44	0.39	8.82 (6.52, 11.94)
EHEC marker	Yes	0.92	0.66	0.78	0.88	6.27	0.25	25.19 (18.27, 34.72)
	No	0.91	0.63	0.78	0.85	5.25	0.26	20.54 (15.07, 27.98)

[a] *Indicates if E. coli level (log_{10}MPN/100 ml water) was included (Yes) or not (No) as a variable in the forest.*
[b] *AUC, area under the receiver operating characteristic curve.*
[c] *Kappa score (a measure of agreement between the observed outcome and the predicted outcome; a value of 1 is indicative of perfect agreement and a value of 0 is indicative of an agreement no greater than that of chance).*
[d] *Measure is biased by the decision threshold used.*
[e] *LR+, positive likelihood ratio (the likelihood of a predicted pathogen presence when a pathogen is present compared to the likelihood of a predicted pathogen presence when a pathogen is absent).*
[f] *LR–, negative likelihood ratio (the likelihood of a predicted pathogen absence when a pathogen is present compared to the likelihood of a predicted pathogen absence when a pathogen is absent).*
[g] *DOR, diagnostic odds ratio (the ratio of the odds of a predicted pathogen presence if the pathogen is present to the odds of a predicted pathogen presence if the pathogen is absent).*
[h] *95% CI, 95% confidence interval.*

TABLE 4 | Differences in performance measures for the conditional forest models displaying the relationship of the pathogen detection outcomes (*Salmonella* presence and EHEC marker detection) with the land use, weather, sampling site characteristics, and water quality predictors with generic *E. coli* level included and excluded as a predictor[a].

Outcome	*E. coli*[b]	Threshold	Sensitivity	Specificity	LR+[c]	LR–[d]
Salmonella presence	Yes	0.5	0.69	0.82	3.83	0.38
		0.4	0.78	0.71	2.70	0.31
		0.3	0.86	0.58	2.07	0.23
	No	0.5	0.69	0.80	3.44	0.39
		0.4	0.78	0.72	2.82	0.30
		0.3	0.87	0.61	2.23	0.21
stx/ eaeA co-detection	Yes	0.5	0.78	0.88	6.27	0.25
		0.4	0.83	0.82	4.72	0.20
		0.3	0.93	0.73	3.40	0.10
	No	0.5	0.78	0.85	5.25	0.26
		0.4	0.88	0.75	3.50	0.16
		0.3	0.95	0.65	2.67	0.08

[a] *Threshold value indicates the predicted probability a sample must be greater than to be labeled as a positive sample.*
[b] *Indicates if E. coli level (log_{10}MPN/100 ml water) was included (Yes) or excluded (No) in the model as a possible predictor.*
[c] *LR+, positive likelihood ratio (the likelihood of a predicted pathogen presence when a pathogen is present compared to the likelihood of a predicted pathogen presence when a pathogen is absent).*
[d] *LR–, negative likelihood ratio (the likelihood of a predicted pathogen absence when a pathogen is present compared to the likelihood of a predicted pathogen absence when a pathogen is absent).*

they provide a conceptual framework on which future work (development of models that can be integrated into on-farm decision-making) can build. Our results can also be used to identify factors important for predicting pathogen presence in southwestern US canal water to guide future data collection to be used to provide maximum value for the refinement of predictive models that can be deployed for industry use.

Salmonella and EHEC Marker Prevalence

Salmonella has been isolated from flowing surface water sources in both this and previous studies (Duffy et al., 2005; Haley et al., 2009; Wilkes et al., 2009, 2011; Benjamin et al., 2013; Strawn et al., 2013a,b; Cooley et al., 2014; Stea et al., 2015; Bradshaw et al., 2016; Falardeau et al., 2017; Tian et al., 2017; Partyka et al., 2018; Truchado et al., 2018; Weller et al., 2020a,b). While *Salmonella* prevalence varied widely between these studies, the *Salmonella* prevalence reported here (36%) falls within the range reported by these previous studies, which was between 3% (6/223) (British Columbia, Canada; Falardeau et al., 2017) and 76% (80/105) (Georgia, USA; Bradshaw et al., 2016).

While several studies have attempted to assess the prevalence of EHEC or different EHEC subgroups (e.g., *E. coli* O157) in surface water, the specific methodologies used can have a considerable impact on prevalence estimates. Some studies reported the percent of culture-confirmed EHEC or STEC positive samples (Wilkes et al., 2009, 2011; Benjamin et al.,

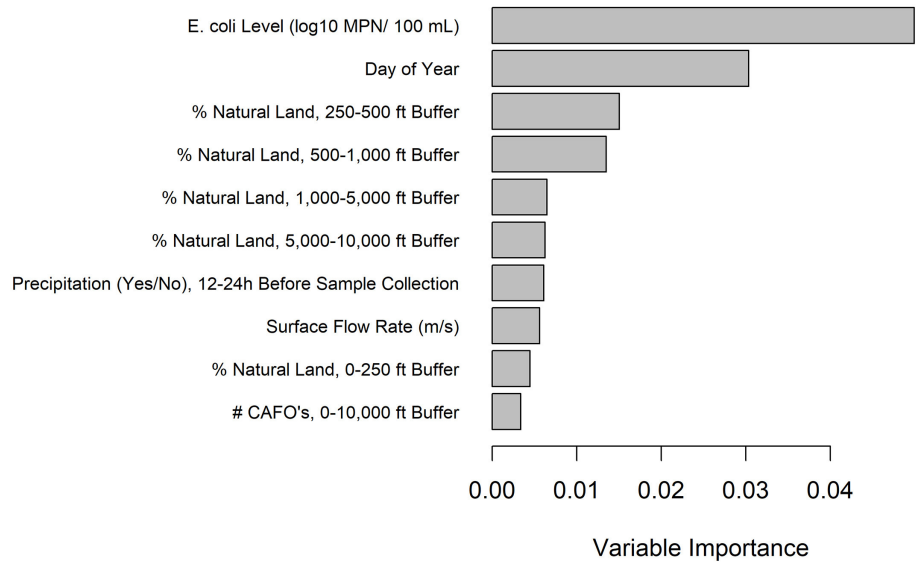


FIGURE 3 | Variable importance values for the EHEC marker conditional forest model with generic *E. coli* included as a predictor. Only the top 10 predictors are included and are listed from most to least important.

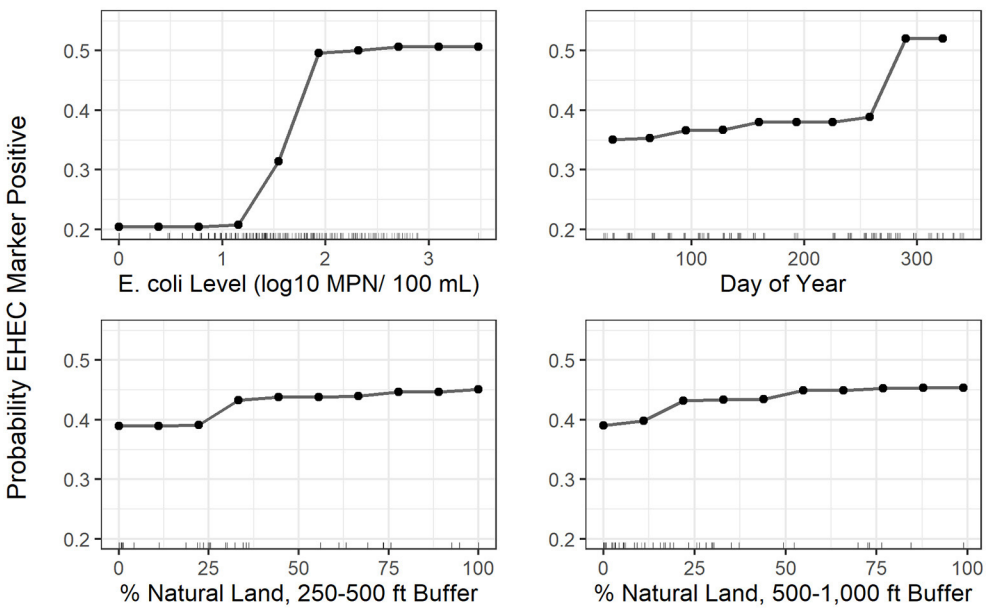


FIGURE 4 | Partial dependence plots for the 4 top-ranked predictors according to variable importance in the EHEC marker conditional forest model with generic *E. coli* included as a predictor; the plots indicate how the predicted probability of a water sample being positive for EHEC marker detection changes as the x-axis variable (predictor) changes. The tick marks along the x-axis indicate values of the predictor variable in samples used to fit the conditional forest model.

2013; Strawn et al., 2013a; Cooley et al., 2014; Tanaro et al., 2014; Nadya et al., 2016; Falardeau et al., 2017; Tian et al., 2017; Partyka et al., 2018; Truchado et al., 2018; Haymaker et al., 2019), others solely relied on PCR screens for either O157 markers (Stea et al., 2015) or EHEC markers (Shelton et al., 2011; Bradshaw et al., 2016; Weller et al., 2020a,b), such as the study reported here. Regardless, most of these previous studies that assessed EHEC in running surface waters have reported lower prevalence than found here (21%). For example, the EHEC prevalence ranged from <1% (5/818) (Ontario, Canada; Wilkes et al., 2009) to 19% (63/330) (British Columbia, Canada; Nadya et al., 2016); both studies utilized culture confirmed EHEC results. While the higher EHEC prevalence in the current study could be due to a truly higher prevalence, the use of a PCR

screen for *stx* and *eaeA* in a single sample without culture confirmation may overestimate the prevalence of EHEC, as this method may (i) detect *stx* and *eaeA* in different organisms and (ii) may detect genetic material from dead organisms. The lower prevalence of EHEC in previous studies that used culture confirmation is also likely explained by the lack of reliable EHEC culture-conformation methods which can underestimate its true prevalence (Muniesa et al., 2006; Bettelheim, 2007; Baker et al., 2019).

The Complexity of Relationships Observed Was Dependent on the Analytical Approach

Two modeling approaches were utilized, including (i) multivariable regression to identify associations between environmental variables and *Salmonella* presence and EHEC marker detection, and (ii) conditional forest to develop models to predict *Salmonella* presence and EHEC marker detection. For both *Salmonella* presence and EHEC marker detection, there were several differences in the variables retained in the regression models and the variables ranked as important by the forests, even though there are some overlaps between important variables between the modeling strategies. This highlights the fact that different modeling strategies are able to detect different relationships in the data (Kuhn and Johnson, 2013). For instance, regression relies on the assumption that there is a linear relationship between independent variables and the log-odds of the dependent variable being detected, and therefore, non-linear relationships cannot be detected. In addition, logistic regression cannot handle a large number of independent variables simultaneously (requiring variable selection before model development), missing data, or correlated variables, and can only account for a limited number of interactions. In comparison, conditional forest utilizes tree-based modeling which does not require the same linear assumption to be met, implicitly accounts for hierarchical relationships and interactions in the data, and can handle missing data, large numbers of independent variables, and correlation. For example, some variables were important by univariable regression but could not be included in final multivariable regression models because they loaded on more than 1 principal component (e.g., percent of natural cover around the sampling site for the *Salmonella* model). PCA is one strategy used for variable selection in regression analysis, as regression analysis is unable to handle overly complex models (see Kuhn and Johnson, 2013 for additional variable reduction strategies). On the other hand, variable selection is incorporated into the conditional forest algorithm and as such is better able to capture the complex relationships inherent to environmental data (Weller et al., 2020b). However, regression-based analysis does have its advantages, especially for hypothesis testing of the relationships between specific, independent variables and the outcome. Regression models are more interpretable than forests, making it easier to understand the relationships in a regression as opposed to forest models (Kuhn and Johnson, 2013).

Only precipitation 1–2 days before sampling was included among the 10 top-ranked predictors in the *Salmonella* forest and retained in the final regression model. Several other studies also found an increased likelihood of *Salmonella* detection following rain events (Haley et al., 2009; Wilkes et al., 2009; Liang et al., 2013; Stea et al., 2015; Weller et al., 2020a). For instance, in a survey of surface water in Georgia, USA, Haley et al. (2009) found significant ($P < 0.005$) positive correlations between *Salmonella* levels, and rainfall 1 and 2 days before sample collection. This relationship may be driven by increases in run-off during rain events, which can transfer *Salmonella* from terrestrial sources to waterways. Unlike rain, developed open space was included in the final *Salmonella* regression model but was not highly ranked in the *Salmonella* forest. According to the regression model, as the percent of developed open space increased, the log-odds of detecting *Salmonella* decreased. Developed open space may act as a proxy for built-landscape features that prevent run-off and microbial contaminants from entering canals, such as vegetative buffers (e.g., in parks) or improved drainage systems. Consistent with these findings, a survey of Central California waterways found a significantly lower prevalence ($P < 0.05$) of *Salmonella* in human-impact areas (47%) compared to animal-impacted areas (74%) (Tian et al., 2017). However, several studies have found a positive association between *Salmonella* presence and variables linked to human presence human presence (Johnson et al., 2003; Weller et al., 2020a). Johnson et al. (2003) speculated this inconsistency between studies could be due to the quality of wastewater removal infrastructure in the sampling area.

For the EHEC models, generic *E. coli* levels and precipitation before sampling were included in both the regression and forest models. The relationship between precipitation and an increased log-odds of EHEC detection was likely also driven by an increase in run-off during rain events, similar to the relationship between *Salmonella* and precipitation discussed above. The relationship between EHEC detection and precipitation is also consistent with past studies (Stea et al., 2015; Nadya et al., 2016). Conversely, there is considerable variability between previous studies in the existence, direction, and strength of the relationship between EHEC detection and generic *E. coli* levels. For example, some studies, like the study presented here, found evidence of a relationship (Holvoet et al., 2014; Stea et al., 2015; Bradshaw et al., 2016; Falardeau et al., 2017; Truchado et al., 2018; Weller et al., 2020b), while others did not (Shelton et al., 2011; Benjamin et al., 2013; Falardeau et al., 2017; Partyka et al., 2018; Weller et al., 2020a).

While there were overall differences in the variables identified as being associated with *Salmonella* presence or EHEC marker detection by regression analysis and those identified as a top ranked predictor by conditional forest analysis, both modeling strategies used together can provide a more complete understanding of the processes that drive pathogen presence. For instance, the variables associated with pathogen presence via regression provide easy to interpret information on associations between a subset of factors and likelihood of pathogen contamination. On the other hand, the top ranked variables in the *Salmonella* presence or EHEC marker detection conditional forests may provide insight into what variables are important for inclusion in models that predict pathogen presence in

agricultural water. This can be used to determine what additional information should be collected to improve the performance of these predictive models so they can be implemented by industry. However, the complex interactions between variables included in the conditional forest models can make it difficult to assess how a change in one variable alone (e.g., occurrence of rainfall) will impact the outcome (i.e., pathogen presence).

Machine-Learning-Based Models Have Potential for Prediction of Pathogen Contamination Likelihood, Including Real Time Prediction That Does Not Require Microbiological Data

Generic *E. coli* is traditionally used as a fecal indicator in agricultural water to indicate potentially unhygienic conditions. However, the high cost, slow turnaround time, uneven distribution of generic *E. coli* in surface waters, and inconsistent relationships between pathogen presence and generic *E. coli* level limit its value and feasibility of routine use (Pachepsky et al., 2016; Wall et al., 2019). Our data here provide further support that conditional forest models are able to predict the presence of *Salmonella* and EHEC markers, as supported by AUC values of 0.84 for the *Salmonella* model and 0.92 for the EHEC marker model (AUC values of 0.8–0.9 are generally indicative of excellent predictive performance; Mandrekar, 2010). Previous studies by Polat et al. (2020) and Weller et al. (2020c) also previously reported that machine learning models show potential as a strategy for identifying contaminated agricultural water in Florida ponds and New York streams, respectively. However, these AUC values from the current study were calculated using cross-validation, as opposed to an independent test dataset, and may be overfit. Regardless, overfitting concerns are mitigated by the fact that the models developed here were developed as a proof of concept and conceptual framework, and not intended for actual use on-farms. If predictive models are going to be developed as an alternative or supplement to indicator-based monitoring, sufficient data is needed, ideally spanning several years and regions, to allow for separate, independent training and test data. Furthermore, additional information is needed to determine if predictive models should be developed for individual waterways, specific regions, or if a standard model can be used across multiple regions.

Importantly, the removal of generic *E. coli* level as a predictor in the *Salmonella* presence and EHEC marker detection conditional forests did not substantially decrease predictive performance of either model. This provides evidence that it is possible to eliminate the use of generic *E. coli* water testing (or other microbial water testing strategies) and replace it with real-time predictive models with limited impact on the accuracy of identifying when water may be contaminated with pathogens and thus at an increased potential risk. These real-time models would be advantageous, as produce growers could estimate the likelihood of pathogen presence in their water sources at the time of water application.

Presenting Predictive Modeling Outcomes as Continuous Risk Measures and Dichotomized Outcomes Have Distinct Advantages

The output of the conditional forest models is the predicted probability that a sample will be positive for *Salmonella* presence or EHEC marker detection. One method for using this predicted probability for making decisions on how to utilize the water would be to dichotomize the outcome (i.e., pathogen is present or absent) based on if the predicted probability is greater than or less than some set threshold value. If this strategy is used, specificity, sensitivity, positive likelihood ratio, and negative likelihood ratio can be calculated. While the specificity is adequate for our *Salmonella* model, the sensitivity is low (0.69). This is particularly concerning as this means the model often calls *Salmonella*-positive samples as negative, and thus may lead to instances where corrective actions (e.g., water treatment) were not performed when they should have been. However, by lowering the threshold value, the sensitivity of the model can be increased, minimizing this risk. A similar phenomenon was observed for the EHEC forest. If predictive models are to be used by produce growers to guide on-farm decision making (e.g., if corrective actions are needed before using water to irrigate crops), optimization of this threshold value is needed. A future quantitative microbial risk assessment would be helpful to identify the risk of illness associated with different threshold values (Uyttendaele et al., 2015; Rock et al., 2019). This information could then be used to optimize what threshold value should be used to balance the predicted number of illnesses vs. the costs associated with different corrective actions (e.g., water treatment). While dichotomizing the outcome of the model, as described above, creates an easier to interpret model, it does cause a loss of information. As an alternative, the predicted probability of a sample being positive could instead be directly used for decision making; however, this would also require quantitative risk assessment to determine how predicted probabilities should be used.

Natural Cover and the Day of Year Are Important for Prediction of *Salmonella* Presence and EHEC Marker Detection

As previously discussed, there were some differences in variables associated with pathogen presence by multivariable regression and the top ranked variables for predicting pathogen presence by conditional forest. However, a specific discussion of the top ranked variables is important for informing what information should be collected for future refinement of models used to predict pathogen presence in agricultural water. Natural cover variables and the day of year were included in the ten top-ranked predictors in the *Salmonella* and EHEC forests. For both *Salmonella* and EHEC, there was a positive monotonic relationship between the percent of natural cover around the sampling site and pathogen presence. Since natural cover may function as habitat for wildlife, this may indicate wildlife is acting as a pathogen source in southwestern US canals; this is supported by the limited number of past studies that examined

the prevalence of enteric pathogens in southwestern wildlife (e.g., Jay et al., 2007; Jay-Russell et al., 2014). For example, an Arizona study found 32% (N = 103 total samples) of coyote fecal samples were positive for *Salmonella*, while none were STEC-positive and 4.9% were enteropathogenic *E. coli*-positive (Jay-Russell et al., 2014). On the other hand, in a study investigating *E. coli* O157:H7 in feral swine in the central California coast, 14.9% (13/87) of samples were positive for *E. coli* O157:H7 (Jay et al., 2007). Since, in the current study, the relationship between natural cover and *Salmonella* presence and EHEC marker detection is weak, additional research is needed to fully characterize the role wildlife plays as a source of enteric pathogen contamination for southwestern canals. Given the need for additional research, and the important ecosystem services provided by natural cover (e.g., water filtration) and wildlife (e.g., pest control, pollination), the authors want to emphasize we are not advocating the removal of natural cover or wildlife from growing areas (Aarons and Gourley, 2012; Allende et al., 2018; Navarro-Gonzalez et al., 2020).

Temporal trends in *Salmonella* presence and EHEC marker detection were accounted for by including the day of year each sample was collected on as a predictor in the forests. The day of year was used instead of season, as seasons are arbitrary periods of time; the end of a season is more similar to the beginning of the subsequent season than the beginning of the season itself. As such, using the day of year as a continuous variable reduces bias in the final model by not forcing the data into arbitrary categories. For both models, the probability of a sample being pathogen-positive remained low until approx. September, after which the probability of a positive increased. This likely indicates some event occurs during early fall that leads to an increased likelihood of pathogen contamination of southwestern US canals. For instance, it is possible the canals are cleaned at this time of year, which causes the sediments at the bottom to re-distribute and re-contaminate the water. However, the current study only spanned 1 year so additional research is needed to determine if this relationship holds across time. Despite this limitation, our finding of intra-annual trends in microbial water quality is consistent with past studies that looked at *Salmonella* (Wilkes et al., 2009; Liang et al., 2013; Cooley et al., 2014; Stea et al., 2015; Tian et al., 2017; Weller et al., 2020b) and EHEC (Shelton et al., 2011; Stea et al., 2015; Nadya et al., 2016; Tian et al., 2017). For instance, Cooley et al. (2014) found a higher *Salmonella* prevalence in the spring and summer compared to the fall and winter in Central California surface water samples. Furthermore, consistent with our current study, Stea et al. (2015) found a higher prevalence of STEC in the later summer and fall compared to all other seasons in Nova Scotia, Canada. Overall, the data collected to date appear to indicate that enteric pathogen contamination of surface water often shows some type of seasonality, although the specific trends appear to differ across locations and studies.

LIMITATIONS

While a large area (∼28,000 km^2) of the produce growing region in the southwestern US is represented here, stratification for certain land-use or sample site factors was not performed during sample site selection. As such, this could have biased the results (i.e., some potentially important factors could have been missed due to underrepresentation of certain variables). Additionally, since this is a proof of concept, only a small number of samples were collected (N = 169) and each site was only visited few times (1–3 times), which could result in combinations of factors associated with an altered likelihood of pathogen presence being missed (e.g., if the greatest likelihood of *Salmonella* contamination is after a rain event next to a dairy farm but no samples were collected after a rainfall from a site next to a dairy farm, this signal would have been missed). Therefore, further studies, with larger sampling efforts and spanning multiple years and geographical locations or growing regions, are needed to yield models appropriate for industry use and to answer key questions such as if a single model can be used from water source to water source, if a single model can be used from climate to climate (or region to region), and if a single model can be used over several years. In addition, there are other factors that may be important for pathogen presence in canals such as difference in elevation between surrounding land and the canal, livestock density surrounding the canals (instead of just CAFO presence), land use along canal flow paths, and relative humidity; future model building efforts should consider collecting these data to include in their models. Furthermore, there were several factors with missing data (e.g., flow rate could not be measured at all sites due to safety concerns). This could lead to information bias (i.e., bias caused by a lack of correct or complete information) in logistic regression, but we expect this to be non-differential (i.e., the bias direction is independent of the model outcome), indicating it should not have impacted the results of the study.

CONCLUSIONS

Machine learning-based predictive models, such as conditional forest models show promise for predicting *Salmonella* presence and EHEC marker detection in southwestern US canals used as sources for agricultural water. The use of machine learning models, in addition to regression analysis, provides a more complete assessment of the relationships between spatial and temporal factors and foodborne pathogen presence in agricultural water due to the complexity in the system. Furthermore, the use of predictive modeling, and real-time predictive models (using no microbiological data), may provide an alternative or supplement to traditional generic *E. coli* testing for fine-tuning when and where food safety hazards may be present in agricultural water and corrective action is needed. The forests developed in the current study specifically indicate that use around the sampling site and day of the year are important predictors for both *Salmonella* presence and EHEC marker detection in southwestern US canal water. Despite the promising results in this and previous studies, these studies were proof of concept. Therefore, before predictive models can be deployed on farms and integrated into on-farm risk management plans additional research is needed to determine if models can predict pathogen presence accurately for regions, water types

(e.g., canal, stream, pond), and years, other than the region(s), water type(s), and year(s) where the training data were collected.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CR, DW, NB, and MW: conceptualized and designed the study. CR and NB: collected the samples. NB, AB, SR, and DW: processed the samples. AB: performed data analysis and model development under the guidance of DW and MW. AB, DW, and MW: wrote the manuscript and all authors assisted in revising the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Center for Produce Safety (2017CPS09). Data analysis, manuscript preparation, and revision were also partially supported by the National Institute of Environmental Health Sciences, National Institutes of Health (NIH) under award number T32ES007271. This work is supported by the Specialty Crop Research Initiative program, Award Number: 2019-51181-30016 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

ACKNOWLEDGMENTS

We would like to thank Maureen Gunderson, Kayla Ferris, Nathanael Henderson, Julia Muuse, and Sriya Sunil for their help on various aspects of this project. We would also like to acknowledge the various irrigation districts within the Southwest growing region for allowing our team to collect agricultural water samples and providing guidance on sampling locations appropriate for the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.693631/full#supplementary-material

REFERENCES

Aarons, S. R., and Gourley, C. J. P. (2012). The role of riparian buffer management in reducing off-site impacts from grazed dairy systems. *Renew. Agric. Food Syst.* 28, 1–16. doi: 10.1017/S1742170511000548

Allende, A., Datta, A. R., Smith, W. A., Adonis, R., MacKay, A., and Adell, A. D. (2018). Implications of new legislation (US FSMA) and guidelines (EC) on the establishment of management systems for agricultural water. *Food Microbiol.* 75, 119–125. doi: 10.1016/j.fm.2017.10.002

Allende, A., and Monaghan, J. (2015). Irrigation water quality for leafy crops: a perspective of risks and potential solutions. *Int. J. Environ. Res. Public. Health.* 12, 7457–7477. doi: 10.3390/ijerph120707457

Baker, C. A., De, J., Bertoldi, B., Dunn, L., Chapin, T., Jay-Russell, M., et al. (2019). Prevalence and concentration of *stx+ E. coli* and *E. coli* O157 in bovine manure from Florida farms. *PLoS ONE.* 14:e0217445. doi: 10.1371/journal.pone.0217445

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beauvais, W., Gart, E. V., Bean, M., Blanco, A., Wilsey, J., McWhinney, K., et al. (2018). The prevalence of *Escherichia coli* O157:H7 fecal shedding in feedlot pens is affected by the water-to-cattle ratio: a randomized controlled trial. *PLoS ONE.* 13:e0192149. doi: 10.1371/journal.pone.0192149

Benjamin, L., Atwill, E. R., Jay-Russell, M., Cooley, M., Carychao, D., Gorski, L., et al. (2013). Occurrence of generic *Escherichia coli*, *E. coli* O157 and *Salmonella* spp. in water and sediment from leafy green produce farms and streams of the Central California Coast. *Int. J. Food Microbiol.* 165, 65–76. doi: 10.1016/j.ijfoodmicro.2013.04.003

Bettelheim, K. A. (2007). The non-O157 shiga-toxigenic (verotoxigenic) *Escherichia coli*; under-rated pathogens. *Crit. Rev. Microbiol.* 33, 67–87. doi: 10.1080/10408410601172172

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., et al. (2016). mlr: Machine Learning in R. *J. Mach. Learn. Res.* 17, 1–5. Available online at: http://jmlr.org/papers/v17/15-066.html.

Bradshaw, J. K., Snyder, B. J., Oladeinde, A., Spidle, D., Berrang, M. E., Meinersmann, R. J., et al. (2016). Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. *Water Res.* 101, 498–509. doi: 10.1016/j.watres.2016.05.014

California Leafy Greens Marketing Agreement (2020). *Commodity Specific Food Safety Guidelines for Production and Harvest of Lettuce and Leafy Greens*. Available online at: https://lgma-assets.sfo2.digitaloceanspaces.com/downloads/CA_LGMA_METRICS_FINAL_VERSION_Accessible_Jan2020.pdf (accessed on July 2, 2020).

Callejon, R. M., Rodriguez-Naranjo, M. I., Ubeda, C., Hornedo-Ortega, R., Garcia-Parrilla, M. C., and Troncoso, A. M. (2015). Reported foodborne outbreaks due to fresh produce in the United States and European Union: trends and causes. *Foodborne Pathog. Dis.* 12, 32–38. doi: 10.1089/fpd.2014.1821

Castro-Ibanez, I., Gil, M. I., Tudela, J. A., Ivanek, R., and Allende, A. (2015). Assessment of microbial risk factors and impact of meteorological conditions during production of baby spinach in the Southeast of Spain. *Food Microbiol.* 49, 173–181. doi: 10.1016/j.fm.2015.02.004

Cooley, M. B., Quinones, B., Oryang, D., Mandrell, R. E., and Gorski, L. (2014). Prevalence of shiga toxin producing *Escherichia coli*, *Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbiol.* 4:30. doi: 10.3389/fcimb.2014.00030

Delaquis, P., Bach, S., and Dinu, L. D. (2007). Behavior of *Escherichia coli* O157:H7 in leafy vegetables. *J. Food Prot.* 70, 1966–1974. doi: 10.4315/0362-028X-70.8.1966

Duffy, E. A., Lucia, L. M., Kells, J. M., Castillo, A., Pillai, S. D., and Acuff, G. R. (2005). Concentrations of *Escherichia coli* and genetic diversity and antibiotic resistance profiling of *Salmonella* isolated from irrigation water, packing shed equipment, and fresh produce in Texas. *J. Food Prot.* 68, 70–79. doi: 10.4315/0362-028X-68.1.70

European Commission (2017). *Commission Notice on Guidance Document on Addressing Microbiological Risks in Fresh Fruits and Vegetables at Primary Production Through Good Hygiene*. Official Journal of the European Union.

Falardeau, J., Johnson, R. P., Pagotto, F., and Wang, S. (2017). Occurrence, characterization, and potential predictors of verotoxigenic *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella* in surface water used for produce irrigation in the Lower Mainland of British Columbia, Canada. *PLoS ONE.* 12:e0185437. doi: 10.1371/journal.pone.0185437

Franz, E., Schijven, J., de Roda Husman, A. M., and Blaak, H. (2014). Meta-regression analysis of commensal and pathogenic *Escherichia coli* survival in soil and water. *Environ. Sci. Technol.* 48, 6763–6771. doi: 10.1021/es501677c

Freshcare (2020). *Food Safety and Quality Standard Edition 4.2*. Available online at: https://www.freshcare.com.au/wp-content/uploads/Freshcare-Food-Safety-Quality-Standard-Edition-4.2-JAN-2021.pdf (accessed May 22, 2021).

Gore, J. A., and Banning, J. (2017). "Discharge measurements and streamflow analysis," in *Methods in Stream Ecology, 3rd ed*, eds F. R. Hauer, and G. Lamberti (Salt Lake City, UT: Academic Press), 49–70.

Greene, S. K., Daly, E. R., Talbot, E. A., Demma, L. J., Holzbauer, S., Patel, N. J., et al. (2008). Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. *Epidemiol. Infect*. 136, 157–165. doi: 10.1017/S095026880700859X

Haley, B. J., Cole, D. J., and Lipp, E. K. (2009). Distribution, diversity, and seasonality of waterborne *Salmonellae* in a rural watershed. *Appl. Environ. Microbiol*. 75, 1248–1255. doi: 10.1128/AEM.01648-08

Hanning, I. B., Nutt, J. D., and Ricke, S. C. (2009). Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention methods. *Foodborne Pathog. Dis*. 6, 635–648. doi: 10.1089/fpd.2008.0232

Haymaker, J., Sharma, M., Parveen, S., Hashem, F., May, E. B., Handy, E. T., et al. (2019). Prevalence of shiga-toxigenic and atypical enteropathogenic *Escherichia coli* in untreated surface water and reclaimed water in the Mid-Atlantic U.S. *Environ. Res*. 172, 630–636. doi: 10.1016/j.envres.2019.02.019

Holvoet, K., Sampers, I., Seynnaeve, M., and Uyttendaele, M. (2014). Relationship among hygiene indicators and enteric pathogens in irrigation water, soil and lettuce and the impact of climatic conditions on contamination in the lettuce primary production. *Int. J. Food Microbiol*. 171, 21–31. doi: 10.1016/j.ijfoodmicro.2013.11.009

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. (2006). Survival Ensembles. *Biostatistics* 7, 355–373. doi: 10.1093/biostatistics/kxj011

Jamieson, R., Gordon, R., Joy, D., and Lee, H. (2004). Assessing microbial pollution of rural surface waters: a review of current watershed scale modeling approaches. *Agric. Water Manag*. 70, 1–17. doi: 10.1016/j.agwat.2004.05.006

Jay, M. T., Cooley, M., Carychao, D., Wiscomb, G. W., Sweitzer, R. A., Crawford-Miksza, L., et al. (2007). *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg. Inf. Dis*. 13, 1908–1911. doi: 10.3201/eid1312.070763

Jay-Russell, M. T., Hake, A. F., Bengson, Y., Thiptara, A., and Nguyen, T. (2014). Prevalence and characterization of *Escherichia coli* and *Salmonella* strains isolated from stray dog and coyote feces in a major leafy greens production region at the United States–Mexico border. *PLoS ONE*. 9:e113433. doi: 10.1371/journal.pone.0113433

Johnson, J. Y. M., Thomas, J. E., Graham, T. A., Townshend, I., Byrne, J., Selinger, L. B., et al. (2003). Prevalence of *Escherichia coli* O157:H7 and *Salmonella* spp. in surface waters of southern Alberta and its relation to manure sources. *Can. J. Microbiol*. 49, 326–335. doi: 10.1139/w03-046

Kim, J. S., Lee, G. G., Park, J. S., Jung, Y. H., Kwak, H. S., Kim, S. B., et al. (2007). A novel multiplex PCR assay for rapid and simultaneous detection of five pathogenic bacteria: *Escherichia coli* O157:H7, *Salmonella*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Vibrio parahaemolyticus*. *J. Food Prot*. 70, 1656–1662. doi: 10.4315/0362-028X-70.7.1656

Kuhn, M. (2020). *caret: Classification and Regression Training*. R package version 6.0-86. Available online at: https://CRAN.R-project.org/package=caret (accessed March 01, 2021).

Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer.

Liang, Z., He, Z., Zhou, X., Powell, C. A., Yang, Y., He, L. M., et al. (2013). Impact of mixed land-use practices on the microbial water quality in a subtropical coastal watershed. *Sci. Total Environ*. 449, 426–433. doi: 10.1016/j.scitotenv.2013.01.087

Liu, H., Whitehouse, C. A., and Li, B. (2018). Presence and persistence of *Salmonella* in water: the impact on microbial quality of water and food safety. *Front. Public Health*. 6:159. doi: 10.3389/fpubh.2018.00159

Lopez-Galvez, F., Allende, A., Pedrero-Salcedo, F., Alarcon, J. J., and Gil, M. I. (2014). Safety assessment of greenhouse hydroponic tomatoes irrigation with reclaimed and surface water. *Int. J. Food Microbiol*. 191, 97–102. doi: 10.1016/j.ijfoodmicro.2014.09.004

Lothrop, N., Bright, K. R., Sexton, J., Pearce-Walker, J., Reynolds, K. A., and Verhougstraete, M. P. (2018). Optimal strategies for monitoring irrigation water quality. *Agric. Water Manag*. 199, 86–92. doi: 10.1016/j.agwat.2017.12.018

Lu, L., Hume, M. E., Sternes, K. L., and Pillai, S. D. (2004). Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. *Water Res*. 38, 3899–3908. doi: 10.1016/j.watres.2004.06.020

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *J Thorac. Oncol*. 5, 1315–1316. doi: 10.1097/JTO.0b013e3181ec173d

Mootian, G., Wu, W. H., and Matthews, K. R. (2009). Transfer of *Escherichia coli* O157:H7 from soil, water, and manure contaminated with low numbers of the pathogen to lettuce plants. *J. Food Prot*. 72, 2308–2312. doi: 10.4315/0362-028X-72.11.2308

Muniesa, M., Jofre, J., Garcia-Aljaro, C., and Blanch, A. R. (2006). Occurrence of *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia coli* in the environment. *Environ. Sci. Technol*. 40, 7141–7149. doi: 10.1021/es060927k

Nadya, S., Delaquis, P., Chen, J., Allen, K., Johnson, R. P., Ziebell, K., et al. (2016). Phenotypic and genotypic characteristics of shiga toxin-producing *Escherichia coli* isolated from surface waters and sediments in a Canadian urban-agricultural landscape. *Front. Cell. Infect. Microbiol*. 6:36. doi: 10.3389/fcimb.2016.00036

Navarro-Gonzalez, N., Wright, S., Aminabadi, P., Gwinn, A., Suslow, T. V., and Jay-Russell, M. T. (2020). Carriage and subtypes of foodborne pathogens identified in wild birds residing near agricultural lands in California: a repeated cross-sectional study. *Appl. Environ. Microbiol*. 86, e01678–e01619. doi: 10.1128/AEM.01678-19

Pachepsky, Y., Shelton, D., Dorner, S., and Whelan, G. (2016). Can *E. coli* or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters? *Crit. Rev. Microbiol*. 42, 384–393. doi: 10.3109/1040841X.2014.954524

Pachepsky, Y., Shelton, D. R., McLain, J. E. T., Patel, J., and Mandrell, R. E. (2011). Irrigation waters as a source of pathogenic microorganisms in produce: a review. *Adv. Agron*. 113, 73–138. doi: 10.1016/B978-0-12-386473-4.00002-6

Park, S., Szonyi, B., Gautam, R., Nightingale, K., Aniciso, J., and Ivanek, R. (2012). Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. *J. Food Prot*. 75, 2055–2081. doi: 10.4315/0362-028X.JFP-12-160

Partyka, M. L., Bond, R. F., Chase, J. A., and Atwill, E. R. (2018). Spatiotemporal variability in microbial quality of western U.S. agricultural water supplies: a multistate study. *J. Environ. Qual*. 47, 939–948. doi: 10.2134/jeq2017.12.0501

Polat, H., Topalcengiz, Z., and Danyluk, M. D. (2020). Prediction of *Salmonella* presence and absence in agricultural surface waters by artificial intelligence approaches. *J. Food Saf*. 40:e12733. doi: 10.1111/jfs.12733

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed March 01, 2021).

Rock, C. M., Brassill, N., Dery, J. L., Carr, D., McLain, J. E., Bright, K. R., et al. (2019). Review of water quality criteria for water reuse and risk-based implications for irrigated produce under the FDA Food Safety Modernization Act, produce safety rule. *Environ. Res*. 172, 616–629. doi: 10.1016/j.envres.2018.12.050

Savichtcheva, O., Okayama, N., and Okabe, S. (2007). Relationships between Bacteroides 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Res*. 41, 3615–3628. doi: 10.1016/j.watres.2007.03.028

Sbodio, A., Maeda, S., Lopez-Velasco, G., and Suslow, T. V. (2013). Modified Moore swab optimization and validation in capturing *E. coli* O157:H7 and *Salmonella enterica* in large volume field samples of irrigation water. *Food Res. Int.* 51, 654–662. doi: 10.1016/j.foodres.2013.01.011

Shelton, D. R., Karns, J. S., Coppock, C., Patel, J., Sharma, M., and Pachepsky, Y. A. (2011). Relationship between *eae* and *stx* virulence genes and *Escherichia coli* in an agricultural watershed: implications for irrigation water standards and leafy green commodities. *J. Food Prot*. 74, 18–23. doi: 10.4315/0362-028X.JFP-10-241

Stea, E. C., Truelstrup Hansen, L., Jamison, R. C., and Yost, C. K. (2015). Fecal contamination in the surface waters of a rural and an urban-source watershed. *J. Environ. Qual*. 44, 1556–1567. doi: 10.2134/jeq2014.11.0459

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Grohn, Y. T., Worobo, R. W., et al. (2013a). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol*. 79, 588–600. doi: 10.1128/AEM.02491-12

Strawn, L. K., Grohn, Y. T., Warchocki, S., Worobo, R. W., Bihn, E. A., and Wiedmann, M. (2013b). Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Appl. Environ. Microbiol*. 79, 7618–7627. doi: 10.1128/AEM.02831-13

Strobl, C., Boulesteix, A., L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinform*. 9:307. doi: 10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A. L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform*. 8:25. doi: 10.1186/1471-2105-8-25

Tam, S., and Petersen, A. (2014). *B.C. Sprinkler Irrigation Manual*. British Columbia Ministry of Agriculture. Available online at: https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/agriculture-and-seafood/agricultural-land-and-environment/water/irrigation/sprinkler-irrigation-manual/563000-1_web_titlepage_preface_tofc.pdf (accessed July 2, 2020).

Tanaro, J. D., Piaggio, M. C., Galli, L., Gasparovic, A. M. C., Procura, F., Molina, D. A., et al. (2014). Prevalence of *Escherichia coli* O157:H7 in surface water near cattle feedlots. *Foodborne Pathog. Dis*. 11, 960–965. doi: 10.1089/fpd.2014.1770

Tian, P., Yang, D., Shan, L., Wang, D., Li, Q., Gorski, L., et al. (2017). Concurrent detection of human norovirus and bacterial pathogens in water samples from an agricultural region in Central California Coast. *Front. Microbiol*. 8:1560. doi: 10.3389/fmicb.2017.01560

Truchado, P., Hernandez, N., Gil, M. I., Ivanek, R., and Allende, A. (2018). Correlation between *E.* coli levels and the presence of foodborne pathogens in surface water: establishment of a sampling program. *Water Res*. 128, 226–233. doi: 10.1016/j.watres,.2017.10.041

U. S. Centers for Disease Control and Prevention (2006). *Multistate Outbreak of E. coli O157:H7 Infections Linked to fResh Spinach (Final Update)*. Available online at: https://www.cdc.gov/ecoli/2006/spinach-10-2006.html (accessed July, 2, 2020).

U. S. Centers for Disease Control and Prevention (2018a). *Multistate Outbreak of Shiga Toxin—Producing Escherichia coli O157:H7 Infections Linked to Leafy Greens (Final Update)*. Available online at: https://www.cdc.gov/ecoli/2017/o157h7-12-17/index.html (accessed July 2, 2020).

U. S. Centers for Disease Control and Prevention (2018b). *Multistate Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce (Final Update)*. Available online at: https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html (accessed July, 2, 2020).

U. S. Food and Drug Administration (2015). *Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption*. Final Rule, Docket No. FDA-2011-N-0921. Available online at: https://www.federalregister.gov/documents/2015/11/27/2015-28159/standards-for-the-growing-harvesting-packing-and-holding-of-produce-for-human-consumption (accessed on July 2 2019).

Uyttendaele, M., Jaykus, L. A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., et al. (2015). Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. *Compr. Rev. Food Sci. Food Saf*. 14, 336–356. doi: 10.1111/1541-4337.12133

Wall, G. L., Clements, D. P., Fisk, C. L., Stoeckel, D. M., Woods, K. L., and Bihn, E. A. (2019). Meeting report: key outcomes from a collaborative summit on agricultural water standards for fresh produce. *Compr. Rev. Food Sci. Food Saf*. 18, 723–737. doi: 10.1111/1541-4337.12434

Weller, D., Belias, A., Green, H., Roof, S., and Wiedmann, M. (2020a). Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Front. Sustain. Food Syst*. 3:124. doi: 10.3389/fsufs.2019.00124

Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., et al. (2020b). Complex interactions between weather, and microbial and physicochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water. *Front. Microbiol*. 11:134. doi: 10.3389/fmicb.2020.00134

Weller, D., Love, T., Belias, A., and Wiedmann, M. (2020c). Predictive models may complement existing *E. coli* water quality testing for managing enteric pathogen contamination of Northeastern streams used for produce production. *Front. Sustain. Food Syst*. 4:561517. doi: 10.3389/fsufs.2020.561517

Wilkes, G., Edge, T., Gannon, V., Jokinen, C., Lyautey, E., Medeiros, D., et al. (2009). Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Res*. 43, 2209–2223. doi: 10.1016/j.watres.2009.01.033

Wilkes, G., Edge, T. A., Gannon, V. P. J., Jokinen, C., Lyautey, E., Neumann, N. F., et al. (2011). Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res*. 45, 5807–5825. doi: 10.1016/j.watres.2011.06.021

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., et al. (2018). A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens*. 146, 108–123. doi: 10.1016/j.isprsjprs.2018.09.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/frwa.2021.710853

Changes in Bacterial Communities During Treatment of Municipal Wastewater in Arctic Wastewater Stabilization Ponds

*Yannan Huang[1], Colin M. Ragush[1], Lindsay H. Johnston[1], Michael W. Hall[2], Robert G. Beiko[2], Rob C. Jamieson[1] and Lisbeth Truelstrup Hansen[1,3]**

[1] Centre for Water Resources Studies, Dalhousie University, Halifax, NS, Canada, [2] Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada, [3] National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

OPEN ACCESS

Edited by:
Jasna Kovac, The Pennsylvania State University (PSU), United States

Reviewed by:
Yun Zhou, Huazhong Agricultural University, China
Stephanie Gora, York University, Canada
Anas Ghadouani, University of Western Australia, Australia

***Correspondence:**
Lisbeth Truelstrup Hansen
litr@food.dtu.dk

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *17 May 2021*
Accepted: *26 July 2021*
Published: *20 August 2021*

Citation:
Huang Y, Ragush CM, Johnston LH, Hall MW, Beiko RG, Jamieson RC and Truelstrup Hansen L (2021) Changes in Bacterial Communities During Treatment of Municipal Wastewater in Arctic Wastewater Stabilization Ponds. Front. Water 3:710853. doi: 10.3389/frwa.2021.710853

Wastewater stabilization ponds (WSPs) are commonly used to treat municipal wastewater in the Canadian Arctic. Bacterial community structure and functionality remain mostly uncharacterized for arctic WSPs, yet are presumed important for treatment outcomes during the 3-month summer treatment season with open water in the WSPs. The objective of this study was to investigate treatment performance and related temporal and spatial changes in the structure and putative function of bacterial communities during treatment of municipal wastewater in the WSPs of Pond Inlet and Clyde River, Nunavut over two consecutive summer treatment seasons. Influent raw wastewater contained a high organic load and large bacterial communities (~9 log 16S rRNA copies/mL) belonging mainly to *Proteobacteria.* Although designed to be facultative ponds, both WSPs remained anaerobic with neutral pH values (7.5–7.8) throughout the summer treatment season. Water quality data showed that nutrients [measured as carbonaceous biological oxygen demand ($CBOD_5$)], total suspended solids, and total ammonia nitrogen were progressively reduced during treatment in the ponds as the summer progressed. The pond bacterial population size and species richness depended on the pond temperature (2–18°C), with 8.5 log 16S rRNA copies/mL and the largest alpha diversities (Shannon-Wiener index of 4-4.5) observed mid-season (late July). While the phylogenetic beta diversity in raw wastewater from the two locations remained similar, pond bacterial communities underwent significant ($p < 0.05$) changes to dominance of *Comamonadaceae, Geobacteracea*, and *Porphyromonadaceae*. Multivariate distance based redundancy analysis and predicted gene functionalities in the microbiota agreed with water quality results that microbial removal of nutrients (e.g., $CBOD_5$) peaked in the middle of the summer coinciding with the treatment period with the highest pond temperatures. Information from this study will be useful for further development of models to predict biological treatment outcomes, which could be used to size and assess the feasibility of WSPs in extreme climates. Higher pond temperatures resulted in

optimal biological processes and nutrient removal in the middle of the summer. While it is challenging to control environmental factors in a passive wastewater treatment system there are some design considerations that could be used to optimize temperature regimes, such as the depth of the pond.

Keywords: pond temperature, bacterial diversity, nutrient removal, cold climate, sewage microbiology, treatment season, remote communities

INTRODUCTION

Wastewater stabilization ponds (WSPs) are commonly used to treat municipal wastewater in remote arctic communities in Canada. The WSPs consist of either naturally occurring lakes with uncontrolled release or engineered structures with a lining and berm for controlled release of the water. Engineered WSPs have typically been designed to be facultative (i.e., aerobic) and to hold the annual volume of wastewater generated within the community (Ragush et al., 2015).

These WSPs are passive (i.e., non-mechanical) treatment systems that depend on physical settling and biological processes and are strongly affected by the climatic conditions including ambient temperature (Heinke et al., 1991). Operations of arctic WSPs are unique in that the content will remain frozen for 9 months of the year limiting the treatment season to the three summer months with open water and temperatures that support biological activity (Balch et al., 2018). Also, due to the low water consumption per capita (∼100 L/person × day) in arctic regions such as Nunavut compared to Southern Canada (average of 274 L/ person × day), the deposited sewage tends to be more concentrated (Daley et al., 2014; Ragush et al., 2015). This can lead to high organic loading rates (OLRs) and anaerobic conditions in the facultative ponds which may affect the microbiological processes negatively (Ragush et al., 2017). Past studies have shown that primary treatment can be attained in arctic WSPs with removal of around 25–40% of nutrients, oxygen demanding material, suspended solids, and 99% of fecal indicator *Escherichia coli* (Marais, 1974; Heinke et al., 1991), even in WSPs with high OLRs and anaerobic conditions (Ragush et al., 2015; Huang et al., 2018).

It is well-known that microbial communities play an important role in biological wastewater treatment (Wagner et al., 2002). Insights in the dynamics of the wastewater microbiome during treatment can therefore be useful in determining which organisms are present and how they relate to the treatment of the wastewater for the benefit of future process optimization.

Studies of full scale mechanical wastewater treatment facilities show that low temperature generally reduces the bacterial diversity and slows down the growth and metabolic activity of functional bacteria to the extent where the treatment is affected. Rodriguez-Caballero et al. (2012) observed that low wastewater temperatures during the winter (10–15°C) caused difficulties in maintaining the required rates of nitrogen removal in biological wastewater treatment plants (WWTPs) operated in Northern Sweden. Another cold-climate study (Ju et al., 2014) found that the relative abundances of functional groups in activated sludge communities were shaped by seasonal temperature changes, i.e., when the temperature dropped during winter season, the abundances of nitrite-oxidizing bacteria (NOB) and hydrolysers decreased, leading to the incomplete nitrification and hydrolysis of organic material. The relative temperature-driven reduction in microbial diversity was also observed in a Finnish study conducted in the seven Northern communities, where wastewater was treated in bioreactors operated at low temperatures of 3–7°C resulting in good nutrient removal but variable performance for nitrogen removal (Gonzalez-Martinez et al., 2018).

In terms of microbial communities related to treatment in arctic WSPs, Gromala et al. (2021) recently examined water samples from WSPs located in Baker Lake (Qamani'tuaq), Cambridge Bay (Iqaluktuuttiaq), and Kugluktuk (Qurluktuk), all located in Western Nunavut, and reported temporal and spatial differences in the abundance of major bacterial phyla that were relatable to specific WSPs and the time after the Spring thaw. However, studies of the structure and function of microbial communities in relation to the treatment performance of arctic WSP systems throughout the summer treatment season have to the best of our knowledge not been performed.

The objective of this study was to investigate the temporal and spatial changes in the structure and putative function of bacterial communities and their link to treatment performance during treatment of municipal wastewater in arctic WSP systems located in the communities of Pond Inlet (Mittimatalik) and Clyde River (Kangiqtugaapik) in Nunavut, Canada. This study was conducted over two summer treatment seasons (June to early-September) and involved monitoring microbial populations in influent raw wastewater and the progressively treated water in the WSPs as well as selected environmental and water quality parameters. Information about the microbiology of these systems can help in understanding passive, biological wastewater treatment systems operated in cold climates.

MATERIALS AND METHODS

Study Sites

From June, 2013 to September, 2014, six and five sampling trips were made to Pond Inlet and Clyde River, Nunavut, Canada. Both communities are located on Qikitaaluk/Baffin Island. The area's polar climate is characterized by long cold winters (∼9 months from mid-September to May) and short cool summers (June to mid-September).

Domestic wastewater is collected daily or several times a week from holding tanks located inside buildings (homes, library, schools, health clinic, municipal offices, etc.) and trucked to

the WSPs for treatment. Wastewater is loaded into the WSPs throughout the year, however, treatment mainly occurs after the Spring thaw and during summer months, where the content of the WSPs is liquid. None of the communities have any industries or agricultural activities that would contribute to the wastewater.

Pond Inlet (72° 41′ 57″ N, 77° 57′ 33″ W) has a population of 1612 (Statistics Canada, 2012). The average daily temperature is −34°C in February (coldest month) and 6°C in July (warmest month, Environment Canada, 2014a). The community's WSP system consists of one engineered pond with a surface area of ∼40,000 m^2 and an average depth of ∼1.9 m during the summer. The treated wastewater (effluent) is discharged once annually in September by pumping the water over the berm after which it goes through a 500 m channel and into the ocean (Eclipse Sound). The annual estimated wastewater generation for the community is 8.0×10^7 L (140 L/person × day).

Clyde River (70° 28′ 26″ N, 68° 35′ 10″ W) has a population of 1004 (Statistics Canada, 2012). The community experiences average daily temperatures of −30°C in February (coldest month) and 5°C in July (warmest month, Environment Canada, 2014b). The WSP system in Clyde River consists of a small primary pond (6,000 m^2) and a larger secondary pond (15,000 m^2). The estimated annual wastewater volume is 3.1×10^7 L (88 L/person × day). This two-cell WSP system is designed to be operated in a staged manner, where the raw wastewater is deposited in the primary pond. The pre-processed wastewater would then be transferred into the secondary pond for further treatment before the annual decant in September, where the water is pumped into a wetland area that leads into the ocean (Patricia Bay). In September 2013, operations deviated as untreated wastewater was deposited in both ponds.

Sampling Strategy

On each visit, grab samples of wastewater were obtained from the wastewater trucks (i.e., raw wastewater, $n \geq 3$), and from different locations ($n = 2$–5) and at varying depths (0.1, 1.1, and 1.9 m) in the ponds to examine spatial variability. Effluent samples were also obtained during the decant ($n = 3$) in Pond Inlet or from the secondary pond ($n = 3$) in Clyde River just prior to the annual decant.

Samples from the Pond Inlet WSP were obtained in the beginning (late June/early July, just after the Spring thaw,), middle (late July/early August), and end (early/middle September, time of the annual decant which is just before to the freeze-up) of the summer treatment season in 2013 and 2014. Sludge samples (one composite sample per visit) were also obtained from the Pond Inlet WSP in 2014.

Samples from Clyde River were collected three times in 2013, while sampling in 2014 took place at the beginning and end of the summer just prior to the decant event.

Wastewater samples were collected in 1-L or 500-mL sterile containers (Nalgene, Fisher Scientific, Nepean, ON, Canada) from an inflatable boat or from the shore of the pond using a subsurface pole sampler and then stored in a cooler (4°C). Within 24 h of the sampling event, samples were flown to the Northern Water Quality Laboratory (NWQL) at the Nunavut Research Institute in Iqaluit, NU, for further processing.

Continuous Monitoring of WSPs and Water Quality Measurements

Deployment of multi-parameter sondes (YSI Inc., Yellow Spring, OH) allowed for *in-situ* hourly measurements of wastewater temperature, pH, and dissolved oxygen (DO). Sondes were installed in the WSPs (0.1 m below the surface) during the first sampling trip and retrieved at the end of each treatment season.

Grab samples were analyzed for the content of carbonaceous biological oxygen demand ($CBOD_5$), total ammonia (TA), total phosphorus (TP), total nitrogen (TN), and total suspended solids (TSS) using standard methods. Briefly, $CBOD_5$ was analyzed in duplicate following the APHA standard method 5210 B (American Public Health Association, 1999). Dissolved oxygen was measured using a probe (Orion 83005MD, Fisher Scientific, Ottawa, ON, Canada) and an Orion StarTM series meters (Fisher Scientific). The APHA standard methods 2540 D (American Public Health Association, 1999) with Whatman TM934-AH 47 mm glass fiber filters (Fisher Scientific) were used for measurements of TSS. Measurements of total ammonia (TA) was done using a Thermo Scientific OrionTM High-Performance Ammonia Electrode with the addition of Ionic Strength Adjuster (Fisher Scientific). The ascorbic acid method *via* Hach® TNT™ or TNT plus™ test kits was used to test TP. TN was analyzed using Hach® TN Test 'N Tubes™ (0.5 to 25.0 mg/L N), according to the manufacturer's procedure.

DNA Extraction

Genomic DNA was extracted from samples immediately upon arrival to the NWQL. DNA was extracted from three biological replicates (separate samples) per sampling site whenever available, with two technical replicates per sample. Ten mL of each sample was centrifuged at 3,200 × *g* for 10 min (min) to obtain a pellet, which was subjected to genomic DNA extraction using the MO BIO PowerSoil DNA isolation kit (MO BIO Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. Within 24–48 h, all DNA samples were transported to the laboratory at Dalhousie University in Halifax, NS in a cooler (4°C) and stored at −20°C until further analysis.

Size of Bacterial Communities (16S rRNA Copy Numbers)

Bacterial 16S rRNA gene copy numbers were quantified in a quantitative PCR (qPCR) assay with the BACT 2 primer set (1369F: 5'-CGGTGAATACGTTCYCGG-3'; 1492R: 5'-GGWTACCTTGTTACGACTT-3') (Suzuki et al., 2000). The qPCR amplification was performed on a Bio-Rad CFX96 Touch™ Real-Time PCR detection system (Bio-Rad, Hercules, CA, USA). Each reaction consisted of: 20-μL total reaction volumes consisting of 4.0 μL of template DNA, 4.4 μL of sterile and nuclease-free water, 0.8 μL of each primer (10 μM), and 10 μL of 2× Power SYBR Green PCR master mix (ThermoFisher Scientific). The thermocycler program consisted of 10 min of initial denaturation at 95°C, followed by 40 cycles of 15 s denaturation at 94°C, 30 s annealing at 55°C, and 30 s extension at 72°C. Melt curve analysis (30 s at 55°C and then increasing by

0.5°C each cycle until reaching 95°C) showed that the melting temperature of the amplicon was 84.5 ± 0.5°C.

A standard curve was constructed using 10-fold dilutions of a plasmid DNA extracted from an *Escherichia coli* DH5α culture containing a positive control plasmid (pCR2.1, TOPO TA PCR 2.1 with an insertion of the 16S rRNA fragment, gift from Dr. C. Yost, University of Regina, Canada), with triplicate measurements of diluted standards. Assay efficiency was calculated to be 102%. The coefficient of determination values (R^2) of the standard curve was 0.999, the limits of detection (LOD) and quantification (LOQ) were 68.6 and 686 copies/reaction, respectively. Absence of PCR inhibitors was confirmed by comparison of the threshold cycle (Ct) values for 10-fold diluted DNA extracts.

Illumina MiSeq Paired-End Sequencing and Bioinformatics Analysis

The V6-V8 regions of the bacterial 16S rRNA gene amplicon library preparation and Illumina MiSeq Paired-End (PE) sequencing were performed following the established protocol (Comeau et al., 2017) at the Dalhousie University Integrated Microbiome Resource (IMR; http://cgeb-imr.ca). The demultiplexed PE reads were obtained in the fastq format from IMR and subjected to the bioinformatics analysis workflow briefly described in the following section.

The raw PE sequencing reads were preprocessed by stitching PE reads and removing low quality reads in PEAR (Zhang et al., 2014), and removing chimeras in VSEARCH (version 1.11.1, 28) in the Microbiome Helper pipeline at IMR (Comeau et al., 2017). As a result, an average of 73.1% of the reads (range 49.6–94.6% for individual samples) was generated as final high-quality sequences. The final high-quality sequences were then clustered into Operational Taxonomic Units (OTUs) at an identity level of ≥97% using the *de novo* UPARSE-OTU algorithm in the USEARCH pipeline (Edgar, 2013). OTUs from UPARSE were deposited in the Microbiome Helper pipeline to filter out spurious OTUs (Comeau et al., 2017), resulting in a total of 2,093 OTUs (from 139 samples). Subsequent analyses were carried out using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline version 1.9.1 (Caporaso et al., 2010a), unless otherwise noted. The taxonomic assignment was performed in the Ribosome Database Project (RDP) classifier version 2.2 (Wang et al., 2007) based on the Greengenes database (version 13_8) at the cut-off value of 60% (Claesson et al., 2009). The rarefied OTU table was generated by subsampling a minimal number of reads (10,539 sequences per sample), resulting in a total of 1,924 OTUs. The sequence alignment was built using the Python Nearest Alignment Space Termination (PyNAST) tool (Caporaso et al., 2010b). This alignment was used to build a phylogenetic tree in FastTree (version 2.1.10, http://www.microbesonline.org/fasttree/) which infers approximately-maximum-likelihood phylogenetic trees (Price et al., 2010). The rarefied OTU table and the phylogenetic tree were used in the downstream analysis, which included determining the core bacterial community that requires the core OTUs to be present in 100% of the samples (Huse et al., 2012), alpha-diversity measures (number of the observed OTUs, Chao1, the Shannon-Wiener (SWI) and the Simpson evenness indexes), and phylogenetic beta-diversity measures, respectively. The phylogenetic beta-diversity was calculated using both the weighted and unweighted UniFrac metrics (Lozupone and Knight, 2005). Principal Coordinate Analysis (PCoA) and an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree were used to visualize the diversity trends based on the weighted UniFrac beta-diversity distance matrix. The trends revealed by the unweighted UniFrac analyses resembled the weighted UniFrac results and are therefore not presented. GenGIS (Parks et al., 2013) was applied to visualize information about bacterial communities representing different locations, times and years in a sampling map.

A distance-based redundancy analysis (dbRDA) was performed on the weighted UniFrac beta-diversity matrix and selected environmental variables for pond samples obtained from each of the two communities over the course of the 2-year study period. The models were assessed using the ANOVA function (R Core Team, 2021) to determine the significance of the environmental predictors (temperature, $CBOD_5$, TSS, TP, TN, and TA) on the beta-diversity among communities. The dbRDA was run using the capscale function in the R package vegan (Oksanen et al., 2020) and ordination plots were generated using the ggord package (Beck, 2020).

Predictions of Functions in Bacterial Communities

The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt, version 1.1.0, Langille et al., 2013) in the Microbiome Helper pipeline (http://github.com/mlangill/microbiome_helper/wiki/PICRUSt-workflow) was used to predict gene contents based on 16S rRNA gene surveys. The PICRUSt analysis works from the observation that there is a good correlation between the phylogenetic information inferred from 16S rRNA marker gene sequences and genomic content when related reference genomes are available. Prior to the analysis, the accuracy of predictions is measured by the Nearest Sequenced Taxon Index (NSTI), with lower values indicating a closer relationship and availability of closely related reference genomes (Langille et al., 2013). Low NSTI values of 0.07 ± 0.01 and 0.08 ± 0.02 were calculated for wastewater samples from Pond Inlet and Clyde River, respectively. For comparison, it has been reported (Langille et al., 2013) that human-associated samples had the lowest (best) NSTI values (0.03 ± 0.02), while higher NSTI values (>0.14) were reported for other mammalian gut and soil samples. Thus, arctic municipal wastewater samples appeared to constitute a suitable data set to derive functional predictions by using PICRUSt and the recommended workflow.

Statistical Analysis

Significant differences ($p < 0.05$) in the log 16S rRNA copy numbers/mL and alpha diversity among sample groups were tested using the non-parametric Kruskal-Wallis test performed in Prism 7 (version 7.0b, Graph Pad Software, Inc., La Jolla, CA, USA). In the study of phylogenetic beta-diversity measures, the nonparametric statistical test through

TABLE 1 | Water quality in the Pond Inlet and Clyde River wastewater treatment systems.

Community	Year	Water type	Season	Temperature (°C)	$CBOD_5$ (mg/L)	TSS (mg/L)	TN (mg/L - N)	TP (mg/L)	TA (mg/L - N)
Pond Inlet	2013	Raw	All	19.5 ± 1.6[a]	591.8 ± 205.3	412.4 ± 177.0	160.4 ± 24.1	62.5 ± 17.1	136.5 ± 11.2
		WSP	Beginning	1.1 ± 0.1	191.4 ± 38.6	46.3 ± 27.3	84.4 ± 5.3	27.3 ± 1.0	82.6 ± 31.5
		WSP	Middle	15.8 ± 0.0	177.3 ± 14.0	21.7 ± 2.7	96.6 ± 3.5	31.1 ± 0.6	101.1 ± 16.3
		WSP	End	0.3 ± 0.0	112.6 ± 9.4	55.6 ± 10.9	101.9 ± 9.0	35.0 ± 0.6	89.9 ± 1.9
		Effluent	End	1.5 ± 0.7	103.3 ± 14.1	64.5 ± 8.9	109.1 ± 9.9	35.1 ± 1.0	87.1 ± 4.6
		WSP Beginning – WSP End		Removal (%)	41.2	−20.1	−20.7	−28.2	−8.8
		Raw to Effluent		Removal (%)	82.5	84.4	32.0	43.8	36.2
	2014	Raw	All	20.5 ± 4.6	347.4 ± 129.8	201.7 ± 91.4	128.6 ± 18.6	43.5 ± 6.1	83.3 ± 33.0
		WSP	Beginning	8.7 ± 0.0	226.7 ± 27.6	42.1 ± 35.6	119.2 ± 11.2	35.3 ± 2.9	106.3 ± 5.9
		WSP	Middle	18.8 ± 0.0	124.3 ± 26.5	76.3 ± 17.5	102.8 ± 7.6	32.0 ± 0.6	128.6 ± 2.5
		WSP	End	6.0 ± 0.0	114.8 ± 16.9	36.3 ± 6.7	87.8 ± 9.0	35.3 ± 0.2	90.8 ± 5.9
		Effluent	End	2.9 ± 0.1	113.3 ± 3.2	19.8 ± 0.4	117.0 ± 8.8	33.8 ± 0.4	86.4 ± 8.0
		WSP Beginning – WSP End		Removal (%)	49.4	13.8	26.3	0.0	14.6
		Raw to Effluent		Removal (%)	67.4	90.2	9.0	22.3	−3.7
Clyde River	2013	Raw	All	20.7 ± 2.6	389.6 ± 155.5	243.0 ± 119.4	129.6 ± 26.4	48.3 ± 8.3	113.0 ± 25.5
		WSP1	Beginning	3.4 ± 0.0	280.8 ± 10.8	64.8 ± 8.7	102.4 ± 8.0	33.6 ± 0.6	89.3 ± 8.4
		WSP1	Middle	11.5 ± 0.0	160.3 ± 31.6	51.0 ± 6.8	67.7 ± 0.8	23.4 ± 0.2	64.8 ± 2.8
		WSP1	End	3.0 ± 0.0	183.8 ± 20.0	61.8 ± 3.5	95.0 ± 3.6	29.0 ± 0.0	108.8 ± 2.1
		WSP2	Beginning	3.4 ± 0.0	105.0 ± 10.4	27.8 ± 2.9	76.1 ± 17.7	19.4 ± 0.2	51.3 ± 0.5
		WSP2	Middle	11.5 ± 0.0	92.3 ± 16.5	27.8 ± 2.9	60.1 ± 5.3	18.4 ± 0.2	45.5 ± 3.1
		WSP 2	End	3.0 ± 0.0	70.3 ± 25.7	37.8 ± 4.0	74.3 ± 1.7	19.6 ± 0.1	82.7 ± 2.4
		WSP1 Beginning – WP2 End		Removal (%)	75.0	41.7	27.4	41.7	7.4
		Raw – WSP2 End		Removal (%)	82.0	84.4	42.7	59.4	26.8
	2014	Raw	All	15.8 ± 1.8	277.7 ± 25.4	243.7 ± 107.4	90.0 ± 10.5	38.8 ± 5.8	104.9 ± 29.0
		WSP1	Beginning	6.4 ± 0.0	315.0 ± 56.8	16.0 ± 4.1	100.0 ± 0.0	43.9 ± 0.0	137.5 ± 8.3
		WSP1	End	4.3 ± 0.0	228.5 ± 87.9	41.8 ± 5.0	95.0 ± 5.4	26.7 ± 0.3	93.7 ± 6.1
		WSP2	Beginning	6.4 ± 0.0	105.3 ± 3.8	15.8 ± 1.3	77.0 ± 2.0	21.9 ± 0.0	62.4 ± 1.2
		WSP2	End	4.3 ± 0.0	69.0 ± 2.9	19.8 ± 5.7	68.1 ± 1.2	17.8 ± 0.1	68.9 ± 6.3
		WSP1 Beginning – WP2 End		Removal (%)	78.1	−23.8	31.9	59.5	49.7
		Raw – WSP2 End		Removal (%)	75.2	91.9	24.3	54.1	34.3

[a]*Average value ± standard deviation (n-1).*

permutations analysis of similarities (ANOSIM) approach was performed in the QIIME pipeline (Caporaso et al., 2010a) to test whether there were statistically significant differences among the trends that were observed in PCoA plots. In the study of inferring functional content in the bacterial communities, significant differences at a 5% level between two groups were tested using Welch's *t*-test with the Benjamini-Hochberg false-discovery rate (FDR) correction, while significant differences at a 5% level among three or more groups were tested using the Kruskal-Wallis *H*-test with FDR correction. These two tests were conducted in the statistical analysis of taxonomic and functional profiles (STAMP) software version 2.1.3 (Parks et al., 2014).

RESULTS

WSP Environment and Removal of Bacteria and Nutrients

Raw wastewater samples from both communities consistently contained an average of 9.0 log 16S rRNA copies/mL, a level which showed no change ($p > 0.05$) over the duration of the study (data not shown). The water quality varied but tended be contain higher levels of $CBOD_5$ and other nutrients in Pond Inlet compared to Clyde River (**Table 1**).

Surface pond temperatures in Pond Inlet gradually increased from 2 to 8°C to mid-season peak temperatures of 18°C followed by a decrease to below 5°C in the month of September (**Figures 1A,B**). The bacterial population size followed the changes in pond temperatures with midseason peaks of 8–8.9 log 16S rRNA copies/mL followed by significant ($p < 0.05$) decrease to 6.5–7 log 16S rRNA copies/mL at the end. The population size in 2014 was significantly ($p < 0.05$) higher than in 2013, which coincided with the higher pond temperatures recorded in 2014. Results from the deployed DO and pH sensors revealed negligible DO levels (<0.2 mg/L), except for two one-week mid-season spikes to measurable levels (see **Supplementary Figure 1**), and stable pH-values from 7.5 to 7.8 (**Supplementary Figure 2**). Absence of marked increases in pond pH and DO indicated that algal blooms were absent in 2013 and 2014.

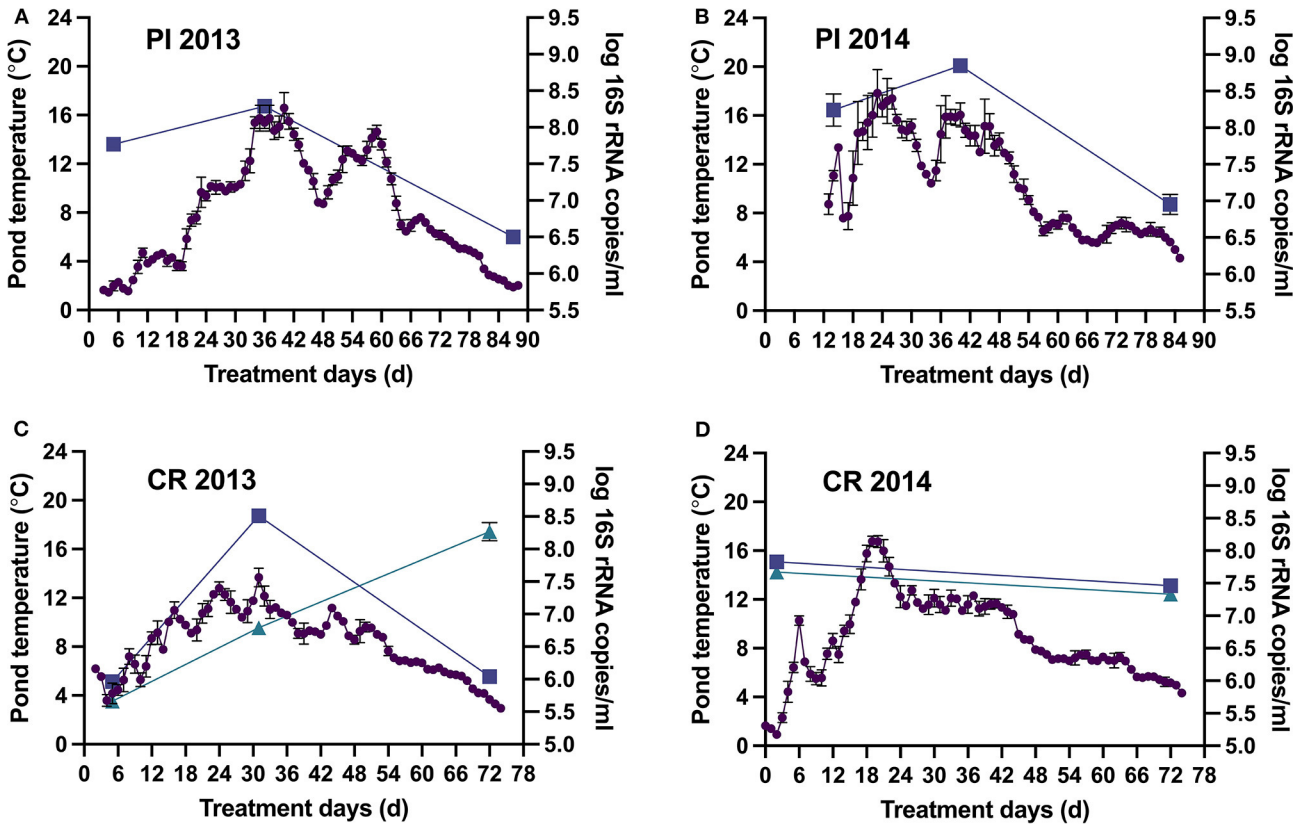


FIGURE 1 | Pond Inlet (PI) and the Clyde River (CR) WSP system surface pond temperatures (•) and bacterial populations [log 16S rRNA copies/mL, ■ primary pond, ▲ secondary pond (only CR), mean ± standard deviation] during the 2013 **(A,C)** and 2014 **(B,D)** treatment seasons. Bacterial populations were determined at the beginning, middle, and end of treatment seasons. Temperatures are represented by daily averages of hourly measurements ($n = 24$, mean ± standard deviation) from the deployed sensors.

In Clyde River, the highest and lowest bacterial levels in the primary pond (∼8.5 and ∼6.0 log 16S rRNA copies/mL, respectively) also coincided with the highest (13.7°C) and lowest (3.7°C) pond temperatures in 2013 (**Figure 1C**). Normal operation of the Clyde River WSP systems in 2014 reduced populations in the secondary pond to 7.4 log 16S rRNA copies/mL at the end of the treatment season (**Figure 1D**). However, in September 2013, the secondary pond contained 8.3 log 16S rRNA copies/mL (**Figure 1C**), which was likely caused by direct discharge of raw wastewater into that pond. DO levels remained negligible (<0.2 mg/L) throughout both treatment seasons, and pH ranged from 7.5 to 7.8, indicating absence of algal blooms in both years (**Supplementary Figures 1, 2**).

Overall, arctic WSP treatment effected a >2 log reduction in bacterial numbers leading to an effluent with a content of 6–7.0 log 16S rRNA copies/mL ready to be released into the receiving environment (**Figure 1**). Water quality data showed that treatment in the WSPs removed 67–83, 84–92, 9–43, 22–59% and none to 36% of the $CBOD_5$, TSS, TN, TP, and TA, respectively, found in the raw wastewater (**Table 1**). The content of $CBOD_5$ in the ponds decreased over the summer treatment season, with the lowest values 70–90 mg/L observed at the end of the summer in the secondary pond in Clyde River as opposed to 103–115 mg/L in Pond Inlet.

Alpha Diversity in Bacterial Communities in Arctic WSP Systems

The alpha diversity decreased during the treatment of the wastewater as shown by the significant ($p < 0.05$) decreases in the SWI from values of ∼5 for the raw wastewater to values ranging between 2.5 and 4.5 in the ponds of both communities (**Figure 2**). Comparison of pond samples across the treatment seasons showed that the alpha diversity was lower at the start and end of the treatment season (2.5–4) while SWI values peaked at 4.3–4.5 during the warmer mid-season. The alpha diversity tended to be lower in Clyde River samples, which coincided with pond temperatures overall being colder in Clyde River compared to Pond Inlet (**Figure 1**). Other alpha-diversity indices (observed_OTUs and chao1) revealed the same trends (**Supplementary Tables 1, 2**). However, the community evenness (Simpson_e), ranged from 0.02 to 0.06 regardless of the sample type or season or year (**Supplementary Tables 1, 2**).

The alpha diversity of bacterial community in Pond Inlet sludge samples, which were only obtained in 2014, increased over the treatment season, a change which occurred simultaneously

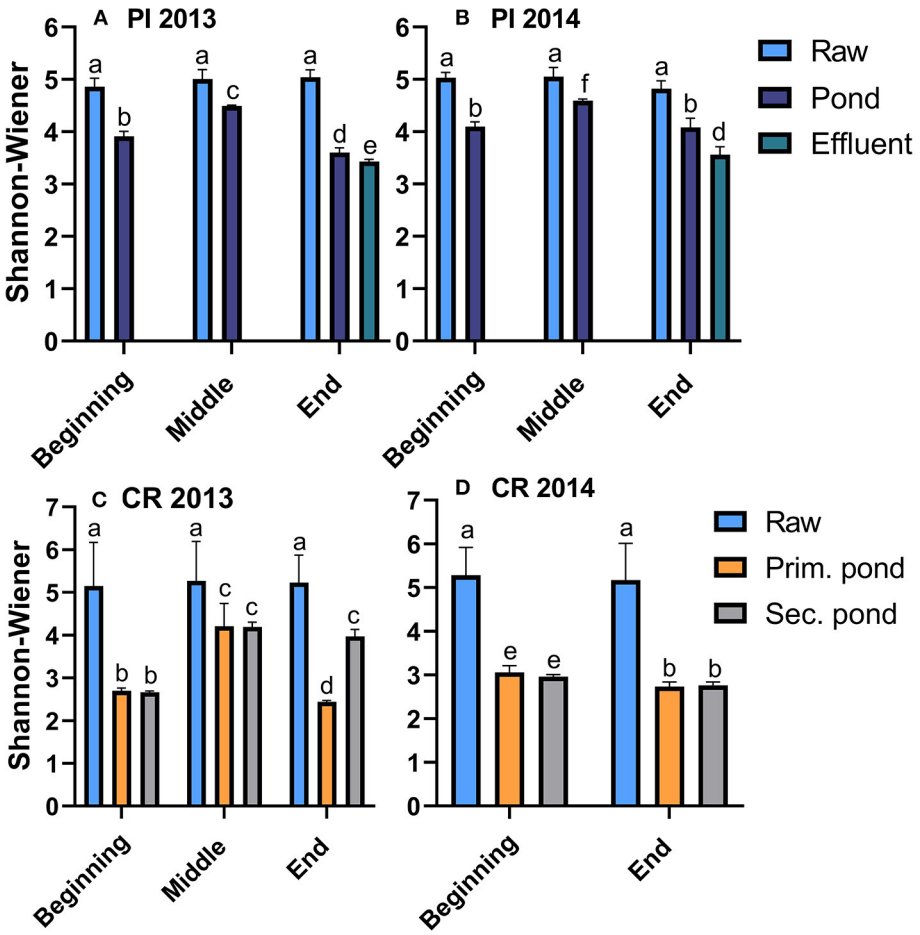


FIGURE 2 | Alpha diversity (Shannon-Wiener Index) in raw and treated wastewater samples from Pond Inlet [PI, **(A,B)**] and Clyde River [CR, **(C,D)**] in 2013 and 2014. Bars with different letters within the same community are significantly ($p < 0.05$) different from each other.

with growing bacterial populations with final levels of 9.9 log 16S rRNA copies/mL (**Supplementary Table 1**).

Beta Diversity and Core Microbiome in Raw Wastewater

The weighted and unweighted UniFrac measures showed that the bacterial beta diversity was not significantly ($p > 0.05$) different in raw wastewater samples obtained from the two communities during the study period. The raw wastewater core microbial community was comprised of 18 bacterial families mainly belonging to the *Proteobacteria* phylum. Dominant *Proteobacteria* families were *Aeromonadaceae* (25.2–41.8%), followed by *Rhodospirillaceae* (8.5–20.9%), *Comamonadaceae* (2.7–18.6%), *Enterobacteriaceae* (3.3–11.5%), *Pseudomonadaceae* (5.4–10.5%), and *Campylobacteraceae* (1.1–6.2%) (**Figure 3**). Within the *Bacteroidetes* phylum and *Bacteroidia* class, two families were present; *Bacteroidaceae* with a prevalence of 4.2–12.3% and *Porphyromonadaceae* with abundances ranging from 2.4 to 8.5%. The genus level composition of the core microbiomes in the raw wastewater samples is shown in **Supplementary Figure 3**.

Treatment and Changes in Phylogenetic Beta Diversity in Arctic WSP Systems

Treatment in the WSP systems in Pond Inlet and Clyde River significantly ($p < 0.05$) altered the phylogenetic beta diversity of bacterial communities. For Pond Inlet, this observation is exemplified by a comparison of raw, pond, sludge and effluent samples from September 2014, which showed different distributions of bacterial families along the treatment process (**Figure 4A**). Here, the *Aeromonadaceae* family (40.0%) dominated in the raw wastewater samples but became a minor group (1.8%) in the pond samples. Sludge samples were dominated by the *Pseudomonadaceae* family (29.7%). Pond and effluent samples resembled each other with the most abundant group being *Comamonadaceae* (∼41%), *Campylobacteraceae* (∼15%), and *Geobacteraceae* (∼13%). The above mentioned families were represented at the genus level by *Rhodoferax*, *Arcobacter*, and *Geobacter* in the pond and effluent samples (**Supplementary Figure 4**). The resulting PCoA plot (**Figure 4B**) from the weighted UniFrac analysis shows that 53.3% of the variation in microbial beta diversity depended on the sample type while another 24.0% of the variation was due

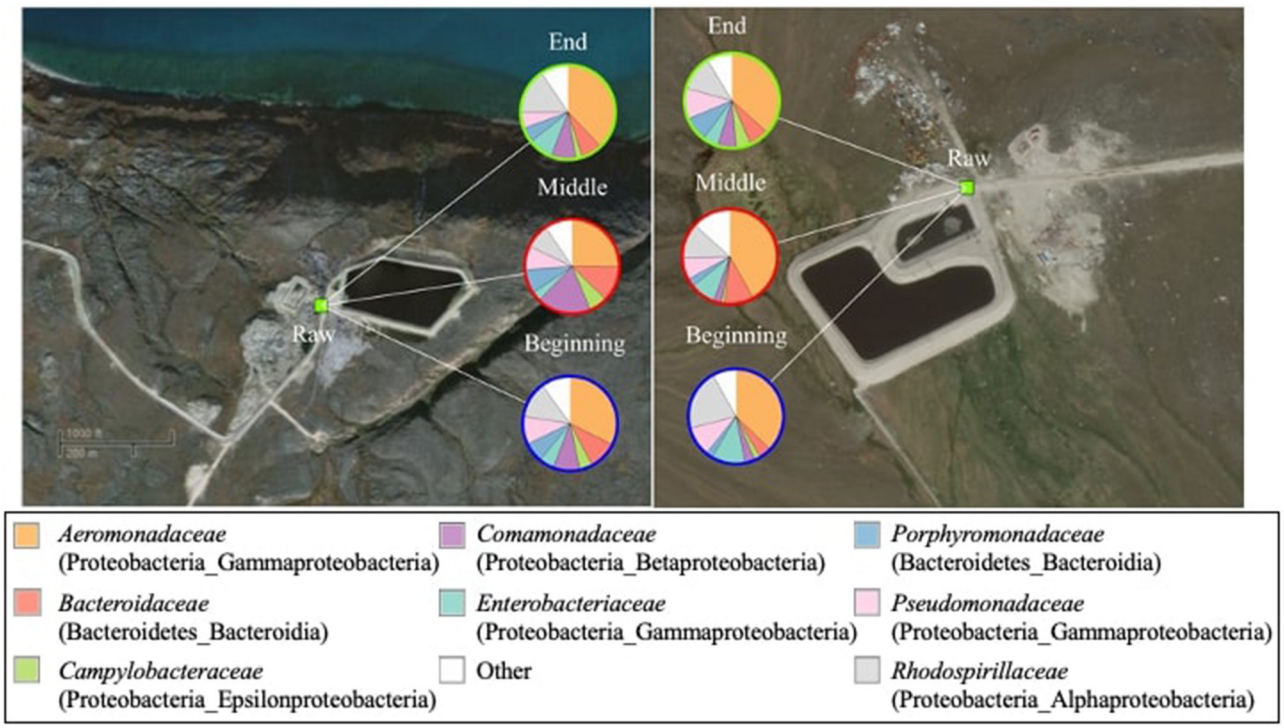


FIGURE 3 | Composition of the core bacterial community (family level) in raw wastewater samples ($n = 34$ in total) from Pond Inlet (left) and Clyde River (right) obtained in 2013–2014. The relative abundances of bacterial families (top eight of 18 families) are shown in the pie charts. Legends on top of the pie charts (Beginning, Middle, and End) indicate the sampling time. Bacterial families and their associated phylum and class names are listed in brackets. Other: sum of bacterial families with relative abundances below 10%.

to the difference between raw wastewater vs. samples from the WSP. Results from Clyde River showed similar changes in the composition of bacterial communities along the WSP treatment train with *Geobacteraceae*, *Campylobacteraceae*, and *Comamonadaceae* making up 50% of the bacterial community in the secondary pond (**Supplementary Figure 5**).

Structure of Bacterial Communities in Relation to Wastewater Treatment in the WSPs

The time of sampling within the treatment season significantly ($p < 0.05$) influenced the phylogenetic beta diversity of bacterial communities in the arctic WSPs as did the sampling year. These seasonal and annual differences in pond bacterial communities are shown for Pond Inlet in **Figure 5**. Results from Clyde River showed similar trends and the results from 2014 are presented in **Figure 6**.

In Pond Inlet in 2013, early season samples were dominated by *Pseudomonadaceae* (49.5%) and *Campylobacteraceae* (13.6%) (**Figure 5A**). Communities then shifted to a mid-season mix dominated by *Comamonadaceae* (36.8%) and *Campylobacteraceae* (24.5%) and finally to an end-season blend with *Comamonadaceae* (31.7%), *Geobacteraceae* (16.8%), and *Porphyromonadaceae* (10%) dominating. For the annual differences, the mid-season dominating family *Comamonadaceae* varied between 36.8 and 70.0% in 2013 and 2014, respectively. When switching to the end-season, the relative abundance of *Comamonadaceae* decreased to 31.7% in 2013 and 44.3% in 2014 (**Figure 5A**).

The dbRDA plot relating the measured environmental and water quality factors to the weighted UniFrac beta diversity among pond communities in Pond Inlet revealed that temperature, $CBOD_5$ and TA were significant factors ($p < 0.05$) driving differences in the bacterial communities (**Figure 5B**). The $CBOD_5$ content and *Pseudomonadaceae* and *Campylobacteraceae* correlated positively with the beginning of the treatment season, while the negative correlation at the middle and end of the treatment season showed the role of the microbiota (i.e., *Comamonadaceae*) in removing $CBOD_5$, a development that correlated with the higher temperatures in the middle of the season and its effect on the bacterial communities. The decrease in TP correlated with increases in *Geobacteraceae,* and members of the anaerobic *Bacteroidales* (incl. *Porphyromonadaceae*) at the end of the treatment season. The dbRDA plot also revealed differences between 2013 and 2014 microbial communities obtained in the middle and end of the season supporting the marked effect of the ambient temperature on the microbiota and wastewater treatment in the Pond Inlet WSP system.

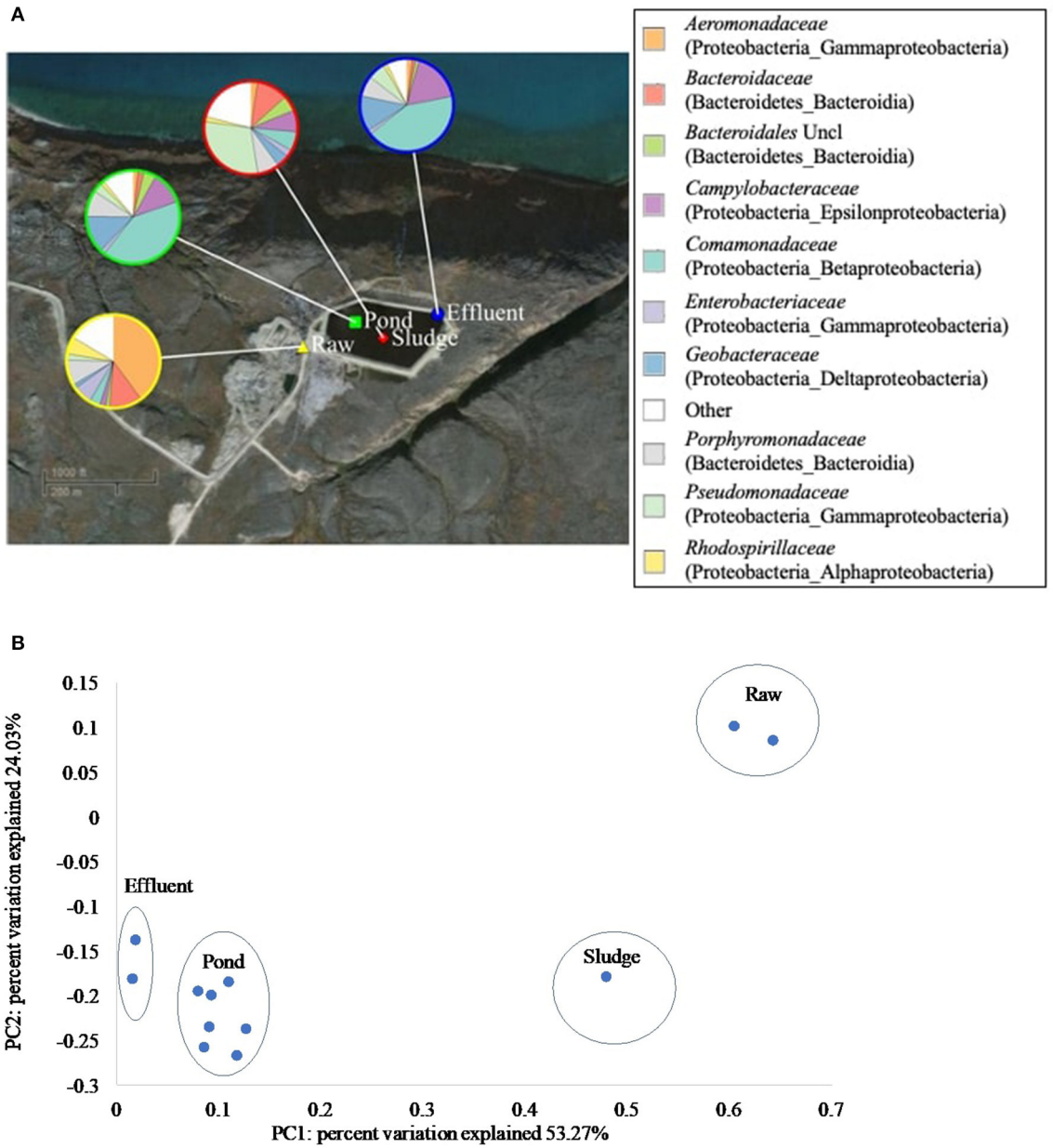


FIGURE 4 | Composition and phylogenetic beta-diversity of bacterial communities along the Pond Inlet WSP treatment train in September 2014. **(A)** Relative abundance of bacterial families (top 10 of 269 families) in wastewater samples collected along the treatment process. The legend shows the bacterial families with the phylum and class names listed in brackets. Uncl: unclassified. Other: sum of bacterial families with relative abundances below 10%. **(B)** PCoA plot presented the phylogenetic beta-diversity of bacterial communities measured by the weighted UniFrac metric.

For Clyde River, *Pseudomonadaceae* dominated in the primary pond at the beginning of the 2014 season, while late season primary pond samples contained a diverse community including *Campylobacteraceae*, *Geobacteraceae*, *Rhodopirillaceae*, and *Comamonadaceae* (**Figure 6A**). A similar change from a *Pseudomonadaceae* dominated (about 50%) community to a highly diverse community at the end of the season was seen in the secondary pond samples, where the proportion of *Porphyromonadaceae* and *Comamonadaceae* increased together with *Geobacteraceae*. These seasonal trends were reflected in the phylogenic UPGMA tree with the first branch splitting at the beginning/end of season followed by a second split into the primary/secondary pond. The dbRDA plot relating the measured environmental and water quality factors to the beta diversity among pond communities in Clyde River revealed that temperature, $CBOD_5$, TA, TSS, TN, and TP were

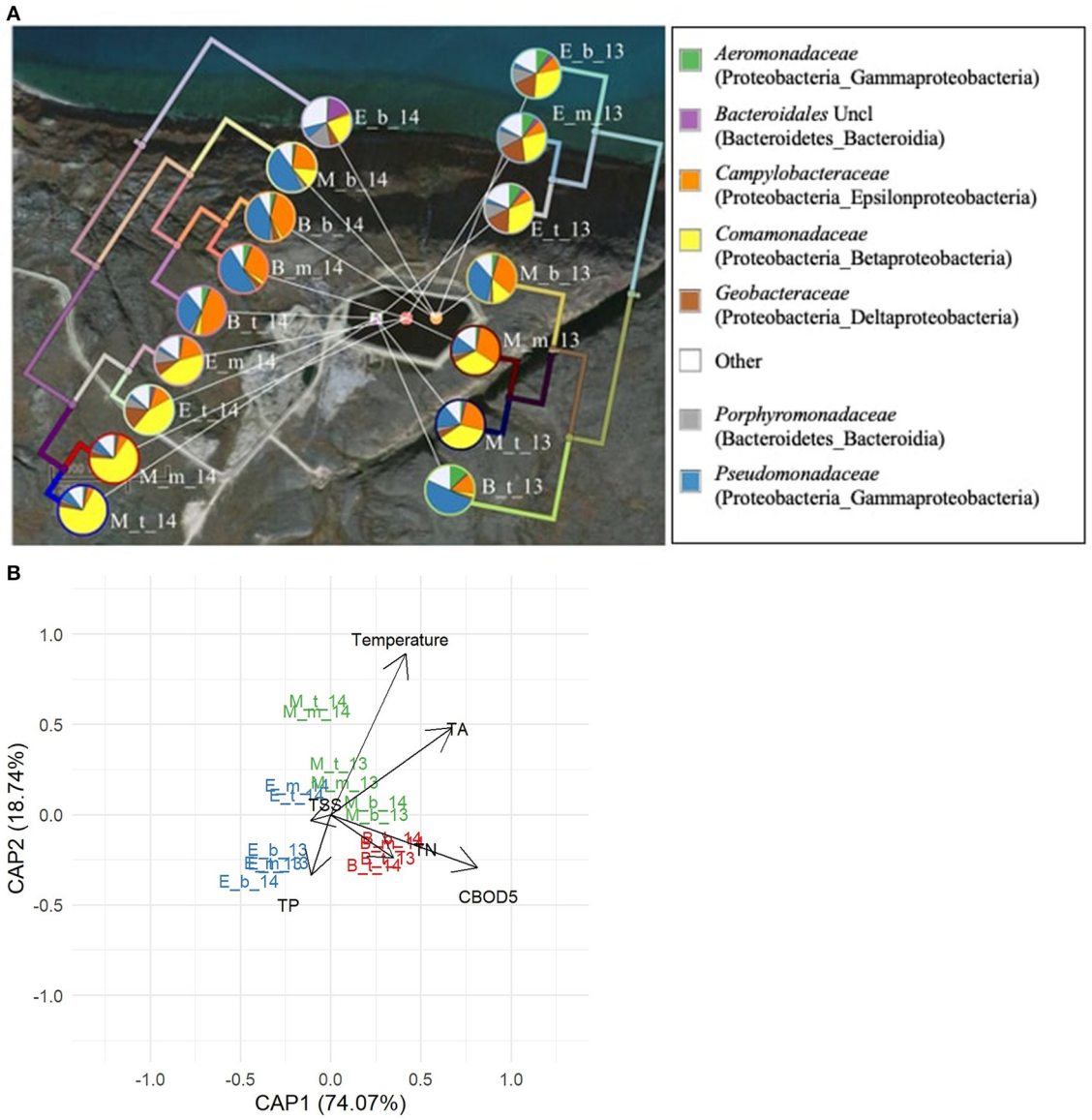


FIGURE 5 | Bacterial communities in the Pond Inlet WSP during the 2013 (right side) and 2014 (left side) treatment seasons. **(A)** The seven most abundant families (relative abundances ≥ 10% among 269 bacterial families) are presented in the pie charts. The UPGMA tree shows the bacterial phylogenetic beta-diversity (weighted UniFrac metric) in each sampling year. Legends beside the pie charts: B – Beginning, M – Middle, E – End of summer treatment season; t: top, m: middle, and b: bottom sampling depths in the WSP, respectively. The legend (right) denotes the bacterial families with the phylum and class names in brackets. Uncl: unclassified. Other: sum of bacterial families with relative abundances below 10%. **(B)** Distance-based RDA plot showing the relationship between temperature, water quality measures and the seasonal and annual phylogenetic bacterial beta-diversity in pond samples. Significant factors ($p < 0.01$) were temperature, $CBOD_5$ and TP.

significantly ($p < 0.05$) correlated to the observed differences (**Figure 6B**).

The UPGMA tree for Pond Inlet revealed that bacterial communities in the bottom layer differed significantly ($p < 0.05$) from the top or middle layers during the 2014 treatment season, but not in 2013 (**Figure 5A**). Interestingly, the different sampling depths in Clyde River two-cell WSP did not affect ($p > 0.05$) microbial community diversity in 2014 (**Figure 6A**).

Functional Content Predictions for Bacterial WSP Communities

The two most significant ($p < 0.05$) KEGG pathways, which were predicted to be present in the bacterial communities in WSP samples, are associated with carbohydrate and energy metabolism. As shown in **Figure 7A**, the predicted proportion of sequences associated with carbohydrate metabolism was significantly higher ($p < 0.05$) during the middle of the season compared to the beginning and end of the 2014 treatment season

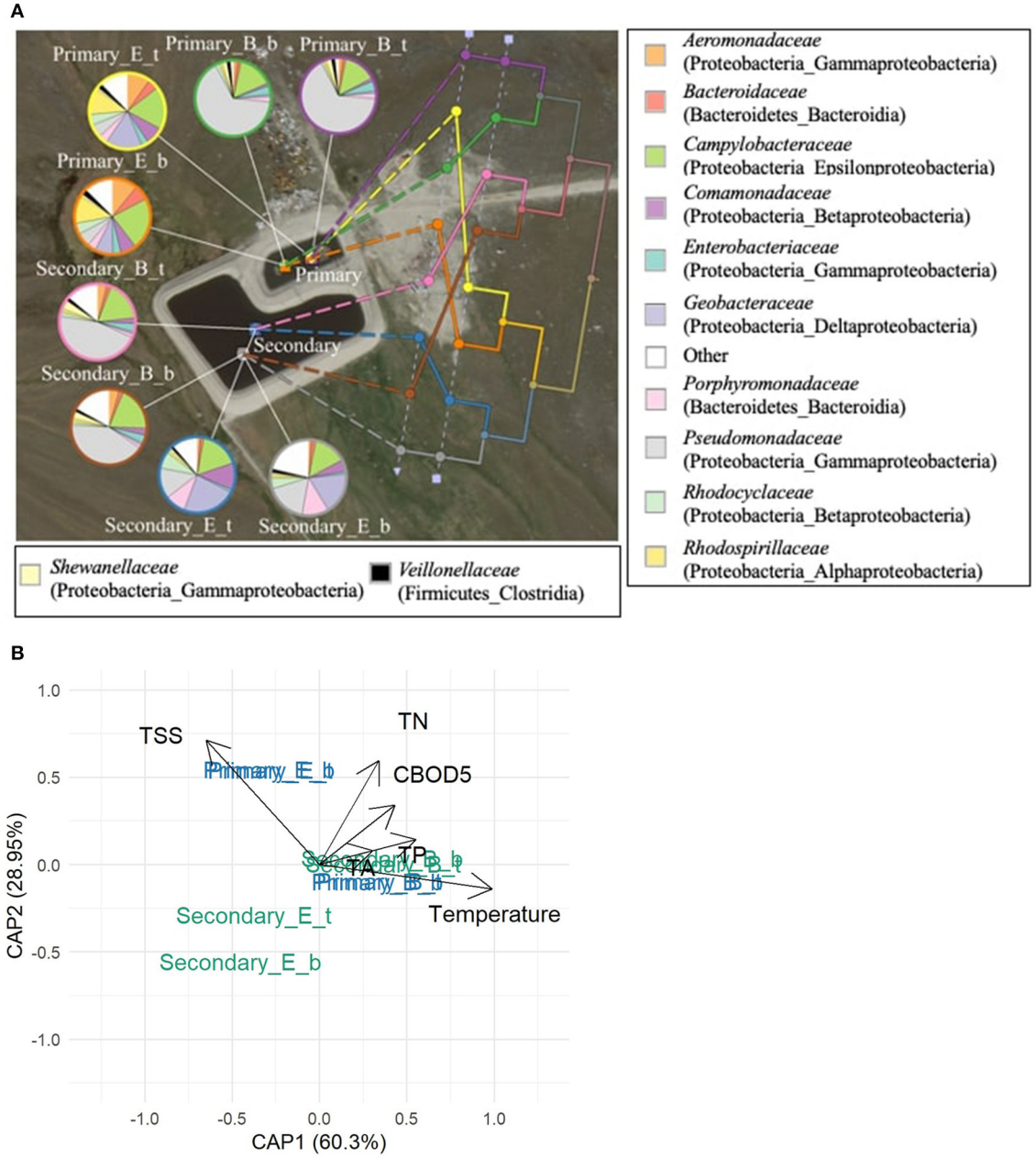


FIGURE 6 | Bacterial communities in Clyde River WSP system during the 2014 treatment season. **(A)** The UPGMA tree shows the phylogenetic beta-diversity of bacterial communities measured by the weighted UniFrac metric. The 12 most abundant bacterial families (relative abundances ≥ 10%, total of 269 bacterial families) are presented in the pie charts. Legends beside the pie charts: Primary – Primary pond, Secondary – Secondary pond; B – Beginning, E – End of summer treatment season; t – top and b – bottom indicates the sampling sites in the WSP, respectively. The legend beside the map shows bacterial families: the phylum and class names are listed in the bracket of each family name. Other: sum of bacterial families with relative abundances <10%. **(B)** Distance-based RDA plot showing the relationship between temperature, water quality measures (TSS, TN, $CBOD_5$, TA, TP, all significant, $p < 0.05$) and the seasonal phylogenetic bacterial beta-diversity in the pond samples.

in Pond Inlet. The carbohydrate metabolism was predicted to be significantly ($p < 0.05$) higher in the bacterial community of the secondary pond in Clyde River compared to the primary pond (**Figure 7B**). The energy metabolism in Pond Inlet and Clyde River bacterial WSP communities followed the same trend (data not shown). The predicted abundance of genes responsible for the carbohydrate and energy metabolism in mid-season pond samples were significantly ($p < 0.05$) higher in 2014 compared to 2013 (data not shown).

In both Pond Inlet and Clyde River WSPs, PICRUSt identified low proportions (0.002–0.006%) of 16S rRNA genes associated with the three KEGG pathways that encode ammonia monooxygenase, an enzyme which is involved in microbial removal of ammonia. This agrees with the relative low abundance

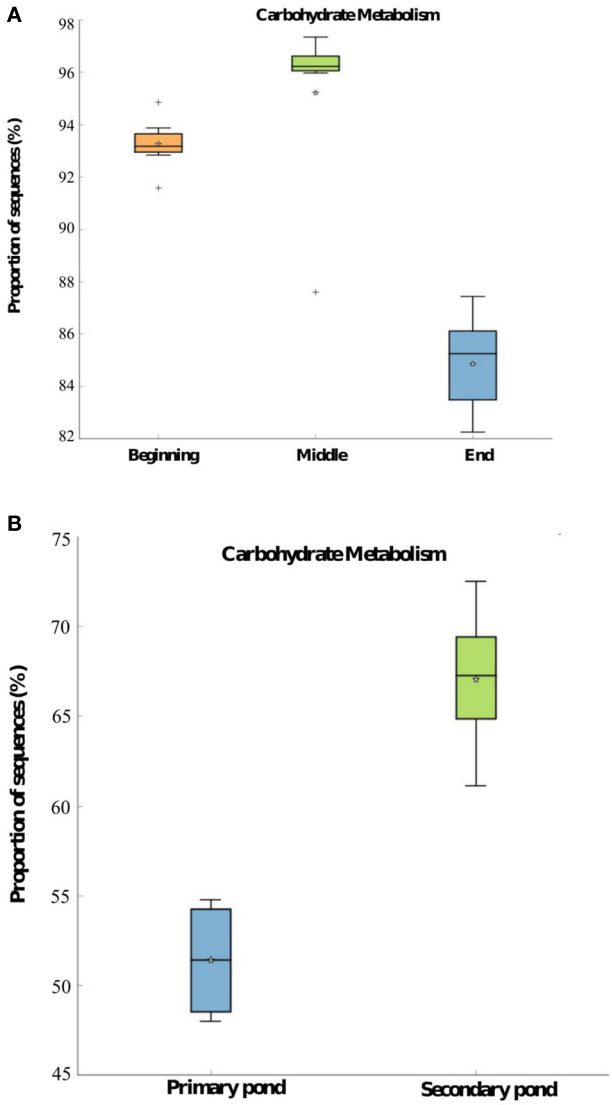


FIGURE 7 | PICRUSt predicted abundance of genes belonging to carbohydrate metabolism with significant differences ($p < 0.05$) based on the mean proportions and identities of 16S rRNA genes in pond samples obtained **(A)** from the beginning to the end of the 2014 treatment season in Pond Inlet, and **(B)** from the primary and secondary pond in Clyde River at the end of the 2014 treatment season. Significant differences ($p < 0.05$) were detected by the Kruskal-Wallis H-test with a multiple test correction Benjamini-Hochberg FDR for the Pond Inlet analysis and the Welch's *t*-test with a multiple test correction Benjamini-Hochberg FDR for the Clyde River analysis. *: mean. +: Outliers.

(0.01%) of both the *Nitrosomonadaceae* and *Crenotrichaceae* families, which are associated with the three pathways, in the WSPs bacterial communities.

DISCUSSION

The treatment of the wastewater resulted in removal of 70–90% of the content of TSS and $CBOD_5$ and reduction of *E. coli* levels to 10^{4-6} MPN/100 mL (Huang et al., 2018) leading to treated water qualities (effluent in Pond Inlet and water in the secondary pond at the end of the season in Clyde River), which complied with the criteria set by the licensing Nunavut Water Board of there being $<10^4$-10^6 *E. coli* MPN/100 mL, 180 mg/L TSS, and 120 mg/L $CBOD_5$ in water released into the recipient (Nunavut Water Board, 2014). These criteria have been set in consideration of the climate, nature of the treatment systems, receiving environment and population density in Nunavut and are different from the Canadian Wastewater Systems Effluent Regulations (WSERs), which require maximum contents of 25 mg/L $CBOD_5$, 25 mg/L TSS, and 1.25 mg/L un-ionized ammonia for wastewater effluents in Southern Canada (Environment Canada, 2015). The observed removal of nutrients is likely due to a combination of biological processes, physical settling processes, dilution (due to snow and rainfalls) and evaporation events following the disposal of the wastewater into the WSPs (Heinke et al., 1991; Ragush et al., 2015).

The bacterial communities loaded into the WSPs were similar as the influent raw wastewater contained a core set of bacteria in spite of coming from the two geographically separated arctic communities. It may be that low per capita consumption of drinking water and absence of industrial or agricultural activities in both communities led to the presence of this core bacterial community dominated by *Proteobacteria* (**Figure 3**). Treatment subsequently led to significant changes to the composition of bacterial communities and a 2-log reduction in 16S rRNA gene copies/ml in the treated water at the end of the treatment season. This reduction was similar to the 2-3 log removal of total coliform bacteria and *Escherichia coli* previously reported for these WSPs (Huang et al., 2018). This suggests a preferential reduction in enteric bacteria, i.e., a disinfection effect, which was supported by a shift from dominance of fecally related *Aeromonadaceae*, *Bacteroidaceae*, *Campylobacteraceae*, and *Enterobacteriaceae* in influent wastewater toward *Comamonadaceae, Porphyromonadaceae* (*Paludibacter*), and *Geobacteraceae* in the treated water (**Figure 4** and **Supplementary Figure 5**). A similar reduction in the content of enteric bacteria and shift in bacterial communities was observed in a comparative study of the influent and effluent of a German WWTP (Numberger et al., 2019). The disinfection effect may be due to pond environmental factors (temperature, DO, pH, etc.), microbial completion and/or antagonism and solar irradiation although the underlying mechanisms remain unknown (Ho and Goethals, 2020).

Monitoring of the WSP environment throughout the 2013 and 2014 summer treatment season showed that the intended facultative ponds stayed anaerobic with very low levels of oxygen, neutral pH-values and therefore no signs of algae growth (**Supplementary Figures 1, 2**). This was different from 2011 where a reduced operational depth in the pond, yielding a lower OLR ($kg/m^2/day$), and an unusual warm summer led to an algal bloom in the Pond Inlet WSP (Ragush et al., 2015). The lack of algae in the WSPs could affect the wastewater treatment as normal WSP treatment is based on growth of algae to supply oxygen to the aerobic decomposing bacteria, sequester phosphorous and disinfection due to the increased alkalinity of the water during algal blooms (Amengual-Morro et al., 2012; Ho and Goethals, 2020). The absence of algae growth may be related to the high OLRs and low temperatures in the

ponds as demonstrated in model experiments (Ragush et al., 2017). Since the WSPs remained anaerobic and with neutral pH throughout the treatment season, temperature became an important driver of changes in the bacterial communities over the treatment season. When the pond temperatures peaked mid-season, the pond bacterial community became more abundant and diverse. This coincided with improved removal of fecal indicator and pathogenic bacteria as well as nutrients (**Table 1**; Huang et al., 2018), suggesting that disinfection and nutrient removal are correlated to bacterial richness and diversity in WSPs. The correlation between temperature, nutrient removal and bacterial richness and alpha-diversity was similarly observed in seasonal studies of sewage batch reactors (Ebrahimi et al., 2010) and WWTPs operated in colder climates (Rodriguez-Caballero et al., 2012; Ju et al., 2014; Kang et al., 2018). The multivariate dbRDA supported that largely anaerobic biological treatment of the wastewater was closely related to bacterial community shifts and temperature in the WSPs (**Figures 5B**, **6B**). This potentially makes the treatment success unpredictable due to annual differences in the ambient summer temperatures. However, when comparing a cold (2013) vs. a warm (2014) year, treatment outcomes in terms of effluent $CBOD_5$, TSS, and TA were still comparable (**Table 1**), indicating some robustness of the biological processes as well as concurrent impact of the physical processes.

Decomposing microorganisms with a wide range of metabolisms are important for removal of $CBOD_5$ and other nutrients in WSP systems (Wagner et al., 2002; Ferrera and Sánchez, 2016; Ho and Goethals, 2020). Bacterial community changes along the treatment process (raw to pond, and season) led to reductions in alpha diversity and increases in *Comamonadaceae, Geobacteraceae*, and *Porphyromonadaceae*. Similar impacts of the treatment process were observed in the study of several full-scale municipal WWTPs (Hu et al., 2012; Gao et al., 2016; Numberger et al., 2019; Xue et al., 2019). The members of the bacterial communities were not evenly composed in the arctic WSP systems as indicated by low Simpson evenness measures (0.02–0.06) in all samples. However, the unevenness of bacterial community composition in arctic WSPs was not unexpected as *Proteobacteria* was by far the most abundant bacterial phylum with a relative abundance of 70–90% in all samples. The same dominance of *Proteobacteria* was also reported by Gromala et al. (2021) for three arctic WSPs located in Western Nunavut and in seven Finnish WWTPs operated above the Arctic Circle (Gonzalez-Martinez et al., 2018). Members of this bacterial phylum are known to be responsible for removal of the organic waste in municipal wastewater (Wagner et al., 2002). *Comamonadaceae* (member of the *Burkholderiales* order in *Betaproteobacteria*) was the most dominant group in pond and treated effluent samples from Pond Inlet and was in the dbRDA related to removal of nutrients ($CBOD_5$). This family of bacteria has previously been reported to be involved in the biological nutrient removal processes, especially carbon and nitrogen containing compounds when the oxygen supply was limited (Sato et al., 2016; Usharani, 2019) as was the case in the present study. At the genus-level, *Rhodoferax* spp. (*Comamonadaceae*) became the dominant group in the pond and effluent samples (**Supplementary Figure 4**). This genus was recently reported to be present in the WSP in Baker Lake (Gromala et al., 2021), indicating its presence in the arctic sewage system, and is known to decompose organic material (Zhang et al., 2012). The *Geobacter* genus [*Geobacteraceae, Deltaproteobacteria* or the proposed *Desulfobacterota* (Waite et al., 2020)], which started as a minor group in the raw wastewater and rose in abundance in the WSP samples, is associated with cold anaerobic soil and aquatic environments and an ability to reduce iron and sulfate, and can complete oxidize organic compounds (Lovley et al., 1987), and was observed in the anaerobic tank of industrial wastewater treatment systems operating at low temperatures (8–16°C) in China (Zhang et al., 2017). *Porphyromonadaceae* and other members of the anaerobic *Bacteroidetes* are known to degrade complex organic substances in WWTPs (Gao et al., 2016; Zhang et al., 2017).

In 2014 samples from the Pond Inlet WSP, the bacterial diversity depended on the sampling depth, which may be due to the solids settling processes, temperature shifts between the top and bottom layer of the pond, DO gradients, and other unmonitored wastewater and operational factors. However, this depth-related difference was not observed in samples from 2013 or from Clyde River, which may indicate variable mixing of the wastewater in the ponds depending on the year and WSP design. Sludge is known to accumulate in arctic WSPs (Miyamoto and Heinke, 1979). An examination of the sludge in Pond Inlet during the 2014 treatment season revealed increases in sludge community size and diversity (**Figure 4** and **Supplementary Table 1**), indicating the role of the sludge community in treatment of the wastewater in arctic WSPs should be the subject of further investigations.

The predicted functional gene content of wastewater bacterial communities pointed to higher levels of carbohydrate and energy metabolism in the middle of the treatment season in Pond Inlet WSP, concurring with the $CBOD_5$ removal observed between the beginning and middle of the season in 2014 (**Table 1**). In the two-cell WSP system in Clyde River, the predicted higher carbohydrate and energy metabolism of the secondary pond bacterial communities similarly coincided with the further reduction of $CBOD_5$ levels following initial treatment in the primary pond (**Table 1**). Gromala et al. (2021) also reported presence of KEGG pathways associated with carbohydrate and energy metabolisms in the Western Nunavut WSPs. Taken together, it seems likely that optimal $CBOD_5$ removal can be related to higher pond temperatures, larger bacterial populations, greater bacterial diversity, and enhanced carbohydrate and energy metabolic activity levels in mid-season WSPs.

The low proportion of genes associated with ammonium oxidizing pathways matched the low relative abundance of ammonia oxidizing bacteria in both WSPs. This finding was supported by the water quality data (**Table 1**), which showed a modest ammonia (TA) removal in both WSPs during the study period. Interestingly, Gromala et al. (2021) reported absence of ammonia oxidation pathways in WSPs in Baker Lake, and Cambridge Bay, while the genes in the pathway were detected in Kugluktuk. These findings overall suggest that the current design of arctic WSPs does not promote growth of microbial communities, which substantially contribute to the removal of ammonia.

CONCLUSIONS

The study explored the composition and putative function of bacterial communities in arctic WSPs treating municipal wastewater during two consecutive summer treatment seasons. The WSPs received raw wastewater which had a high nutrient content and similar bacterial communities despite the geographical separation. Although designed to be facultative, the ponds stayed anaerobic with neutral pH-values, and temperature in the WSPs became a major determinant of the size, composition and diversity of the bacterial communities. The multivariate dbRDA showed correlations between temperature, nutrient removal and increased levels of *Comamonadaceae, Geobacteraceae*, and *Porphyromonadaceae*. The predicted gene functions (KEGG pathway study) supported that increased microbial removal of nutrients occurred mid-season and in the secondary pond of the two-cell Clyde River WSP system.

While it is challenging to control environmental factors in a passive wastewater treatment system there are some design considerations that could be used to optimize temperature regimes, such as the depth of the pond. Also, use of ponds in sequence would improve treatment. Information from this study would also be useful for further development of models that predict biological treatment, which could be used to size and assess the feasibility of WSPs in extreme climates.

RESEARCH PERMITS

The permit for this work is registered with the Nunavut Research Institute under the following title:

"Northern Municipal Wastewater Effluent Discharge Quality Objectives in the Context of Canadian Council of Ministers of the Environment Strategy and Environment Canada's Wastewater Systems Effluent."

DATA AVAILABILITY STATEMENT

The sequencing dataset presented in this study has been deposited at NCBI as a Bioproject PRJNA630784. The remaining data can be found in the **Supplemental Material**.

AUTHOR CONTRIBUTIONS

YH: conception, execution of field work and experiments, data analyses, bioinformatics analysis, and writing of manuscript. CR: execution of field work, analysis of wastewater quality, and review of manuscript. LJ: statistical analyses, graphing, and review of manuscript. MH: bioinformatics analysis and review of manuscript. RB: conception, supervision, bioinformatics analysis, and review of manuscript. RJ: funding acquisition, conception, supervision, and writing and review of manuscript. LT: funding acquisition, conception, supervision, and writing and review of manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The Dalhousie University Northern Wastewater Research team would like to thank the Government of Nunavut, Natural Science and Engineering Research Council of Canada (NSERC): Systems Training and Education in Water Assets Research and Development (STEWARD) and the Canadian Water Network for providing funding in support of this study.

ACKNOWLEDGMENTS

Our team would like to extend our gratitude to the communities of Pond Inlet and Clyde River for their support of this study. Our gratitude also goes to the Nunavut Research Institute for providing laboratory space in Iqaluit. Finally, our team members (Amy Jackson, Emmalina Corriveau, Joanna Poltarowicz, Justine Lywood, Mark Greenwood, and Meggie Letman) are sincerely thanked for their help in both laboratory and field work. The funding agencies were not involved in conducting the research or preparing this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.710853/full#supplementary-material

REFERENCES

Amengual-Morro, C., Niell, G. M., and Martínez-Taberner, A. (2012). Phytoplankton as bioindicator for waste stabilization ponds. *J. Env. Man.* 95, S71–S76. doi: 10.1016/j.jenvman.2011.07.008

American Public Health Association (1999). *Standard Methods for the Examination of Water and Wastewater*. Baltimore: United Book Press.

Balch, G., Hayward, J., Jamieson, R., Wooton, B., and Yates, C. N. (2018). "Recommendations for the use of tundra wetlands for treatment of municipal wastewater in Canada's far North," in *Multifunctional Wetlands*, eds N. Nagabhatla and C. Metcalfe (Cham: Springer Nature), 83–120. doi: 10.1007/978-3-319-67416-2_3

Beck, M. W. (2020). *ggord: Ordination PlotsWith ggplot2. R Package Version 1.1.5*. Available online at: https://zenodo.org/record/3828862#.YQqirS271pQ (accessed July 18, 2021).

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010b). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinform* 26, 266–267. doi: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010a). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J. R., Smidt, H., et al. (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* 4:e6669. doi: 10.1371/journal.pone.0006669

Comeau, A. M., Douglas, G. M., and Langille, M. G. I. (2017). Microbiome helper: a custom and streamlined workflow for microbiome research. *mSystems* 2, e00127–e00116. doi: 10.1128/mSystems.00127-16

Daley, K. D., Castleden, H., Jamieson, R., Furgal, C., and Ell, L. (2014). Municipal water qualities and health in Nunavut households: an exploratory case study in Coral Harbour, Nunavut, Canada. *Int. J. Circumpolar Hlth.* 73:23843. doi: 10.3402/ijch.v73.23843

Ebrahimi, S., Gabus, S., Rohrbach-Brandt, E., Hosseini, M., Rossi, P., Maillard, J., et al. (2010). Performance and microbial community composition dynamics of aerobic granular sludge from sequencing batch bubble column reactors operated at 20°C, 30°C, and 35°C. *Appl. Microbiol. Biotechnol.* 87, 1555–1568. doi: 10.1007/s00253-010-2621-4

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Environment Canada (2014a). *Canadian Climate Normals 1981-2010 Station Data. Pond Inlet A Station, Nunavut*. Available online at: http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=1774&autofwd=1 (accessed May 3, 2014).

Environment Canada (2014b). *Canadian Climate Normals 1981-2010 Station Data. Clyde River A Station, Nunavut*. Available online at: http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=1743&lang=e&dCode=0&province=NU&provBut=Go&month1=0&month2=12 (accessed July 18, 2021).

Environment Canada (2015). *Wastewater Systems Effluent Regulations*. Available online at: http://www.ec.gc.ca/eu-ww/ (accessed December 15, 2016).

Ferrera, I., and Sánchez, O. (2016). Insights in microbial diversity in wastewater treatment systems: How far have we come? *Biotechnol. Adv.* 34, 790–802. doi: 10.1016/j.biotechadv.2016.04.003

Gao, P., Xu, W., Sontag, P., Li, X., Xue, G., Liu, T., et al. (2016). Correlating microbial community compositions with environmental factors in activated sludge from four full-scale municipal wastewater treatment plants in Shanghai, China. *Appl. Microbiol. Biotechnol.* 100, 4663–4673. doi: 10.1007/s00253-016-7307-0

Gonzalez-Martinez, A., Sihvonen, M., Muñoz-Palazon, B., Rodriguez-Sanchez, A., Mikola, A., and Vahala, R. (2018). Microbial ecology of full-scale wastewater treatment systems in the Polar Arctic Circle: archae, bacteria and fungi. *Sci. Rep*. 8:2208. doi: 10.1038/s41598-018-20633-5

Gromala, M., Neufeld, J. D., and McConkey, B. J. (2021). Monitoring microbial population and antibiotic reistance gene enrichment associated with arctic waste stabilization ponds. *Appl. Environ. Microbiol.* 87, e02914–e02920. doi: 10.1128/AEM.02914-20

Heinke, G. W., Smith, D. W., and Finch, G. R. (1991). Guidelines for the planning and design of wastewater lagoon systems in cold climates. *Can. J. Civ. Eng.* 18, 556–567. doi: 10.1139/l91-068

Ho, L., and Goethals, P. L. M. (2020). Municipal wastewater treatment with pond technology: Historical review and future outlook. *Eco. Eng.* 148:105791. doi: 10.1016/j.ecoleng.2020.105791

Hu, M., Wang, X., Wen, X., and Xia, Y. (2012). Microbial community structures in different wastewater treatment plants as revealed by 454-Pyrosequencing analysis. *Bioresour. Technol.* 117, 72–79. doi: 10.1016/j.biortech.2012.04.061

Huang, Y., Truelstrup Hansen, L., Ragush, C. M., and Jamieson, R. C. (2018). Disinfection and removal of human pathogenic bacteria in arctic waste stabilization ponds. *Environ. Sci. Pollut. Res.* 25, 32881–32893. doi: 10.1007/s11356-017-8816-9

Huse, S. M., Ye, Y., Zhou, Y., and Fodor, A. A. (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* 7:e34242. doi: 10.1371/journal.pone.0034242

Ju, F., Guo, F., Ye, L., Xia, Y., and Zhang, T. (2014). Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ. Microbiol. Rep.* 6, 80–89. doi: 10.1111/1758-2229.12110

Kang, X., Wu, X., Wang, H., Zeng, X., Leng, Y., and Li, S. (2018). Illumina sequencing reveals bacterial community shift and its role in a full-scale A_2O sewage treatment process at low temperatures. *Desalin. Water Treat.* 109, 193–204. doi: 10.5004/dwt.2018.22142

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lovley, D. R., Stolz, J. F., Nord, G. L., and Phillips, E. J. P. (1987). Anaerobic production of magnetite by a dissimilatory iron-reducing microorganism. *Nature* 330: 252–254. doi: 10.1038/330252a0

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Marais, G. V. R. (1974). Fecal bacterial kinetics in stabilization ponds. *J. Environ. Eng.* 100, 119–139. doi: 10.1061/JEEGAV.0000138

Miyamoto, H. K., and Heinke, G. W. (1979). Performance evaluation of an arctic sewage lagoon. *Can. J. Civ. Eng.* 6, 324–328. doi: 10.1139/l79-033

Numberger, D., Ganzert, L., Zoccarato, L., Mülhdorfer, K., Sauer, S., Grossart, H.-P., et al. (2019). Characterization of bacterial communities in wastewater with enhance taxonomic resolution by full-length 16S rRNA sequencing. *Sci. Rep*. 9:9673. doi: 10.1038/s41598-019-46015-z

Nunavut Water Board (2014). *Public Registry FTP. Nunavut Water Board*. Available online at: https://www.nwb-oen.ca/content/public-registry (accessed January 5, 2014).

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). *vegan: Community Ecology Package. R package version 2.5-7*. Available online at: https://CRAN.R-project.org/package=vegan (accessed July 18, 2021).

Parks, D. H., Mankowski, T., Zangooei, S., Porter, M. S., Armanini, D. G., Baird, D. J., et al. (2013). GenGIS 2: geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. *PLoS ONE* 8:e69885. doi: 10.1371/journal.pone.0069885

Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *J. Bioinform.* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed July 18, 2021).

Ragush, C. M., Poltarowicz, J. M., Lywood, J., Gagnon, G. A., Truelstrup Hansen, L., and Jamieson, R. C. (2017). Environmental and operational factors affecting carbon removal in model arctic waste stabilization ponds. *Ecol. Eng.* 98, 91–97. doi: 10.1016/j.ecoleng.2016.10.031

Ragush, C. M., Schmidt, J. J., Krkosek, W. H., Gagnon, G. A., Truelstrup Hansen, L., and Jamieson, R. C. (2015). Performance of municipal waste stabilization ponds in the Canadian Arctic. *Ecol. Eng.* 83, 413–421. doi: 10.1016/j.ecoleng.2015.07.008

Rodriguez-Caballero, A., Hallin, S., Påhlson, C., Odlare, M., and Dahlquist, E. (2012). Ammonia oxidizing bacterial community composition and process performance in wastewater treatment plants under low temperature conditions. *Water Sci. Technol.* 65, 197–204. doi: 10.2166/wst.2012.643

Sato, Y., Hori, T., Navarro, R. R., Habe, H., Yanagishita, H., and Ogata, A. (2016). Fine-scale monitoring of shifts in microbial community composition after high organic loading in a pilot-scale membrane bioreactor. *J. Biosci. Bioeng.* 121, 550–556. doi: 10.1016/j.jbiosc.2015.10.003

Statistics Canada (2012). *Census Profile. 2011 Census. Statistics Canada Catalogue no. 98-316-XWE. Ottawa*. Available online at: http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/index.cfm?Lang=E (accessed December 30, 2016).

Suzuki, M. T., Taylor, L. T., and DeLong, E. F. (2000). Quantitative analysis of small-subunit rRNA genes in mixed microbial populations *via* 5′-nuclease assays. *Appl. Environ. Microbiol.* 66, 4605–4614. doi: 10.1128/AEM.66.11.4605-4614.2000

Usharani, B. (2019). Metagenomics study of the microbes in constructed wetland system treating sewage. *Int. Lett. Nat. Sci.* 74, 26–48. doi: 10.18052/www.scipress.com/ILNS.74.26

Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., and Daims, H. (2002). Microbial community composition and function in wastewater treatment plants. *Ant. Leeuwenhoek* 81, 665–680. doi: 10.1023/A:1020586312170

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the proteobacterial class *Deltaproteobacteria* and *Oligoflexia* and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* 70, 5972–6016. doi: 10.1099/ijsem.0.004213

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Xue, J., Schmitz, B. W., Caton, K., Zhang, B., Zabaleta, J., Garai, J., et al. (2019). Assessing the spatial and temporal variability of bacterial communities in two Bardenpho wastewater treatment systems *via* Illumina MiSeq sequencing. *Sci. Tot. Environ.* 657, 1543–1552. doi: 10.1016/j.scitotenv.2018.12.141

Zhang, B., Xu, X., and Zhu, L. (2017). Structure and function of the microbial consortia of activated sludge in typical municipal wastewater treatment plants in winter. *Sci. Rep.* 7:17930. doi: 10.1038/s41598-017-17743-x

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *J. Bioinform.* 30, 614–620. doi: 10.1093/bioinformatics/btt593

Zhang, T., Shao, M. F., and Ye, L. (2012). 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J.* 6, 1137–1147. doi: 10.1038/ismej.2011.188

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/frwa.2021.711477

Rapid qPCR-Based Water Quality Monitoring in New York State Recreational Waters

*Cristina P. Fernández-Baca[1], Catherine M. Spirito[1], Justin S. Bae[1], Zsofia M. Szegletes[1], Nathan Barott[1], Desiree J. Sausele[1], Yolanda M. Brooks[1], Daniel L. Weller[2] and Ruth E. Richardson[1]**

[1] Department of Civil and Environmental Engineering, Cornell University, Ithaca, NY, United States, [2] Department of Environmental Science and Forestry, State University of New York College of Environmental Science and Forestry, Syracuse, NY, United States

Public swimming beaches often rely on culture-based methods to determine if fecal indicator bacteria (FIB) levels are greater than health risk-based beach action values (BAV). The slow turnaround time of culture-based assays can prevent effective beach closure and reopening decisions. Faster testing methods that can be completed on-site are needed. Additionally, beach closures are currently based on high FIB levels, but at-present there are no tools to examine the health risks to bathers from myriad pathogens (e.g., bacteria, viruses, protozoa) that may be present in recreational waters. Twelve New York State beaches ($n = 9$ freshwater and $n = 3$ marine) were monitored over the course of summer 2018, and two of the freshwater beaches were monitored in fall 2017 as part of a preliminary study. A rapid, in-field workflow for detecting fecal enterococci in water samples was tested using four assays on two Biomeme handheld devices. All Biomeme-based workflows involved in-field DNA extractions and qPCR using portable devices. Beach water samples were also analyzed using EPA-approved or EPA-based qPCR methods: two culture-based methods, Enterolert (targeting enterococci at freshwater and marine beaches) and Colilert (targeting *E. coli* at freshwater beaches); and one qPCR method based on EPA 1611.1. For low abundance pathogen quantification, nanoscale-qPCR was conducted in 2018 using the Pathogen Panel which targeted 12 viral, bacterial, and protozoal pathogens. In fall 2017, the qPCR-based methods performed similarly to Enterolert (r^2 from 0.537 to 0.687) and correctly classified 62.5–75.0% of water samples for a BAV of 104 MPN per 100 ml. In summer 2018, the correlation between *Enterococcus* levels based on Biomeme qPCR and Enterolert varied substantially between the 12 beaches. Inclusion of diverse regions and beach types may have confounded the Biomeme qPCR results. The EPA 1611.1-based method showed a weak, significant correlation ($r^2 = 0.317$, $p = 0.00012$) with Enterolert. Nanoscale-qPCR showed low-levels of pathogens present at all beach sites; but only three showed up with any substantial frequency, *E. coli eae* (25% of samples), norovirus (31.4%), and *Giardia lamblia* (11.4%). Preliminary studies to establish beach-specific correlation curves between rapid qPCR and Enterolert methods are needed before any qPCR assay is used to inform beach decisions.

Keywords: water quality monitoring, fecal indicator bacteria (FIB), nanoscale qPCR, pathogen abundance, enterococci, beach action value, qPCR (quantitative PCR), *E. coli*

OPEN ACCESS

Edited by:
Constantinos V. Chrysikopoulos, Technical University of Crete, Greece

Reviewed by:
Vasiliki Syngouna, University of Patras, Greece
Patrick C. Hanington, University of Alberta, Canada

***Correspondence:**
Ruth E. Richardson
rer26@cornell.edu

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *18 May 2021*
Accepted: *09 September 2021*
Published: *08 October 2021*

Citation:
Fernández-Baca CP, Spirito CM, Bae JS, Szegletes ZM, Barott N, Sausele DJ, Brooks YM, Weller DL and Richardson RE (2021) Rapid qPCR-Based Water Quality Monitoring in New York State Recreational Waters. Front. Water 3:711477. doi: 10.3389/frwa.2021.711477

INTRODUCTION

In the United States, federal law requires recreational swimming waters be monitored for the presence of fecal indicator bacteria (FIB) to maintain public health and safety. The EPA recommends using fecal enterococci and *Escherichia coli* for freshwater testing, and enterococci for marine water testing (US EPA, 2013). While FIB themselves are not necessarily pathogenic, elevated FIB levels may indicate fecal contamination, and thus the presence of fecally-associated pathogens. Typically, fecally-associated pathogens are present at much lower levels than FIB making them more difficult to detect. Additionally, because there are many species of pathogenic bacteria, viruses, and protozoa, there is currently no single method to measure the diverse microorganisms that can lead to illness. Thus, FIB are used as proxies to estimate the risk of waterborne pathogens. Recently, the use of FIBs to estimate public health risk has come under scrutiny. A statistical analysis of 540 published studies of indicator-pathogen correlations found that no single indicator, including enterococci or *E. coli*, was strongly correlated with pathogen presence (Wu et al., 2011). This suggests our current methods for estimating water quality to reduce public health risks may need to be re-evaluated.

Current water testing methods vary by location but are either molecular or culture-based, with the latter being the conventional method for quantifying FIBs. Examples of culture-based methods include Colilert®, Enterolert®, and EPA methods such as US EPA Method 1603 (for *E. coli*) and US EPA Method 1600 (for enterococci). Although these methods are relatively inexpensive and easy to perform, they all require samples to be transported to a centralized lab for testing and necessitate an overnight incubation step resulting in a significant turnaround time for results. Specialized equipment for sealing and incubating test kits as well as trained lab technicians and reliable electricity, are all necessary for culture-based testing. Crucially, the long incubation period and reliable electricity source required by these methods limits their feasibility for routine, rapid, on-site monitoring. Although research has been done to eliminate the need for electricity by using homemade incubators, these systems are not widely used (Bernardes et al., 2020). In recent years, new culture-based methods have been developed including TECTA (Bramburger et al., 2015) and Compartment Bag Tests (CBTs; Brooks et al., 2017; Wang et al., 2017). However, these methods require processing times ranging from 7 to 48 h to obtain results.

Molecular methods, such as quantitative polymerase chain reaction (qPCR) are culture independent and can be used to specifically test samples for the presence of genes of interest (i.e., FIB indicator genes). qPCR-based methods can achieve results in under 4 h after water samples arrive at a laboratory. Although the EPA has approved two qPCR methods for *Enterococcus* 23S rRNA genes, EPA 1611.1 and 1609 (US EPA, 2013, 2015), they have not yet promulgated the methods in the Federal Register (https://www.federalregister.gov/agencies/environmental-protection-agency). Thus, although some states have started to implement these methods [e.g., Michigan (Dorevitch et al., 2017; Byappanahalli et al., 2018)], states like New York, do not recognize EPA 1611.1 and 1609 as suitable *Enterococcus* water quality testing methods for determining beach closures and openings. The EPA has not yet approved any qPCR-based methods for *E. coli* monitoring, although there are several published qPCR assays that are both sensitive and specific to *E. coli*, including assays targeting the 23S rRNA, 16S rRNA, *uidA*, and *rodA* genes (Huijsdens et al., 2002; Frahm and Obst, 2003; Chern et al., 2011; Haugland et al., 2021). Several other molecular methods have been developed for water quality monitoring, such as isothermal amplification for on-site water testing (Mauk et al., 2015), including for *Enterococcus* spp. (Kolm et al., 2017); targeting cells with qPCR or a propidium iodide treatment (Bae and Wuertz, 2012), digital PCR (Wang et al., 2016), and high throughput sequencing (Schang et al., 2016). Additionally, many different qPCR-based microbial source tracking (MST) tools have been used with varying success to test for source-specific markers (Harwood et al., 2014). Among these, nanoscale qPCR (ns-qPCR) using OpenArray® technology was recently used in a "Pathogen Panel" to test for *Enterococcus* spp. and *E. coli* as well as waterborne pathogens and MST markers in the Hudson River watershed (Brooks et al., 2020). Compared to culture-based testing, current molecular methods to quantify FIB are more sensitive and rapid, resulting in faster turnaround times for water quality monitoring (Griffith and Weisberg, 2011; Dorevitch et al., 2017). However, molecular methods generally require specialized equipment, centralized laboratories, trained technicians, and electricity. Additionally, these methods are currently more expensive than standard culture-based methods (Schang et al., 2016).

Water testing protocols (e.g., *E. coli* vs. enterococci quantification) vary in New York State (NYS) by locality, but across the state, if a beach exceeds the Beach Action Value (BAV) for the concentration of FIB for that location, the beach will issue an advisory and close immediately until the issue is resolved—i.e., a new water sample is tested and confirmed to contain FIB levels below the BAV (Water Quality Monitoring, 2004). Although time-consuming, culture-based methods are the standard for quantifying FIBs at NYS beaches. qPCR-based methods are comparatively more rapid and can be more sensitive than culture-based methods; however, they can also be more technically challenging and at-present, have not been approved for statewide testing. Biomeme, a Philadelphia based company, has developed a portable qPCR platform for rapid on-site testing to address this issue. Both the original two3 (now discontinued) and the newer model, Franklin (originally named the three9), developed to replace the two3, are portable thermal cycling devices created for quick, easy on-site FIB quantification via qPCR. Biomeme has also designed extraction kits for field processing samples that can be analyzed on their qPCR devices allowing for test results to be known within a few hours. However, similar to other qPCR-based methods, their protocol has not yet been EPA-approved to replace culture-based methods for water quality monitoring of recreational waters in NYS.

In this study, four different methods (**Figure 1**) including both culture-based (i.e., Enterolert and Colilert) and qPCR-based (i.e., EPA 1611.1-based, three Biomeme methods, and ns-qPCR) were used to test water samples from twelve different public swimming areas across NYS (see **Figure 2** for locations). Our objectives were

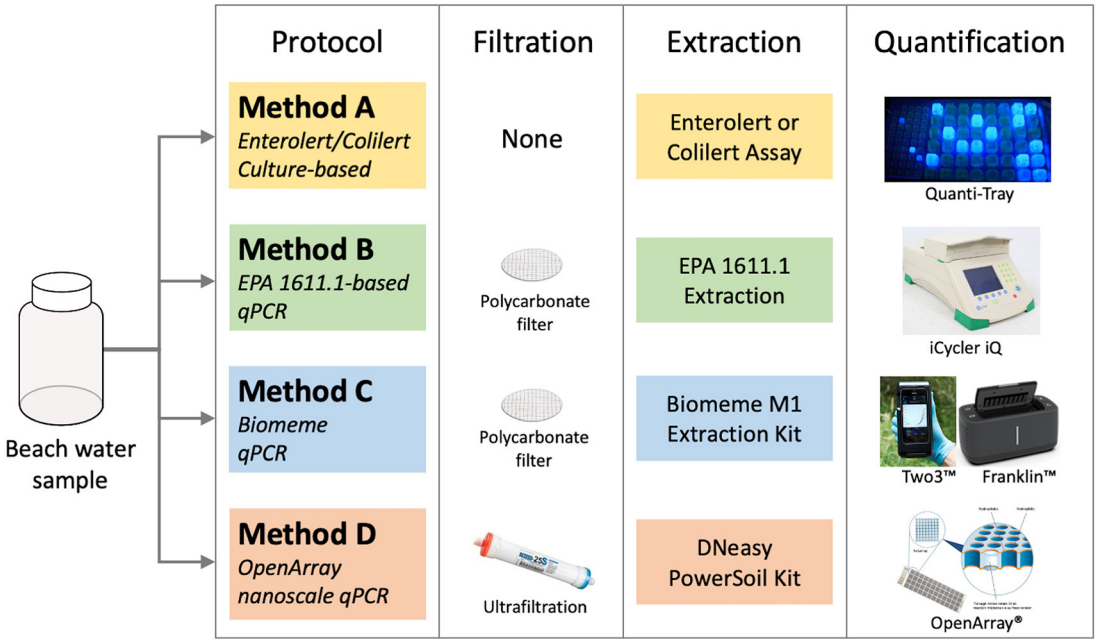


FIGURE 1 | Schematic of the culture-based (Method A) and qPCR-based protocols (Methods B–D) show the workflow from raw beach water samples to comparable results. Method A (Enterolert or Colilert) is the standard method used by NYS beaches. Method C includes four qPCR variants (C1, C2, C2-IPC, and C3). Details on the qPCR assays and which methods were conducted in each year are described in **Table 1**.

to determine: (1) how well the different qPCR methods (i.e., EPA 1611.1-based and various Biomeme methods) correlated with commonly used culture-based methods; (2) how well new qPCR methods developed by Biomeme compare with a qPCR method based on EPA 1611.1 (i.e., using EPA 1611.1 primers and probe); (3) if the ns-qPCR based Pathogen Panel (targeting 12 viral, bacterial, and protozoal pathogens) can be successfully deployed as a water quality indicator across various site locations (four geographic regions) and water types (fresh and marine water).

MATERIALS AND METHODS

Preliminary Study in Fall of 2017

In the fall of 2017, a preliminary study focused on testing water samples from two creek swimming sites both located in the Finger Lakes region (Buttermilk Falls State Park and Robert H. Treman State Park) (see map, **Figure 2**). Water samples were collected weekly beginning September 13 and ending October 10, 2017 and were analyzed using a qPCR method based on EPA 1611.1 (hereafter referred to as EPA-based or EPA 1611.1-based) and a culture-based (Enterolert) method, as well as three Biomeme methods (Methods C1, C2, and C3) (**Figure 1**; **Table 1**). In total 16 samples were taken in 2017 across the two beaches.

Twelve Beach Study in Summer of 2018

In the summer of 2018, 10 additional NYS parks swimming areas were added (**Figure 2**). For this second study, we collaborated with the NYS Department of Parks, Recreation, and Historical Preservation to coordinate collection of beach water samples to be synoptic with their routine sampling for FIB monitoring. Water samples were collected from swimming areas in the 12 NYS parks (**Figure 2**) including three Great Lakes beaches (i.e., Woodlawn Beach State Park and Evangola State Park on Lake Erie, and Fair Haven State Park on Lake Ontario); four inland beaches in the Finger Lakes region (i.e., Treman State Park, Buttermilk Falls State Park, Taughannock Falls State Park, Fillmore Glen State Park); two inland beaches in the Palisades (Lake Tiorati Beach and Lake Welch Beach); and three marine beaches on Long Island (i.e., Heckscher State Park, Jones Beach State Park, and Wildwood State Park). During the 2018 swimming season from June 11 to July 30, Buttermilk and Treman state parks were sampled weekly between 8:00 and 9:00 a.m. At the remaining 10 sites, morning samples were collected two to three times (Fair Haven was sampled only once and intensive sampling sites were sampled three times, see below) between 8:00 and 9:00 a.m.

A three-day intensive sampling campaign was done to gauge FIB and pathogen fluctuations throughout the day at four beaches across the state (Welch, Buttermilk, Heckscher and Woodlawn). Intensive sampling campaigns involved three consecutive days of sampling during which seven samples were collected. Sampling started and ended with a sample taken at 8:00 a.m. on the 1st and 3rd day. On the second day, five samples were taken beginning at 8:00 a.m. at 3 h intervals as follows: 8:00 a.m., 11:00 a.m., 2:00 p.m., 5:00 p.m., 8:00 p.m. In total, 61 samples were taken in 2018.

Samples collected in 2018 were analyzed using one culture-based EPA method (Enterolert) and one qPCR EPA-based method (EPA 1611.1-based assay), as well as one Biomeme

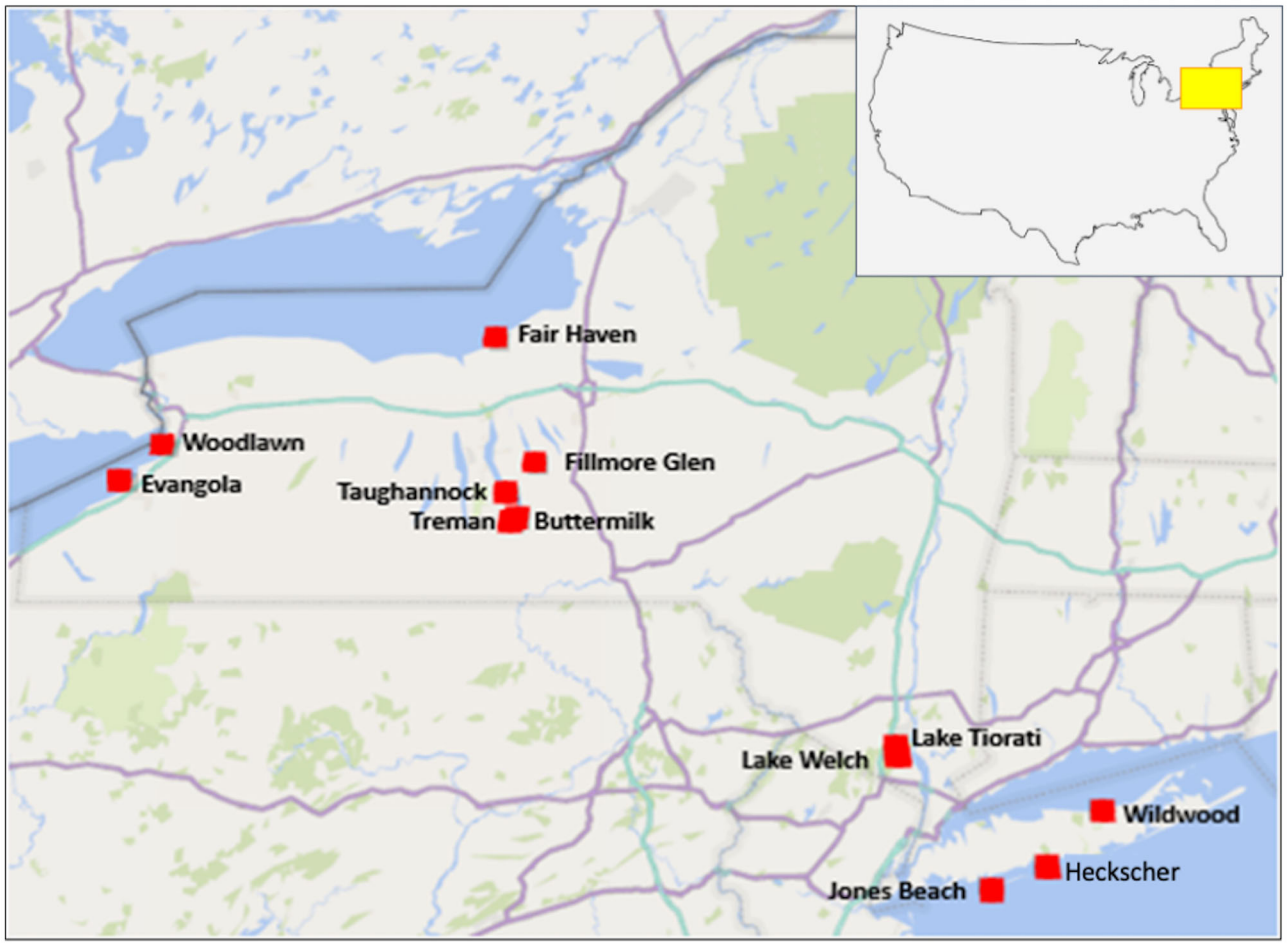


FIGURE 2 | Sampling locations of swimming areas for New York State parks study.

method, Franklin-IPC, which included an internal positive control and was not available in 2017 (**Figure 1**; **Table 1**). The two3 device was discontinued and replaced by the Franklin device, thus, the two3 was not used again in 2018. The BioPoo assay was available in 2018, however we used the Franklin-IPC method as this method is more comparable to the EPA-based qPCR method (as well as the two3 methods). The Franklin-IPC (and both assays run on the two3) use the same primers and probe as the EPA 1611.1 method to target the *Enterococcus* 23S rRNA gene [see methods section *EPA-based qPCR method (Method B)*], while the BioPoo assay targets the *tuf* gene [see methods section *Biomeme Franklin with BioPoo in 2017 (Method C3)* for more details].

Water Sampling and Environmental Data Collection

In both years, for each sample, a closed, autoclaved 1 L polypropylene or glass bottle was held 15 to 30 cm below the water surface, opened, filled with water, and closed again below the surface using gloved hands. Duplicate water samples were taken for each timepoint. All samples were stored on ice during transit to the laboratory where they were processed within 6 h of collection. In 2018, at each location, water temperature and turbidity (Hach 2100Q Portable Turbidimeter), were recorded. Rainfall over the previous 24 and 96 h was also collected in 2018 from the nearest weather station (see **Supplementary Data File 1** for the full dataset). For the nanoscale OpenArray qPCR used in 2018, an additional 10 L sample was taken concurrently with the 1 L sample at each location, as described above, using a 10 L carboy in place of the 1 L bottle. Each carboy was bleach-sterilized overnight and rinsed 3 times with sterile water prior to sampling.

Culture-Based Quantification (Method A)

Enterolert and Colilert Quanti-Tray/2000 (IDEXX Laboratories, Westbrook, ME) were used to quantify culturable concentrations of enterococci and *E. coli*, respectively. Both methods are EPA-approved for freshwater beach monitoring. At all sites Enterolert was used, as this method can be used for either beach type, while

TABLE 1 | Overview of the qPCR methods used in this study.

qPCR method	Assay abbreviation	qPCR device	Target gene for *Enterococcus*	DNA template volume (μl)	Sample volume (ml)	Internal positive control	Sample years analyzed
B	EPA 1611.1-based	BioRad iCycler	23S rRNA gene EPA 1611.1 primer/probes	5	100–1000	yes[†]	2017 + 2018
C1	two3-5ul	Biomeme two3	23S rRNA gene EPA 1611.1 primer/probes	5	100–250	no	2017
C2	two3-15ul	Biomeme two3	23S rRNA gene EPA 1611.1 primer/probes	15	100–250	no	2017
C2-IPC	Franklin-IPC	Biomeme Franklin	23S rRNA gene EPA 1611.1 primer/probes	18	100–1,000	yes	2018
C3	BioPoo	Biomeme Franklin	*tuf* gene Biomeme primer/probe	20	100–1,000	yes	2017
D	Pathogen Panel	Thermo OpenArray	23S rRNA gene EPA 1611.1 primer/probes	6.25	10	yes	Subset of 2018

[†]*EPA 1611.1 prescribes the use of a sample process control, however in this study an internal amplification control (IAC) was used. EPA method 1611.1-based (Method B), Biomeme method variants (Methods C1, C2, C2-IPC, C3), and OpenArray, nanoscale-qPCR (Method D).*

Colilert was only used at a subset of sites. For marine beaches, only Enterolert assays were used.

Enterolert and Colilert measurements were performed according to the manufacturers' instructions. Briefly, samples were poured into trays using gloved hands, sealed using the Quanti-Tray Sealer (IDEXX Laboratories, Westbrook, ME), and placed in an incubator at 41°C (±0.5°C) for 24 h. After incubation, the trays were read by illuminating the wells with a 6-watt, 365 nm UV light and recording the number of large and small wells that were positive for blue fluorescence. The number of illuminated wells was converted to MPN per 100 ml using the IDEXX 51-Well Quanti-Tray MPN Table (provided by the manufacturer). The IDEXX-QC Enterococci Quality Controls "*Enterococcus faecalis*" and "*Streptococcus bovis*" positive and negative controls (IDEXX Laboratories, Westbrook, ME), respectively, were run to ensure the specificity and accuracy of the testing method. Controls were run concurrently with samples and matched expected results. For all beaches, where possible, we obtained official State Park results for culture-based methods, both Enterolert and Colilert, that were performed at certified labs and used to determine BAV exceedances (See **Supplementary Data File 1**).

Water Sample Filtration for qPCR-Based Methods

Membrane Filtration to Concentrate Biomass for qPCR Methods (Methods B and C)

Samples used for qPCR-based methods B and C (i.e., EPA 1611.1-based and the Biomeme methods) were membrane filtered using white polycarbonate filters (diameter, 47 mm, pore size between 0.4 and 0.6 μm) in accordance with EPA method 1611.1 (US EPA, 2015). Each sample was mixed by shaking for 30 s, and 100 to 1,000 ml of sample (exact volumes recorded for each sample) were vacuum filtered through the membrane filter. This was followed by one rinse with distilled water. Filters with captured biomass were stored in sterile, screw cap, 15 ml polypropylene tubes at −80°C until DNA extraction. For 1 L samples, filter clogging was an issue for two of the samples, in these cases multiple membrane filters were used to process one sample and all filters were combined into a single composite sample for DNA extraction.

Ultrafiltration to Concentrate Low Abundance Pathogens for Nanoscale qPCR (Method D)

For ns-qPCR quantification of low-abundance pathogens, samples were processed using tangential flow ultrafiltration to concentrate biomass from large 10 L water samples using a hollow-fiber ultrafilter, Rexseed 25S (Asahi, USA), as described previously (Rhodes et al., 2012; Brooks et al., 2020). Negative controls for each sampling date were also processed and consisted of 10 L of deionized water. All samples were concentrated to approximately 250 ml of retentate. Based on Polaczyk et al. (2008), 0.001% (w/w) bovine serum albumin, 0.9 M sodium chloride, and 0.012% (w/w) polyethylene glycol were added to the retentate followed by an overnight precipitation at 4°C to aid in virus precipitation. The samples were then centrifuged at 10,000 × g for 30 min at 4°C and the supernatant was decanted. The pelleted biomass was resuspended in 5 ml of sterile PBS, distributed into microcentrifuge tubes in 1 ml aliquots, then stored at −80°C until DNA extraction.

DNA Extraction

EPA-Based qPCR Method (Method B)

For Method B, DNA extractions were performed according to the EPA 1611.1 method with the modification that no sample processing control was used. Filters were transferred to 2 ml screw cap centrifuge tubes prepacked with 0.3 g of 212 to 300 μm autoclaved, glass beads and AE buffer (10 mM Tris-Cl 0.5 mM EDTA; pH 9.0) using sterile tweezers. The tubes were then bead beaten at maximum speed for 60 s and centrifuged at 12,000 × g for 60 s. The supernatant was transferred to a clean 2 ml microcentrifuge tube and centrifuged a second time at 12,000 × g for 5 min. The supernatant was then transferred to a new 2 ml tube and was either immediately used for qPCR or frozen at −80°C until use.

Biomeme In-Field Extraction Kit (Method C)

For Method C, DNA extractions from filters were done using the Biomeme Field Sample Prep Kit (M1) following the manufacturer's instructions (Biomeme, Philadelphia, PA). Briefly, sterile tweezers were used to transfer the filters into the kit extraction tubes prefilled with a lysis buffer. The tubes were shaken vigorously to resuspend the filtrate. The subsequent protein, salt, and drying wash followed by air drying and DNA elution were carried out according to the Biomeme protocol (**Figure 3**). The M1 kit was designed to be used in the field, and in our study the kit was used in the field to immediately process the majority of samples. Samples taken back to laboratory for processing were those taken on the third day of an intensive sampling campaign (4 samples total). All DNA extracts were stored at −80°C.

OpenArray Extractions (Method D)

DNeasy Powersoil DNA extraction kits (Qiagen, Valencia, CA) were used to extract DNA from the 1 ml aliquots of the resuspended biomass pellets following manufacturer's instructions. Prior to DNA extraction, bacteriophage MS2 (ATCC 15597-B1), was added to the resuspended biomass pellets and served as an extraction and amplification control (Dreier et al., 2005; Brooks et al., 2020). The DNA extracts were stored at −80°C.

qPCR Assays

EPA 1611.1-Based Assay (Method B)

Enterococcus quantification targeting the large subunit ribosomal ribonucleic acid gene (lsrRNA, 23S rRNA) was done using iTaq Universal SYBR Green Supermix (BioRad, Hercules, CA) and the primers/probe and thermal cycling protocol from the EPA 1611.1 method (US EPA, 2015). TaqMan qPCR assays were performed in triplicate on a BioRad iCycler (BioRad, Hercules, CA). Total reaction volumes of 25 μl were used with 5 μl of template (standards, blanks, and/or samples), 10 μl iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA) and a mixture of the *Enterococcus*-specific primers and probe. The thermal protocol had an initial holding period of 2 min at 50°C then a 10 min period at 95°C, followed by 40 cycles of denaturing for 15 s at 95°C and annealing/elongation for 60 s at 60°C.

Standards for *E. faecalis* 23S rRNA were created using gBlocks (Integrated DNA Technologies (IDT), Coralville, IA). All standards were made in AE buffer using a dilution series from 8.0×10^3 copies μl^{-1} down to 1×10^0 copy μl^{-1}. For all qPCR methods copies μl^{-1} were converted to copies per 100 ml of water sampled. To ensure specificity of amplification, a subset of PCR products were run on a 1% agarose gel to confirm the correct sized product (93 bp for *Enterococcus*).

To assess inhibition in samples an internal amplification control (IAC) was used retroactively on 85% of the samples (all samples that remained). The IAC was based on the US EPA 1696 method (US EPA, 2019) using the primers and probe from this method as well as the same IAC sequence which was created using gBlocks (IDT). Six no template controls (NTCs) were run per plate and the average Cq of the NTCs (Cq = 27.52) plus three times the standard deviation (Cq + 3 Standard deviations = 28.23) was used as the Cq cutoff for inhibition. If the Cq of 2 of the 3 triplicates run for each sample was below this cutoff Cq then the sample was not considered to be inhibited. The IAC results showed that no samples were inhibited when the same sample dilutions were used as those in the original analysis.

Biomeme Workflows (Method C1, C2, C2-IPC, C3)

Four different methods were run on Biomeme devices (see **Table 1**), which varied in the amount of template used (from 5 to 20 μl), the presence/absence of an internal positive control (IPC), and the Biomeme device used (i.e., two3 or Franklin/three9). The Biomeme two3 (now discontinued) had two sampling channels of different excitation and emission wavelengths (allowing duplex reactions) and could analyze three samples at a time (Sepulveda et al., 2018). The replacement for the two3, the Biomeme Franklin/three9 (hereafter referred to as Franklin) can analyze up to nine samples at a time and has three different fluorescence channels, allowing for triplex reactions. Biomeme qPCR was performed using the same thermal profile across both devices with an initial denature step of 60 s at 95°C followed by 45 cycles each with 1 s at 95°C and 20 s at 60°C. The total run time for one run was 45 min. Standards for all Biomeme workflows were dilutions of either gBlocks (Integrated DNA Technologies (IDT), Coralville, IA) (for Methods C1, C2 and C2-IPC) or genomic pure culture DNA from *Enterococcus faecilis* (for Method C3). Standards dilutions were created from 1×10^5 copies μl^{-1} to 1×10^1 copies μl^{-1}.

Biomeme Two3 With 5 μl Sample Volume in 2017 (Method C1)

Biomeme's LyoDNA Mix (Biomeme, Philadelphia, PA) was re-suspended by adding a 12% aqueous glycerol solution as per manufacturer's instructions to obtain a 2X concentration master mix. Each 25 μl qPCR reaction contained 12.5 μl of 2X LyoDNA master mix, 5 μl of water, 1 μl of 10 μM forward and reverse primers, 0.5 μl of 10 μM probe, and 5 μl of template DNA. The primers and probes used and their concentrations (400 and 200 nM, respectively) in Method C1 were the same as those of EPA 1611.1-based method (Method B).

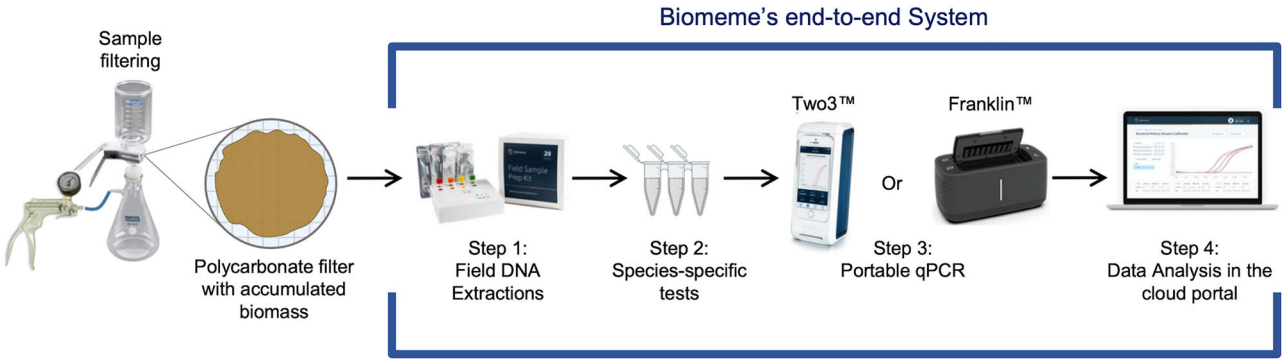


FIGURE 3 | Summary of Biomeme qPCR workflow. Water is membrane-filtered then (1) DNA is extracted using the field DNA sample prep kit; (2) *Enterococcus* qPCR assays are set up in reagent strips; (3) qPCR assays are run on the smartphone-controlled handheld thermocycler; (4) data is retrieved for analyses via the cloud portal or exported as raw data from the smartphone.

Biomeme Two3 With 15 μl Sample Volume in 2017 (Method C2)

Method C2 was designed to increase sensitivity of the original Method C1 to low levels of enterococci. A 5X LyoDNA master mix concentration was used in place of the 2X LyoDNA master mix used in Method C1 which allowed for the addition of more template DNA (15 μl instead of 5 μl). Each 20 μl reaction was composed of 4 μl of 5X LyoDNA master mix, 0.44 μl of nuclease-free water, 0.08 μl of 100 μM forward and reverse primers, 0.4 μl of 10 μM probe, and 15 μl of template DNA. The greater template volume was used to increase the concentration of initial copies per reaction; all other assay concentrations (e.g., primers, probes, etc.) in this method were the same as those used in Method C1.

Biomeme Franklin With IPC in 2018 (Method C2-IPC)

Method C2-IPC was similar to method C2 with two main differences, first the Franklin device was used in place of the discontinued two3 device and second the LyoDNA master mix was replaced with the LyoDNA+IPC master mix prepackaged in Go-Strips (Biomeme, Philadelphia, PA). Go-Strips include proprietary lyophilized primers, probe, and DNA for an IPC assay in addition to the lyophilized master mix. For this method, 2 μl of primer/probe mix for *Enterococcus* spp. (same as in EPA 1611.1-based and C1 and C2 methods) and 18 μl of sample were added to the IPC Go-Strips. The IPC was used to assess inhibition and after initial testing indicated inhibition across many samples (i.e., no IPC amplification), all samples were run at 1:10 dilution. The delta Ct method was used to calculate gene copy numbers, using the difference between IPC and *Enterococcus* Cts to create a standard curve specific for the C2-IPC method.

Biomeme Franklin With BioPoo in 2017 (Method C3)

Method C3, which was run on the Franklin device, uses Biomeme's BioPoo Go-Strips. These Go-Strips contain the lyophilized, proprietary primers and probes for three targets: an IPC, the translation elongation factor (*tuf*) gene of *Enterococcus*, and Human Fecal Marker 183 (HF183), a human-associated *Bacteroides* spp. The Biomeme BioPoo assay tests for the *tuf* gene rather than the 23S rRNA gene targeted by the EPA 1611.1 primers and probe. For this method, 20 μl of sample was loaded into a well and used to resuspend the master mix. Sample inhibition was assessed by examining IPC results, if the IPC did not amplify, samples were run again at a 1:10 dilution.

Pathogen Panel ns-qPCR (Method D)

As described in Brooks et al. (2020) a "Pathogen Panel" was designed for OpenArray plates (printed by Thermo-Fisher in an 18 x 3 format to allow for 18 targets with three technical replicates per sample), to determine the presence and abundance of 12 waterborne pathogens (*Campylobacter* spp. unknown target, *Shigella* spp./enteroinvasive *E.coli ipaH*, *Salmonella* spp. unknown target; *E. coli stx1*; *E. coli eae*, eaeF2/R; *E. coli O157 rfbE*; norovirus group II, QNIF2d/Cog2R; adenovirus type 40/41, VTB1-HAdVF; enterovirus highly conserved 5′ untranslated region; hepatitis A highly conserved 5′non-coding region, HAV240/HAV H68; rotavirus A NSP3; *Giardia lamblia* 18S rRNA, Giardia80F/127R), two general FIB [*E. coli* unknown gene and *Enterococcus* spp. 23S rRNA (EPA Method 1611.1)], and three source-specific FIB (i.e., human, *Bacteroides* spp. 16S rRNA, HumM2; cow, *Bacteroides* spp. 16S rRNA, CowM3; and poultry, *Brevibacterium* spp. LA35 16S rRNA) as well as an internal positive control (MS2) (Brooks et al., 2020). Standards for the 17 targets in the Pathogen Panel were either genomic DNA (*Salmonella* spp. control strain from the Cornell Animal Health Diagnostic Center; *Campylobacter jejuni* from ATCC, 33560D-5; *Shigella flexneri* from ATCC, 29903D-5), plasmid DNA (*E. coli* 23S rRNA (Thermo Fisher ABI part no. 4460366), *Giardia lamblia* (Genscript, Piscataway, NJ), and norovirus (IDT, Coralville, IA)], or synthetic gBlocks purchased from IDT. The detailed ns-qPCR workflow is described in Goodman et al. (2016). Samples with qPCR inhibition were diluted 1:10 for ns-qPCR.

Beach Action Value Classification

For all FIB qPCR-based methods, we developed correlation curves between cellular equivalents (CE) per 100 ml as

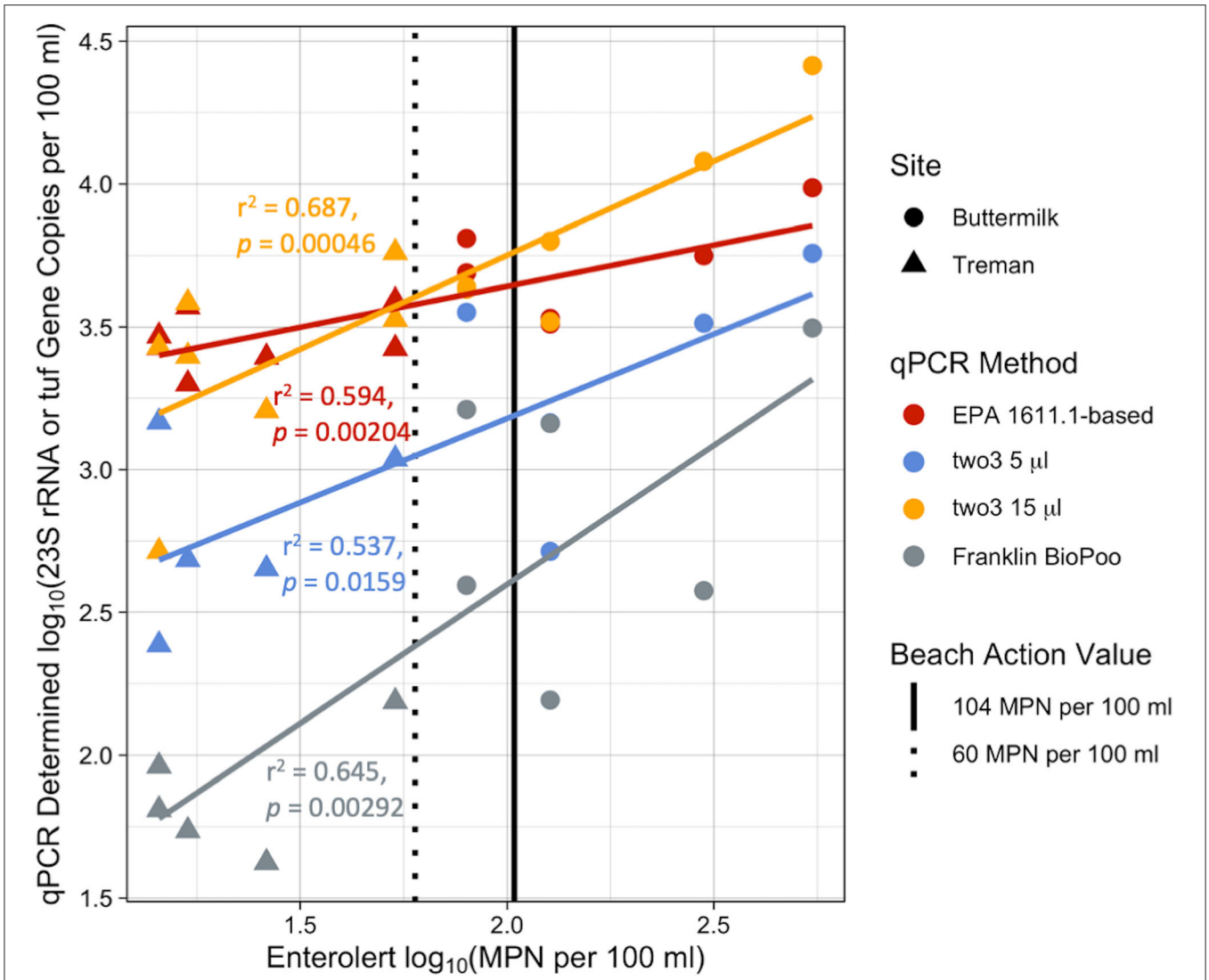


FIGURE 4 | Comparison of qPCR methods to Enterolert. $\log_{10}$-$\log_{10}$ comparison of the Biomeme methods C1, C2, and C3 (two3-5 µl, two3-15 µl, and BioPoo, respectively) and EPA-based method with the standard culture-based method (Enterolert) currently used by beaches for fall 2017 Buttermilk (circle) and Treman (triangle) samples. The Biomeme method C1 and C2 and EPA-based methods quantify the 23S rRNA gene for *Enterococcus*, while the Biomeme method C3 (Franklin BioPoo) quantifies the *tuf* gene. Vertical lines denote beach action values (BAVs) used by different localities, the dotted black line represents a BAV of 60 MPN per 100 ml and the solid line represents a BAV of 104 MPN per 100 ml.

determined by qPCR vs. most probable number (MPN) of enterococci per 100 ml as determined by the culture-based method Enterolert. From these correlation curves we determined the qPCR value corresponding to BAVs used by different localities in NYS. The two BAVs used were 60 MPN per 100 ml for freshwater and 104 MPN per 100 ml for marine beaches. Using these values, we calculated whether any given sample was in exceedance of or below the BAV (indicating an acceptable level of risk, "safe") based on their qPCR-determined enterococci gene concentrations. For each sample, we then compared the qPCR-based classification to the culture-based, Enterolert, classification and categorized every qPCR measurement as either correctly classified (correct above or below the BAV) or incorrectly classified (false negative or false positive) using Enterolert as the standard.

Statistical Analyses

All statistical analyses were conducted in R (version 3.6.1) (R Core Team, 2013). Data normality was tested in linear and $\log_{10}$-$\log_{10}$ space and all data showed normality in $\log_{10}$-$\log_{10}$ space thus all plots and calculations are presented using the transformed data. Linear regression analyses were conducted using a significance cutoff of $p = 0.05$. Correlations of environmental variables with qPCR- and culture-based FIB concentrations as well as pathogen concentrations were made with the *cor()* function in R using the Pearson method with pairwise comparisons. Heat

maps were created in R using the Euclidean distance for the dissimilarity measure.

RESULTS

Preliminary Beach Study (Fall 2017)

In the preliminary fall 2017 study, two riverine state park swimming areas were monitored (i.e., Buttermilk and Treman State Parks) using four different qPCR methods (i.e., EPA 1611.1-based assay and three Biomeme workflows) for two months. The EPA 1611.1-based assay and all three Biomeme methods correlated positively with the culture-based Enterolert method currently used by the two beaches (**Figure 4**). All qPCR-based values are reported as gene copies per 100 ml to allow for easy comparisons across methods and years. The in-field and lab benchtop qPCR devices performed relatively similarly with stronger correlations between qPCR and culture-based results for the two3-15 μl ($r^2 = 0.687$, $p = 0.00046$) and Franklin BioPoo ($r^2 = 0.645$, $p = 0.00292$) method compared to the EPA-based assay ($r^2 = 0.594$, $p = 0.00204$), with the two3-5 μl method having the weakest correlation ($r^2 = 0.537$, $p = 0.0159$). The slopes of the fit-lines also varied by qPCR method with the Franklin BioPoo having a near 1-to-1 correlation with Enterolert results (slope = 0.974, intercept = 0.657), while the other methods had fit-lines with slopes <1 (two3-5 μl slope = 0.591, intercept = 1.99; two3-15 μl slope = 0.657, intercept = 2.44; EPA 1611.1-based method slope = 0.287, intercept = 3.07). No amplification of HF183 was found in any of the fall 2017 samples with the BioPoo assay. Treman State Park had lower FIB counts overall than Buttermilk State Park. All Enterolert results for Treman were below the 60 MPN per 100 ml BAV, whereas Buttermilk results were all above this threshold and most samples were above the 104 MPN per 100 ml BAV (**Figure 4**).

Percent correct classifications for each qPCR method were calculated using culture-based Enterolert results as the standard for "correct" quantification. For example, if the qPCR method results agreed with the Enterolert results (e.g., both showed FIB counts were above the BAVs and the beach should be closed) then the qPCR method results was considered correctly classified, if the two methods did not agree then the qPCR method was deemed incorrectly classified, using the Enterolert method result as the standard reference (**Table 2**). Different localities use different BAVs, thus both were used for these calculations (i.e., 60 MPN per 100 ml and 104 MPN per 100 ml). For the BAV of 60 MPN per 100 ml, the two3-5 μl and BioPoo methods performed the best (78.6% and 81.3% correct classifications, respectively), while the EPA 1611.1-based assay and two3-5 μl performed similarly (68.8% correct classifications for both). For the BAV of 104 MPN per 100 ml, correct classification was best for EPA 1611.1-based assay (75%) while the two3-5 μl and BioPoo Biomeme methods performed similarly (71.4 and 68.8%, respectively) and the two3-15 μl performed worse (62.5%). False negatives were highest for the BioPoo method (18.8% and 25% for BAVs of 60 and 104 MPN per 100 ml, respectively), indicating that this qPCR method would underestimate FIBs more frequently than the other qPCR methods. However, BioPoo's false positive rate is the lowest among the qPCR methods (0 to 6.3% for BAVs of 60 and 104 MPN per 100 ml, respectively).

There was a strong correlation between the results of the various Biomeme qPCR methods with the EPA 1611.1-based method for fall 2017 samples (**Figure 5**). The correlation was strongest between the EPA 1611.1-based method and the Franklin BioPoo ($r^2 = 0.647$, $p = 0.00283$) and two3-5 μl ($r^2 = 0.651$, $p = 0.0048$) and was slightly lower for the two3-5 μl ($r^2 = 0.404$, $p = 0.0196$). The slopes of the fit-lines were all >1 (two3-5 μl slope = 1.35, intercept = −1.24; two3-15 μl slope= 1.97, intercept = −3.99; Franklin BioPoo slope = 2.49, intercept = −6.47).

Twelve Beach Study (Summer 2018)

The summer 2018 study included more beaches that were also more diverse (i.e., both freshwater and marine across various regions) than the fall 2017 dataset. The Biomeme qPCR results were variable for the Franklin-IPC method and did not correlate ($p < 0.05$) with the culture-based Enterolert results (**Figure 6A**). Similar to fall 2017 results, the EPA 1611.1-based method significantly ($p = 0.000126$) positively correlated with Enterolert results (slope = 0.577, intercept = 3.67, $r^2 = 0.317$), however the correlation was not as strong in 2018 as compared to 2017 (**Figure 6A**). The Franklin-IPC method did

TABLE 2 | Accuracy of qPCR methods for predicting beach action value (BAV) exceedances as compared to the culture-based Enterolert method in fall 2017 and summer 2018 samples.

Year	Method	Prediction of Beach Action Value (BAV) exceedances							
		BAV = 60 MPN/100 ml				BAV = 104 MPN/100 ml			
		% Correct classification	% Incorrect classification	% False negative	% False positive	% Correct classification	% Incorrect classification	% False negative	% False positive
2017	EPA 1611.1-based (B)	68.8	31.3	12.5	18.8	75.0	25.0	12.5	12.5
	Two3 5μl (C1)	78.6	21.4	14.3	7.1	71.4	28.6	21.4	7.1
	Two3 15μl (C2)	68.8	31.3	12.5	18.8	62.5	37.5	18.8	18.8
	BioPoo (C3)	81.3	18.8	18.8	0.0	68.8	31.3	25.0	6.3
2018	EPA 1611.1 (B)	69.6	30.4	7.84	23.5	65.2	34.7	11.8	23.5
	Franklin-IPC (C2-IPC)	35.6	64.4	24.1	37.0	35.6	64.4	24.1	37.0

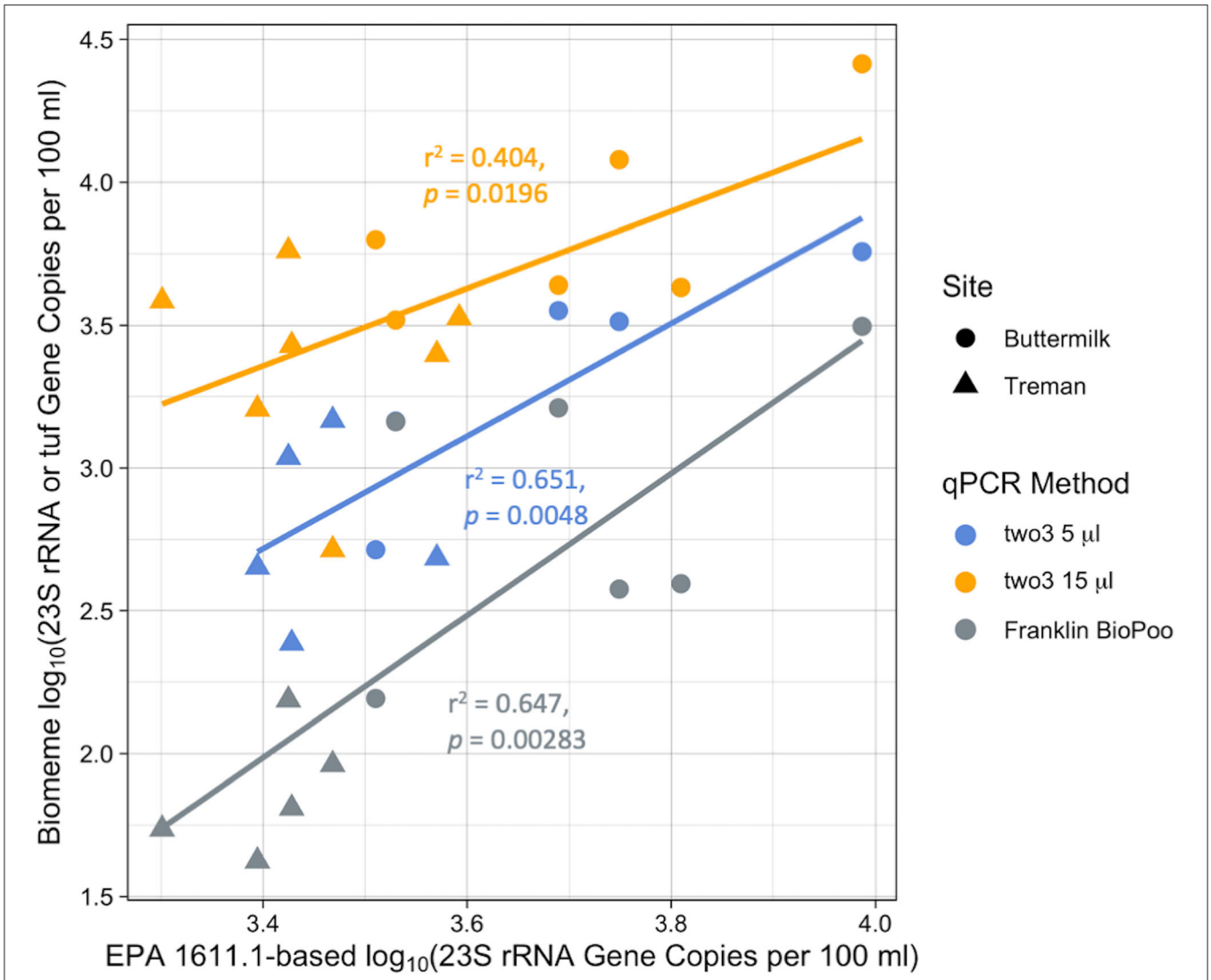


FIGURE 5 | Comparison of novel qPCR methods to the EPA 1611.1-based method for fall 2017 beach samples. log_{10}-log_{10} plots of Biomeme method (C1, two3-5 µl; C2, two3-15 µl; and C3, BioPoo) results as compared to EPA-based method. The BioPoo assay quantifies the *tuf* gene while all other assays quantify the 23S rRNA gene for *Enterococcus*.

not appear to correlate with EPA 1611.1-based method results (**Figure 6B**). Interestingly, when only the Finger Lakes beaches were examined, there was a weak, significant correlation between the two qPCR methods ($r^2 = 0.237$, $p = 0.0105$), however other beach locations did not show any correlations between the two qPCR methods (**Supplementary Figure 1**). Although there were only three marine sites, all were below the BAV of 60 MPN per 100 ml whereas the freshwater beaches ranged from 0 MPN per 100 ml up to 2420 MPN per 100 ml, the upper limit of the Enterolert kit.

Examining the results by geographic region (**Supplementary Figures 2A,B**) showed that the positive correlation between the EPA 1611.1-based and Enterolert methods was a result of the Finger Lakes', Great Lakes', and Palisades' beach results, whereas the Long Island beaches were more variable. The Finger Lakes, Great Lakes, and Palisades beaches are all freshwater beaches, thus we grouped results by beach type (freshwater vs. marine) (**Supplementary Figures 2C,D**). When considering only freshwater beaches the correlation between the EPA 1611.1-based assay and Enterolert was stronger ($r^2 = 0.408$, $p < 0.0001$, **Supplementary Figure 2D**) than when all sites were grouped together. This grouping also highlighted the lack of correlation between the EPA 1611.1-based assay and Enterolert measurements for marine sites ($r^2 = 0.204$, $p = 0.262$). The beach type analysis did not change the relationship between Franklin-IPC and Enterolert results which showed no correlation for either freshwater ($r^2 = 0.000297$, $p = 0.921$) or marine beaches ($r^2 = 0.00114$, $p = 0.931$) (**Supplementary Figures 2C,D**, respectively).

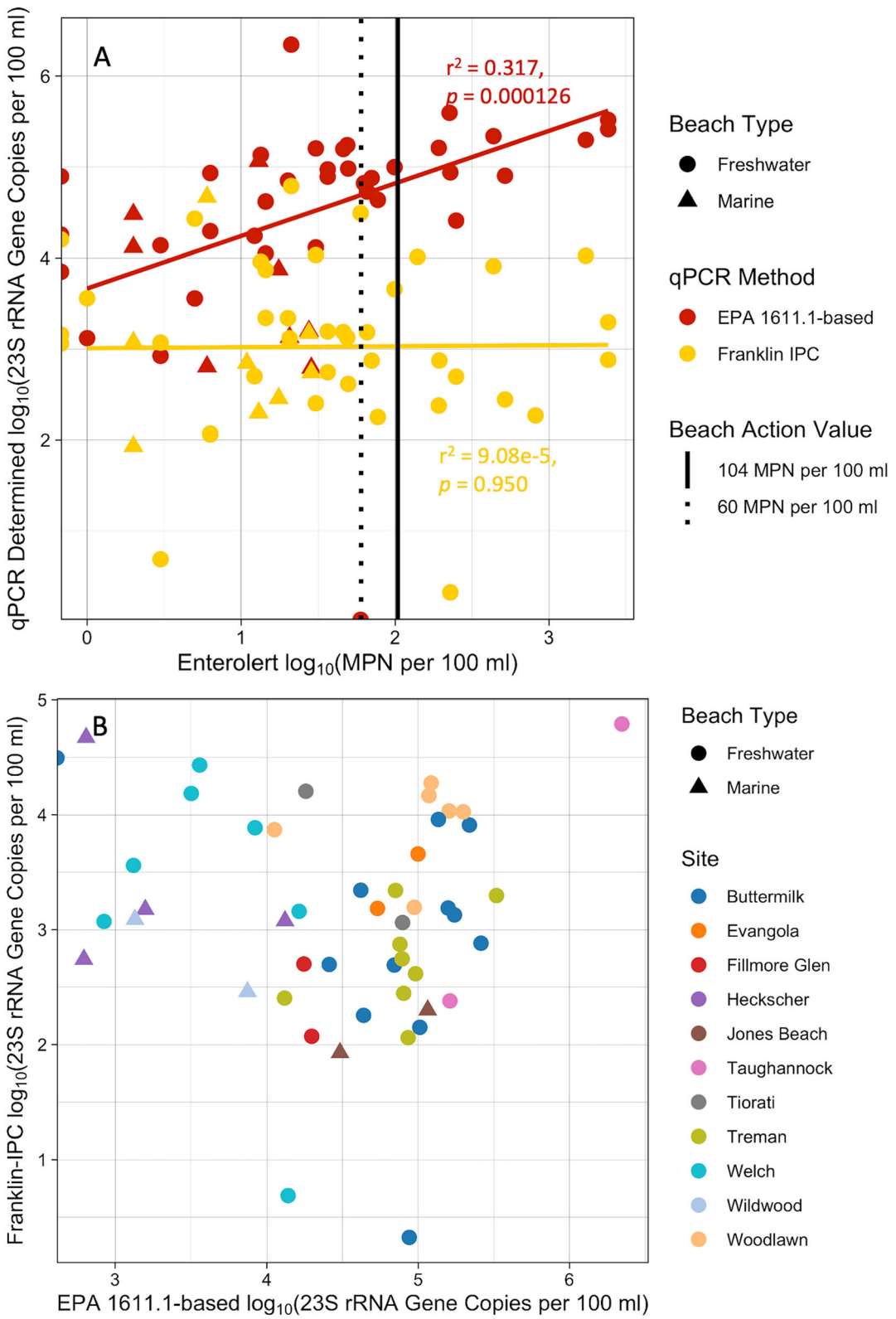


FIGURE 6 | Summer 2018 water sample results from 12 New York State beaches. **(A)** Comparison of qPCR methods EPA-based and Biomeme Franklin-IPC (Methods B and C2-IPC) to Enterolert. Vertical lines denote beach action values (BAVs) used by different localities, the dotted black line represents a BAV of 60 MPN per 100 ml and the solid line represents a BAV of 104 MPN per 100 ml. **(B)** Comparison of qPCR methods: Biomeme Franklin-IPC with EPA 1611.1.

Percent correct classifications were again calculated for the summer 2018 dataset (**Table 2**). The EPA 1611.1-based method performed similarly over the two years with correct classifications of the majority of samples (69.9%) and a relatively low false negative rate (7.84%). In contrast, the Franklin-IPC method correctly classified only 35.6% of samples and incorrectly classified the rest, with 24.1% of those incorrect classifications being false negatives. The EPA 1611.1-based method performed slightly worse for the 104 MPN per 100 ml BAV (65.2% correct classifications) while the Franklin-IPC method performed the same under both BAVs.

Correlations between environmental conditions (i.e., turbidity, water temperature, rainfall over previous 24 h and rainfall over previous 4 d) and measured FIBs (e.g., culture-based and qPCR-based methods) were also examined (**Supplementary Figure 3**). Enterolert and Colilert were moderately positively correlated ($r = 0.25$) and both methods were positively correlated with rainfall ($r = 0.46$ and 0.18 for previous 24 h rainfall, respectively). Water temperature showed weak to moderate correlations with turbidity ($r = 0.53$), Franklin-IPC ($r = 0.36$), and EPA 1611.1-based ($r = 0.18$) results. Turbidity was moderately positively correlated with Colilert ($r = 0.43$) and Enterolert ($r = 0.36$) measurements.

Comparison Across Years (Buttermilk and Treman)

Two beaches, Buttermilk and Treman, were examined in both fall 2017 and summer 2018 using the EPA 1611.1-based and Enterolert methods. The summer 2018 EPA 1611.1-based method showed a weaker positive correlation to Enterolert results in 2018 ($r^2 = 0.208$) as compared to 2017 ($r^2 = 0.59$) (**Supplementary Figure 4**). Unlike in fall 2017, there were several Treman Enterolert samples in 2018 that were above both BAVs. Treman samples in 2018 ranged from as low as 0 up to 2420 MPN per 100 ml, the upper detection limit of the test (average of 324 ± 712 MPN per 100 ml), showing much more variability than in 2017 where samples averaged 28.0 ± 18.0 MPN per 100 ml. Buttermilk Enterolert samples likewise showed more variability in 2018 compared to 2017 (average 369 ± 688 MPN per 100 ml and 210 ± 184 MPN per 100 ml, for 2018 and 2017, respectively). Some of this variability may be explained by having more samples in 2018 (15 and 12 samples for Buttermilk and Treman, respectively) than in 2017 (eight each for Buttermilk and Treman). Additionally, the 2017 samples were taken in the fall during which the beaches were always closed and there were no rain events over the 4-week sampling period. In contrast, in 2018 samples included several rain events and the sampling period spanned 2 months in the summer.

Intensive Beach Study (Summer 2018)

In 2018, four beaches were selected for "intensive" sampling, meaning samples were taken at 8:00 a.m. for 3 consecutive days, and on the second day samples were also taken at 3 h intervals starting from the 8:00 a.m. sample until 8:00 p.m. (i.e., 11:00 a.m., 2:00 p.m., 5:00 p.m., and 8:00 p.m.). All samples were analyzed using three methods: Enterolert, EPA 1611.1-based, Franklin-IPC. The results were mixed, gene copies per 100 ml determined by Franklin-IPC and EPA 1611.1-based methods were similar at Woodlawn but were markedly different at other sites including Buttermilk (**Supplementary Figure 5**). At each of the four intensive sampling locations, there were rain events during the sampling campaigns; however, they varied in intensity and duration. At Heckscher and Buttermilk rain was light ≤ 2.54 mm over <4 h. Heckscher had rain before the third day's morning sample (6:00 to 9:00 a.m.) and Buttermilk during the morning of the second day (9:00 to 12:00 p.m.). While Welch had 12.7 mm rain over 5 h (3:00 to 9:00 a.m.) and Woodlawn had 8.89 mm overnight for 12 h (10:00 to 10:00 a.m.). Heckscher was the only marine site included in the intensive sampling campaigns and showed no overlapping trends between the qPCR methods or between the qPCR methods and culture-based methods. Woodlawn was the only site that showed similar trends over time between the methods and did appear to show a spike in FIB concentrations across all three methods following the rain event (**Supplementary Figure 5**).

Pathogen Panel Results (Summer 2018)

The Pathogen Panel results showed there were few pathogens across morning samples from the New York State beaches (**Figure 7A**). One Treman sample (6/25) and two Fillmore Glen samples showed the highest pathogen counts for *Enterococcus* spp. and the various *E. coli* strains, while the other beach samples were relatively low for all pathogens (**Figure 7A**). Creek samples had higher pathogen counts than other watershed types, and with one exception (Taughannock 6/18) all samples with *Bacteroides* HumM2 were found in creek samples (**Figure 7A**). Grouping by geographic regions showed that the Finger Lakes samples had higher pathogen counts overall than the Great Lakes, Long Island, or Palisades regions, however the sampling regime from the Finger Lakes included more rain events (**Figure 7B**; **Supplementary Table 1**). Although there were only three marine beach sites, they had much lower pathogen counts overall, and more non-detects than the freshwater sites (**Supplementary Figure 6**). Several pathogens were never detected and thus were not included in the analysis, they were adenovirus, *Brevibacterium* spp. LA35 (avian marker), hepA, and *Shigella* spp. The internal spike-in positive control (MS2 virus) was detected in 20% of samples and 2 of 3 sterile water processing blanks with Cts ranging from 20.75 to 23.43. The majority of samples (80%) had MS2 below detection, possibly due to sample dilution.

Samples from the intensive sampling campaign for two Buttermilk, Welch, Heckscher, and all Woodlawn intensive samples were also examined using the Pathogen Panel and only seven pathogens were found across the sites (**Supplementary Figure 7**). All of the intensive sites captured a rain event, but a spike in pathogen concentrations was only seen at Woodlawn following an overnight rain. At Buttermilk, pathogen levels stayed elevated during an ongoing light morning rain event (**Supplementary Figure 7**). Pathogens were detected at both Heckscher (enterovirus and *Campylobacter* spp.) and Welch (Bacteroides HumM2) 2:00 p.m. samples but were not detected in the 8:00 a.m. samples on the same day (Welch) or the day after (Heckscher) with no rain event occurring during

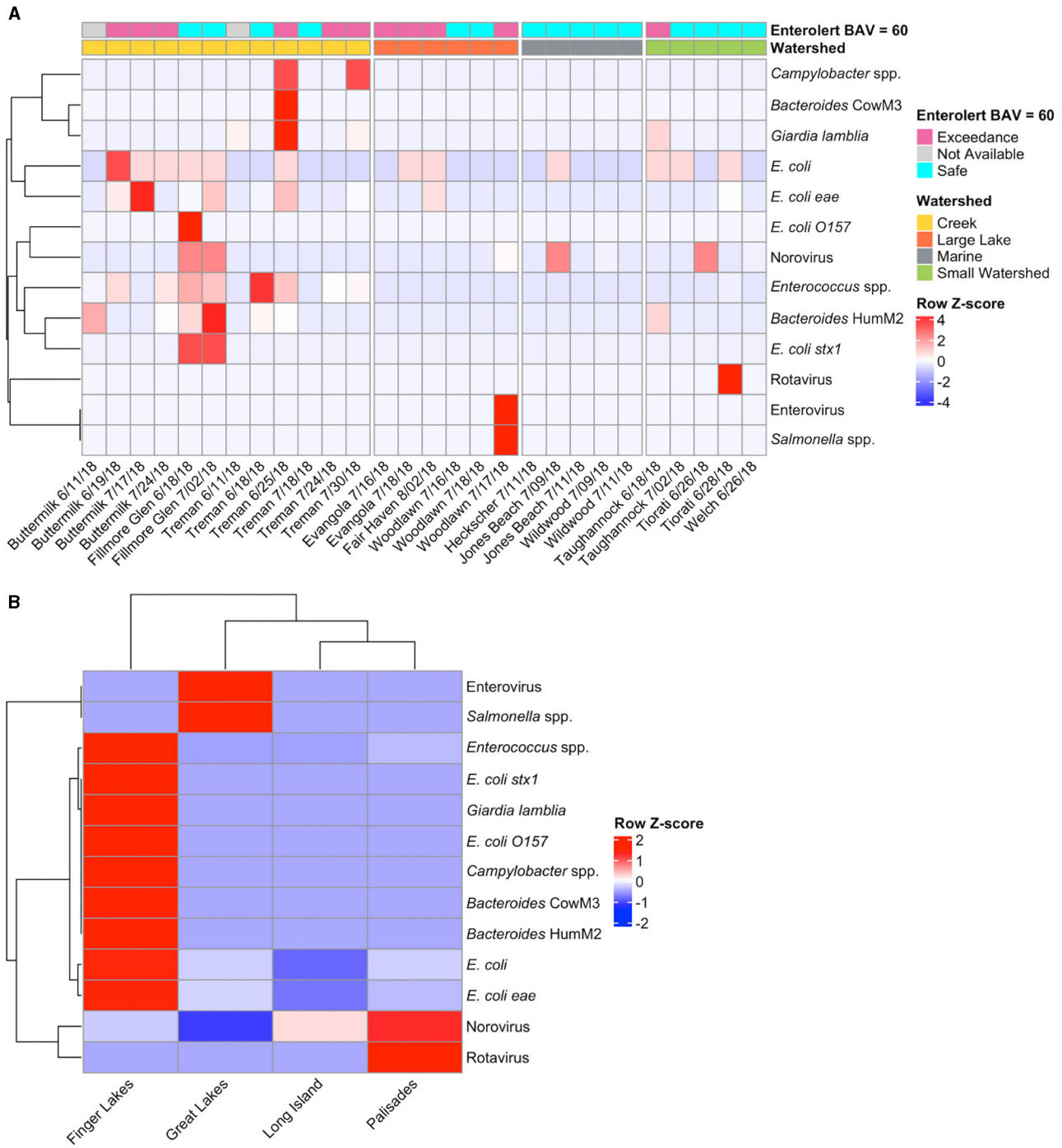


FIGURE 7 | Heat map of Pathogen Panel OpenArray-detected pathogens in the 2018 New York State beach morning samples for **(A)** individual samples by site and date and **(B)** samples grouped by region. For each sample, the watershed type (i.e., Creek, Large Lake, Marine, or Small Watershed) and the Enterolert results are shown. Samples were marked as "Safe" if Enterolert results were below a BAV of 60 MPN per 100 ml; "Exceedance" if results exceeded this cutoff and the beach was closed; or "Not Available" if Enterolert results were unavailable at this site.

this time period. Correlations between detected pathogens and environmental conditions (i.e., turbidity, water temperature, rainfall over previous 24 h and rainfall over previous 4 d) were examined (**Supplementary Figure 8**). Strong positive correlations were found between *Bacteroides* CowM3 and *Campylobacter* spp. ($r = 0.7$) and *Giardia lamblia* ($r = 0.73$).

Campylobacter spp. and *Giardia lamblia* ($r = 0.75$), and HumM2 with *E. coli stx1* and *Enterococcus* spp. ($r = 0.59$ and 0.57, respectively).

DISCUSSION

Preliminary Study Results (Fall 2017)

In 2017, all of the qPCR methods performed similarly when compared to the culture-based Enterolert method. There were significant correlations of the Enterolert results with the EPA 1611.1-based assay ($p = 0.00204$) and all three Biomeme methods (two3-5 µl, $p = 0.0159$; two3-15 µl, $p = 0.00046$; BioPoo, $p = 0.00292$). The correlation was strongest for the two3-15 µl ($r^2 = 0.687$) and the BioPoo assay ($r^2 = 0.645$), followed by EPA 1611.1-based assay ($r^2 = 0.594$) and the two3-5 µl ($r^2 = 0.537$). Although the qPCR results correlated well overall with the Enterolert data, they had variable performance for correctly predicting BAV exceedances (ranging from 68.8 to 81.3% and 62.5 to 75.0% correct classifications for BAVs of 60 and 104 culturable FIB per 100 ml, respectively). Occurrences of false negatives were higher (12.5 to 18.8% and 12.5 to 25% for BAVs of 60 and 104 MPN per 100 ml, respectively) than false positives (0 to 18.8% and 6.3 to 18.8% for BAVs of 60 and 104 MPN per 100 ml, respectively). Based on these results, adopting any of the qPCR methods as the standard for beach monitoring could increase the potential of swimmer pathogen exposure as we saw in some cases, the qPCR methods would have resulted in a beach opening while the culture-based method would have led to a beach closure. Noble et al. (2010) also found qPCR-based methods underestimated *Enterococcus* spp. and *E. coli* concentrations in recreational waters, possibly due to inhibition of the qPCR assays (Noble et al., 2010). On the other hand, qPCR false positives indicate an overestimation of FIBs and studies have likewise shown that qPCR-based methods can overestimate FIB abundance (Lavender and Kinzelman, 2009; Raith et al., 2014). The mixed incidence of both false negatives and false positives in our dataset suggests that qPCR methods may require individualized, assay-specific BAVs based on site-specific studies. The EPA has suggested a qPCR BAV of 640 CCE per 100 ml for the EPA 1611.1 method, however the Biomeme methods currently lack recommended BAVs (US EPA, 2015).

Neither the two3 methods nor the EPA 1611.1-based method used in this study included a sample processing control, thus, we were unable to assess DNA extraction efficiency in these samples. The BioPoo method did include an IPC which informed our sample dilution to reduce or eliminate IPC inhibition and, presumably, *tuf* assay inhibition. The BioPoo assay had the highest correct BAV classification of all the methods (81.3 and 68.8% for BAVs of 60 and 104 MPN per 100 ml, respectively). Interestingly, the BioPoo method targets a different gene, i.e., the *tuf* gene, than the two3 and EPA 1611.1 primers and probe, which target the 23S rRNA gene of *Enterococcus*. The BioPoo assay showed the lowest intercept (0.657) when compared to the Enterolert method out of all the qPCR methods (ranging from 1.99 to 2.44 for the two3 methods and up to 3.07 for the EPA-based method), suggesting the use of the *tuf* gene could lower the background signal compared to the 23S rRNA gene. Biomeme's research team selected assays with the design objective of reducing the range of the background signal and increasing resolution of results. However, this specificity comes with a cost with respect to sensitivity; the BioPoo assay was more likely than the other qPCR methods to give a false negative, underscoring the importance of creating a different BAV for each qPCR chemistry.

All of the Biomeme methods correlated significantly ($p < 0.05$ for all) with the EPA 1611.1-based method in 2017 (**Figure 5**). This relationship was partly driven by differences in FIB concentrations between the two sites. While Treman had relatively low culturable enterococci (average of 24.5 ± 19.4 MPN per 100 ml) Buttermilk had much higher levels (average of 184 ± 182 MPN per 100 ml) which strengthened the correlation between qPCR methods and between qPCR-based methods and Enterolert. The differences observed at these two sites is likely due to land use differences within their respectively watersheds. Buttermilk's watershed has more residential housing and agricultural land compared to Treman's watershed which has more forest cover and is more rural. These differences in land use result in less runoff originating from human-impacted areas in Treman's watershed compared to Buttermilk which could ultimately be reflected in the lower FIB counts observed in Treman's water samples as compared to Buttermilk's.

qPCR-Based FIB Monitoring (Summer 2018)

In summer of 2018, we expanded the monitoring sites from two local, freshwater beaches to 12 beaches across New York State, including nine freshwater and three marine beaches, from four geographic areas (i.e., Great Lakes, Finger Lakes, Palisades, and Long Island), and four watershed types (creek, small watershed, large lake, marine). The diversity of sampling sites likely led to weakening of the correlation between the EPA 1611.1-based assay and Enterolert results from 2017 ($r^2 = 0.59$, **Figure 4**) to 2018 ($r^2 = 0.317$, **Figure 6A**). When examining only the two sites studied in both years the correlation was still weaker in 2018 compared to 2017 ($r^2 = 0.208$, **Supplementary Figure 4**). This variability in correlation may be due to seasonal differences. In summer 2018 the water quality monitoring spanned 8 weeks, during which beaches were in use and rainfall was recorded, while in fall 2017 monitoring spanned 4 weeks, during which the beaches were closed, and no rainfall was recorded. Haugland et al. (2021) found good correlation between qPCR and culture-based methods for quantification of *E. coli* among freshwater sites in recreational water across Michigan. Because of this we chose to group the 2018 data by beach type (freshwater vs. marine) and found there was an improvement in the correlation between the EPA 1611.1-based assay and Enterolert results for freshwater beaches ($r^2 = 0.408$, $p < 0.0001$, **Supplementary Figure 2D**). In contrast, the marine beaches showed no significant correlation ($p = 0.262$, **Supplementary Figure 2D**). Differences including tides, salinity, and wrack could all impact the presence, concentration, and quantification of FIBs at marine beaches (Sinton et al., 2002; Boehm, 2007; Imamura et al., 2011; Byappanahalli et al., 2012). These results highlight the importance of creating site-specific correlations to account for the differing characteristics

of freshwater and marine beaches, including salinity, which are known to influence FIB quantification (Sinton et al., 2002; Imamura et al., 2011; Byappanahalli et al., 2012).

Unlike in 2017, the Biomeme results of summer 2018 did not correlate with the culture-based (**Figure 6A**) or EPA-based qPCR (**Figure 6B**) methods. Even when examining results by beach location or beach type, there did not appear to be a correlation between the Franklin-IPC and Enterolert measurements (**Supplementary Figures 2A,C**). However, the two qPCR methods (i.e., Franklin-IPC and EPA 1611.1-based assay) showed a moderate, significant correlation when grouping by beach location but only for the Finger Lakes region ($r^2 = 0.273$, $p = 0.0105$, **Supplementary Figure 1**). In general, the EPA 1611.1-based method appears to be more robust across different sites than the Franklin-IPC method, however, the Biomeme methods that showed strong correlations in fall of 2017 (i.e., C1, C2, and C3) were not tested in 2018. One of these methods, the BioPoo assay, may hold promise for future testing as it performed well in 2017, simultaneously targets HF183, and includes an IPC to test for inhibition. In this study, an IAC was used in the EPA 1611.1-based assay to assess inhibition in samples and results from the IAC analysis showed no amplification inhibition. However, an IAC cannot account for inefficiencies during DNA extraction. The EPA 1611.1 method prescribes the use of a sample processing control to address this and, additionally, recommends the use of a matrix spike to evaluate and correct for sample site variations. However, in this study we were only able to include the IAC analysis and not the prescribed sample processing control of matrix spikes in our EPA-based qPCR method. Including these additional analyses most likely would have resolved some of the site variability observed. Indeed, including both matrix spikes and sample processing controls may eliminate this issue, not only in the EPA-based assay but also the Biomeme assays.

We examined the impact of rain events, turbidity, and water temperature on the measured FIB concentrations and found rainfall to correlate moderately positively with Enterolert ($r = 0.59$ for previous 24 h rainfall); whereas only one of the qPCR methods, the EPA 1611.1-based assay, correlated with rainfall ($r = 0.37$ and -0.09 for EPA 1611.1-based and Franklin-IPC methods, respectively). Lavender and Kinzelman (2009) found that non-consensus between culture- and qPCR-based methods was associated with rainfall among other environmental conditions, thus it is possible that rainfall inhibited one or both of our qPCR measurements.

Indeed, it is likely the Franklin-IPC measurements were inhibited despite the use of sample dilution and IPC analysis. Differential amplification responses to inhibitors commonly found in environmental samples, including humic acids, have previously been described (Huggett et al., 2008; Green and Field, 2012). Green and Field (2012) found qPCR inhibition sensitivity to be assay dependent when comparing HF183 and *E. coli* amplification in the same sample and suggested it may be due to differing assay sensitivities to humic acids. Cao et al. (2012) suggested current internal control methods to assess *Enterococcus* inhibition are ineffective and should not be relied upon to determine enterococci inhibition. They recommended using dilutions and robust qPCR chemistries (i.e., those designed for environmental samples) to overcome sample inhibition, instead of relying on internal control assessments. Huggett et al. (2008) found inhibition variability between control and target reactions in which the control showed no inhibition, yet the target was inhibited. Thus, even though we were able to account for IPC inhibition in the Franklin-IPC method it is possible the *Enterococcus* assay itself was still inhibited, due to differences in assay sensitivities (Huggett et al., 2008; Green and Field, 2012). Huggett et al. (2008) suggest that internal controls should be assessed with their targets to ensure similar sensitivities to inhibitors likely to be present in a given sample. Future testing for the Franklin-IPC method should include some analysis of target and control inhibition as well as analysis of the chemistry's robustness to different environmental samples.

Challenges for Implementation of qPCR-Based FIB Monitoring

There are many challenges to the implementation of any new technology and/or assay for water quality monitoring. In our study we found some inconsistencies in the correlation between qPCR-based and culture-based methods for FIB quantification across a range of diverse sites. Kinzelman et al. (2011) found there were varying degrees of correlation between the culture- and qPCR-based methods for FIB quantification depending on site (i.e., geographic location), water type (marine vs. fresh), as well as qPCR inhibition, and differences in viable vs. non-viable cells. Additionally, environmental conditions such as recent rainfall and flow rate can impact the background DNA levels (e.g., fragmented DNA or dead cells) and we found greater discrepancies between culture and qPCR methods in 2018 as compared to 2017, possibly due to the multiple rain events recorded in 2018 (Lavender and Kinzelman, 2009). Sunlight is also known to play a role in modulating *E. coli* and enterococci concentrations, however photoinactivation of these organisms is not well-understood under varying environmental conditions and we did not measure irradiance in our study (Boehm et al., 2009).

The correlation between culture-based and qPCR-based methods was not expected to be 1-to-1 at any given site. qPCR-based methods do not discriminate between live and dead cells in a sample while culture-based methods exclusively quantify viable/culturable cells. However, previous studies have shown strong correlations between culture- and qPCR-based methods (Schang et al., 2016; Haugland et al., 2021). Thus, even though these methods may not necessarily result in the same absolute values for FIB quantification, a strong correlation between the two measurements suggest there is potential for qPCR methods to be used in place of culture-based methods for beach closure and reopening determinations.

Our study included many different sites with unique characteristics, and we were not able to create a universal correlation curve between any qPCR method and Enterolert that would work for all sites. Thus, site-specific studies would need to be completed before the implementation of any new qPCR method. In fact, the EPA recommends that any qPCR method, including EPA 1611.1, be evaluated on a site-specific

basis, mainly to assess the extent and frequency of PCR inhibition at the site (US EPA, 2012). As suggested by Ferretti et al. (2013) sampling locations need to be analyzed to determine if inhibition is an issue as well as to measure the variability of qPCR results on a spatial and temporal basis. Future studies of qPCR-based methods, including the rapid in-field Biomeme method, would benefit from extensively studying a few sites rather than studying a number of sites with diverse characteristics as well-including both sample processing controls (to assess DNA extraction variability) and matrix spikes (to assess site differences). Interestingly, Lavender and Kinzelman (2009), successfully applied site-specific corrections to increase correlations between qPCR and culture-based methods. Applying these site-specific correction factors to established correlation curves may be a promising solution to this issue.

For future testing of the Biomeme system, more locations across New York State, including marine beaches, would need to be studied. Additionally, more frequent sampling at each location, as well as method evaluation under different environmental conditions (e.g., rain) would need to be completed. This would allow for a better understanding of the overall performance of the Biomeme system and its applicability for NYS beach safety monitoring in the future. The creation of site-specific correlation curves between qPCR and culture-based FIB measurements would require the commitment and training of staff at each beach site to conduct these studies. Although time-consuming, implementation of qPCR methods necessitates the creation of robust, site-specific correlation curves which, once established, would result in the ability to rapidly test water (under 4 h) using the Biomeme system allowing for more timely closing and reopening of beaches. Of the Biomeme methods tested, the BioPoo assay holds the most promise because of its ability to be deployed on-site, the inclusion of an internal positive control (for inhibition assessment), and quick turnaround time for results. The results of this study suggest there is promise for qPCR-based methods for FIB monitoring, including the rapid, on-site methods tested herein. However, it is essential that future testing of qPCR-based methods rigorously examine site variations before they are employed. It cannot be overstated, that a method which can successfully quantify FIB (from filtration to DNA extraction and quantification) on-site would lead to safer swimming conditions. The ability to test samples without the delay caused by travel times to a lab as well as the ease-of-use of a method such as Biomeme, make more frequent water testing feasible and could eliminate the need for highly technical personnel to conduct testing.

Pathogen Panel Results (Summer 2018)

The Pathogen Panel used in this study was previously tested on water samples taken from the Hudson River watershed (Brooks et al., 2020). Here we used the panel to analyze a subset of samples from the 12 New York State beaches examined in summer 2018. We aimed to test the robustness of the panel on a wide range of sample locations which included a minimum of one sample per location, different water types (i.e., marine and freshwater), and rain events. Of the 12 viral, bacterial and protozoal pathogens monitored via the Pathogen Panel only three showed up with any substantial frequency: *E. coli eae* (25% of samples), norovirus (31.4%), and *Giardia lamblia* (11.4%), with others (including *Campylobacter* spp., *E. coli* O157, *E. coli stx1*, enterovirus, rotavirus, and *Salmonella* spp.) present in $< 6\%$ of the 35 samples. Three pathogens were not detected in any sample: *Shigella* spp., adenovirus, and hepatitis A.

Fillmore Glen had a surprisingly high Pathogen Panel gene count for the FIB, *Enterococcus* spp., (average 7,733 $\pm$ 1,283 23S rRNA genes per 100 ml) given the low Enterolert FIB counts (average of 9.25 $\pm$ 4.17 culturable enterococci per 100 ml). This suggests that there may be a significant source of FIB upstream of the swimming area, but that most of the intact FIB are not culturable. The human-specific FIB marker, HumM2, was also detected in these samples as were pathogenic *E. coli* (strains *stx1* and *eae*). The Pathogen Panel cannot distinguish between viable and dead or deactivated cells; thus, the high signal could be due to an upstream wastewater treatment facility that is inactivating, but not effectively removing the pathogens. In addition to Fillmore Glen, HumM2 was found at two other beaches, Buttermilk and Treman, on multiple dates, as well as Taughannock and Welch on one date each (HumM2 was found in 9 of the 35 samples). However, no sample had HumM2 levels above the median concentrations resulting in 30 GI illnesses per 1,000 swimmers (2,800 gene copies per 100 ml), suggesting any potential wastewater contamination is minimal (Boehm et al., 2015). Interestingly, we found weak to moderate positive correlations between the culture-based measurements (Enterolert and Colilert) and some pathogens, including *Giardia lamblia* and *Salmonella* spp. (**Supplementary Figure 8**). However other pathogens and MST markers, such as HumM2, showed no or negative correlations with the culture-based methods. These results indicate that the current water monitoring methods which measure enterococci and *E. coli* may not be indicative of pathogen presence (and human health risk) as previously suggested by Wu et al. (2011).

At the intensive sampling sites there were changes in the pathogen profile over time (**Supplementary Figure 7**). This is consistent with previous studies showing concentrations of FIB and pathogens can change within hours (Kim et al., 2009; Converse et al., 2012a; Desai and Rifai, 2013; Dorevitch et al., 2017). A study by Kim et al. (2009) found combined sewer overflow systems showed different diurnal patterns for FIB and pathogens (including noroviruses and enteroviruses), where FIB typically peak in the morning and remain relatively stable through the afternoon before decreasing overnight, whereas pathogens showed more variability. Here, changes in pathogen concentrations were observed over the course of a day at the Woodlawn site. Most notably, concentrations changed between the morning and afternoon samples, which also followed an overnight rain event, potentially amplifying diurnal changes in FIB and pathogen concentrations (**Supplementary Figures 5**, **7**). There was a spike in *Enterococcus* spp., norovirus, and enterovirus in the first sample following the rain event, whereas later in the day *E. coli eae* and *Campylobacter* spp. were detected. If the beach sites in this study are impacted by combined sewer overflows, wastewater treatment plant outfalls, or onsite wastewater treatment systems such as septic systems, diurnal FIB and pathogen patterns could be expected.

The poultry-associated MST marker (*Brevibacterium* spp. LA35 16S rRNA) was never detected, consistent with a lack of poultry farms around the study beaches. However, the lack of detection of this MST marker does not necessarily suggest poultry are not contributing to water quality issues. As stated in Brooks et al. (2020), the large amplicon size for the poultry associated MST marker (571 bp) may have impacted the detection of this marker. Previous studies have found that the presence of waterfowl significantly negatively impacts water quality and is associated with increased detection of *E. coli*, *Enterococcus* spp., and potential human pathogens (human-specific *Bacteroides* HF183 and *Campylobacter* spp.) (Hansen et al., 2011; Converse et al., 2012b); however due to the lack of poultry MST marker detection in our study we were unable to find any similar correlation at the NYS beaches. The cow MST marker was only found at one beach on one date (Treman 6/25). This was a date preemptively closed for heavy rain (>17 mm) and showed the presence of other pathogens including *Giardia lamblia*. The infrequent detection of *Bacteroides* CowM3 suggests cow feces did not have a major impact on beach water quality at these parks.

The Pathogen Panel positive control, MS2 virus, was only detected in 20% of water samples and was below detection in all other samples. Because the qPCR cycle threshold values for the MS2 virus were large, it is possible that even a small amount of partial inhibition could result in a non-detect of the MS2 target. Huggett et al. (2008) suggested that inhibition compatibility, between controls and targets, is especially important for qPCR array methods which use one control for myriad targets that may have differential inhibitor responses. Although we were able to detect pathogens at low levels in our samples even when MS2 was not detected, there still appeared to be some partial inhibition in samples which could lead to an underestimation of actual target levels. Future ns-qPCR array-based methods with multiple targets should consider the use of several positive controls exhibiting inhibition responses similar to the different target genes.

In general, the Finger Lakes sites had the highest pathogen detection of any of the regions, with Long Island having the least (**Figure 7B**). This may be due in part to capturing more rain events in the Finger Lakes region than any other region (average rainfall of 8.7 mm over 24 h period prior to sampling and 51.6% of samples capturing a rain event, **Supplementary Table 1**). There were fewer pathogens detected at marine sites as compared to freshwater sites, and any pathogen detected in marine samples was also detected in freshwater samples (**Supplementary Figure 6**). Previous studies have found differential decay of the HF183 marker between marine and freshwater samples and although this marker was not included in our Pathogen Panel it does suggest that pathogen marker concentrations can vary substantially by beach type (Shanks et al., 2006; Green et al., 2011). Nine of the 35 samples run on the Pathogen Panel had no detection of any pathogens. This included one Evangola State Park sample which had confirmed good water quality, the MS2 control was positive for both sampling dates while all other targets were negative for the first sample date and had only *E. coli* present on the second date.

CONCLUSIONS

The qPCR method based on EPA 1611.1 in this study yielded *E. faecalis* quantities that correlated positively with culture-based values in both years, while Biomeme methods showed strong correlations in 2017 and mixed results in 2018. The BioPoo assay (Method C3) had the highest agreement with the Enterolert method on exceedance of the 60 MPN per 100 ml BAV (81.3% agreement of the samples) but also had one of the higher false negative rates (18.8%). The BioPoo assay had the lowest background signal of the qPCR methods tested, demonstrating its potential for beach monitoring, but would require a site-specific correlation curve to avoid the incidences of false negatives.

In 2018, the Biomeme Franklin-IPC results did not correlate with Enterolert results, however they did moderately correlate with the EPA-based qPCR results when considering only the Finger Lakes region ($r^2 = 0.273$, $p = 0.0105$), indicating site-specific correlation curves are particularly important for this assay. The EPA 1611.1-based assay results correlated weakly with the Enterolert results, suggesting this method is more robust to site differences ($r^2 = 0.317$, $p = 0.00012$), however this relationship may be improved by the use of a sample processing control and a matrix spike analysis for each site. The Pathogen Panel successfully detected low levels of MST and pathogen markers across a broad range of sites in 2018. Results revealed Finger Lakes creek beaches had greater pathogen detection than other regions while marine beaches had the lowest pathogen detection. A quarter of the samples showed no FIB or pathogens detected with the Pathogen Panel; however, it is possible there was inhibition in some samples.

Of the Biomeme methods tested, the BioPoo assay holds promise for rapid on-site testing due to its ease-of use and the ability to assess qPCR inhibition using the internal positive control. Additionally, it can measure HF183 which could help inform beach sites of potential human fecal contamination. However, prior to implementation of any qPCR-based methods for FIB monitoring it is critical to assess inhibition of qPCR runs to avoid false negatives. Further testing to assess qPCR inhibition of sample controls and sample targets is needed to ensure similar inhibition responses in any qPCR assay for FIB monitoring. Beach-specific correlations must also be established between qPCR and culture-based assays. More testing is needed to evaluate the robustness of the Biomeme workflow for rapid, on-site monitoring of FIB.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CF-B, RR, CS, and YB conceived and designed the experiments. CF-B, ZS, NB, DS, and JB performed the experiments. CF-B, CS, DW, and RR analyzed the data. CF-B wrote

the initial manuscript. All authors approved the manuscript for publication.

FUNDING

Funding for this work was provided by Cornell's Atkinson Center for a Sustainable Future (ACSF) via the Academic Venture Fund and the Department of the Interior's US Geological Survey administered through the New York State Water Resources Institute.

ACKNOWLEDGMENTS

We thank the NYS Office of Parks, Recreation & Historic Preservation for access to samples from swimming areas at state beaches. We would also like to thank Susan Foster who ran the internal amplification control in our samples. We would like to acknowledge Cornell's Atkinson Center for a Sustainable Future (ACSF) for funding via its Academic Venture Fund, the Department of the Interior's US Geological Survey administered through the New York State Water Resources Institute. These funding agencies neither directed research nor reviewed this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.711477/full#supplementary-material

REFERENCES

Bae, S., and Wuertz, S. (2012). Survival of host-associated bacteroidales cells and their relationship with Enterococcus spp., Campylobacter jejuni, Salmonella enterica, Serovar Typhimurium, and Adenovirus in freshwater microcosms as measured by Propidium Monoazide-Quantitative PCR. *Appl. Environ. Microbiol.* 78, 922–932. doi: 10.1128/AEM.05157-11

Bernardes, C., Bernardes, R., Zimmer, C., and Dorea, C. C. (2020). A simple off-grid incubator for microbiological water quality analysis. *Water* 12:240. doi: 10.3390/w12010240

Boehm, A. B. (2007). Enterococci concentrations in diverse coastal environments exhibit extreme variability. *Environ. Sci. Technol.* 41, 8227–8232. doi: 10.1021/es071807v

Boehm, A. B., Soller, J. A., and Shanks, O. C. (2015). Human-associated fecal quantitative polymerase chain reaction measurements and simulated risk of gastrointestinal illness in recreational waters contaminated with raw sewage. *Environ. Sci. Technol. Lett.* 2, 270–275. doi: 10.1021/acs.estlett.5b00219

Boehm, A. B., Yamahara, K. M., Love, D. C., Peterson, B. M., McNeill, K., and Nelson, K. L. (2009). Covariation and photoinactivation of traditional and novel indicator organisms and human viruses at a sewage-impacted marine beach. *Environ. Sci. Technol.* 43, 8046–8052. doi: 10.1021/es9015124

Bramburger, A. J., Brown, R. S., Haley, J., and Ridal, J. J. (2015). A new, automated rapid fluorometric method for the detection of Escherichia coli in recreational waters. *J. Great Lakes Res.* 41, 298–302. doi: 10.1016/j.jglr.2014.12.008

Brooks, Y. M., Collins, S. M., Mbullo, P., Baoteng, G. O., Young, S. L., and Richardson, S. L. (2017). Evaluating human sensory perceptions and the compartment bag test assays as proxies for the presence and concentration of Escherichia coli in drinking water in Western Kenya. *Am. J. Trop. Med. Hyg.* 97, 1005–1008. doi: 10.4269/ajtmh.16-0878

Brooks, Y. M., Spirito, C. M., Bae, J. S., Hong, A., Mosier, E. M., Sausele, D. J., et al. (2020). Fecal indicator bacteria, fecal source tracking markers, and pathogens detected in two Hudson River tributaries. *Water Res.* 71:115342. doi: 10.1016/j.watres.2019.115342

Byappanahalli, M. N., Nevers, M. B., Korajkic, A., Staley, Z. R., and Harwood, V. J. (2012). Enterococci in the environment. *Microbiol. Mol. Biol. Rev.* 76, 685–706. doi: 10.1128/MMBR.00023-12

Byappanahalli, M. N., Nevers, M. B., Shively, D. A., Spoljaric, A., and Otto, C. (2018). Real-time water quality monitoring at a Great Lakes National Park. *J. Environ. Qual.* 47, 1086–1093. doi: 10.2134/jeq2017.11.0462

Cao, Y., Griffith, J. F., Dorevitch, S., and Weisberg, S. B. (2012). Effectiveness of qPCR permutations, internal controls and dilution as means for minimizing the impact of inhibition while measuring Enterococcus in environmental waters. *J. Appl. Microbiol.* 113, 66–75. doi: 10.1111/j.1365-2672.2012.05305.x

Chern, E. C., Siefring, S., Paar, J., Doolittle, M., and Haugland, R. A. (2011). Comparison of quantitative PCR assays for Escherichia coli targeting ribosomal RNA and single copy genes: qPCR assays for E. coli targeting rRNA and single copy genes. *Lett. Appl. Microbiol.* 52, 298–306. doi: 10.1111/j.1472-765X.2010.03001.x

Converse, R. R., Griffith, J. F., Noble, R. T., Haugland, R. A., Schiff, K. C., and Weisberg, S. B. (2012a). Correlation between quantitative PCR and culture-based methods for measuring Enterococcus spp. over various temporal scales at three California marine beaches. *Appl. Environ. Microbiol.* 78, 1237–1242. doi: 10.1128/AEM.07136-11

Converse, R. R., Kinzelman, J. L., Sams, E. A., Hudgens, E., Dufour, A. P., Ryu, H., et al. (2012b). Dramatic improvements in beach water quality following gull removal. *Environ. Sci. Technol.* 46, 10206–10213. doi: 10.1021/es302306b

Desai, A. M., and Rifai, H. S. (2013). Escherichia coli concentrations in urban watersheds exhibit diurnal sag: implications for water-quality monitoring and assessment. *J. Am. Water Resour. Assoc.* 49, 766–779. doi: 10.1111/jawr.12039

Dorevitch, S., Shrestha, A., DeFlorio-Barker, S., Breitenbach, C., and Heimler, I. (2017). Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: implications for public notification. *Environ. Health* 16:45. doi: 10.1186/s12940-017-0256-y

Dreier, J., Störmer, M., and Kleesiek, K. (2005). Use of bacteriophage MS2 as an internal control in viral reverse transcription-PCR assays. *J. Clin. Microbiol.* 43, 4551–4557. doi: 10.1128/JCM.43.9.4551-4557.2005

Ferretti, J. A., Tran, H. V., Peterson, S. J., and Loftin, V. (2013). Rapid method demonstration project at four New Jersey marine beaches using real time quantitative Polymerase Chain Reaction (qPCR). *Mar. Pollut. Bull.* 71, 51–63. doi: 10.1016/j.marpolbul.2013.03.038

Frahm, E., and Obst, U. (2003). Application of the fluorogenic probe technique (TaqMan PCR) to the detection of Enterococcus spp. and Escherichia coli in water samples. J. Microbiol. Methods 52, 123–131. doi: 10.1016/S0167-7012(02)00150-1

Goodman, L. B., Anderson, R. R., Slater, M., Ortenberg, E., Renshaw, R. W., Chilson, B. D., et al. (2016). High-throughput detection of respiratory pathogens in animal specimens by nanoscale PCR. *J. Vis. Exp.* 117:e54781. doi: 10.3791/54781

Green, H. C., and Field, K. G. (2012). Sensitive detection of sample interference in environmental qPCR. *Water Res.* 46, 3251–3260. doi: 10.1016/j.watres.2012.03.041

Green, H. C., Shanks, O. C., Sivaganesan, M., Haugland, R. A., and Field, K. G. (2011). Differential decay of human faecal Bacteroides in marine and freshwater. *Environ. Microbiol.* 13, 3235–3249. doi: 10.1111/j.1462-2920.2011.02549.x

Griffith, J. F., and Weisberg, S. B. (2011). Challenges in implementing new technology for beach water quality monitoring: lessons from a California demonstration project. *Mar. Technol. Soc. J.* 45, 65–73. doi: 10.4031/MTSJ.45.2.13

Hansen, D. L., Ishii, S., Sadowsky, M. J., and Hicks, R. E. (2011). Waterfowl abundance does not predict the dominant avian source of beach Escherichia coli. *J. Environ. Qual.* 40, 1924–1931. doi: 10.2134/jeq2011.0111

Harwood, V. J., Staley, C., Badgley, B. D., Borges, K., and Korajkic, A. (2014). Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* 38:1–40 doi: 10.1111/1574-6976.12031

Haugland, R., Oshima, K., Sivaganesan, M., Dufour, A., Varma, M., Seifring, S., et al. (2021). Large-scale comparison of E. coli levels determined by culture and a qPCR method (EPA Draft Method C) in Michigan towards the implementation of rapid, multi-site beach testing. *J. Microbiol. Methods* 184:106186. doi: 10.1016/j.mimet.2021.106186

Huggett, J. F., Novak, T., and Garson, J. A., Green, C., Morris-Jones, S. D., Miller, R. F., et al. (2008). Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. *BMC Res. Notes* 1, 1–9. doi: 10.1186/1756-0500-1-70

Huijsdens, X. W., Linskens, R. K., Mak, M., Meuwissen, S. G. M., Vandenbroucke-Grauls, C. M. J. E., and Savelkoul, P. H. M. (2002). Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. *J. Clin. Microbiol.* 40, 4423–4427. doi: 10.1128/JCM.40.12.4423-4427.2002

Imamura, G. J., Thompson, R. S., Boehm, A. B., and Jay, J. A. (2011). Wrack promotes the persistence of fecal indicator bacteria in marine sands and seawater. *FEMS Microbiol. Ecol.* 77, 40–49. doi: 10.1111/j.1574-6941.2011.01082.x

Kim, W. J., Managaki, S., Furumai, H., and Nakajima, F. (2009). Diurnal fluctuation of indicator microorganisms and intestinal viruses in combined sewer system. *Water Sci. Technol.* 60, 2791–2801. doi: 10.2166/wst.2009.732

Kinzelman, J. L., Bushon, R. N., Dorevitch, S., and Noble, R. T. (2011). Comparative Evaluation of Molecular and Culture Methods for Fecal Indicator Bacteria for Use in Inland Recreational Waters. London: Water Intelligence Online. doi: 10.2166/9781843395454

Kolm, C., Martzy, R., Brunner, K., Mach, R. L., Krska, R., Heinze, G., et al. (2017). A complementary isothermal amplification method to the U.S. EPA quantitative polymerase chain reaction approach for the detection of enterococci in environmental waters. *Environ. Sci. Technol.* 51, 7028–7035. doi: 10.1021/acs.est.7b01074

Lavender, J. S., and Kinzelman, J. L. (2009). A cross comparison of qPCR to agar-based or defined substrate test methods for the determination of Escherichia coli and enterococci in municipal water quality monitoring programs. *Water Res.* 43, 4967–4979. doi: 10.1016/j.watres.2009.08.010

Mauk, M. G., Liu, C., Sadik, M., and Bau, H. H. (2015). "Microfluidic devices for Nucleic Acid (NA) isolation, isothermal NA amplification, and real-time detection," in Mobile Health Technologies, eds A., Rasooly and K. E. Herold (New York, NY: Springer New York), 15–40. doi: 10.1007/978-1-4939-2172-0_2

Noble, R. T., Blackwood, A. D., Griffith, J. F., McGee, C. D., and Weisberg, S. B. (2010). Comparison of rapid quantitative PCR-based and conventional culture-based methods for enumeration of Enterococcus spp. and Escherichia coli in recreational waters. Appl. Environ. Microbiol. 76, 7437–7443. doi: 10.1128/AEM.00651-10

Polaczyk, A. L., Narayanan, J., Cromeans, T. L., Hahn, D., Roberts, J. M., Amburgey, J. E., et al. (2008). Ultrafiltration-based techniques for rapid and simultaneous concentration of multiple microbe classes from 100-L tap water samples. *J. Microbiol. Methods* 73, 92–99. doi: 10.1016/j.mimet.2008.02.014

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raith, M. R., Ebentier, D. L., Cao, Y., Griffith, J. F., and Weisberg, S. B. (2014). Factors affecting the relationship between quantitative polymerase chain reaction (qPCR) and culture-based enumeration of Enterococcus in environmental waters. *J. Appl. Microbiol.* 116, 737–746. doi: 10.1111/jam.12383

Rhodes, E. R., Villegas, L. F., Shaw, N. J., Miller, C., and Villegas, E. N. (2012). A modified EPA method 1623 that uses tangential flow hollow-fiber ultrafiltration and heat dissociation steps to detect waterborne Cryptosporidium and Giardia spp. *J. Vis. Exp.* 65:e4177. doi: 10.3791/4177

Schang, C., Henry, R., Kolotelo, P. A., Prosser, T., Crosbie, N., Grant, T., et al. (2016). Evaluation of techniques for measuring microbial hazards in bathing waters: a comparative study. *PLoS ONE* 11:e0155848. doi: 10.1371/journal.pone.0155848

Sepulveda, A., Hutchins, P., Massengill, R., and Dunker, K. (2018). Tradeoffs of a portable, field-based environmental DNA platform for detecting invasive northern pike (Esox lucius) in Alaska. Manage. Biol. Invasions 9, 253–258. doi: 10.3391/mbi.2018.9.3.07

Shanks, O. C., Nietch, C., Simonich, M., Younger, M., Reynolds, D., and Field, K. G. (2006). Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Appl. Environ. Microbiol.* 72, 5537–5546. doi: 10.1128/AEM.03059-05

Sinton, L. W., Hall, C. H., Lynch, P. A., and Davies-Colley, R. J. (2002). Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. *Appl. Environ. Microbiol.* 68, 1122–1131. doi: 10.1128/AEM.68.3.1122-1131.2002

US EPA (2013). *Method 1609: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay*. Washington, DC: Office of Water.

US EPA (2015). *Method 1611: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay*. Washington, DC: Office of Water.

US EPA (2019). *Method 1696: Characterization of Human Fecal Pollution in Water by HF183/BacR287 TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay*. Washington, DC: Office of Water.

US EPA, OW (2012). "Microbial (Pathogen)/Recreational Water Quality Criteria." Reports and Assessments. Washington, DC: US EPA. Available online at: https://www.epa.gov/wqc/microbial-pathogenrecreational-water-quality-criteria (accessed August 31, 2021).

Wang, A., McMahan, L., Rutstein, S., Sobsey, M. D., Stauber, C., and Reyes, J. (2017). household microbial water quality testing in a peruvian demographic and health survey: evaluation of the compartment bag test for Escherichia coli. *Am. J. Trop. Med. Hyg.* 96, 970–975. doi: 10.4269/ajtmh.15-0717

Wang, D., Yamahara, K. M., Cao, Y., and Boehm, A. B. (2016). Absolute quantification of enterococcal 23S rRNA gene using digital PCR. *Environ. Sci. Technol.* 50, 3399–3408. doi: 10.1021/acs.est.5b05747

Water Quality Monitoring (2004). *Codes Rules and Regulations*. New York, NY: Water Quality Monitoring.

Wu, J., Long, S. C., Das, D., and Dorner, S. M. (2011). Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *J. Water Health* 9, 265–278. doi: 10.2166/wh.2011.117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 01 November 2021
doi: 10.3389/frwa.2021.724025

Temporal Stability of Phytoplankton Functional Groups Within Two Agricultural Irrigation Ponds in Maryland, USA

Jaclyn E. Smith[1,2,3]*, *Jennifer L. Wolny*[4†], *Matthew D. Stocker*[1,2,3], *Robert L. Hill*[2] and *Yakov A. Pachepsky*[1]

[1] Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD, United States, [2] Department of Environmental Science and Technology, University of Maryland, College Park, MD, United States, [3] Oak Ridge Institute for Science and Education, Oak Ridge, TN, United States, [4] Maryland Department of Natural Resources, Annapolis, MD, United States

OPEN ACCESS

Edited by:
Jasna Kovac,
The Pennsylvania State University (PSU), United States

Reviewed by:
Maria Emilia Cunha,
Instituto Português do Mar e Atmosfera, Portugal
Raymond James Ritchie,
Prince of Songkla University, Thailand

***Correspondence:**
Jaclyn E. Smith
jackie.smith@usda.gov

†Present address:
Jennifer L. Wolny,
Office of Regulatory Science, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, College Park, MD, United States

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: 11 June 2021
Accepted: 30 September 2021
Published: 01 November 2021

Citation:
Smith JE, Wolny JL, Stocker MD, Hill RL and Pachepsky YA (2021) Temporal Stability of Phytoplankton Functional Groups Within Two Agricultural Irrigation Ponds in Maryland, USA. Front. Water 3:724025. doi: 10.3389/frwa.2021.724025

Phytoplankton functional groups and their influence on water quality have been studied in various types of water bodies but have yet to be studied in agricultural irrigation ponds. Freshwater sources (e.g., lakes, rivers, and reservoirs) have been previously shown to exhibit high spatial and temporal variability in phytoplankton populations. Improvements in the monitoring of phytoplankton populations may be achieved if patterns of stable spatial variability can be found in the phytoplankton populations through time. The objective of this work was to determine if temporally stable spatial patterns in phytoplankton communities could be detected in agricultural irrigation ponds using a functional group approach. The study was performed at two working agricultural irrigation ponds located in Maryland, USA over two summer sampling campaigns in 2017 and 2018. Concentrations of four phytoplankton groups, along with sensor-based and fluorometer based water quality parameters were measured. Temporal stability was assessed using mean relative differences between measurements in each location and averaged measurements across ponds on each sampling date. Temporally stable spatial patterns of three phytoplankton functional groups were found for both ponds over the two sampling seasons. Both ponds had locations where specific phytoplankton functional group concentrations were consistently higher or lower than the pond's average concentration for each sampling date. Zones of consistently higher or lower than average concentrations were associated with flow conditions, pond morphology, and human activities. The existence of temporally stable patterns of phytoplankton functional group concentrations can affect the outcome of a water quality assessment and should be considered in water quality monitoring designs.

Keywords: agricultural irrigation pond, mean relative difference analysis, phytoplankton functional groups, spatial patterns, water quality

INTRODUCTION

Phytoplankton are commonly found members of microbial populations within many diverse water bodies including agricultural irrigation ponds. These primary producers are an important component of the food web within aquatic ecosystems. Previous research has shown that phytoplankton may be an effective bio-indicator of water quality and also a reflection of ecosystem health (Wang et al., 2015; Su et al., 2017; Adloff et al., 2018).

Freshwater phytoplankton populations are typically divided into functional groups based on morphology, physiology, adaptations, and ecological attributes (Reynolds et al., 2002; Varol, 2019; Jin et al., 2020). Three major phytoplankton functional groups are diatoms (Bacillariophyta), green algae (Chlorophyta), and cyanobacteria (Cyanophyta; also commonly referred to as blue-green algae), each of which possess different qualities that may influence and be indicative of water quality (Shi et al., 2012, 2015; Xiao et al., 2013). The richness and uniformity of the phytoplankton community may also indicate different water properties and a range of water qualities from pristine to degraded water quality conditions. Phytoplankton communities have been utilized as an indication of the trophic state of a water body (Hu et al., 2012; Ren et al., 2016; Rimet and Druart, 2018), to confirm eutrophication (Ren et al., 2016; Varol, 2019), pollution and/or other anthropogenic effects (Shi et al., 2015; Feki-Sahnoun et al., 2018). The use of phytoplankton functional groups in more complex assessments, such as understanding biogeochemical models (Shimoda and Arhonditsis, 2016) and in the development of remote sensing technologies (Wolanin et al., 2016; Vandermeulen et al., 2017; Xi et al., 2017) continues to be a growing research area in large water bodies or on broad scales, but less is known about the temporal stability of these groups on smaller scale irrigation water systems (e.g., irrigation ponds, retention ponds, and aquaculture ponds).

Agricultural irrigation water has been shown to play a substantial role in the microbial contamination of fresh produce and foodborne illness outbreaks (World Health Organization, 2008; Uyttendaele et al., 2015; Jongman and Korsten, 2018). Certain groups of phytoplankton can form large proliferations or "blooms" and release toxins into the environment (Wood, 2016; Bouma-Gregson et al., 2017) which can be biotransported into the food supply (Bittencourt-Oliveira et al., 2016; Buratti et al., 2017). This presents both environmental and human health risks. Monitoring of irrigation water quality is important to avoid the transport of degraded and potentially contaminated waters to nearby crops.

Research on phytoplankton communities has previously been conducted across numerous water body types to determine spatial and temporal population trends, assess species composition, and community responses to changes in water quality. Within the Chesapeake Bay watershed, long-term phytoplankton data sets have been used to augment and support water quality guidelines in lakes, rivers, and estuaries (Marshall et al., 2006, 2009; Marshall, 2013, 2014; Hernandez Cordero et al., 2020), but not specifically for agricultural irrigation waters.

Although phytoplankton may be used as water quality bio-indicators, attempts to integrate phytoplankton community assessments to agricultural irrigation water quality seemingly have been limited to laboratory studies (DeLorenzo et al., 2002). The objective of this study was to determine if temporally stable spatial patterns of phytoplankton functional groups exist within temperate agricultural irrigation ponds and if these groups could be correlated to easily measured water quality parameters which could lead to potential improvements in on-farm water quality monitoring and aid with the prediction and mitigation of food-safety issues.

METHODS

Pond Monitoring

Sampling was conducted at two working farms for two consecutive growing seasons (2017–2018). These ponds were chosen because water was routinely drawn for irrigation of co-located crop fields. Each pond was sampled six times during the May through October growing season, with an exception to Pond 2 in 2017 with only five sampling dates. For 2017 sampling occurred from May to August and for 2018 from June to October. This resulted in a total of 276 and 242 phytoplankton samples collected for Pond 1 and Pond 2, respectively. Both ponds were located within a one hour drive from the USDA-ARS laboratory, so samples were maintained at ambient temperature and processed the same day as collection.

Site Descriptions

Pond 1 is a 1.01-acre man-made embankment pond located in Germantown, MD, USA with an average depth of 2.7 m (**Figure 1**-P1). Vegetation surrounding Pond 1 embankments consisted of deciduous trees and shrubs along the northern and eastern banks with the remaining embankments having a grass cover. The pond is surrounded by crop fields. When the water level in this pond gets low, the farm operators will occasionally pump water into Pond 1 from another pond which is stream-fed. The inflow and outflows are both located near sampling location 15. The irrigation pump intake is located near location 12 and is ∼2-3 feet below the water's surface. The photic zone in Pond 1, as determined by Secchi depth, averages 0.8 m. In 2017 and 2018, the algicide copper sulfate was commonly used to treat the water in Pond 1.

Pond 2 is located at the University of Maryland Wye Research Center in Wye Mills, MD, USA (**Figure 1**-P2). This pond is a 1.05-acre excavated pond with an average depth of 2.7 m and most of the bank areas are covered with grass and dense shrubs. Large trees are also present along the perimeter but are ∼20 m from the water's edge. This pond is surrounded by crop fields, farm buildings, and one residential property. In March of each year, the surrounding crop fields receive chemical fertilizers, but no animal manures are applied. This pond is primarily fed through rainfall which typically enters through an ephemeral creek that leads into a culvert near location 12. This culvert tends to have a substantial inflow only when precipitation has recently occurred. On the south end of the pond, there is a water-level dependent outflow drain near location 24. The irrigation pump

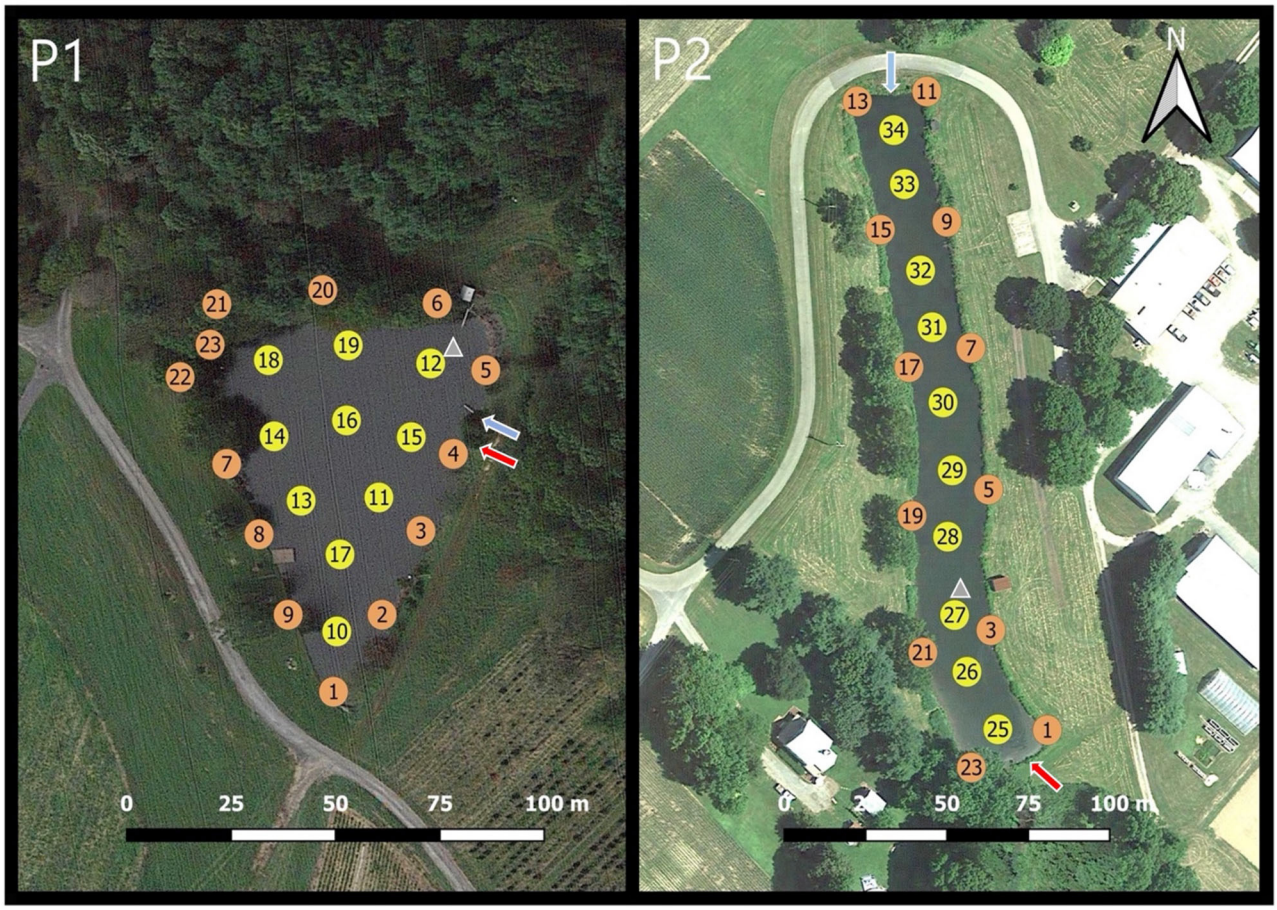


FIGURE 1 | Sampling locations for both Pond 1 (P1) and Pond 2 (P2). Station location number is stated inside the circle. Yellow circles indicate interior water sampling locations and orange circles indicate nearshore sampling locations. Blue arrows represent inflow points and outflows are represented by red arrows. Irrigation intake is represented by a gray triangle.

intake is near location 27 and is ∼2-3 feet below the water's surface. The depth of the photic zone, determined by Secchi depth, for Pond 2 averages 0.5 m.

Sample Collection, Handling, and Storage

Pond 1 had 23 sampling locations and Pond 2 had 22 sampling locations (**Figure 1**). Surface water samples were taken at a depth of 0–15 cm. Nearshore samples were taken with a 500 mL hand grab sampler at ∼1.5 m from the shoreline. Interior samples were taken from a boat with GPS tracking used to provide consistency of sampling locations between different sampling dates. Sampling locations remained the same for every sampling date over both years. After collection, samples were immediately placed into a cooler without ice to help maintain the original ambient water temperature. Samples were then transported to the lab for analysis.

In-Field Measurements

In-situ water quality measurements were taken concurrently with sample collection using a YSI Exo-2 sonde (YSI Inc., Yellow Springs, OH). The YSI sonde was used to measure temperature (°C), dissolved oxygen (DO mg L^{-1}), pH, fluorescent dissolved organic matter (*f*DOM, RFU), chlorophyll-*a* (CHL YSI, RFU), phycocyanin (Phyco YSI, RFU), and turbidity (NTU). A Secchi disk was used to measure water transparency, approximating the photic zone depth (m). Precipitation data was obtained from weather stations located within 3 km of each pond.

Laboratory Measurements

Water samples were measured for colored dissolved organic matter (CDOM, μg L^{-1}), *in-vivo* or whole-cell chlorophyll-*a* (CHL RFU, RFU), and phycocyanin (Phyco LAB, μg L^{-1}) using an Aquafluor fluorometer (Turner Designs, San Jose, CA). Samples were also processed and measured for extracted chlorophyll (CHL EXT, μg L^{-1}) following EPA method 445 (EPA, 1997) using an Aquafluor fluorometer. For the extraction process, ∼100 mL of pond water was vacuum filtered using 0.7 μm glass fiber filters (Whatman, Maidstone, United Kingdom) and steeped in a 90% acetone and 10% deionized water solution overnight at 4°C before being analyzed with the fluorometer. A

subsample of ∼50 mL was taken for phytoplankton identification and enumeration. This subsample was preserved with Lugol's iodine solution at a 1% final concentration. Samples were stored at 4°C and in the dark to prevent phytoplankton cell degradation until microscopic analysis could be completed.

Microscope Analysis

During examination and enumeration of the preserved phytoplankton samples each phytoplankton was identified to the lowest taxon possible using John et al. (2011) and Bellinger and Sigee (2015). To assess the phytoplankton community at the group level species data was recorded as cell abundance (cells L^{-1}) and then classified into one of four major phytoplankton functional groups: diatoms, dinoflagellates, chlorophytes (including motile and non-motile species), and cyanobacteria as done for corresponding long-term, regional datasets (Lamlou, 1977; Marshall et al., 2006; Marshall, 2013, 2014). Because of the infrequent occurrence of dinoflagellate species in both ponds over the 2 years these data were not included in the final analysis but are available in **Supplementary Figure 1**. The cell abundance data for potentially toxic cyanobacteria species were compared with cell abundances presented in national and regional action guidelines (VDH, 2015; EPA, 2019).

All phytoplankton samples were examined using a Nikon Ts2R inverted microscope (Nikon Instruments Inc., Melville, NY) and a modified Utermöhl method as described in Marshall and Alden (1990). A 2- or 3-mL Lugol's iodine preserved sample was pipetted into a chambered covered glass slide (Thermo Scientific, Rochester, NY), and allowed to settle for 30 min to 1 h. After settling, enumeration started in the upper left-hand corner of the chambered slide. After the first frame was counted, the next frame would be moved down and to the right to avoid frame overlap and possible double counting of algal cells. This movement of the field of view created a diagonal pattern across the cover glass slide. The frames were counted in this pattern until either a 200-cell minimum or 20 frames were examined.

Statistics and Graphics

To assess spatio-temporal stability of phytoplankton functional groups, mean relative difference method (MRD) was applied. The mean relative difference indicates how an individual location compares to the pond average over multiple sampling dates and reveals areas that are consistently higher or lower than the pond's average for a measured parameter. This method follows those reported in other spatial pattern studies (Pachepsky et al., 2017; Stocker et al., 2018). The relative difference RD_{ij} between the observation of variable x at location i at time j (x_{ij}), and the spatial average of x at the same time ($<x>_j$), is defined as:

$$RD_{ij} = \frac{x_{ij} - \langle x \rangle_j}{\langle x \rangle_j}$$

The MRD for location i then becomes

$$MRD_i = \frac{1}{N_t} \sum_{j=1}^{j=N_t} RD_{ij}$$

Where N_t is the number of sampling days, and $i = 1, 2, \ldots, N_i$, where N_i is the total number of locations.

The coefficient of variation (CV) was computed for each phytoplankton functional group for each date and pond. The calculation for CV is defined as:

$$CV = \frac{\sigma_{ij}}{\mu_{ij}}$$

Where σ_{ij} is the population standard deviation of phytoplankton functional group i on sampling date j and μ_{ij} is the population mean of phytoplankton functional group i on sampling date j.

Mean relative differences and Spearman rank correlations were computed in RStudio. Correlations were considered moderate if $r \geq 0.400$ (p-values, $P1 = 0.059$ $P2 = 0.065$) and considered strong if $r \geq 0.600$ (p-values, $P1 < 0.001$ $P2 < 0.001$). Sigmaplot v. 13 (SYSTAT, Chicago, IL, USA) and QGIS (OSGeo, Switzerland) were used to create visual representations of the data.

RESULTS

Data Summary

Weather Data

Daily ambient air temperature and precipitation data for both ponds and years are displayed in **Figure 2**. In 2017, there was an increase in the air temperatures at both ponds from May to July. In 2018, the initial increase in temperature was less pronounced than in 2017. Pond 1 experienced more rainfall in 2018 compared to 2017. Information on the number of days following the last rainfall event from sampling dates and total rainfall accumulations may be seen in **Supplementary Table 1**. Over the two years, sampling at Pond 1 was performed six times with a rainfall event occurring the day before sampling, three times with a rainfall event occurring one to three days before sampling, and three times when a rainfall event was four or more days before sampling. At Pond 2, sampling was done twice with a rainfall event occurring the day before sampling, six times with a rainfall event occurring one to three days before sampling, and three times with a rainfall event occurring four or more days before sampling. Major precipitation events (>6 cm) at Pond 1 occurred on Jul-28-17 and Jul-21-18 with daily rainfall accumulations of 10.04 and 14.10 cm, respectively. Sampling near both major precipitation dates was avoided, and sampling was not conducted for three days following a major event. At Pond 2, major precipitation events occurred on Jul-28-17, Jul-29-17, Aug-7-17, and Jul-21-18 with daily rainfall accumulations of 8.28, 8.40, 16.76, and 8.03 cm, respectively. Sampling was avoided within three days of these rainfall events with the exception of the Aug-7-17 event. Sampling occurred on Aug-8-17 which was one day following a major rainfall event.

Water Quality Parameters

Time series data of water quality parameters measured for 2017 and 2018 are presented in **Supplementary Tables 2, 3**. Mean values of all measurements related to phytoplankton pigments (Phyco YSI, CHL YSI, EXT CHL, LAB CHL, and Phyco LAB) were generally higher at Pond 2 than at Pond 1 for both 2017

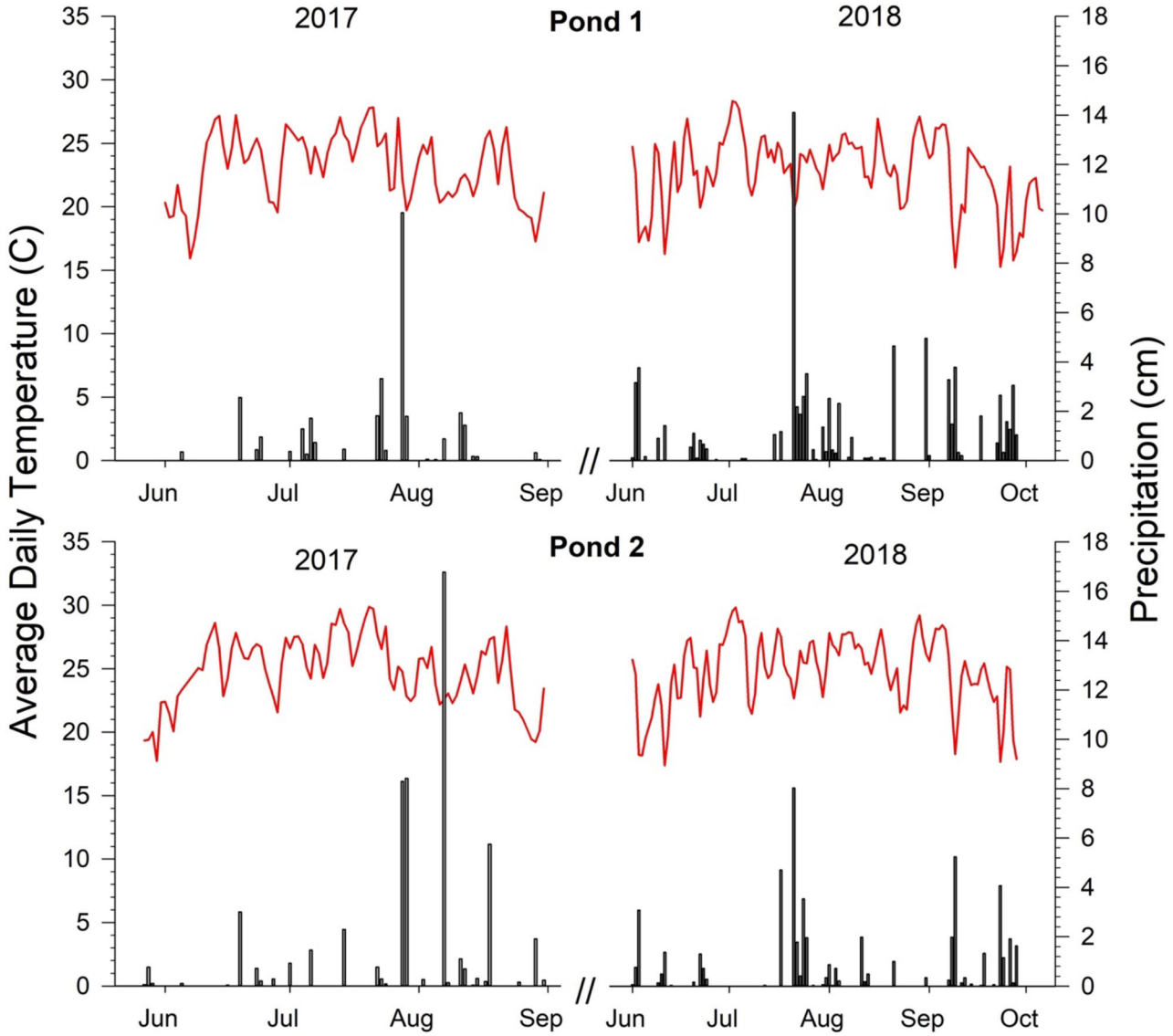


FIGURE 2 | Precipitation and ambient air temperature data for both ponds. Precipitation data is represented in bars and temperature data are shown with red lines.

and 2018. Specific conductance and pH measurements were lower for both ponds in 2018 compared to 2017. The positive relationship between higher pH and higher DO concentrations was more pronounced for Pond 2, compared to Pond 1. Algicide was applied to Pond 1 after the first sampling date on Jul-1-18. Consequently, in Pond 1 all measurements related to phytoplankton pigments (Phyco YSI, CHL YSI, CHL EXT, LAB CHL, and Phyco LAB) displayed large decreases on sampling date Jul-5-18. Phycocyanin measurements remained low for the remainder of the sampling season (Phyco YSI, Phyco LAB), while chlorophyll measurements recovered after two sampling dates (CHL YSI, CHL EXT, LAB CHL). A decrease in Phyco YSI, CHL YSI, and CHL EXT measurements was seen in Pond 2 in 2017 following a 16.8 cm rainfall event. Phycocyanin measurements (Phyco YSI and Phyco LAB) for both ponds in 2018 indicated a cyanobacteria bloom was present during the first sampling dates. Phyco LAB measurements on the first sampling dates were 114 and 110 μg L^{-1} for Pond 1 and Pond 2, respectively. Furthermore, these blooms were also visually identified by the appearance of green surface scums and confirmed *via* microscopy analysis of phytoplankton samples. Phycocyanin measurements remained approximately the same in Pond 2 during the entire sampling season.

Phytoplankton Functional Groups

The time series data of log concentrations of green algae, diatoms, and cyanobacteria for both ponds and years are presented in the box plot graphs of **Figure 3**. Descriptive statistics for all phytoplankton groups, both ponds, and both sampling years are reported in **Supplementary Table 4**. Green algae displayed

the lowest variability and cyanobacteria displayed the highest variability among the phytoplankton functional groups for both ponds. The coefficients of variation (CV) values are presented in **Supplementary Table 5**. In 2017 the CVs for Pond 1 ranged from 0.024 to 0.066 for green algae, 0.064 to 0.124 for diatoms, and 0.074 to 0.214 for cyanobacteria. The CVs followed a similar pattern in Pond 1 during 2018 with green algae CVs ranging from 0.023 to 0.051, diatoms from 0.044 to 0.132, and cyanobacteria from 0.040 to 0.262. The green algae CVs were generally lower in Pond 2 than the values for diatoms and cyanobacteria. The CVs for diatoms and cyanobacteria did not follow the same pattern as found for Pond 1, the overall ranges of the diatom CVs were less than the cyanobacteria CVs for each respective sampling season. Green algae and diatoms had a similar intra-seasonal (May–August) trend during 2017 at both ponds wherein population growth occurred from May to June followed by a period of stabilization for the remainder of the sampling season. Diatoms and green algae in Pond 1 exhibited similar trends in 2018 (June–October) displaying a period of stabilization from June to July followed by a drop in concentrations for the remainder of the sampling season. Cyanobacteria trends were drastically different from 2017 to 2018 for both ponds. A cyanobacteria bloom was observed within both ponds during June 2018. During this study, copper sulfate was applied to Pond 1 on Jul-1-18 and impacted the total phytoplankton concentrations, particularly decreasing the abundance of cyanobacteria species.

Temporally Stable Patterns of Phytoplankton Functional Groups

Temporal stability was assessed by considering the standard errors of the mean relative differences for each location. The mean relative differences along with standard error bars are displayed in **Supplementary Figure 2**. Small standard errors indicate that a location has minimal phytoplankton variation between each sampling date and large standard errors indicate substantial phytoplankton variation between sampling dates. Green algae, diatoms, and cyanobacteria displayed temporally stable spatial patterns in both ponds and over the entire 2-year study period.

Pond 1

The MRD values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the 2-years of observations at Pond 1 are shown in **Supplementary Figure 1**. Visual representations of the locations with consistently higher and consistently lower concentrations of each phytoplankton functional groups are displayed in **Figures 4–6**. The same patterns were observed for all three phytoplankton functional groups in Pond 1. The MRDs of each group tended to be lower for the interior sampling locations, and higher for the nearshore sampling locations. For green algae (**Figure 4**-P1), zones with consistently lower concentrations were all interior locations, except for location 6 where the irrigation pump is located. Zones of high concentrations of green algae were seen at the southern shoreline of the pond (locations 1, 8, 10), as well as locations 5 and 23. The southern shoreline of the pond is very shallow and located adjacent and downhill from crop fields. Location 5 is the site of an inflow pipe which pumps water from a nearby stream-fed pond. Location 23 is a very shallow area with aquatic vegetation and has an inflow from an ephemeral stream. Locations of consistently higher and lower concentrations of diatoms (**Figure 5**-P1) were very similar to those of green algae. Low concentrations of diatoms were exclusively observed at interior sampling locations. High concentrations of diatoms were observed for locations 1, 5, 6, 23 (previously described), and 2. Locations with consistently higher and lower cyanobacteria concentrations are displayed in **Figure 6**-P1. Low cyanobacteria concentrations were found at all interior sampling locations except for location 20. High cyanobacteria concentrations were found at all nearshore sampling locations except for location 17. Consistently high concentrations of cyanobacteria were also seen close to the ephemeral stream inflow at locations 21 and 22; and at location 1 near the crop fields.

Pond 2

The MRD values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the 2-year period for Pond 2 are shown in **Supplementary Figure 1**. Visual representations of MRDs for green algae, diatoms, and cyanobacteria are displayed in **Figures 4–6**. Low MRDs for green algae (**Figure 4**-P2) were all nearshore sampling locations of the pond, although there was a somewhat dispersed distribution with similar values not observed within one specific area of the pond. Locations with consistently higher concentrations of green algae were mostly interior sampling locations except for sampling location 21, which is a shallow location with aquatic vegetation present. High concentrations of diatoms were found at nearshore sampling locations and within a small zone on the southeastern shoreline of the pond where there is a water level dependent outflow drain (**Figure 5**-P2). Cyanobacteria MRDs displayed a zonal pattern (**Figure 6**-P2) with higher cyanobacteria concentrations located in the northern portion of the pond apart from the observations at location 21 (as previously described). Consistently low concentrations of cyanobacteria formed a zone in the middle of the pond containing both interior and nearshore sampling locations.

Water Quality Patterns

The mean relative difference values of measured water quality parameters (Temp, DO, SPC, pH, NTU, Phyco YSI, CHL YSI, *f*DOM, and CHL EXT) for both ponds are shown in **Supplementary Figures 3A–I**. Within Pond 1, low MRDs were observed for temperature, DO, and pH for nearshore locations. For turbidity, Phyco YSI, CHL YSI, *f*DOM, and extracted chlorophyll, high MRDs were associated with nearshore locations and low MRDs were associated with interior locations within Pond 1. Within Pond 2, similar trends were observed with high MRD values for temperature, DO, and pH being observed at interior locations and low MRD values found at the nearshore locations. An inverse distribution was seen for turbidity, Phyco YSI, CHL YSI, *f*DOM, and extracted chlorophyll within Pond 2. Low MRD values were typically observed for the interior locations and high MRD values were observed for the nearshore locations. Therefore, both ponds exhibited differences in the

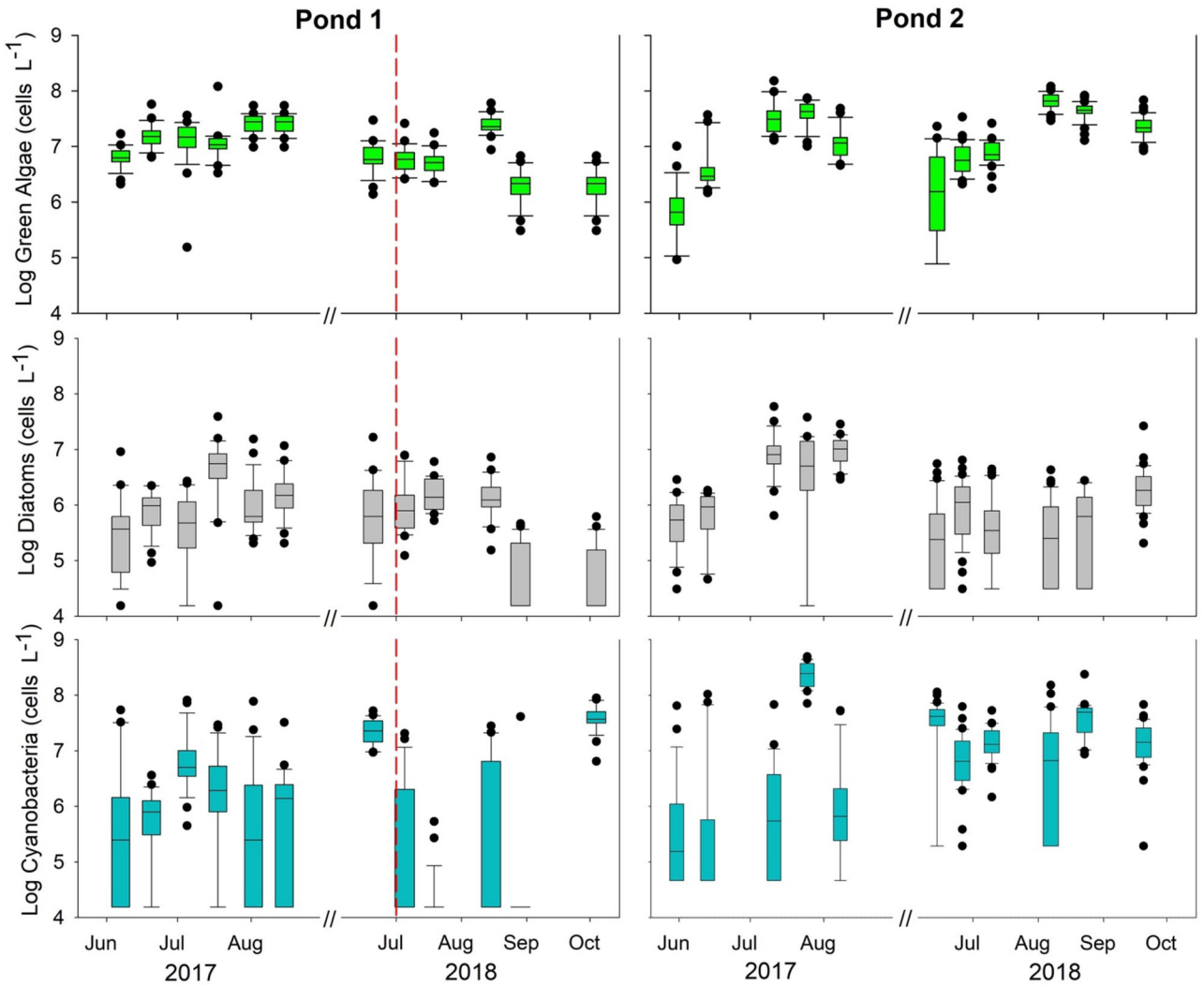


FIGURE 3 | Time series data of log concentrations of the major phytoplankton functional groups (green algae, diatoms, and cyanobacteria) for both years. The mid-point line of the box plots represents the median for each date. Outliers (10 and 90th percentiles) are represented by black circles. The red dashed line represents an algicide application that took place on Jul-1–18 at Pond 1.

water quality parameters between the interior and the nearshore sampling locations.

Phytoplankton and Water Quality MRD Correlations

The Spearman rank correlations between the mean relative differences of the water quality parameters and the mean relative differences of phytoplankton groups are displayed in **Table 1**. Moderate correlations were defined as $r \geq 0.400$ (p-values, $P1 = 0.059$ $P2 = 0.065$) and are highlighted in yellow. Strong correlations were defined as $r \geq 0.600$ (p-values, $P1 < 0.001$ $P2 < 0.001$) and are highlighted in blue. Moderate and strong correlations were observed within Pond 1 for the green algae MRDs and most of the water quality MRDs (DO, SPC, pH, NTU, Phyco YSI, CHL YSI, and CHL EXT). Lower correlations were observed for diatom MRDs and cyanobacteria MRDs within Pond 1. There were no moderate or strong correlations observed for diatom MRDs in Pond 1. The cyanobacteria MRDs were moderately correlated with the MRDs of the SPC, pH, and NTU parameters. Pond 2 differed from Pond 1 regarding MRD correlations. Within Pond 2, green algae MRDs were characterized with fewer moderate correlations than diatom MRDs and cyanobacteria MRDs. Green algae MRDs within Pond 2 had a strong correlation with the MRDs of extracted chlorophyll. Diatom MRDs correlated strongly with most water quality MRDs (Temp, DO, SPC, pH, and NTU) and moderately with CHL EXT. Cyanobacteria MRDs correlated moderately with Phyco YSI and CHL YSI and strongly with most water quality MRDs (Temp, SPC, pH, and CHL EXT).

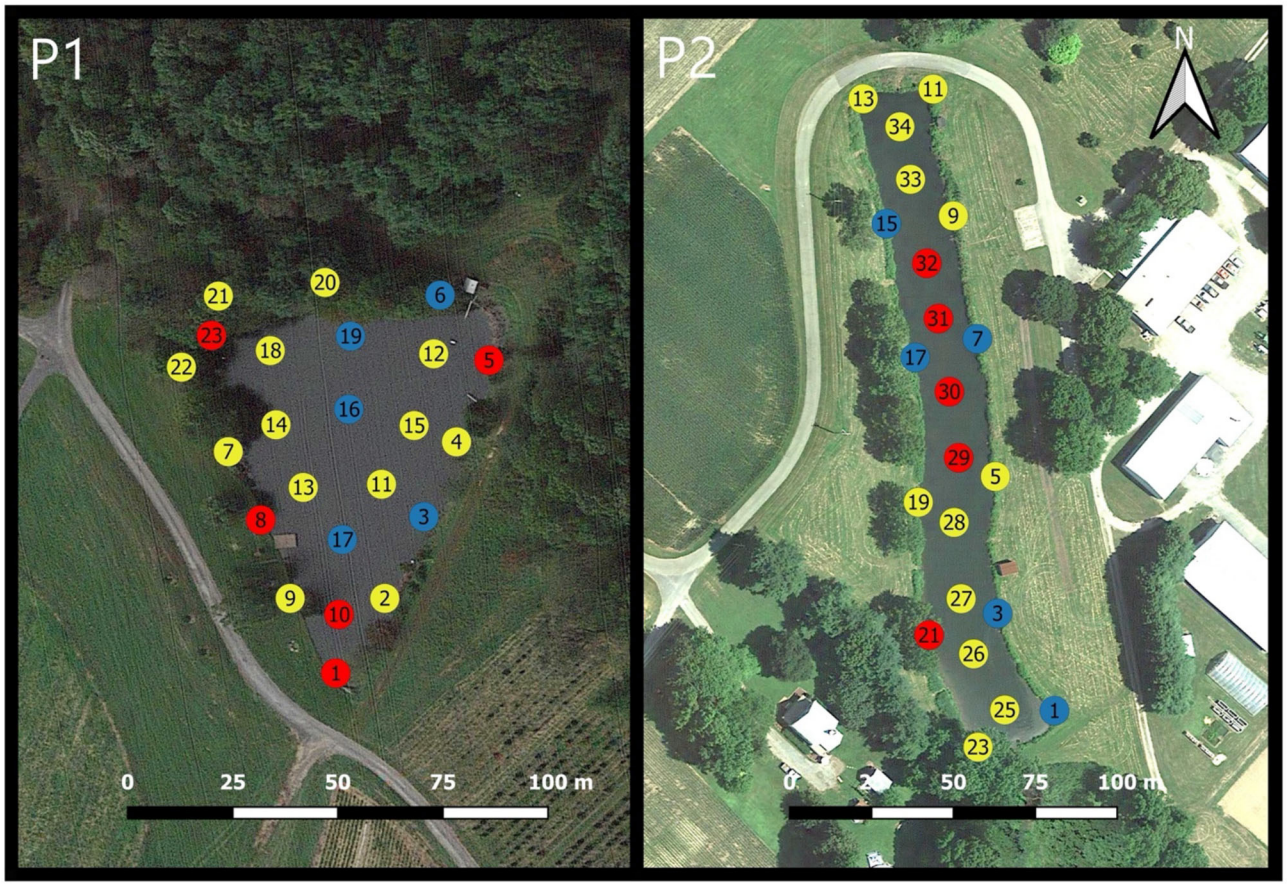


FIGURE 4 | Mean relative differences (MRD) of green algae over the 2-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red, and locations between the 25 and 75th percentile are displayed in yellow.

DISCUSSION

The abundance and species distribution of phytoplankton taxa has been used as a bioindicator of water quality across freshwater and marine systems for decades (Patrick, 1973; Smith, 2003; Reynolds et al., 2012). Decadal phytoplankton datasets have proven to be useful when examining the seasonal periodicity and long-term trends in coastal water quality (Marshall et al., 2009; Nishikawa et al., 2010; Hernandez Cordero et al., 2020), however, similar longitudinal datasets are lacking for agricultural irrigation waters despite the fact it has been reported that land-use and nutrient loading can impact phytoplankton biodiversity in agricultural waters (Zhang et al., 2020). Smith et al. (2020) demonstrated that within agricultural irrigation ponds there was a relationship between easily measured environmental co-variates, such as CDOM and NTU, and cyanobacteria (phycocyanin) concentrations. However, the temporal and spatial stability of the cyanobacteria, or other phytoplankton functional groups, in these ponds was not examined. Here, an assessment of the phytoplankton community present during the May to October growing season, when agricultural irrigation water is used most frequently and the risk due to cyanotoxins is greatest, is presented.

The agricultural irrigation ponds examined in this study, located on working farms in Maryland, did not exhibit drastically different phytoplankton populations during the growing seasons of 2017 and 2018. Diatom concentrations did not differ within the two ponds and were comparable with concentrations found in other temperate freshwater lakes and reservoirs (Rollwagen-Bollens et al., 2013; Gorokhova and Zinchenko, 2019; Jia et al., 2019) including lakes studied by Marshall (2013, 2014) in Virginia, located south of this study area. Mean concentrations of green algae were similar within the two ponds, but Pond 1 had a smaller overall range of concentrations than Pond 2. Concentrations of green algae were comparable to values reported for other freshwater systems (Dembowska et al., 2018; Gorokhova and Zinchenko, 2019; Khaliullina and Fazlieva, 2019). Pond 2 had slightly higher concentrations of green algae and cyanobacteria. These higher values may potentially be explained by the absence of an algicide application for Pond 2. Concentrations of cyanobacteria within Pond 1 were similar to those previously reported within temperate lakes (Dembowska et al., 2018; Jia et al., 2019), including those studied

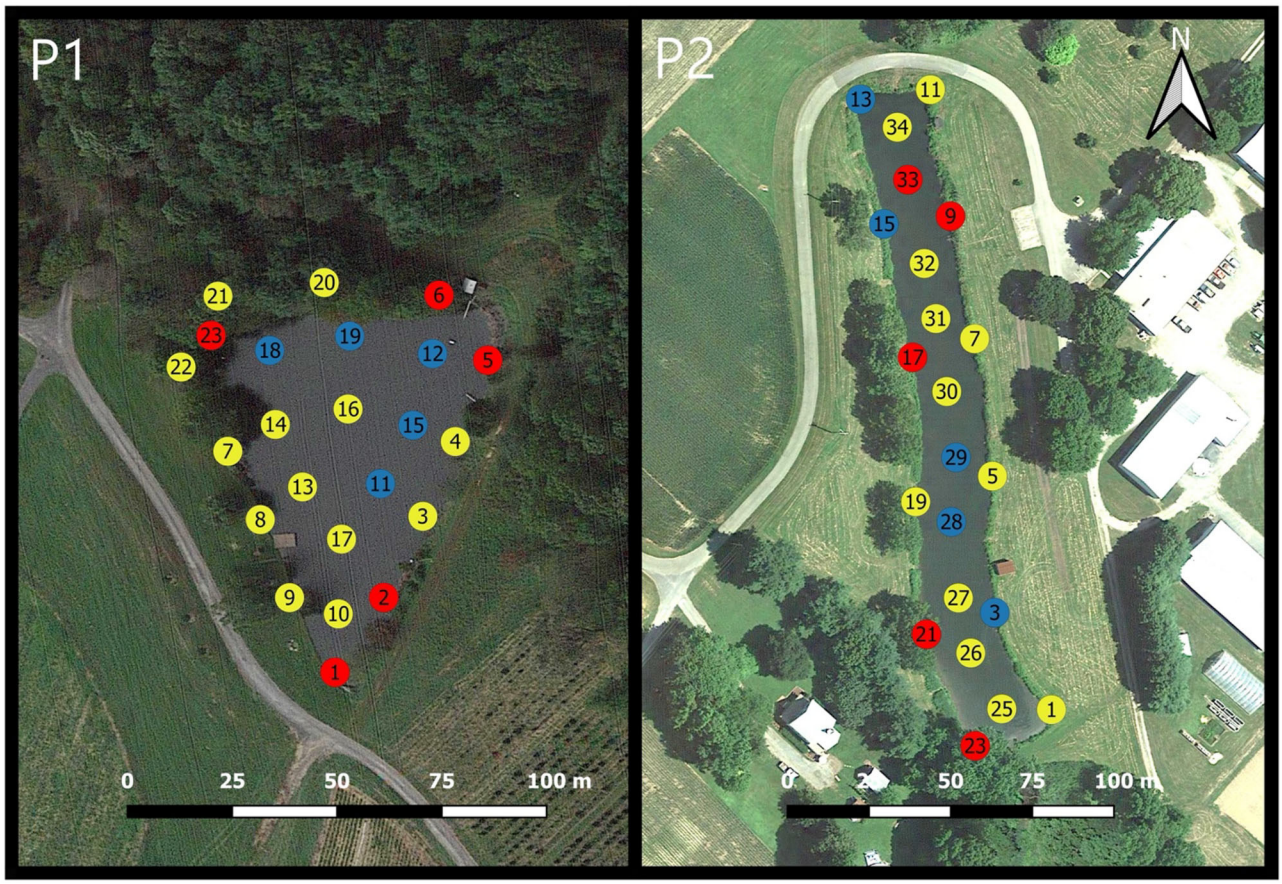


FIGURE 5 | Mean relative differences (MRD) of diatoms over the 2-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red, and locations between the 25 and 75th percentile are displayed in yellow.

locally by Marshall (2013, 2014). However, due to recurrent cyanobacteria blooms composed mainly of *Aphanizomenon* spp. and *Microcystis wesenbergii*, in Pond 2, cell concentrations were comparable with concentrations reported within small temperate lakes in which cyanobacteria blooms frequently occur (Lee et al., 2015; Woodhouse et al., 2016), including other Maryland lakes (Tango and Butler, 2008; J. Wolny, unpublished data), but were greater than those recorded by Marshall (2013, 2014).

Both ponds displayed spatial and temporal variability of major phytoplankton functional groups during the two growing seasons. Spatio-temporal variations of phytoplankton communities have been documented within freshwater lakes (Wu et al., 2014; Naselli-Flores and Padisák, 2016; Xiao et al., 2018), wetland ponds (Soininen et al., 2007), reservoirs (Alexander and Imberger, 2009), rivers (Marshall et al., 2009), and estuaries (Marshall et al., 2006). While the phytoplankton community temporal trends noted in this study were similar to those reported by Marshall (2013, 2014) for Virginia lakes, comparisons between these earlier studies and spatial variation are not possible due to the limited spatial variance in the Virginia lakes dataset. The heterogeneity or homogeneity of phytoplankton communities should not be an assumed trait within a water body. As explained by Lewis (1978), not all species or groups of phytoplankton continuously exhibit heterogenous distributions, but rather homogenous and heterogenous distributions may synchronously exist within a water body. Additionally, exogenous forces, such as wind, water flow, and lake morphology, have all been documented to attribute to the spatial variation of green algae, diatoms, and cyanobacteria (Li et al., 2013).

Two forms of spatial trends were observed within the two ponds during this study. The predominant spatio-temporal trends that were present within both ponds appeared to be a contrast between interior and nearshore sampling locations. Within Pond 1, this trend was displayed for all groups. Pond 1 had consistently higher concentrations of green algae, cyanobacteria, and diatoms at nearshore sampling locations; and consistently lower concentrations of green algae, cyanobacteria, and diatoms at interior locations. This pattern of higher concentrations at nearshore sampling locations vs. interior sampling locations was reported for both ponds in a preceding study using average quartile ranks of phycocyanin concentrations (Smith et al., 2020). Higher concentrations of phytoplankton being closer to the shoreline of shallow water bodies has been

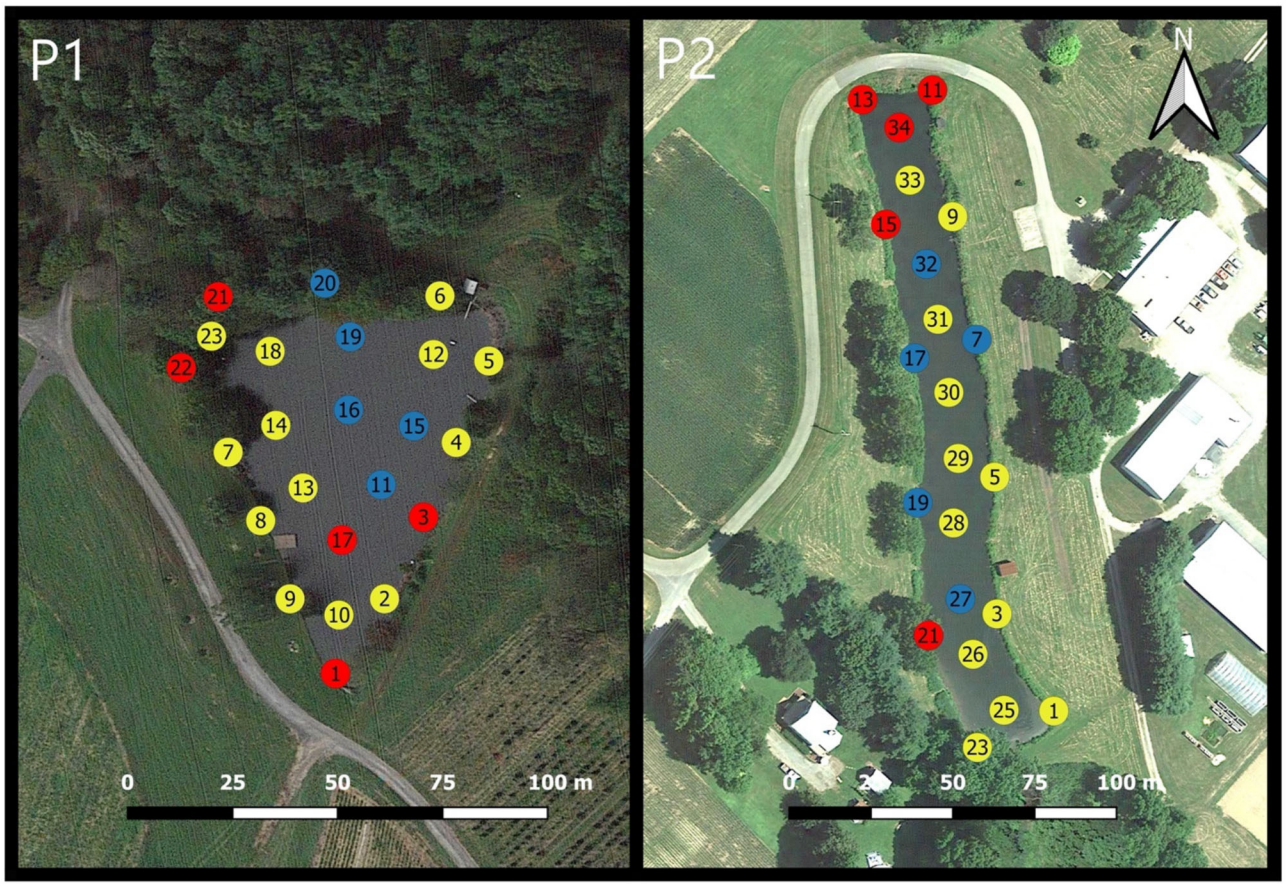


FIGURE 6 | Mean relative differences (MRD) of cyanobacteria over the 2-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red and locations between the 25 and 75th percentile are displayed in yellow.

attributed to several different concepts. Bondarenko et al. (1996) stated that spatial distribution of phytoplankton was related to water depth, with shallow waters being richest in phytoplankton. Both ponds in this study were only 2.7 m deep, thus indicating that even in shallow environments depth-dependent gradients can be set up within the phytoplankton community. In other studies, greater abundances of phytoplankton were found in stands of *Phragmites australis* and other aquatic plants, due to the creation of favorable water quality conditions, including increased phosphorus concentrations (Celewicz-Goldyn and Kuczynska-Kippen, 2008, 2017) and in zones with elevated nutrient concentrations and water temperatures (Chen et al., 2003). Aquatic vegetation was noted at both ponds and future work will look to correlate the spatial patterns of the phytoplankton community with the characteristics of the resident aquatic vegetation.

Similarly, within Pond 2 consistently higher concentrations of diatoms were found at nearshore locations. The opposite trend was observed for green algae within Pond 2 wherein consistently higher concentrations of green algae were observed for interior sampling locations, and consistently lower concentrations of green algae were observed for nearshore sampling. While this is a difference from co-located Pond 1, this trend has been previously documented by Celewicz-Goldyn and Kuczynska-Kippen (2008) who indicated that the greatest abundance of small chlorophytes was found in open waters where the potential threat of predation from zooplankton was less.

The major spatio-temporal patterns observed for Pond 2 were the formation of zones in which cyanobacteria were the dominant taxa. Consistently higher concentrations of cyanobacteria were found at the northern portion of the pond (sampling locations: 11, 13, 15, and 34) near a culvert, which following precipitation events provides inflow of potentially nutrient-rich waters to the pond. This high cyanobacteria biomass zone was also established in the preceding study on quartile ranks of phycocyanin concentrations (Smith et al., 2020). Other studies have documented spatial trends of cyanobacteria among other phytoplankton species due to either nutrient-rich runoff or river inflow (Powell et al., 1975; Marshall et al., 2006; Woodhouse et al., 2016). Other potential explanations for the formation of these cyanobacteria-rich zones could be wind or wind-driven water flow as noted by Cloern et al. (1992) and Fragoso et al. (2008) or microhabitats set up through thermal stratification as noted by Vasas et al. (2013).

TABLE 1 | Spearman rank correlations between the mean relative differences of water quality parameters and MRD values of phytoplankton functional groups.

2017 + 2018	Pond 1			Pond 2		
	Green algae	Diatoms	Cyanobacteria	Green algae	Diatoms	Cyanobacteria
Spearman rank correlations between water quality MRDs and phytoplankton MRDs						
Temp	0.186	<0.001	0.163	<0.001	0.771	0.639
DO	0.498	0.134	0.161	0.024	0.616	0.317
SPC	0.906	0.142	0.703	0.221	0.713	0.842
pH	0.895	0.173	0.931	0.004	0.849	0.684
NTU	0.213	0.104	0.625	0.233	0.903	0.196
Phyco YSI	0.174	0.003	0.103	0.005	0.387	0.589
CHL YSI	0.220	0.001	0.231	0.007	0.297	0.467
*f*DOM	0.017	<0.001	0.029	0.125	0.043	0.390
CHL EXT	0.668	0.012	0.033	0.692	0.471	0.875

Moderate correlations were defined as R ≥ 0.400 and are highlighted in yellow (p-values, P1 = 0.059 P2 = 0.065). Strong correlations were defined as R ≥ 0.600 and are highlighted in blue (p-values, P1 < 0.001 P2 < 0.001).

There was a zone of consistently low cyanobacteria concentrations within Pond 2 that was located near the middle of the pond (sampling locations: 17, 27, 32, 7, 19, 5). The pump house and water intake pipe for the farm irrigation system is in this area. The location of the irrigation intake pipe has important implications for food safety. It has been well-established that irrigation waters with toxigenic cyanobacteria can contaminate crops (Miller and Russell, 2017), remain in soils for extended periods of time (Machado et al., 2017), and may even be taken up by the root system of the produce (Lee et al., 2017). Thus, placing an irrigation intake system in a location with consistently higher concentrations of cyanobacteria may increase produce contamination risks. It appears that the pump and intake infrastructure in Pond 2 is located in a low-risk zone, as cell concentrations of potentially toxic cyanobacteria species never exceeded EPA or regional guidelines (VDH, 2015; EPA, 2019). However, future research and monitoring efforts should focus on determining the prevalence of cyanotoxins in these irrigation waters.

Of note, a copper sulfate algicide was applied to Pond 1 midway through the study, on Jul-1-18, to mitigate a bloom of *Microcystis*, a potentially toxigenic cyanobacteria species. While the concentration of copper sulfate used is unknown, all measured water quality parameters decreased significantly following this application. These reductions were comparable to values reported by Schrader et al. (2000) and Song et al. (2011) within other inland waters that were assessed during and after treatments with copper sulfate. Average concentrations of DO, pH, CDOM, and *f*DOM returned to pre-application levels about 1 month after application. For the algal pigments (CHL RFU, CHL YSI, CHL EXT, and phycocyanin), all concentrations decreased after the copper sulfate application and slowly recovered to either pre-application levels or higher by the end of August. The return of chlorophyll-a readings to previous values was also reported by Dia (2016) and Effler et al. (1980) following low-level algicide treatments (8–14 $\mu g\ L^{-1}$) in freshwater lakes. Elder and Horne (1978) reported the recovery of pre-treatment algal populations in as little as five days after treatment and attributed this recovery to copper sulfate possibly being beneficial for biological activity if applied in very low concentrations (5–10 $\mu g\ L^{-1}$). The effect of algicide on cyanobacteria concentrations was more pronounced than the effect on green algae and diatoms concentrations. Similar responses were reported by Padovesi-Fonseca and Philomeno (2004) and XiaoLi et al. (2009) wherein cyanobacteria concentrations, including *Microcystis aeruginosa, Cylindrospermopsis raciborskii,* and *Anabaena flos-aquae,* decreased and green algae and diatoms became the dominant taxa after an algicide application, with the subsequent population changes attributed to cyanobacteria species sensitivity to copper. The rate of recovery of the phytoplankton community to the application of copper sulfate, or other algicides, should be monitored if the algicide application is meant to act as a safeguard to crops from cyanotoxin exposure *via* irrigation waters. It should be mentioned that the treatment of water bodies with algicides can immediately release large quantities of cyanotoxins, if toxin-producing algal species are highly concentrated (Zhou et al., 2013; Greenfield et al., 2014). An assessment of phytoplankton community composition should be performed prior to an algicide application if the water is to be used for crop irrigation or as drinking water for livestock to safeguard against the introduction of concentrated biotoxins.

Although not similar, both ponds expressed moderate and strong correlations between the spatial patterns of phytoplankton functional groups and water quality parameters. Pond 1 had strong water quality correlations with green algae, while Pond 2 had strong correlations with diatoms and cyanobacteria. The correlations between water quality and phytoplankton spatial trends provides helpful insights for irrigation pond monitoring. The examination of phytoplankton community structure using microscopy is an intensive analysis which requires extensive laboratory infrastructure and highly trained personnel (Lawton et al., 1999). However, if strong correlations exist among water quality parameters and optical properties associated with distinct phytoplankton functional groups, the option of using less specialized monitoring methods, such as *in-situ* sensors or drone-based imagery could be employed for routine resource management. These technologies would be efficient and cost-effective methods capable of being used by a

broader group of personnel to safeguard against irrigating crops with degraded water drawn from agricultural irrigation waters. While identifying water quality covariates and the use of optical techniques to assess the phytoplankton functional groups present in water will not identify toxic vs. non-toxic phytoplankton species it can provide the information necessary to make better informed decisions about when and where to conduct toxin risk assessments.

CONCLUSIONS

Using a mean relative difference analysis to assess spatio-temporal stability, it was determined that phytoplankton functional groups exhibited stable spatio-temporal trends in the two agricultural irrigation ponds evaluated in this study. Temporally stable spatial patterns of the three phytoplankton functional groups studied here were found within both ponds over the two sampling years. Both ponds had locations where phytoplankton group concentrations were consistently higher or lower than the pond's average concentrations. Typically, these patterns could be classified into two categories: nearshore or interior sampling locations or zones. These distributions indicate the importance of sampling locations for water quality monitoring purposes. If sampling is performed in areas of consistently higher or lower concentrations of phytoplankton, that sample may not be an accurate representation of the phytoplankton community within the entire waterbody. Because of the correlation between water quality parameters and certain phytoplankton functional groups it may be possible to employ broad-based technologies to routinely monitor irrigation waters for potentially harmful cyanobacteria instead of relying on labor intensive microscopy methods. However, it is important to note that there are other types of agricultural ponds, such as aquaculture ponds and retention ponds. While this study can provide a framework for assessing agricultural ponds no extrapolation should be made to these other water sources from the finding presented here.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JS: conducted sample collection, microscopy assays, and data analysis and wrote the original draft. JW: aided with microscopy and provided substantial assistance with manuscript editing. MS: assisted with sample collection and manuscript editing and created GIS-based figures. RH: provided guidance during the writing process. YP: provided guidance and insight during sample collection design, data analysis, and writing process. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.724025/full#supplementary-material

Supplementary Figure 1 | The mean relative difference values of the logarithms of dinoflagellate concentrations computed over the 2-year period for both Pond 1 and Pond 2.

Supplementary Figure 2 | The mean relative difference values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the 2-year period for Pond 1 and Pond 2.

Supplementary Figure 3 | (A) The mean relative difference values of temperature over the 2-year period for Pond 1 and Pond 2. **(B)** The mean relative difference values of dissolved oxygen over the 2-year period for Pond 1 and Pond 2. **(C)** The mean relative difference values of specific conductance over the 2-year period for Pond 1 and Pond 2. **(D)** The mean relative difference values of pH over the 2-year period for Pond 1 and Pond 2. **(E)** The mean relative difference values of turbidity over the 2-year period for Pond 1 and Pond 2. **(F)** The mean relative difference values of phycocyanin from the YSI sonde over the 2-year period for Pond 1 and Pond 2. **(G)** The mean relative difference values of chlorophyll pigment from the YSI sonde over the 2-year period for Pond 1 and Pond 2. **(H)** The mean relative difference values of fluorescent dissolved organic matter over the 2-year period for Pond 1 and Pond 2. **(I)** The mean relative difference values of extracted chlorophyll over the 2-year period for Pond 1 and Pond 2.

Supplementary Table 1 | Sampling dates from 2017 and 2018 with corresponding rainfall event information for Pond 1 and Pond 2.

Supplementary Table 2 | Time series data of 2017 water quality parameters for Pond 1 and Pond 2.

Supplementary Table 3 | Time series data of 2018 water quality parameters for Pond 1 and Pond 2.

Supplementary Table 4 | Descriptive statistics of phytoplankton functional groups.

Supplementary Table 5 | Coefficient of variation values for Pond 1 and Pond 2.

REFERENCES

Adloff, C. T., Bem, C. C., Reichert, G., and Azevedo, J. C. R. de. (2018). Analysis of the phytoplankton community emphasizing cyanobacteria in four cascade reservoirs system of the Iguazu River, Paraná, Brazil. *Brazil. J. Water Resourc.* 23:318170050. doi: 10.1590/2318-0331.0318170050

Alexander, R., and Imberger, J. (2009). Spatial distribution of motile phytoplankton in a stratified reservoir: the physical controls on patch formation. *J. Plankton Res.* 31, 101–118. doi: 10.1093/plankt/fbn101

Bellinger, E. D., and Sigee, D. C. (2015). *Freshwater Algae, 1st Edn.* Hoboken, NJ: John Wiley & Sons, Ltd. doi: 10.1002/9781118917152

Bittencourt-Oliveira, M. do C., Cordeiro-Araújo, M. K., Chia, M. A., Arruda-Neto, J. D. de, T., Oliveira, Ê. T. de, Santos, F., et al. (2016). Lettuce irrigated with contaminated water: photosynthetic effects, antioxidative response and bioaccumulation of microcystin congeners. *Ecotoxicol. Environ. Saf.* 128, 83–90. doi: 10.1016/j.ecoenv.2016.02.014

Bondarenko, N. A., Guselnikova, N. E., Logacheva, N. F., and Pomazkina, G. V. (1996). Spatial distribution of phytoplankton in Lake Baikal, Spring 1991. *Freshw. Biol.* 35, 517–523. doi: 10.1111/j.1365-2427.1996.tb01765.x

Bouma-Gregson, K., Power, M. E., and Bormans, M. (2017). Rise and fall of toxic benthic freshwater cyanobacteria (*Anabaena* spp.) in the Eel river: buoyancy and dispersal. *Harmful Algae* 66, 79–87. doi: 10.1016/j.hal.2017.05.007

Buratti, F., Manganelli, M., Vichi, S., Stefanelli, M., Scardala, S., Testai, E., et al. (2017). Cyanotoxins: producing organisms, occurrence, toxicity, mechanism of action and human health toxicological risk evaluation. *Archiv. Toxicol.* 91, 1049–1130. doi: 10.1007/s00204-016-1913-6

Celewicz-Goldyn, S., and Kuczynska-Kippen, N. (2008). *Spatial Distribution of Phytoplankton Communities in Small Water Body*. Roczniki Akademii

Rolniczej w Poznaniu. Botanika-Steciana, 12. Available online at: http://agro.icm.edu.pl/agro/element/bwmeta1.element.agro-09ba7268-dd81-45ba-993f-5bff48b31977 (accessed September 17, 2021).

Celewicz-Goldyn, S., and Kuczynska-Kippen, N. (2017). Ecological value of macrophyte cover in creating habitat for microalgae (diatoms) and zooplankton (rotifers and crustaceans) in small field and forest water bodies. *PLoS ONE* 12:e0177317. doi: 10.1371/journal.pone.0177317

Chen, Y., Qin, B., Teubner, K., and Dokulil, M. T. (2003). Long-term dynamics of phytoplankton assemblages: microcystis-domination in Lake Taihu, a large shallow lake in China. *J. Plankton Res*. 25, 445–453. doi: 10.1093/plankt/25.4.445

Cloern, J. E., Alpine, A. E., Cole, B. E., and Heller, T. (1992). Seasonal changes in the spatial distribution of phytoplankton in small, temperate-zone lakes. *J. Plankton Res*. 14, 1017–1024. doi: 10.1093/plankt/14.7.1017

DeLorenzo, M. E. M., Taylor, L. A., Lund, S. A., Pennington, E. D., Strozier, E. D., Fulton, M. H. (2002). Toxicity and bioconcentration potential of the agricultural pesticide endosulfan in phytoplankton and zooplankton. *Archiv. Environ. Contaminat. Toxicol.* 42, 173–181. doi: 10.1007/s00244-001-0008-3

Dembowska, E. A., Mieszczankin, T., and Napiórkowski, P. (2018). Changes of the phytoplankton community as symptoms of deterioration of water quality in a shallow lake. *Environ. Monitor. Assess*. 190:95. doi: 10.1007/s10661-018-6465-1

Dia, S. A. (2016). *Efficacy of Three Algaecides on Algal Blooms in Hypereutrophic Lakes*. (Master's Thesis), American University of Beirut, Beirut, Lebanon.

Effler, S. W., Litten, S., Field, S. D., Tong-Ngork, T., Hale, F., Meyer, M., et al. (1980). Whole lake responses to low level copper sulfate treatment. *Water Res*. 14, 1489–1499. doi: 10.1016/0043-1354(80)90015-9

Elder, J. F., and Horne, A. J. (1978). Copper cycles and $CuSO_4$ algicidal capacity in two California lakes. *Environ. Manag*. 2, 17–30. doi: 10.1007/BF01866443

EPA (1997). *In vitro Determination of Chlorophyll a and Pheophytin a in Marine and Freshwater Algae by Fluorescence. EPA-600R15006* (Washington, DC), 15022.

EPA (2019). *Recommendations for Cyanobacteria and Cyanotoxin Monitoring in Recreational Waters. EPA-823-R-19-001* (Washington, DC), 16.

Feki-Sahnoun, W., Hamza, A., Béjaoui, B., Mahfoudi, M., Rebai, A., and Bel Hassen, M. (2018). Multi-table approach to assess the biogeography of phytoplankton: ecological and management implications. *Hydrobiologia* 815, 229–251. doi: 10.1007/s10750-018-3566-3

Fragoso, C. R., Marques, D. M. L. M., Collischonn, W., Tucci, C. E. M., and van Nes, E. H. (2008). Modelling spatial heterogeneity of phytoplankton in Lake Mangueira, a large shallow subtropical lake in South Brazil. *Ecol. Model*. 219, 125–137. doi: 10.1016/j.ecolmodel.2008.08.004

Gorokhova, O. G., and Zinchenko, T. D. (2019). Phytoplankton of the Usa River and its tributaries (Kuibyshev Reservoir Basin). *Biol. Bullet*. 46, 1382–1389. doi: 10.1134/S1062359019100121

Greenfield, D. I., Duquette, A., Goodson, A., Keppler, C. J., Williams, S. H., Brock, L. M., et al. (2014). The effects of three chemical algaecides on cell numbers and toxin content of the cyanobacteria *Microcystis aeruginosa* and *Anabaenopsis* sp. *Environ. Manag*. 54, 1110–1120. doi: 10.1007/s00267-014-0339-2

Hernandez Cordero, A. L., Tango, P. J., and Batiuk, R. A. (2020). Development of a multimetric water quality indicator for tracking progress towards the achievement of Chesapeake Bay water quality standards. *Environ. Monitor. Assess*. 192:94. doi: 10.1007/s10661-019-7969-z

Hu, C., Ou, Y., Zhang, D., Zhang, H., Yan, C., Zhao, Y., et al. (2012). Phytoremediation of the polluted Waigang River and general survey on variation of phytoplankton population. *Environ. Sci. Poll. Res*. 19, 4168–4175. doi: 10.1007/s11356-012-0931-z

Jia, J., Gao, Y., Song, X., and Chen, S. (2019). Characteristics of phytoplankton community and water net primary productivity response to the nutrient status of the Poyang Lake and Gan River, China. *Ecohydrology* 12:e2136. doi: 10.1002/eco.2136

Jin, Y., Yu, R., Zhang, Z., Zhang, Q., Li, M., Cao, Z., et al. (2020). Spatiotemporal variability of phytoplankton functional groups in a shallow eutrophic lake from cold, arid regions. *Environ. Monitor. Assess*. 192:371. doi: 10.1007/s10661-020-08349-4

John, D. M., Whitton, B. A., and Brook, A. J. (2011). *The Freshwater Algal Flora of the British Isles: An Identification Guide to Freshwater and Terrestrial Algae*. London: Cambridge University Press, 724.

Jongman, M., and Korsten, L. (2018). Irrigation water quality and microbial safety of leafy greens in different vegetable production systems: a review. *Food Rev. Int*. 34, 308–328. doi: 10.1080/87559129.2017.1289385

Khaliullina, L. Y., and Fazlieva, A. R. (2019). The composition and seasonal dynamics of planktonic green algae in the mouth of the Kazanka River (Republic of Tatarstan). *J. Environ. Treat. Techniq*. 5, 993–998.

Lamlou, V. W. (1977). *Distribution of Phytoplankton in Maryland Lakes*. Las Vegas, NV: Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Support Laboratory, Monitoring Operations Division.

Lawton, L., Marsalek, B., Padisak, J., and Chorus, I. (1999). "Determination of cyanobacteria in the laboratory," in *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management*, eds I. Chorus and J. Bartram (New York, NY: E & FN Spon).

Lee, S., Jiang, X., Manubolu, M., Riedl, K., Ludsin, S. A., Martin, J. F., et al. (2017). Fresh produce and their soils accumulate cyanotoxins from irrigation water: implications for public health and food security. *Food Res. Int*. 102, 234–245. doi: 10.1016/j.foodres.2017.09.079

Lee, T. A., Rollwagen-Bollens, G., and Bollens, S. M. (2015). The influence of water quality variables on cyanobacterial blooms and phytoplankton community composition in a shallow temperate lake. *Environ. Monitor. Assess*. 187:315. doi: 10.1007/s10661-015-4550-2

Lewis, W. M. (1978). Spatial distribution of the phytoplankton in a tropical lake (Lake Lanao, Philippines). *Int. Rev. Ges. Hydrobiol. Hydrogr*. 63, 619–635. doi: 10.1002/iroh.19780630508

Li, F., Zhang, H., Zhu, Y., Xiao, Y., and Chen, L. (2013). Effect of flow velocity on phytoplankton biomass and composition in a freshwater lake. *Sci. Tot. Environ*. 447, 64–71. doi: 10.1016/j.scitotenv.2012.12.066

Machado, J., Campos, A., Vasconcelos, V., and Freitas, M. (2017). Effects of microcystin-LR and cylindrospermopsin on plant-soil systems: a review of their relevance for agricultural plant quality and public health. *Environ. Res*. 153, 191–204. doi: 10.1016/j.envres.2016.09.015

Marshall, H., and Alden, R. (1990). A comparison of phytoplankton assemblages and environmental relationships in three estuarine rivers of the lower Chesapeake Bay. *Estuaries*. 13, 287–300. doi: 10.2307/1351920

Marshall, H. G. (2013). Phytoplankton in Virginia lakes and reservoirs. *Virginia J. Sci.* 64, 3–15. doi: 10.25778/f2rh-zd44

Marshall, H. G. (2014). Phytoplankton in Virginia lakes and reservoirs: part II. *Virginia J. Sci*. 65, 3–8. doi: 10.1353/wvh.2014.0003

Marshall, H. G., Lacouture, R. V., Buchanan, C., and Johnson, J. M. (2006). Phytoplankton assemblages associated with water quality and salinity regions in Chesapeake Bay, USA. *Estuar. Coast. Shelf Sci*. 69, 10–18. doi: 10.1016/j.ecss.2006.03.019

Marshall, H. G., Lane, M. F., Nesius, K. K., and Burchardt, L. (2009). Assessment and significance of phytoplankton species composition within Chesapeake Bay and Virginia tributaries through a long-term monitoring program. *Environ. Monitor. Assess*. 150, 143–155. doi: 10.1007/s10661-008-0680-0

Miller, A., and Russell, C. (2017). Food crops irrigated with cyanobacteria-contaminated water: an emerging public health issue in Canada. *Environ. Health Rev*. 60, 58–63. doi: 10.5864/d2017-021

Naselli-Flores, L., and Padisák, J. (2016). Blowing in the wind: how many roads can a phytoplanktont walk down? A synthesis on phytoplankton biogeography and spatial processes. *Hydrobiologia* 764, 303–313. doi: 10.1007/s10750-015-2519-3

Nishikawa, T., Hori, Y., Nagai, S., Miyahara, K., Nakamura, Y., Harada, K., et al. (2010). Nutrient and phytoplankton dynamics in Harima-Nada, Eastern Seto Inland Sea, Japan during a 35-year period from 1973 to 2007. *Estuar. Coast*. 33, 417–427. doi: 10.1007/s12237-009-9198-0

Pachepsky, Y., Kierzewski, R., Stocker, M., Sellner, K., Mulbry, W., Lee, H., et al. (2017). Temporal stability of *Escherichia coli* concentrations in waters of two irrigation ponds in Maryland. *Appl. Environ. Microbiol.* 84:e01876-17. doi: 10.1128/AEM.01876-17

Padovesi-Fonseca, C., and Philomeno, M. G. (2004). Effects of algicide (copper sulfate) application on short-term fluctuations of phytoplankton in Lake Paranoá, Central Brazil. *Brazil. J. Biol*. 64, 819–826. doi: 10.1590/S1519-69842004000500011

Patrick, R. (1973). Use of algae, especially diatoms, in the assessment of water quality. *Biol. Method. Assess. Water Qual* 76–95. doi: 10.1520/STP34718S

Powell, T. M., Richerson, P. J., Dillon, T. M., Agee, B. A., Dozier, B. J., Godden, D. A., et al. (1975). Spatial scales of current speed and phytoplankton biomass fluctuations in Lake Tahoe. *Science* 189, 1088–1090. doi: 10.1126/science.189.4208.1088

Ren, X., Yang, K., Che, Y., Wang, M., Zhou, L., and Chen, L. (2016). Spatial and temporal assessment of the initial pattern of phytoplankton

population in a newly built coastal reservoir. *Front. Earth Sci.* 10, 546–559. doi: 10.1007/s11707-015-0543-2

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L., and Melo, S. (2002). Towards a functional classification of the freshwater phytoplankton. *J. Plankton Res.* 24, 417–428. doi: 10.1093/plankt/24.5.417

Reynolds, C. S., Maberly, S. C., Parker, J. E., and Ville, M. M. D. (2012). Forty years of monitoring water quality in Grasmere (English Lake District): separating the effects of enrichment by treated sewage and hydraulic flushing on phytoplankton ecology. *Freshw. Biol.* 57, 384–399. doi: 10.1111/j.1365-2427.2011.02687.x

Rimet, F., and Druart, J.-C. (2018). A trait database for phytoplankton of temperate lakes. *Ann. Limnol. Int. J. Limnol.* 54:18. doi: 10.1051/limn/2018009

Rollwagen-Bollens, G., Bollens, S. M., Gonzalez, A., Zimmerman, J., Lee, T., and Emerson, J. (2013). Feeding dynamics of the copepod *Diacyclops thomasi* before, during and following filamentous cyanobacteria blooms in a large, shallow temperate lake. *Hydrobiologia* 705, 101–118. doi: 10.1007/s10750-012-1385-5

Schrader, K. K., Duke, S. O., Kingsbury, S. K., Tucker, C. S., Duke, M. V., Dionigi, C. P., et al. (2000). Evaluation of ferulic acid for controlling the musty-odor cyanobacterium, *Oscillatoria perornata*, in aquaculture ponds. *J. Appl. Aquacult.* 10, 1–16. doi: 10.1300/J028v10n01_01

Shi, X., Liu, X., Liu, G., Sun, Z., and Xu, H. (2012). Application of phytoplankton communities for monitoring water quality in the Hangzhou section of Jing-Hang Grand Canal, southern China. *Fundamental Appl. Limnol.* 2012, 1–11. doi: 10.1127/1863-9135/2012/0238

Shi, X., Sun, Z., Liu, G., and Xu, H. (2015). Insights into community-based discrimination of water quality status using an annual pool of phytoplankton in mid-subtropical canal systems. *Environ. Sci. Poll. Res.* 22, 1199–1206. doi: 10.1007/s11356-014-3391-9

Shimoda, Y., and Arhonditsis, G. B. (2016). Phytoplankton functional type modelling: running before we can walk? A critical evaluation of the current state of knowledge. *Ecol. Model.* 320, 29–43. doi: 10.1016/j.ecolmodel.2015.08.029

Smith, J. E., Stocker, M. D., Wolny, J. L., Hill, R. L., and Pachepsky, Y. A. (2020). Intraseasonal variation of phycocyanin concentrations and environmental covariates in two agricultural irrigation ponds in Maryland, USA. *Environ. Monitor. Assess.* 192:706. doi: 10.1007/s10661-020-08664-w

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environ. Sci. Poll. Res.* 10, 126–139. doi: 10.1065/espr2002.12.142

Soininen, J., Kokocinski, M., Estlander, S., Kotanen, J., and Heino, J. (2007). Neutrality, niches, and determinants of plankton metacommunity structure across boreal wetland ponds. *Écoscience* 14, 146–154. doi: 10.2980/1195-6860(2007)14[146:NNADOP]2.0.CO;2

Song, L., Marsh, T. L., Voice, T. C., and Long, D. T. (2011). Loss of seasonal variability in a lake resulting from copper sulfate algaecide treatment. *Phys. Chem. Earth A/B/C* 36, 430–435. doi: 10.1016/j.pce.2010.04.016

Stocker, M. D., Penrose, M., and Pachepsky, Y. A. (2018). Spatial patterns of *Escherichia coli* concentrations in sediment before and after high-flow events in a first-order creek. *J. Environ. Qual.* 47, 958–966. doi: 10.2134/jeq2017.11.0451

Su, X., Steinman, A. D., Xue, Q., Zhao, Y., Tang, X., and Xie, L. (2017). Temporal patterns of phyto- and bacterioplankton and their relationships with environmental factors in Lake Taihu, China. *Chemosphere* 184, 299–308. doi: 10.1016/j.chemosphere.2017.06.003

Tango, P. J., and Butler, W. (2008). Cyanotoxins in tidal waters of Chesapeake Bay. *Northeastern Natural.* 15, 403–416. doi: 10.1656/1092-6194-15.3.403

Uyttendaele, M., Jaykus, L.-A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., et al. (2015). Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. *Comprehens. Rev. Food Sci. Food Saf.* 14, 336–356. doi: 10.1111/1541-4337.12133

Vandermeulen, R. A., Mannino, A., Neeley, A., Werdell, J., and Arnone, R. (2017). Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color. *Opt. Expr.* 25, A785–A797. doi: 10.1364/OE.25.00A785

Varol, M. (2019). Phytoplankton functional groups in a monomictic reservoir: seasonal succession, ecological preferences, and relationships with environmental variables. *Environ. Sci. Poll. Res.* 26, 20439–20453. doi: 10.1007/s11356-019-05354-0

Vasas, G., Farkas, O., Borics, G., Felföldi, T., Sramkó, G., Batta, G., et al. (2013). Appearance of *Planktothrix rubescens* bloom with [D-Asp3, Mdha7]MC-RR in gravel pit pond of a shallow lake-dominated area. *Toxins* 5, 2434–2455. doi: 10.3390/toxins5122434

VDH (2015). *Virginia Recreational Water Guidance for Microcystin and Microcystis Blooms Provisional Guidance*. Available online at: https://townhall.virginia.gov/L/ViewGDoc.cfm?gdid=5479 (accessed September 17, 2021).

Wang, L., Wang, C., Deng, D., Zhao, X., and Zhou, Z. (2015). Temporal and spatial variations in phytoplankton: correlations with environmental factors in Shengjin Lake, China. *Environ. Sci. Poll. Res.* 22, 14144–14156. doi: 10.1007/s11356-015-4640-2

Wolanin, A., Soppa, M. A., and Bracher, A. (2016). Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. *Remote Sens.* 8:871. doi: 10.3390/rs8100871

Wood, R. (2016). Acute animal and human poisonings from cyanotoxin exposure—a review of the literature. *Environ. Int.* 91, 276–282. doi: 10.1016/j.envint.2016.02.026

Woodhouse, J. N., Kinsela, A. S., Collins, R. N., Bowling, L. C., Honeyman, G. L., Holliday, J. K., et al. (2016). Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. *ISME J.* 10, 1337–1351. doi: 10.1038/ismej.2015.218

World Health Organization (2008). *Microbiological Hazards in Fresh Leafy Vegetables and Herbs: Meeting Report (Vol. 14)*. Washington, DC: World Health Organization.

Wu, Z., He, H., Cai, Y., Zhang, L., and Chen, Y. (2014). Spatial distribution of chlorophyll a and its relationship with the environment during summer in Lake Poyang: a Yangtze-connected lake. *Hydrobiologia* 732, 61–70. doi: 10.1007/s10750-014-1844-2

Xi, H., Hieronymi, M., Krasemann, H., and Röttgers, R. (2017). Phytoplankton group identification using simulated and *in situ* hyperspectral remote sensing reflectance. *Front. Mar. Sci.* 2017:272. doi: 10.3389/fmars.2017.00272

Xiao, L.-J., Zhu, Y., Yang, Y., Lin, Q., Han, B.-P., and Padisák, J. (2018). Species-based classification reveals spatial processes of phytoplankton meta-communities better than functional group approaches: a case study from three freshwater lake regions in China. *Hydrobiologia* 811, 313–324. doi: 10.1007/s10750-017-3502-y

Xiao, R., Su, S., Ghadouani, A., and Wu, J. (2013). Spatial analysis of phytoplankton patterns in relation to environmental factors across the southern Taihu basin, China. *Stochastic Environ. Res. Risk Assess.* 27, 1347–1357. doi: 10.1007/s00477-012-0670-1

XiaoLi, Z., LiRong, S., and XiaoMing, Z. (2009). Effects of copper sulfate treatment on eutrophic urban lake phytoplankton communities. *Acta Hydrobiol. Sin.* 33, 596–602. doi: 10.3724/SP.J.1035.2009.40596

Zhang, Z., Gao, J., and Cai, Y. (2020). The direct and indirect effects of land use and water quality on phytoplankton communities in an agriculture-dominated basin. *Environ. Monitor. Assess.* 192:760. doi: 10.1007/s10661-020-08728-x

Zhou, S., Shao, Y., Gao, N., Deng, Y., Qiao, J., Ou, H., et al. (2013). Effects of different algaecides on the photosynthetic capacity, cell integrity and microcystin-LR release of *Microcystis aeruginosa*. *Sci. Tot. Environ.* 463–464, 111–119. doi: 10.1016/j.scitotenv.2013.05.064

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 01 November 2021
doi: 10.3389/frwa.2021.741653

Development of an Irrigation Water Quality Database to Identify Water Resources and Assess Microbiological Risks During the Production of Fresh Fruits and Vegetables

Elizabeth A. Bihn[1], Kurt J. Mangione[2], Bill Lyons[3†], Annette L. Wszelaki[4], John J. Churey[1†], Don M. Stoeckel[1] and Randy W. Worobo[1]*

[1] Department of Food Science, Cornell Agritech, Geneva, NY, United States, [2] New York State Department of Agriculture and Markets Food Laboratory, Albany, NY, United States, [3] New York State Department of Agriculture and Markets, Albany, NY, United States, [4] Department of Plant Sciences, University of Tennessee, Knoxville, Knoxville, TN, United States

OPEN ACCESS

Edited by:
Daniel Lowell Weller, SUNY College of Environmental Science and Forestry, United States

Reviewed by:
Achyut Adhikari, Louisiana State University, United States
Cameron Bardsley, Virginia Tech, United States

***Correspondence:**
Elizabeth A. Bihn
Eab38@cornell.edu

† Retired

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *15 July 2021*
Accepted: *23 September 2021*
Published: *01 November 2021*

Citation:
Bihn EA, Mangione KJ, Lyons B, Wszelaki AL, Churey JJ, Stoeckel DM and Worobo RW (2021) Development of an Irrigation Water Quality Database to Identify Water Resources and Assess Microbiological Risks During the Production of Fresh Fruits and Vegetables. Front. Water 3:741653. doi: 10.3389/frwa.2021.741653

An Irrigation Water Quality Database was developed to help assess the microbiological quality of irrigation water used in fruit and vegetable production in 15 counties in New York (NY) State. Water samples from Tennessee (TN) were also included in the database. Four water quality parameters, quantified generic *Escherichia coli*, specific conductance, pH, and turbidity, were tested. Ground, reservoir, and running water were sampled over 2 years (2009 and 2010), covering three seasons each year (spring, summer, and fall). TN data are for all three seasons in 2010 only. Overall in NY (254 total samples), ground water had a geometric mean of 1 most probable number (MPN)/100 ml, reservoir water had a geometric mean of 8 MPN/100 ml, and running water had a geometric mean of 52 MPN/100 ml. Overall in TN (63 total samples), ground water had a geometric mean of 1 colony forming unit (CFU)/100 ml, reservoir water had a geometric mean of 5 CFU/100 ml, and running water had a geometric mean of 38 CFU/100 ml. These values are all below the 126 MPN/100 ml United States Environmental Protection Agency's Ambient Water Quality Standards (AWQS) standard for fresh water. The presence of *E. coli* had very weak but sometimes statistically signficiant correlatation with water specific conductance, pH, and turbidity, depending on the water source but the *r*-squared effect was not strong enough to make the other measurements a substitute for testing specifically for *E. coli* in water.

Keywords: irrigation water, *E. coli*, microbiological risks, fresh produce, pH, specific conductance, turbidity

INTRODUCTION

Fresh fruit and vegetable production is a multi-billion dollar industry in the United States (U.S.) (Kaufman et al., 2000). These commodities are often irrigated with surface water throughout the U.S. (Suslow et al., 2003; Pachepsky et al., 2011). Surface water is more likely to be exposed to human and animal fecal contamination than ground water, and is expected to pose greater

risk to human health than irrigation with water from deep aquifers with properly constructed and protected wells (Brackett, 1999; Steele and Odumeru, 2004). Contamination of surface water used for the production of fresh fruits and vegetables by human foodborne pathogens such as *Salmonella, Escherichia coli* O157:H7, *Giardia*, *Listeria monocytogenes*, and *Cryptosporidium* has been documented (Steele and Odumeru, 2004; Chaidez et al., 2005; Duffy et al., 2005; Izumi et al., 2008; Jones et al., 2014; Weller et al., 2019).

Initially studies of irrigation waters were concerned primarily with chemical rather than microbiological water-quality parameters (Seiler and Skorupa, 2001). As a result, there was a nationwide knowledge gap regarding sanitary quality of irrigation waters but many studies in the last decade have been focused on on-farm irrigation water quality (Strawn et al., 2013; Jones et al., 2014; Draper et al., 2016; Weller et al., 2019). Recreational water criteria set forth by United States Environmental Protection Agency's Ambient Water Quality Standards (AWQS) (United States Environmental Protection Agency, 2012) were developed for water used for recreation that results in full body contact by people, and it accepts that there is a baseline of individuals who are expected to contract gastrointestinal illness regardless of recreational water quality. The AWQS estimates that 36 illnesses per 1,000 swimmers will result at fresh-water beaches, even when standards are met. The illness rate expectations implicit in the AWQS may not be directly transferrable to irrigation water, but in the absence of water data related to fresh produce production, the AWQS structure and values have been used as the basis for agricultural water quality criteria. The Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens (CSFSG) is one industry guideline that has modified the recreational water standards for use in fresh produce production. In addition to industry adoption of these standards, the Food Safety Modernization Act (FSMA) Produce Safety Rule, released on November 27, 2015, included AWQS-based water quality criteria for water used during the production of fresh produce when the water is intended to, or likely to, contact the harvestable portion of the crop (United States Food and Drug Administration, 2015).

The research outlined in this paper was conducted prior to the release of the FSMA Produce Safety Rule and prior to updated CSFSG water requirements (CSFSGPHLLG, 2020). This research resulted in the development of an Irrigation Water Quality Database to help growers better understand regional water quality and provide an example of how water test results could be collected to expand our understanding of water used for fruit and vegetable production. Preliminary research data gathered prior to the beginning of this project from surface water sources used to overhead irrigate fresh produce crops indicated that if growers were required to adopt the AWQS criteria, they would either have to discontinue the use of some of their water sources, or implement mitigation strategies, such as treatment, to reduce the microbiological load because surface water quality can vary over the season (Bihn et al., 2013). In fact, this concern has already partly come to fruition with the Leafy Greens Marketing Agreement April 2019 revisions to water quality metrics that remain in place today (CSFSGPHLLG, 2020). The revised metrics (for both California and Arizona) incorporate treatment as one risk reduction strategy for member organizations when using surface water or ground water that does not meet stringent water quality metrics for overhead irrigation within 21 days of harvest.

This project aimed to provide an objective assessment of the sanitary quality of surface water used for irrigation in New York (NY) State with a comparison of water from Tennessee at the time of the research through the collection of water quality parameters including quantified generic *E. coli*, specific conductance, pH, and turbidity. Investigating water quality over two years, through three seasons each year, was expected to provide useful insights into variation in water quality and identify ways to improve sampling strategies to further our understanding of surface water quality. Harvest seasons are known to impact the presence and abundance of pathogens in irrigation water, though these fluctuations are not always reflected in the microbial load on seasonal fresh produce (Selma et al., 2007). Since generic *E. coli* is only an indicator of fecal contamination and not the presence of pathogens, it would be valuable to know if growers could use other water quality parameters to predict the presence and abundance of generic *E. coli* besides testing specifically for *E. coli* since this process is expensive and it is often difficult to find laboratories in close proximity to farms. Overall, the database could improve our understanding of current water quality by identifying when fecal contamination is present, help growers meet industry and regulatory expectations, and also provide strategies for implementing an effective water-testing program. In addition, this database was developed to facilitate participation from others interested in providing water quality data so that a nationwide representation could result. Many individual growers and groups, including federal and state organizations, are collecting water samples and generating results, sometimes for the same water sources. Sharing data through a database, such as the one developed in this project, could allow the data to be used more effectively for water management decision making.

MATERIALS AND METHODS

Irrigation Water Database Development

The irrigation water quality database utilized the FileMaker Pro 10 (Santa Clara, CA) data collection system. The database was comprised of six independent, relational files hosted on a FileMaker server. It was designed to facilitate data gathering from multiple sources, such as independent laboratories or researchers at other land-grant universities. Two of the files (data entry and grower address) were designed to be web-accessible, allowing an individual with proper permission to access the data entry file via a web browser without the need to own the FileMaker software. The remaining files had administrator privileges and were not viewable to those entering data.

Each grower who participated in the program was assigned a nine-digit, numeric grower code in an effort to ensure participant confidentiality. The "Grower" file stored the grower's contact information, commodities grown, and water source(s). State and county information for each grower was also recorded. This information was used to generate a water sample collection form,

which included the unique nine-digit grower ID that appeared on each form.

The collection forms were given to those collecting water samples and in turn were submitted to the laboratory for analysis. The completed laboratory forms were then given to the data entry operator for entry into the FileMaker database. Upon registering with the program to enter data, individuals were provided with a database tutorial to instruct them how to access the data entry portal (**Appendix A** in Supplementary Material). They were then provided with an individualized password and login to gain entry to the database.

To enter data, the data entry operator was required to log into the database and enter the grower ID that appeared on each form. The operator entered all of the laboratory data from the water collection form, and submitted it for each grower. Multiple forms were required if an individual grower had more than one water sample source or sampling location. Each sample was entered separately. After all collection forms had been entered, the data entry operator would log out of the system.

When the database administrator logged into the water collection file, all grower collection data was imported automatically from the data entry file into the collection file. Once the import process was complete, all records in the data entry file were erased, thereby preventing the possibility of further access to the entered data. Only those with access to the water sample collection file could view the collected data. The data imported into the collection file was "as entered" by the data entry operator. If there was a mistake, the data entry operator was required to contact the database administrator to correct the data. If an administrator needed to correct entered data, there was a standard protocol in place for making corrections that included entering the user's ID, date, and reason for the correction. This information became part of the permanent record so that any data modifications were tracked.

Water Collection

A standardized water sampling protocol was developed, tested, and used as a training protocol (**Appendix B** in Supplementary Material). Four Department of Agriculture and Markets personnel and a summer intern were trained with the standardized water sampling protocol to facilitate sample collection across the state.

One liter of water was collected from each site into bottles that were either cleaned or purchased, then decontaminated following EPA protocol "B" (United States Environmental Protection Agency, 1992). Water was collected either at the source near the intake or for some limited samples, at the end of water distribution system in the field. For samples collected at the end of the distribution system, it was necessary to ensure the irrigation pump was operating and that it had been running long enough so that the sample was representative of the source. When samples were collected at the source, it was not necessary to ensure the water distribution system was operating. Collected samples were placed in a cooler containing ice packs, and either delivered directly or sent via overnight delivery for analysis at either the New York State Food Laboratory (Albany, NY) or the New York State Agricultural Experiment Station (Geneva, NY). Overnight delivery caused sample analysis to fall outside an 8-h analysis window, but according to Pope et al. (2003), samples held at 10°C and not frozen for up to 48 h generated comparable results to samples processed within the recommended 8-h hold time (American Public Health Association, 1998). TN samples were delivered directly to Microbac Laboratories (Maryville, TN or Nashville, TN) and all samples were analyzed for generic *E. coli*, specific conductance, turbidity, and pH.

Sample Analysis

Samples were analyzed for quantified generic *E. coli,* specific conductance, turbidity, and pH. Other data points related to water collection included date of sample collection, type of water source, name of water source (if a named body of water, such as a stream or lake), and code number for the grower.

Generic *E. coli* Quantification

IDEXX Quanti-Tray 2000 cards (Westbrook, ME) and IDEXX Quanti-Tray sealer (Westbrook, ME) were used according to the manufacturer's instructions for quantification of generic *E. coli* in water samples. Yellow wells and wells fluorescing at 366 nm were counted, and the numbers of positive wells for each test on each sample were recorded. Final most probable number (MPN) per 100 ml was determined by referring to the IDEXX Quanti-Tray 2000 MPN table with an upper limit of detection of 2,419 MPN per 100 ml (Kinzelman et al., 2005). Tennessee water samples were analyzed using modified mTEC (EPA 1603) protocol for quantification of generic *E. coli* resulting in colony-forming units (CFU) per 100 ml.

Specific Conductance

Specific conductance measures water's ability to conduct electricity, normalized to a temperature of 25°C. The parameter was included in this project as an indicator of run-off events that could impact water quality, as run-off events typically cause a decrease in the specific conductance of a particular water source. Specific conductance was measured with a conductivity meter (Oakton Instruments, Vernon Hills, IL), used in accordance with manufacturer's instructions.

Turbidity

Turbidity was measured in Nephelometric Turbidity Units (NTU) using a Hach 2100P Portable Turbidimeter (Loveland, CO) according to manufacturer's instructions.

pH

The pH of water samples was measured using a BeckmanΦ720 pH meter (Brea, CA). All results were recorded on the sample submission form.

Data Analysis

In NY, there were a total of 270 samples collected from 15 counties (**Figure 1**). These samples were taken from wells, rivers, streams, canals, swamps, lakes, and ponds. To aid in analysis, this larger data set was divided into three categories: ground water representing wells; running water representing rivers, streams, and canals; and reservoir water representing lakes, ponds, and swamps. Wells were expected to represent ground water that is

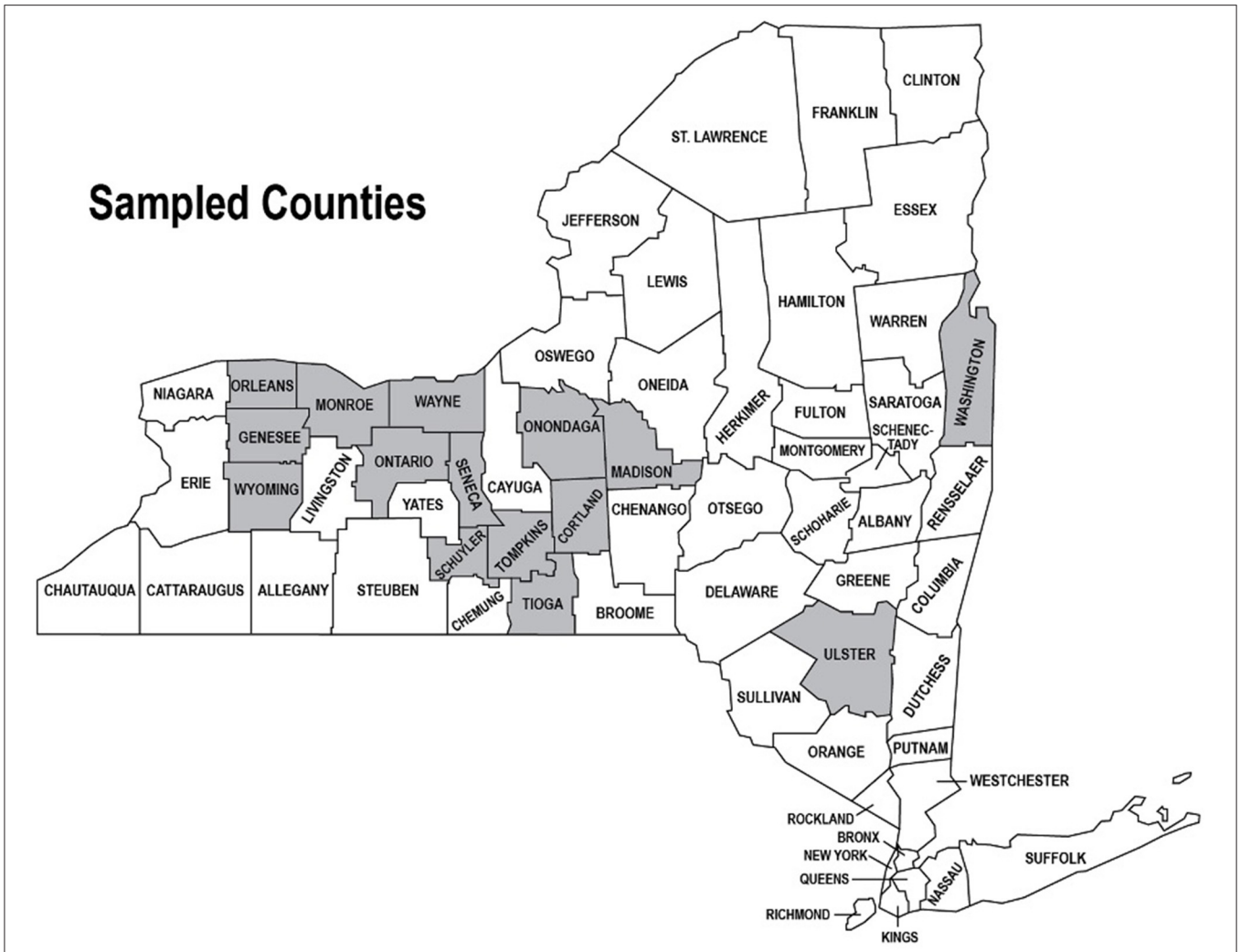


FIGURE 1 | Water samples were collected from 15 counties throughout New York State. Counties with participating farms are highlighted in gray and data generated from the sampling were used to initially populate and test the Irrigation Water Quality Database.

protected from fecal contamination since they are not open to the environment. Though not all wells may be properly capped or protected, they were viewed as a distinct data set. A total of 254 samples were divided into running water (94 samples), reservoir water (137 samples), and ground water (23 samples) groups. Of the remaining samples, one sample was not properly analyzed for generic *E. coli* due to an incubator malfunction, and 15 were sampled directly from irrigation equipment and were removed from the predominant data set.

The database contains NY data as well as data from TN (62 samples). Tennessee data were collected only in 2010, but utilized similar collection parameters to those in NY and were included for comparison in this paper. There were 65 samples collected in TN and entered into the database, including 19 ground water samples, 27 reservoir samples, 16 running water samples, and three municipal samples. The municipal TN samples were not included in the statistical analysis because there were no comparable data in NY. Several samples from TN were taken from irrigation equipment. In the NY data set these samples were removed because in many cases there were samples taken from the source water at the same time. Since there were fewer TN samples, the samples taken from irrigation water equipment were included in the analysis. Tennessee water samples were analyzed for the same parameters as the NY samples but the modified mTEC (EPA 1603) protocol was used for quantification of generic *E. coli* resulting in CFU per 100 ml. Our analysis is supported in the literature by Cho et al. (2010), who reported a positive correlation between CFU and MPN estimates. These methods also both appear on the list of approved microbiological methods for ambient water for measurement of *E. coli* (40 CFR §136.6). Since these protocols are considered equivalent, NY and TN results were compared on a one-to-one basis.

At the time of the study, reporting the geometric mean was the method used by both the AWQS and the CSFSG, so it is the standard used in this work (Dufour and Schaub, 2007; CSFSGPHLLG, 2010). Medians and averages were also calculated

(data not shown). Median calculations aligned very closely to the geometric mean in most cases, while averages were usually much higher since calculating the average does not manage extreme data points in the data set as does the geometric mean calculation. Analysis was performed by year or by season, to determine if trends were present or to otherwise evaluate variability in the data.

All data were log-transformed prior to statistical analysis as is typically recommended for microbiological data that follow a log-normal distribution (American Public Health Association, 1998). In order to allow log transformation, data with "less than" values were assigned a value of half the detection limit. Two data points with "greater than" values were set at the detection limit (2420 MPN/100 ml) with the expectation that rare incidence would not skew statistical outcomes. One of these samples was a pond in New York in 2010, the other was from a drip irrigation line in New York. The central tendency for these data was approximated using the geometric mean. Confidence intervals were calculated using z-scores. Narrative descriptions of differences between the means of data sets are based on visual overlap of 95-percent confidence intervals. In addition, a *t*-test was used to confirm significance of differences ($p < 0.05$; with settings two-tail and unequal variance). Ability of each parameter to explain the patterns in generic *E. coli* concentration was tested by simple linear regression. Correlation coefficients were calculated using the linear regression function in Excel and correlation significance was evaluated at the level $p < 0.05$.

RESULTS

Two hundred seventy surface water samples from NY State were collected and analyzed for pH, specific conductance, turbidity, and quantified generic *E. coli*. Overall for NY State, ground water had a geometric mean of 1 MPN/100 ml, reservoir water had a geometric mean of 8 MPN/100 ml, and running water had a geometric mean of 52 MPN/100 ml. Every ground water sample was below the 126 MPN/100 ml AWQS standard that is also used by the CSFSG. Reservoir water met the 126 MPN/100 ml standard 96% of the time, with only 3% of the samples exceeding 235 MPN/100 ml. The 235 MPN/100 ml value was the single sample upper limit standard for water intended for foliar applications to the edible portion of the crop as set forth in the CSFSG when this study was completed (CSFSGPHLLG, 2010). Similar to the single-sample value of 235 MPN/100 ml, other benchmarks that came into use after the timeframe of sample collection and analysis also originate with the USEPA Recreational Water Standards (United States Environmental Protection Agency, 2012). Specifically, the FSMA Produce Safety Rule, as published in 2015, used a statistical threshold value (STV) of 410 CFUs (equivalent to MPN) in 100 ml water and the August 2020 update to the LGMA metric revision used a single-sample maximum of 576 MPN/100 ml (United States Food and Drug Administration, 2015; CSFSGPHLLG, 2020). Running water samples met the 126 MPN/100 ml standard 74% of time, with only 15% of the samples being higher than the 235 MPN/100 ml standard. Though the rates of exceedance compared to the 235 MPN/100 ml standard (3% for reservoir water and 15% for running water) might appear low, there were samples that exceeded the limit, and therefore, would require mitigation, including retesting or abandoning the water source under contemporary rules during the sample collection time period.

Further analysis of water quality data from NY by year revealed consistent results from year to year and source to source

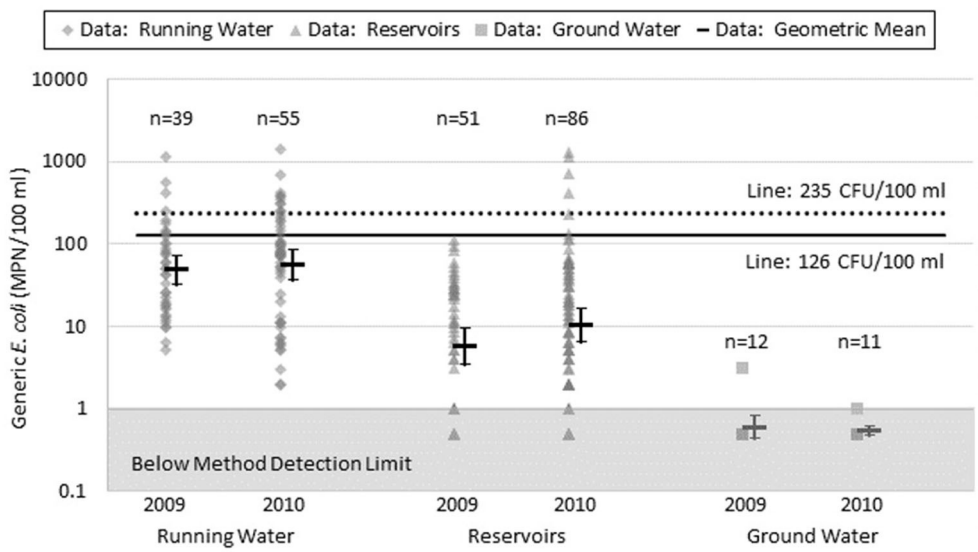


FIGURE 2 | Range of values, geometric mean and confidence interval around the mean for quantified generic *E. coli* (most probable number per 100 ml) for each New York State water source group in 2009 and 2010. Number of samples below detection limit: Running Water 2009 $n = 0$, 2010 $n = 0$; Reservoirs 2009 $n = 13$, 2010 $n = 15$; Wells 2009 $n = 11$, 2010 $n = 10$.

(**Figure 2**). Ground water had the lowest quantified generic *E. coli* counts (with 21 of 23 samples below detection limit) while running water had the highest counts. The geometric mean values for reservoir water samples were almost an order of magnitude lower than running water sources, even though reservoir water sources are open to the environment like running water sources. Unlike reservoir sources, running water sources are subject to a multitude of variables due to the nature of running water sources having miles of banks with potential access by wildlife, septic systems, drainage tiles, run-off, and/or human recreational activity.

Comparison by State and Season

Collaborators in TN began participating in the Irrigation Water Database in 2010. Overall, TN ground water had a geometric mean of 1 CFU/100 ml, reservoir water had a geometric mean of 5 CFU/100 ml, and running water had a geometric mean of 38 CFU/100 ml. Each ground water sample was below the 126 MPN/100 ml standard. Reservoir water met the 126 MPN/100 ml standard 93% of the time, with only 4% of the samples exceeding the 235 MPN/100 ml single upper limit standard. These percentages were consistent with those seen in NY water sources. Tennessee running water samples met the 126 MPN/100 ml standard 75% of the time, with only 6% of the samples being higher than the 235 MPN/100 ml standard. The percentage of samples achieving the 126 MPN/100 ml standard is comparable to NY, while the percentage of samples that exceed the 235 MPN/100 ml limit is approximately half of what was observed in NY. Sample size differences may help to account for this difference. Accurate estimation of exceedance rates was more likely for the larger NY sample set, with 94 observations (14 exceedances), compared with the smaller TN sample set, with only 16 observations (1 exceedance). The exceedance rates may also be affected by regional variation in water quality.

Analysis of TN running water data in the fall (22 September−20 December) indicated an excessively high geometric mean of 255 CFU/100 ml, but the geometric mean was calculated with only three samples and therefore the 95% confidence interval was broad (58–1,100 MPN/100 ml; see **Figure 3**). In general, a five-sample minimum is preferred for calculating a more precise geometric mean. The three samples were 860, 160, and 120 CFU/100 ml, all taken from the same water source, but at different, non-contiguous, locations along the water source, on the same day. Interestingly, two of the three samples did not exceed the 235 MPN/100 ml limit. Absent the highest value (860 MPN/100 ml) results from this data set likely would not have caused the grower to take any action, yet the 860 CFU/100 ml sample exceeds even the most liberal EPA recreational water standards of 575 MPN/100 ml for "Infrequently Used Full Body Contact" water, and it also exceeds the upper limit in the CSFSG standards (576 MPN/100 ml) for water that does not contact the edible parts of the plant such as water delivered through a drip system (United States Environmental Protection Agency (USEPA), 1986; CSFSGPHLLG, 2010).

With a clear understanding of these issues, the three data points from TN fall sampling of running water were included in the comparison of seasons and states found in **Figure 3**. Unlike NY, TN geometric mean concentrations of *E. coli* in running water samples were not always higher than the TN reservoir samples. In NY, *E. coli* geometric mean concentrations in both running and reservoir water were highest in the summer (21 June−21 September); in TN, spring (20 March−20 June) showed the highest counts in both reservoir and running water if the fall

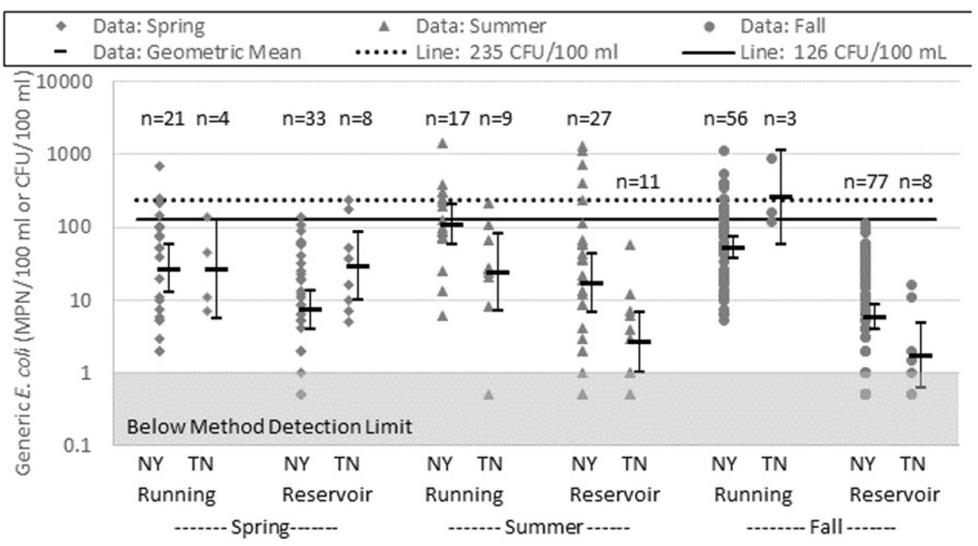


FIGURE 3 | New York (NY) and Tennessee (TN) data were analyzed by season. Spring is represented by samples taken between 20 March and 20 June, summer is represented by samples taken between 21 June and 21 September, and fall is represented by samples taken between 22 September and 20 December. No samples were taken in winter (between 21 December and 19 March). NY samples from 2009 and 2010 were combined for this analysis. TN data are from 2010 only. The shaded gray box depicts the general method detection limit of 1 MPN or CFU per 100 ml.

TABLE 1 | Correlations among *E. coli* presence and three other water quality parameters, specific conductance, turbidity, and pH.

NY surface water sources	*E. coli*/Specific conductance	*E. coli*/Turbidity	*E. coli*/pH
Reservoir	0.20*	0.05	−0.28*
Running	0.00	0.21*	−0.05
Reservoir and running combined	0.13*	0.05	−0.18*

**Statistically significant correlations.*

data from TN running water was discounted due to the issues discussed earlier in this section. To be confident in analysis of the TN data, more samples are needed for better representation of water quality by source type and season. The NY data showed seasonal trends in both running and reservoir water with the changes more pronounced in the running water from season to season. While the database was in use, TN continued to add data, as well as new collaborators in Texas and Maryland. Additional participation in the database from more states would have allowed additional data analysis and may have revealed trends that are regionally and/or seasonally dependent.

Correlations

In an attempt to determine if there were other, less expensive ways to monitor microbial water quality, correlations between *E. coli* counts and the other parameters tested, including specific conductance, turbidity, and pH (**Table 1**) were examined. These correlations were determined only for surface water sources sampled at the source water. There was a very weak statistically signficant negative correlation observed between *E. coli* levels and pH in all surface water samples over the 2 years ($r = -0.18$; $p = 0.01$), and a very weak positive correlation between *E. coli* and turbidity ($r = 0.05$; $p = 0.55$) and statistically significant very weak positive correlation for specific conductance ($r = 0.13$; $p = 0.05$). When running water and reservoir water were analyzed separately, some of the correlations were stronger. The strongest correlations identified were between *E. coli* and specific conductance ($r = 0.20$; $p = 0.02$) and *E. coli* and pH ($r = -0.28$; $p = 0.00$) in reservoir water, and *E. coli* and turbidity in running water ($r = 0.21$; $p = 0.05$). As noted above, several of these correlations were statistically significant ($p < 0.05$) but the r-squared values indicate a very low effect (Cohen, 1992). Based on the analysis these correlations would not be dependable indicators of the presence or quantity of *E. coli*.

DISCUSSION

Fresh produce growers who participated in this project benefited by having baseline water sampling completed for their on-farm water sources. They were supplied with a water testing protocol and had direct access to extension personnel to discuss water quality issues on their farms. Other NY farmers utilized trainings to help them understand the value of water testing, how to test their water sources, and how to locate laboratories that can provide the required analyses.

Ponds are a very common source of surface water in NY, and in this study the geometric mean of *E. coli* counts in ponds was 6 MPN/100 ml. In one particular case, a farm pond had an unusually high *E. coli* water test (>2,420 MPN/100 ml), so contact was made with the grower to discuss what could have been the cause. One of the farm owners participates in the Community Collaboration Rain, Hail, and Snow Network (CoCoRaHS), so they were able to share that the night before the sample was taken, they had 0.2 inch of rain. A review of the data showed all the indications of a rain event in that the turbidity was higher and the pH and specific conductance were lower, indicating dilution by rain. The other factor for this sample was that it was taken directly from the irrigation equipment, which in other instances in this study has also resulted in an unusually high *E. coli* count. A follow-up sample had a much lower *E. coli* count (3.1 MPN/100 ml), and was not a concern to the grower. This farm's participation in CoCoRaHS gave them a monitoring tool they did not know existed. In addition, being able to review the turbidity, specific conductance, and pH data provided a level of confidence in the determination that the water sample was affected by the rain event. Understanding the impact rain has on water quality attributes and the possible influence of irrigation equipment on water tests is important for making management decisions regarding water use. Rainfall data was not collected in this study, but growers should be made aware of options like CoCoRaHS to add to their arsenal of information that could help make water use management decisions.

The purpose of sharing this example is to highlight the fact that growers need to be aware of factors that can impact water quality and be able to interpret water-testing data. This project collected parameters outside the current recommendations for generic *E. coli* testing, yet adding these parameters does not significantly increase water testing cost and may provide growers with additional information that can assist them in determining the source and cause of contamination.

Impact of Sampling Location

For all water collections, sample location was noted. Most samples were taken directly from the source water, but there were several opportunities in the 2010 NY sampling season to collect water samples directly from irrigation equipment. The rationale behind sampling from irrigation equipment was to collect water closest to the point of use, and to obtain data that represent the microbiological quality of water that actually contacted the plant. Our results indicate that this may not be the best way to determine source water quality, but it is more representative of the use of the water and may be relevant to understanding overall risk to produce safety. During the 2010 season, nine pairs of samples were collected from irrigation equipment (running or flushed prior to sampling) and the source water that supplied the irrigation equipment. Two out of the nine sample pairs demonstrated that the irrigation equipment sample had higher *E. coli* counts than the source water sample. There were no cases where the source water sample had higher *E. coli* compared with the associated irrigation equipment sample. Observed differences were beyond the 95% confidence limits of the MPN test method (**Table 2**). For the other seven pairs, the data pairs had overlapping 95% confidence limits in the

TABLE 2 | Comparison of source water samples and irrigation line samples analyzed for *E. coli*.

Sample location[a]	MPN[b]/100 ml	95% confidence[c] lower limit	95% confidence upper limit
Source water	13.5	7.80	21.8
Drip irrigation	25.6	17.2	35.7
Source water	111	81.0	149
Drip irrigation	90.9	66.6	122
Source water	1.00	0	3.70
Drip irrigation	6.30	2.50	12.7
Source water	727	476	1,050
Overhead irrigation	1,200	811	1,750
Source water	**126**	102	152
Drip irrigation	**299**	207	423
Source water	74.3	53.0	98.8
Drip irrigation	72.3	51.5	96.4
Source water	**<1**	0	3.70
Drip irrigation	**>2,420**	1,440	Infinite
Source water	42.6	28.7	60.7
Overhead irrigation	49.6	35.4	67.8
Source water	16.1	9.60	24.9
Overhead irrigation	16.0	9.50	25.1

[a]*Paired samples were taken from both source water and irrigation lines from the same water source.*
[b]*MPN, most probable number.*
[c]*95% confidence limits from manufacturer's table of MPN results.*
Bolded data pairs show different results.

IDEXX Quanti-Tray 2000 results tables (indicating no credible difference in the results), though in three of those seven cases the overlap was one-sided. Concentrations of *E. coli* can vary significantly, making it difficult for growers to know the most representative location from which to collect water samples to determine water quality.

Although there were only two samples from irrigation equipment that differed from their source water samples, the observation highlights several things a grower may want to consider. The first issue is that irrigation equipment could incorporate risks into the production of fresh produce through the addition of microbial contamination that is multiplied from, or did not exist in, the source water. This could be from intrusion of fecal contamination somewhere in the distribution system or growth of *E. coli* and possibly pathogenic bacteria in the distribution system from year to year. Two of the samples were from farms using drip irrigation, and the increase in microbial counts could have been related to soil conditions including the use of amendments that increase *E. coli* counts since the drip tape is in direct contact with the soil. The overall risk to the food production system should be low because the water is applied at or just below the soil line and primarily wets the soil, not the edible portion of the plant. There are two exceptions to this assessment, one is root crops that would be contacted by drip irrigation and the other is a malfunction in the drip irrigation system that results in overhead spray from an emitter. One of the samples was collected from overhead irrigation equipment and had high *E. coli* counts (1200 MPN/100 ml) when compared to the source water (727 MPN/100 ml). The grower identified that this water was used to irrigate leafy greens. *E. coli* counts in both source water and irrigation equipment significantly exceeded the single-sample limit of 235 MPN/100 ml [United States Environmental Protection Agency (USEPA), 1986; CSFSGPHLLG, 2010]. Since this water was identified as being applied by overhead irrigation, it represented the riskiest application method due to wetting of the edible portion of the crop. It is not known whether the grower was actually irrigating with this water, or how close to harvest it was applied, but the data clearly represent concerns that the grower should review. In assessing risks, growers using both systems should review the siphon system that feeds the irrigation pipe to be certain it is suspended in the water and not siphoning bottom sediment, as this is known to increase microbial content (Badgley et al., 2011).

A second important issue is that sampling from the irrigation equipment may not accurately reflect the quality of the source water, and as mentioned above, which should cause a grower to review other aspects of production such as the integrity of water delivery systems. This is important because financial and time resources are limited and growers should be targeting resources to mitigate risk. Modifying a siphon float, flushing, and sanitizing the irrigation lines, or reviewing soil amendment application timelines may be easier, less expensive, and reduce risks more effectively compared to treating the entire water source, but without testing both the source water and the irrigation equipment, the grower would not know where the risk was introduced. The data collected in this study did not support determination of which process explains the observation, but it highlights the value of sampling from each location on the same day to see if there is a difference. Each possibility could lead to different corrective actions to reduce risk but without data indicating an issue, a grower may not know there is a problem.

The CSFSGPHLLG (2010) instructs farm operators to "sample sources as close to the point-of-use as practical, as determined by the sampler to ensure the integrity of the sample." The U.S. Food and Drug Administration (2015) instructs farm operators that samples "must be representative of (the) use of the water." Based on the inconsistent data generated when sampling irrigation equipment, it may be that the sampler could determine that in order to ensure the integrity of the sample, it should be sampled from source water, not the equipment. An alternative solution would be to collect two samples, one from the source water and one from the irrigation equipment, but even this has its challenges. Although irrigation systems may feed off the same main pump line, most systems branch and thus have multiple end points, so that type of testing strategy would increase the number of samples and, subsequently, the cost substantially. If there was concern regarding an overhead irrigation system, growers may consider developing a sampling rotation to identify microbial risks in irrigation lines. This would minimize the number of samples but still enable the sampling of lines over time.

Recommendations for Growers

In drawing conclusions from this research, it is important to consider whether our results impact current Good Agricultural Practices recommendations. Based on weak correlational data,

it is clear that the alternate measurements (pH, specific conductance, and turbidity) do not consistently allow estimation of the quality of a surface water source, based on *E. coli* as the indicator organism, as well as testing the water source for *E. coli*. Though there is much discussion about the value of generic *E. coli* as an indicator organism, it is the best available at the moment and there are laboratories that can complete the testing for growers (Suslow, 2010).

One recommendation for testing surface waters, based on the results of this research, would be to add additional parameters beyond quantified generic *E. coli* that do not substantially increase the cost of the test, but that provide the growers with additional information that might allow them to interpret their test results. In particular, specific conductance ($10) and turbidity ($8) would allow growers to understand if run-off is influencing their test results. Although the value of pH for water quality issues was not viewed as significant enough to recommend, some growers who signed up to participate in this study did so in order to determine the pH of their water sources because they were interested in the information relative to how it impacts their spray mixes. Since it is a relatively inexpensive test ($5) that can provide useful information to growers, including it in the recommended tests seems reasonable. Prices quoted for tests were provided by Certified Environmental Services Inc., a commercial water-testing laboratory in Syracuse, NY and are expected to be representative of prices at similar laboratories.

The amount and timing of testing is another area that warrants discussion. The 1986 EPA Ambient Water Quality Standard Criteria (EPAAWQSC) states that sampling frequency and testing to determine the quality of the water should be "based on a statistically sufficient number of samples (generally not <5 samples equally spaced over a 30-day period)." At the time when this research was conducted, this was adapted for use in the Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens (CSFSGPHLLG, 2010) as part of the California Leafy Greens Marketing Agreement (LGMA). As appeared in the LGMA decision tree for pre-harvest water used for foliar applications that result in the edible portion of the crop being contacted by water (i.e., overhead irrigation, topical protective sprays, frost protection), it stated: "Sampling Frequency: One sample per water source shall be collected and tested prior to use if >60 days since last test of the water source. Additional samples shall be collected at intervals of no <18 h and at least monthly during use. Geometric means, including rolling geometric means, shall be calculated using the five most recent samples." In August, 2020 these standard were updated to require that no untreated surface water be applied to leafy green crops within 21 days of harvest without additional pathogen testing required on the harvested crop (CSFSGPHLLG, 2020). The FSMA Produce Safety Rule, as published in 2015, states that "a minimum total of 20 samples of agricultural water…over a minimum period of 2 years, but not >4 years" is required for long-term water use decision making based on an expected standard deviation of 0.4 among log-transformed concentration data. Data collection and analysis in this project was collected prior to the release of the FSMA Produce Safety Rule so did not conform to these parameters. On March 18, 2019, FDA extended the compliance dates for meeting water quality requirements outline in FSMA Produce Safety Rule Subpart E "to address questions about the practical implementation of compliance with certain provisions and to consider how we might further reduce the regulatory burden or increase flexibility while continuing to protect public health." (United States Food and Drug Administration, 2019). A release of an updated Subpart E is anticipated in 2021.

The sampling strategy set forward by the LGMA at the time this research was conducted was more reasonable for agriculture than five samples spaced over 30 days as in the AWQS, but the calculation of the rolling geometric mean of five samples increases the testing requirements. Some growing seasons are particularly short in NY, so in order to get five water samples over the season, growers would need to sample at least once per week and even at that frequency, they would not be able to calculate their first geometric mean until the season was almost over. The single sample standards are more relevant in short production seasons, but because they are based on the recreational water standards, they accept that some people will get ill. Of note, FDA in the FSMA Produce Safety Rule declined to provide a single-sample standard, opting instead for a STV that "is a measure of variability of (the) water quality distribution, derived as a model-based calculation approximating the 90th percentile using the lognormal distribution."

Understanding and accepting that there are risks associated with using surface water sources during production is key to assessing the safety of surface water sources. Using water quality standards such as the EPA recreational water standards that clearly accept some illness as an outcome may seem unwise but requires additional consideration. Unlike full body contact water that assumes individuals will be directly exposed to the water and ingest some of it, surface water used for irrigation may never be ingested by those who consume the fresh fruits and vegetables. Irrigation water is often applied days before harvest so the water is dried by the time the fresh produce is picked. Exposure to sun and desiccation promote the reduction of microbial populations that may be deposited by the irrigation water, so the risk is likely less than in recreational waters because of the absence of conditions under which bacterial pathogens could subsequently reproduce on the produce (Steele and Odumeru, 2004).

Another consideration that supports the use of surface water for fresh produce production is the world-wide shortage of clean drinking water. If growers are driven to use water that has no detectable fecal-indicator bacteria, they will either have to treat the water or move to ground or municipal water. The use of ground or municipal water in the U.S. may not seem like such a dire option at the moment, but already in states like Florida and California, municipalities have established reclaimed water distribution systems to encourage people to not waste drinking water on watering lawns, gardens, and crops. When countries outside the U.S. are considered, the notion of using a clean drinking water source to irrigate crops is unrealistic in some cases. The World Health Organization sets the standard for wastewater used to irrigate crops that will be eaten raw at <1,000 fecal coliforms per 100 ml water (Blumenthal et al., 2000). Reclaimed water is usually treated, but because it likely originated

from human sewage sources, can still contain human pathogens (Sadovski et al., 1978; Bastos and Mara, 1995; Oron et al., 2010). Treatment of reclaimed water may vary by municipality and location, so the quality of reclaimed water may not be consistent and could easily exceed microbial counts found in surface water sources. Conserving drinking water sources and using surface water sources responsibly is important to the management of water as a natural resource and should be a consideration for growers in assessing the risks of using surface water in relationship to other management issues. It is also important to investigate effective treatment options that do not have negative plant and environmental consequences.

Key Extension Points

This project provided several insights into irrigation water quality and on-farm management practices related to water use. First, it was noted that on-farm water testing was not standard practice on NY farms at the time of the study (2009–2010) despite the existence of voluntary programs (e.g, GAPs, LGMA) that recommend water testing. With increased pressure from buyers and the 2015 publication of the FSMA Produce Safety Rule, interest in water testing is increasing, but much extension training needs to be done to adequately educate farmers on best water testing practices and the implications of those results. Training should include water sampling protocols, how to monitor surface water sources through environmental assessment and water testing, and finally how to interpret results. Growers will need to implement water-testing practices for surface water sources they are using, and conduct inspections that should include assessing risks of their water use practices and implementation of corrective actions that reduce any identified risks when water analysis results exceed expectation or indicate increased risks. Data from this project indicate that surface water quality can vary dramatically over the growing season and there are not always clear factors to indicate why these variations occur. Although in this project we sampled on-farm water sources only three times during each growing season, it may be more practical to recommend that growers test their surface water sources prior to the start of the season and *at least* once a month during use or more often if there are concerns about the quality of the water. Testing should be targeted at times just prior to use so growers have an understanding of quality prior to application, particularly if water is applied directly to the crop or close to harvest. Reviewing water testing results on a per-farm basis and incorporating other available information, such as rainfall data, will allow growers to gain valuable information in assessing the risks their surface water sources may represent.

To implement water-testing practices, growers will need access to water testing laboratories that can provide testing of surface water sources. Furthermore, not every commercial water-testing laboratory may be prepared to handle surface water samples and provide the testing services farmers may need and request. Not all water-testing laboratories are prepared to present water-testing results in a format that is valuable to growers. The FSMA Produce Safety Rule includes the need to calculate a STV to set an upper limit for variability of generic *E. coli* concentrations. This uncommon calculation requirement creates a barrier for on-farm water management decisions. First, most growers do not know how to perform the calculation. Second, if the calculation is done incorrectly, growers may use water that has significant fecal contamination which could increase risk or avoid using water that has low fecal contamination and risk crop loss due to lack of irrigation. Either outcome has serious impacts on crop health and safety. Several groups developed educational materials including fact sheets and water calculators to assist growers with doing these calculations and understanding these requirements (Bihn et al., 2017; Harris and Rock, 2017; Rock and Harris, 2017; Stoeckel et al., 2017).

Water samples also require analysis within a certain time period. Some protocols require analysis within 8 h, while others allow up to 30 h of hold time prior to analysis. This is important, because some farms are not located near water testing laboratories, so they would be forced to use overnight mail delivery for sample submission and would not meet the 8-h requirement. Unfortunately, some farms are in rural areas that do not have access to overnight mail, so this creates another barrier to establishing a water testing strategy. Research has indicated there can be impacts as a result of both hold times and laboratories conducting the analysis on water sample results, including both increases and decreases in *E. coli* counts, depending on these variables (Pope et al., 2003; Selvakumar et al., 2004; United States Environmental Protection Agency, 2006; Harmel et al., 2016). Our analysis method and study design were selected based on research as well as a process that growers could actually implement in order to have samples kept at or below 10°C and not frozen and delivered to a laboratory within 48-h (Pope et al., 2003). As water testing becomes required for produce growers, it is important that farmers be provided with a standard they can meet and that resources, such as access to water testing laboratories, are available.

Lastly, a very important outcome of this research was the development of a database that could capture water testing results and share them in a confidential way to the county level. It is common in agricultural production areas the different farms are using the same water sources, including streams, rivers, and canals. Some government agencies, such as the United States Geological Survey, and many university researchers are sampling the same water sources that growers are using and generating valuable data. As was seen with our database, others were interested in using the database, but funding to maintain and expand it was not available. Regardless of regulatory or audit requirements, it would be great to collect water testing data so that it is available for those that could use it to make better informed food safety management decisions. In many instances, the expense of taking the sample, transporting it to the laboratory, and doing the analysis is already done. The last step of cataloging and sharing the data is missing. The database developed for this project demonstrated that data can be collected in a secure and private way so that water testing information could be shared and benefit the agricultural community as a whole.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

EB was responsible for research design, water sampling, sample analysis at Cornell Agritech, and the writing the manuscript. KM was responsible for sample analysis at the New York State Food Laboratory. BL was responsible for supervision of state personnel that collected water samples. AW was responsible for all data collection in Tennessee and manuscript review. JC was responsible for guidance and assistance in the Cornell Agritech laboratory. DS was responsible for research design guidance and manuscript review, and RW was responsible for research design guidance, sampling analysis guidance, and manuscript review. All authors contributed to the article and approved the submitted version.

FUNDING

The authors acknowledge research funding support from the New York State Cabbage Research and Development Program, USDA National Integrated Food Safety Initiative grant #2008-51110-0688 and the USDA Specialty Crop Block Grant Program in TN.

ACKNOWLEDGMENTS

The authors would like to acknowledget the many growers who allowed us to sample water from their farms. We would also like to thank Mark Kogut for technical assistance on developing the database and for assistance in creating Figures and Tables, Elizabeth Buerman, Bernie DiBattista, VJ DeSantis, and Wendy Ingersoll for assistance in collecting water samples on farms throughout New York State. We thank Donna P. Clements for valuable feedback on the manuscript. We also acknowledge Cheryl D. Galvani and Toria Melville for editorial assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.741653/full#supplementary-material

REFERENCES

American Public Health Association (1998). *Standard Methods for the Examination of Water and Wastewater, 20th Edn*. Washington DC: American Public Health Association, American Water Works Association and Water Environmental Federation.

Badgley, B. D., Thomas, F. I. M., and Harwood, V. J. (2011). Quantifying environmental reservoirs of fecal indicator bacteria associated with sediment and submerged aquatic vegetation. *Environ. Microbiol.* 13, 932–942. doi: 10.1111/j.1462-2920.2010.02397.x

Bastos, R. K. X., and Mara, D. D. (1995). The bacterial quality of salad crops drip and furrow irrigated with waste stabilization pond effluent: an evaluation of the WHO guidelines. *Water Sci. Technol.* 31, 425–430. doi: 10.2166/wst.1995.0511

Bihn, E. A., Fick, B. J., Pahl, D. M., Stoeckel, D., M., Woods, K. L., et al. (2017). *Geometric Means, Statistical Threshold Values, and Microbial Die-Off Rates*. Produce Safety Alliance, Cornell University.

Bihn, E. A., Smart, C. D., Hoepting, C. A., and Worobo, R. W. (2013). Use of surface water in the production of fresh fruits and vegetables: a survey of fresh produce growers and their water management practices. *Food Prot. Trends*. 33, 307–314.

Blumenthal, U. J., Mara, D. D., Peasey, A., Ruiz-Palacios, G., and Stott, R. (2000). Guidelines for the microbiological quality of treated wastewater used in agriculture: recommendations for revising WHO guidelines. *Bull. World Health Org.* 78, 1104–1116. Available online at: https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/bwho/v78n9/v78n9a06.pdf

Brackett, R. E. (1999). Incidence, contributing factors, and control of bacterial pathogens in produce. *Postharvest Biol. Technol.* 15, 305–311. doi: 10.1016/S0925-5214(98)00096-9

Chaidez, C., Soto, M., Gortares, P., and Mena, K. (2005). Occurrence of *Cryptosporidium* and *Giardia* in irrigation water and its impact on the fresh produce industry. *Int. J. Environ. Health Res.* 15, 339–345. doi: 10.1080/09603120500289010

Cho, K. H., Han, D., Park, Y., Lee, S. W., Cha, S. M., J.-H., et al. (2010). Evaluation of the relationship between two different methods for enumeration fecal indicator bacteria: colony-forming unit and most probable number. *J. Environ. Sci.* 22, 846–850. doi: 10.1016/S1001-0742(09)60187-X

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

CSFSGPHLLG (2010). *Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens*. California Leafy Greens Marketing Agreement.

CSFSGPHLLG (2020). *Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens*. California Leafy Greens Marketing Agreement. Downloaded from https://lgmatech.com/resources/

Draper, A. D., Doores, S., Gourama, H., and LaBorde, L. F. (2016). Microbial survey of Pennsylvania surface water used for irrigating produce crops. *J. Food Prot.* 79, 902–912. doi: 10.4315/0362-028X.JFP-15-479

Duffy, E. A., Lucia, L. M., Kells, J. M., Castillo, A., Pillai, S. D., and Acuff, G. R. (2005). Concentration of *Escherichia coli* and genetic diversity and antibiotic resistance profiling of *Salmonella* isolated from irrigation water, packing shed equipment, and fresh produce in Texas. *J.Food Prot.* 68, 70–79. doi: 10.4315/0362-028X-68.1.70

Dufour, A., and Schaub, S. (2007). "The evolution of water quality criteria in the United States, 1922-2003," in *Statistical Framework for Recreational Water Quality Criteria and Monitoring*, ed J. Larry Wymer (Ames, IA: John Wiley and Sons, Ltd.), 1–12. doi: 10.1002/9780470518328.ch1

Harmel, D., Wagner, K., Martin, E., Smith, D., Wanjugi, P., Gentry, T., Gregory, L., and and, T., Hendon (2016). Effects of field storage method on *E. coli* concentrations measured in storm water runoff. *Environ. Monit. Assess.* 188, 170. doi: 10.1007/s10661-016-5183-9

Harris, L. J., and Rock, C. (2017). *Surface Water MWQP Calculator. Western Center for Food Safety, Version 6.0*. Available online at: https://ucfoodsafety.ucdavis.edu/pre-post-harvest/produce-preharvest/agricultural-water (accessed July 14, 2021).

Izumi, H., Tsukada, Y., Poubol, J., and Hisa, K. (2008). On-farm sources of microbial contamination of persimmon fruit in Japan. *J.Food Prot.* 71, 52–59. doi: 10.4315/0362-028X-71.1.52

Jones, L. A., Worobo, R. W., and Smart, C. D. (2014). Plant-pathogenic oomycetes, *Escherichia coli* strains, and Salmonella spp. frequently found in surface water used for irrigation of fruit and vegetable crops in New York State. *Appl. Environ. Microbiol.* 80, 4814–4820. doi: 10.1128/AEM.01012-14

Kaufman, P. R., Handy, C. R., McLaughlin, E. W., Park, K., and Green, G. M. (2000). *Understanding the Dynamics of Produce Markets: Consumption and Consolidation Grow*. United States Department of Agriculture, Economic Research Service Agriculture Information, Bulletin No. 758. Available online at: https://www.ers.usda.gov/webdocs/publications/42294/32086_aib758_002.pdf?v$=$6930.2 (accessed June 11, 2021).

Kinzelman, J. L., Singh, A., Ng, C., Pond, K. R., Bagley, R. C., and Gradus, S. (2005). Use of IDEXX Colilert-18® and Quanti-Tray/2000 as a rapid and simple enumeration method for the implementation of recreational water monitoring and notification programs. *Lake Reserv. Manag.* 21, 73–77. doi: 10.1080/07438140509354414

Oron, G., Gillerman, L., Lael, A., Manor, Y., Braude, E., and Bick, A. (2010). Minimizing health risks during secondary effluent application via subsurface drip irrigation. *Water Sci. Technol.* 62, 2330–2337. doi: 10.2166/wst.2010.502

Pachepsky, Y., Shelton, D. R., McLain, J. E. T., Patel, J., Mandrell, R., et al. (2011). Irrigation waters as a source of pathogenic microorganisms in produce: a review. *Adv. Agron*. 113, 76–141. doi: 10.1016/B978-0-12-386473-4.00002-6

Pope, M. L., Bussen, M., Feige, M. A., Shadix, L., Gonder, S., Rodgers, C., et al. (2003). Assessment of the effects of holding time and temperature on *Escherichia coli* densities in surface water samples. *Appl. Environ. Microbiol*. 69, 6201–6207. doi: 10.1128/AEM.69.10.6201-6207.2003

Rock, C., and Harris, L. J. (2017). *FSMA Produce Safety Rule OnlineCalculator (English and Spanish) University of Arizona Cooperative Extension*. Avaialble online at: https://agwater.arizona.edu/onlinecalc/default.aspx (accessed July 14, 2021).

Sadovski, A. Y., Fattal, B., Goldberg, D., Katzenelson, E., and Shuval, H. I. (1978). High levels of microbial contamination of vegetables irrigated with wastewater by the drip method. *Appl. Environ. Microbiol*. 36, 824–830. doi: 10.1128/aem.36.6.824-830.1978

Seiler, R. L., and Skorupa, J. P. (2001). *National Irrigation Water Quality Program Data-Synthesis Data Base 00-513*. Avaialble online at: http://pubs.usgs.gov/of/2000/ofr00513/ofr00513.pdf (accessed June 11, 2021).

Selma, M. V., Allende, A., Lopez-Galvez, F., Elizaquivel, P., Aznar, R., and Gill, M. I. (2007). Potential microbial risk factors related to soil amendments and irrigation water of potato crops. *J. Appl. Microbiol*. 103, 2542–2549. doi: 10.1111/j.1365-2672.2007.03504.x

Selvakumar, A., Borst, M., Boner, M., and Mallon, P. (2004). Effects of sample holding time on concentrations of microorganisms in water samples. *Water Environ. Res.* 76, 67–72 doi: 10.2175/106143004X141591

Steele, M., and Odumeru, J. (2004). Irrigation water as source of foodborne pathogens on fruit and vegetables. *J.Food Prot*. 67, 2839–2849. doi: 10.4315/0362-028X-67.12.2839

Stoeckel, D. M., Pahl, D. M., Wall, G. L., Woods, K. L., and Bihn, E. A. (2017). *The Water Analysis Method Requirement in the FSMA Produce Safety Rule*. Produce Safety Alliance, Cornell University.

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Grohn, Y. T., Y. T., et al. (2013). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol.* 79, 588–600. doi: 10.1128/AEM.02491-12

Suslow, T. (2010). *Standards for Irrigation and Foliar Contact Water*. Available online at: https://www.pewtrusts.org/$\sim$/media/Assets/2009/PSPWaterSuslow1pdf.pdf (accessed October 04, 2021).

Suslow, T. V., Oria, M. P., Beuchat, L. R., Garrett, E. H., Parish, M. E., Harris, L. J., et al. (2003). Production practices as risk factors in microbial food safety of fresh and fresh-cut produce. *Compr. Rev. Food Sci. Food Saf.* 2(Suppl s1), 38–77. doi: 10.1111/j.1541-4337.2003.tb00030.x

U.S. Food and Drug Administration (2015). *Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption*. Available online at: https://www.regulations.gov/document/FDA-2011-N-0921-18558 (accessed October 04, 2021).

United States Environmental Protection Agency (1992). *Specifications and Guidance for Contaminant-Free Sample Containers. EPA 540-R-93-051. Publication 9240.0-05A*. Available online at: https://nepis.epa.gov/Exe/ZyPDF.cgi/2001266X.PDF?Dockey=2001266X.PDF (accessed August 19, 2021).

United States Environmental Protection Agency (2006). *Assessment of the Effects of Holding Time and Enterocci Concentrations in Fresh and Marine Recreational Waters and Escherichia coli Concentrations in Fresh Recrational Waters*. EPA-821-R-06-019. Washington, DC. Available online at: ///Users/eab38/Downloads/EPA-HQ-OW-2010-0192-0035_content.pdf (accessed September 15, 2021).

United States Environmental Protection Agency (2012). *Recreational Water Quality Criteria. EPA 820-F-12-058*. Available online at: https://www.epa.gov/wqc/2012-recreational-water-quality-criteria-documents (accessed August 22, 2021).

United States Environmental Protection Agency (USEPA) (1986). *Ambient Water Quality Criteria for Bacteria*. EPA 440/5-84-002. Available online at: https://www.epa.gov/sites/production/files/2019-03/documents/ambient-wqc-bacteria-1986.pdf (accessed June 11, 2021).

United States Food and Drug Administration (2015). *Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption*.

United States Food and Drug Administration (2019). *Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption; Extension of Compliance Dates for Subpart Federal Register E*. 84, 9706–9714.

Weller, D. L., Belias, A., Green, H., Wiedmann, M., and Roof, S. (2019). Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Front. Sustain. Food Syst*. 3:124. doi: 10.3389/fsufs.2019.00124

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

BRIEF RESEARCH REPORT
published: 03 December 2021
doi: 10.3389/frwa.2021.730124

Preliminary Investigation of Microbial Community in Wastewater and Surface Waters in Sri Lanka and the Philippines

Andre Freire Cruz[1]*, *R. G. S. Wijesekara*[2], *K. B. S. N. Jinadasa*[3], *Benjamin J. Gonzales*[4], *Takeshi Ohura*[5] and *Keerthi S. Guruge*[6,7,8]*

[1] Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan, [2] Department of Aquaculture and Fisheries, Faculty of Livestock, Fisheries and Nutrition, Wayamba University of Sri Lanka, Makandura, Sri Lanka, [3] Department of Civil Engineering, University of Peradeniya, Peradeniya, Sri Lanka, [4] College of Fisheries and Aquatic Sciences, Western Philippines University Puerto Princesa Campus, Puerto Princesa, Philippines, [5] Graduate School of Agriculture, Meijo University, Nagoya, Japan, [6] National Institute of Animal Health, National Agriculture and Food Research Organization, Ibaraki, Japan, [7] Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Osaka, Japan, [8] National Institute of Fundamental Studies, Kandy, Sri Lanka

OPEN ACCESS

Edited by:
Jingqiu Liao, Columbia University, United States

Reviewed by:
Aeriel Belk, Colorado State University, United States
Jiazhang Chen, Chinese Academy of Fishery Sciences, China

***Correspondence:**
Andre Freire Cruz
andre@kpu.ac.jp
Keerthi S. Guruge
guruge@affrc.go.jp

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *24 June 2021*
Accepted: *26 October 2021*
Published: *03 December 2021*

Citation:
Cruz AF, Wijesekara RGS, Jinadasa KBSN, Gonzales BJ, Ohura T and Guruge KS (2021) Preliminary Investigation of Microbial Community in Wastewater and Surface Waters in Sri Lanka and the Philippines. Front. Water 3:730124. doi: 10.3389/frwa.2021.730124

In this study, the composition and richness of bacterial communities in treated and untreated wastewater from hospitals, commercial, and non-commercial fish farming sites, sewage effluents, and surface waters, which included seawater and fresh water in Sri Lanka and the Philippines, were investigated through 16S rRNA gene amplicon sequence analysis. *Firmicutes* were found predominantly in Sri Lankan hospital wastewaters, while *Cyanobacteria* and *Acidobacteria* were typically detected in fish culture sites and the waste canal in Sri Lanka, respectively. The Shannon–Weaver index (SW) and number of Operational Taxonomic Units (OTUs) were higher in the Philippines than in Sri Lanka. The bacterial richness in the university non-commercial fish pond and sewage effluent displayed greater than that in hospital wastewaters. In addition, the bacterial richness was higher in the untreated wastewater compared to that in the treated wastewater in hospitals. These results indicate the differences among water types in terms of bacterial community, especially influenced by their source.

Keywords: bacterial community, hospital wastewater, fish culture sites, surface water, 16S rRNA (16S rDNA)

INTRODUCTION

Water is one of the most important and essential resources for living organisms in the world. However, uncontaminated water is a basic necessity for humans. Contamination of water resources due to anthropogenic activities is common throughout the world (Khatri and Tyagi, 2015), particularly by chemicals and microorganisms. Integrating knowledge from multiple fields such as hydrology, microbiology, and ecology would increase the understanding of pollution levels and potential causes of pollution (Pandey et al., 2014). Surface, groundwater, and chlorinated urban water in some cities are contaminated with bacteria levels regarded as unsafe as per the standards for potable water (Onyango et al., 2018). There is a growing need to develop a strategy for recognizing potential emerging waterborne pathogens. An understanding of disinfectant action and microbial resistance to treatment processes is required to better identify those pathogens likely

to be of greatest concern (Ashbolt, 2015). Likewise, it is important to have a more relevant and faster indicator for microbial contamination detection in water (Jung et al., 2014).

In a typical tropical country like Sri Lanka, bacterial diseases are frequently linked to the consumption of drinking water contaminated with *Shigella* spp., *Salmonella* spp., and *Campylobacter* spp. A recent study revealed that water in one of the major river basins in Sri Lanka (Kelani river basin) is not suitable for direct consumption as drinking water without proper treatment (Mahagamage et al., 2016). The Kelani river water in Sri Lanka is not only contaminated with *Escherichia coli* but also other species that are resistant to more than one antibiotic. Antibiotic resistance is probably due to the significant seasonal variations and environmental changes (Kumar et al., 2020). The majority (90%) of public water sources (well-water) in the Northern Province of Sri Lanka are microbiologically unsuitable for consumption due to microbial contamination (Arulnesan et al., 2015). The coastal areas of the Ampara district of the eastern area of the country face a lot of challenges induced by water-borne diseases due to pathogenic contamination (Ameer, 2017). Recently, Guruge et al. (2021) reported a wide range of antimicrobials with high concentrations of clarithromycin, sulfamethoxazole, and sulfapyridine in the hospital wastewaters in Sri Lanka. Approximately 61% of the examined *E. coli* isolates in those samples were categorized as multidrug-resistant bacteria.

Practices of open defecation, unhygienic practices, livestock feces, and latrine sources have a significant correlation to contaminated water sources (Gwimbi et al., 2019). Pathogens that are present in the aquatic environment may cause various diseases to people by ingestion of contaminated drinking water. The concentration of selected antimicrobials, the occurrence of resistant *E. coli*, and resistance genes in hospital wastewater and adjacent surface water were previously reported (Guruge et al., 2021). However, details regarding the bacterial community in such wastewaters in Sri Lanka are not well-documented. Therefore, the main objective of this study was to obtain preliminary information on the microbial community structure in wastewaters, urban waterways, lakes, and aquaculture facilities in Sri Lanka. Several wastewaters and freshwater samples previously studied for antimicrobials were included in the present study. For comparison, few samples from similar environmental settings from the Philippines were also investigated.

MATERIALS AND METHODS

Water Samples

Several types of water samples were obtained from Sri Lanka such as untreated and treated wastewaters from three hospitals, five commercial ornamental fish culture sites, one non-commercial university fishpond, one lake, and an urban waste canal receiving treated effluents from the above hospitals. The samples from the Philippines were included with three sewage effluents and seawater from a bay that receives those effluents. Grab water samples were collected in clean 500-ml polypropylene bottles between 2018 and 2019 (Guruge et al., 2021). Details about the sampling are described in **Table 1**. Unfiltered samples were maintained in a freezer (-18°C) before the DNA extraction. A total of 17 samples were analyzed.

Amplicon Gene Analysis

A 0.25-ml aliquot was taken from each sample to extract the DNA by the extraction buffer method (Kageyama et al., 2003) and purified using the Promega PCR purification Kit (Promega Co., USA). The DNA concentration was verified by nanodrop, where the minimum of 5 ng/μl of DNA concentration was established to proceed with the amplicon PCR. The 16S rRNA gene sequences were analyzed by amplicon PCR using a primer pair V3/V4 (Klindworth et al., 2013). The forward primer sequence was V3 (5′-CCTACGGGNGGCWGCAG-3′) and the reverse was V4 (5′-GACTACHVGGGTATCTAATCC-3′) with the overhang adapter added (Fluidigm Co., USA) followed by library construction with Fluidigm barcodes. Then, the library was sequenced at Genome Quebec, Co., Canada using Illumina MiSeq 250 bp, and the data on amplicon sequencing analysis were performed using the Qiime2-2019.10 pipeline (https://qiime2.org/) (Bolyen et al., 2018). This was used to assess the relative abundance of bacterial phyla, richness, and principal coordinate analysis (PCoA) of water samples. For the metataxonomic classification, bacterial DNA was assessed using the classifier gg-13-8-99-515-806-nb-classifier.qza from the Green genes database, whose assignment was carried out by the Basic Local Alignment Search Tool (BLAST). All the metagenome sequences were registered in the DNA Data Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp), accession number DRA012530.

Statistical Analysis

One-way ANOSIM (analysis of similarities based on number of OTUs) was performed based on non-parametrical tests, to evaluate the effects of treatments and location on bacterial community structure in the PCoA. The correlation between the concentration of antimicrobials and the prevalence of the bacterial community was determined by Pearson analysis. The antimicrobial concentrations in the samples from Sri Lanka (SL-1 to SL-8) were retrieved from our previous study (Guruge et al., 2021).

RESULTS AND DISCUSSION

Bacterial Community Analysis

The total number of sequences obtained from all water samples was 2,391,573 with 3,466 features and 780,768 of total frequency, and the overall quality trim length was standardized at 230 bp, using the training process of SKLEARN based on k-mers, value 7, as it is the default balanced QIIME 2 parameter. The results among the bacterial community indicated on a scale from highest to lowest that *Proteobacteria* and *Actinobacteria* were the most common phyla among the samples based on relative abundance (RA). Lower RA of *Proteobacteria* was found in the Sri Lanka lake, whereas *Actinobacteria* was detected in a wide range of sources. *Actinobacteria* showed the lowest RA in samples from the three Sri Lanka hospitals and in the Philippine waste effluents. Furthermore, *Firmicutes* were found mostly in the Sri Lanka hospital sample while *Cyanobacteria* and *Acidobacteria* were

TABLE 1 | Details of samples.

Country	Sample no	Type of water	Sampling period
Philippine	PH 1	Sewage effluent	2019 January
	PH 2	Sewage effluent	2019 January
	PH 3	Bay (sea water)	2019 January
Sri Lanka	SL 1	Urban waste canal	2018 September
	SL 2	Urban lake	2018 September
	SL 3	Hospital wastewater (before treatment)	2018 September
	SL 4	Hospital wastewater (after treatment)	2018 September
	SL 5	Hospital wastewater (before treatment)	2018 September
	SL 6	Hospital wastewater (after treatment)	2018 September
	SL 7	Hospital wastewater (before treatment)	2018 September
	SL 8	Hospital wastewater (after treatment)	2018 September
	SL 9	Commercial ornamental fish culture facility	2019 March
	SL 10	Commercial ornamental fish culture facility	2019 March
	SL 11	Commercial ornamental fish culture facility	2019 March
	SL 12	Commercial ornamental fish culture facility	2019 March
	SL 13	Commercial ornamental fish culture facility	2019 March
	SL 14	Non-commercial university fish culture facility	2019 March

typically present in the Sri Lanka fish farming sites and waste canals, respectively. *Chloroflexi* was found in almost all sources except for the hospitals. *Planctomycetes* were strongly detected in the Sri Lanka lake sample as well (**Figure 1**).

Proteobacteria, *Actinobacteria*, and *Acidobacteria* are ubiquitous in several environments, such as soils, food, and animal feces; thus, it is expected to occur in wastewaters, too. The lakes, fish culture ponds, and hospitals in Sri Lanka have low *Proteobacteria*, and this may be due to the low biochemical oxygen demand (BOD) of these environments. In addition, the high abundance of *Cyanobacteria* in the Sri Lanka fish culture pond could indicate the strong photosynthetic activity in these waters, which could later lead to a healthy and thriving fish community. Each environment could be characterized by a "fingerprint"; i.e., specific bacteria phyla is unique to that particular location. For example, *Verrucomicrobia* was unique to hospital waste effluents in Sri Lanka and the Philippines, while *Gemmatimonadetes* was unique to the waste canal and university fish pond samples in Sri Lanka. The abundance of the *Methylotenera* genus in wastewater is usually associated with the presence of methanol and nitrate from industrial sources (Kalyuhznaya et al., 2009).

Nitrospirae and *Chloroflexi* are the main bacteria indicators for nitrite oxidation and denitrification (Qin et al., 2018). The increase in salt concentration of water bodies may lead to lower bacterial richness because of the selection pressure of this element. A similar effect may be attributed to heavy metals such as mercury, arsenic, and cadmium in soil and marine sediments (Li et al., 2017; Ou et al., 2018). The differences in RA could be clearly noted on *Alphaproteobacteria* and *Betaproteobacteria* phyla. These phyla could accumulate polyhydroxyalkanoate, in wastewater, which could have a strong effect on niche speciation (Oshiki et al., 2013). This chemical has been detected in other environments as well such as compost, river biofilm, and freshwater; thus, similar effect patterns could be extended to other environments, including those in the current research. Hospital wastewater could have distinguished bacterial species more as compared to domestic ones (Ahn and Choi, 2016), with particular regard to *Bacterioidetes* and *Proteobacteria*.

Usually, the source of wastewater determines the bacterial community, and the presence of antibiotics can have a strong influence in this regard (Guruge et al., 2021). In the case of hospital samples from Sri Lanka, the orders *Bifidobacteriales* and *Coriobacteriales* had a positive correlation with chloramphenicol. Additionally, *Lactobacillales*, *Clortridiales*, and *Victivalles* had correlations with the fluoroquinolone antibiotics such as norfloxacin, levofloxacin, and ciprofloxacin (**Table 2**). This could suggest the strong effect of these bacterial groups on water chemical composition. The bacterial community could be distinct in the influent and effluent of wastewater treatment systems. Some harmful ones may be reduced during treatment, but treatment might still release other highly resistant pathogenic bacteria into the environment (Numberger et al., 2019). Moreover, other species such as *Cyanobacteria* are controlled by pH and nitrate (Wei et al., 2014) while *Microcystis* is controlled by temperature (Ji et al., 2018). Therefore, monitoring the nutrient input ccould provide a way to control the bacterial community in lakes (Zhang et al., 2019). Such factors could explain the highest abundance of *Cyanobacteria* in lakes, fish cultures, and bays as compared to hospital wastes.

Previous studies have shown the influence of environmental parameters such as pH, temperature, salinity, and nutrient status on the microbial community in these waters (Swan et al., 2010; Ganzert et al., 2014). The bacterial community could be responsive to environmental changes in Lake Chaohu, China, especially with regard to the differences between water and sediments (Zhang et al., 2019). Microorganisms have potential roles in nutrient biodynamics, pollutant degradation,

TABLE 2 | Correlations between relative abundance of bacteria (Order level) and concentrations (ng/L) of antimicrobials[a] in SL1–SL 8 samples.

Antimicrobials	Bacteria										
	Bifidobacteriales	Coriobacteriales	Stramenopiles	Elusimicrobiales	Lactobacillales	Clostridiales	Victivallales	Rhizobiales	Sphingomonadales	Campylobacterales	Verrucomicrobiales
Sulfapyridine (SPR)[a]	n.s.	n.s.	n.s.	n.s.	0.824*	0.963**	0.868**	n.s.	n.s.	0.886**	n.s.
Sulfamethazine (SMT)	n.s.	n.s.	n.s.	0.826*	n.s.	n.s.	n.s.	n.s.	n.s.	0.583	n.s.
Sulfamethoxazole (SMXZ)	n.s.	n.s.	n.s.	0.863**	n.s.	0.824*	n.s.	n.s.	n.s.	0.979**	n.s.
Trimethoprim (TRI)	n.s.	n.s.	n.s.	n.s.	n.s.	0.893**	0.776*	n.s.	n.s.	0.908**	n.s.
Clarithromycin (CLA)	n.s.	n.s.	n.s.	n.s.	0.888**	0.777*	0.748*	n.s.	n.s.	n.s.	n.s.
Roxithromycin (ROX)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.909**	n.s.	n.s.	0.899**
Chloramphenicol (CHLP)	0.978**	0.914**	−0.201	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Norfloxacin (NORF)	n.s.	n.s.	n.s.	n.s.	0.762*	0.966**	0.870**	n.s.	n.s.	0.931**	n.s.
Levofloxacin (LEV)	n.s.	n.s.	n.s.	n.s.	0.828*	0.979**	0.911**	n.s.	n.s.	0.885**	n.s.
Ciprofloxacin (CIP)	n.s.	n.s.	n.s.	n.s.	0.859**	0.966**	0.888**	n.s.	n.s.	0.849**	n.s.
Triclocarban (TCC)	n.s.	n.s.	−0.792*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Triclosan (TCS)	n.s.	0.720*	−0.728*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Ethyl paraben (EtP)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.857**
Butyl paraben (BuP)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.738*
N,N-diethyl-3-toluamide (DEET)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.784*	n.s.	n.s.
Total antimicrobial concentration	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.808*	n.s.	n.s.	0.846**	n.s.

**Correlation is significant at the 0.05 level (2-tailed).*

***Correlation is significant at the 0.01 level (2-tailed).*

n.s., Non-significant.

[a]*Antimicrobial data were retrieved from Guruge et al. (2021).*

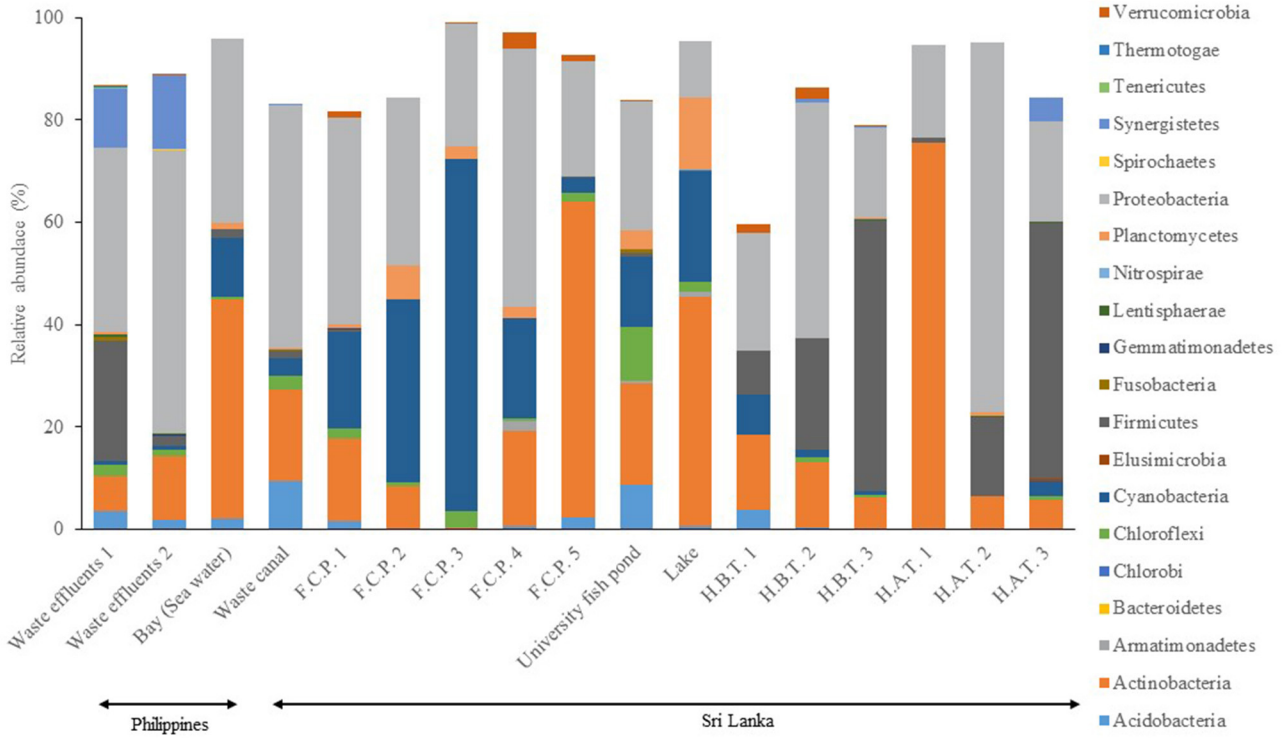


FIGURE 1 | Distribution (relative abundance) of bacterial phylum in water samples. FCP, Commercial Fish Culture Pond; HAT, Hospital Wastewater (After Treatment); HBT, Hospital Wastewater (Before Treatment). The information of sample type is available in **Table 1**.

and transformation in organic matter; hence, they can be used as indicators of watershed quality in the ecosystems (Wei et al., 2008; Chen et al., 2018).

Richness Estimation

The diversity estimation was according to the Shannon–Weaver index (SW), which compared two countries and types of water. The SW was greater in the Philippines than in Sri Lanka (**Figure 2**), whereas such values were found higher in the University fish pond and Waste Effluents and lower at Hospital 1 After treatment and Fish culture 1. In general, the SW of water from the hospital is greater than from fish farming sites; additionally, within the hospital samples, those before treatment have greater SW. However, some variations between these two treatments were apparent. The order of SW from highest to lowest follows the sequence: University fishpond > Waste effluents > H3AT > H3BT > Waste canal > Lake > H2BT > FC4 > FC1 > H2BT > Bay > H1BT > FC5 > FC2 > FC3 = H1AT (AT—After Treatment; BT—Before Treatment; FC—Fish Culture) (**Figure 2**). A similar pattern for the number of Operational Taxonomic Units (OTUs), standardized at 6,000 sequences, was found with a greater number in the Philippines than in Sri Lanka (**Figure 2**). In addition, the comparison between the types of water indicated the highest number of OTUs in fish ponds and waste effluents and lowest in H1AT and FC2. The number of OTUs followed this sequence: University fish pond > Waste effluents > H3BT > Waste canal > H3AT > Lake > Bay > FC1 = FC3 > H2BT > FC5 > FC4 > H2AT > H1BT > H1AT = FC2 (**Figure 2**). The PCoA graph clearly indicated a significant separation between Philippine and Sri Lanka samples. The fish culture samples could also be distinguished from the three samples from hospitals (**Figure 3**). Such differences were confirmed by the ANOSIM data.

The richness data indicated some low variation among environments (hospital, fish culture). The higher diversity in fish farming sites as compared to hospital wastewater could be explained in terms of the contribution of the fish community in maintaining the local microflora, although the biochemical oxygen demand might be higher in these areas. Fish culture demands a significant oxygen requirement, but they could improve the bacterial diversity in these environments.

Normally, the industrial wastewater exhibits lower bacterial richness than other sources, with particular concern to *Nitrospira* populations. This suggests the effect of these influents on nitrification and denitrification (Yang et al., 2020).

Bacterial richness is considered a crucial factor to monitor the quality of water since the diversity is thought to improve the defense and recovery ability of the ecosystem against disturbances (Zinger et al., 2012). Hospital wastewaters are full of antibiotic-resistant bacteria as compared to others, suggesting that the presence of pharmaceuticals could affect and determine the bacterial community in these waters (Akiba et al., 2015; Guruge et al., 2021). Wastewater can affect the sediment bacterial community by the nutrient and organic loading they carry,

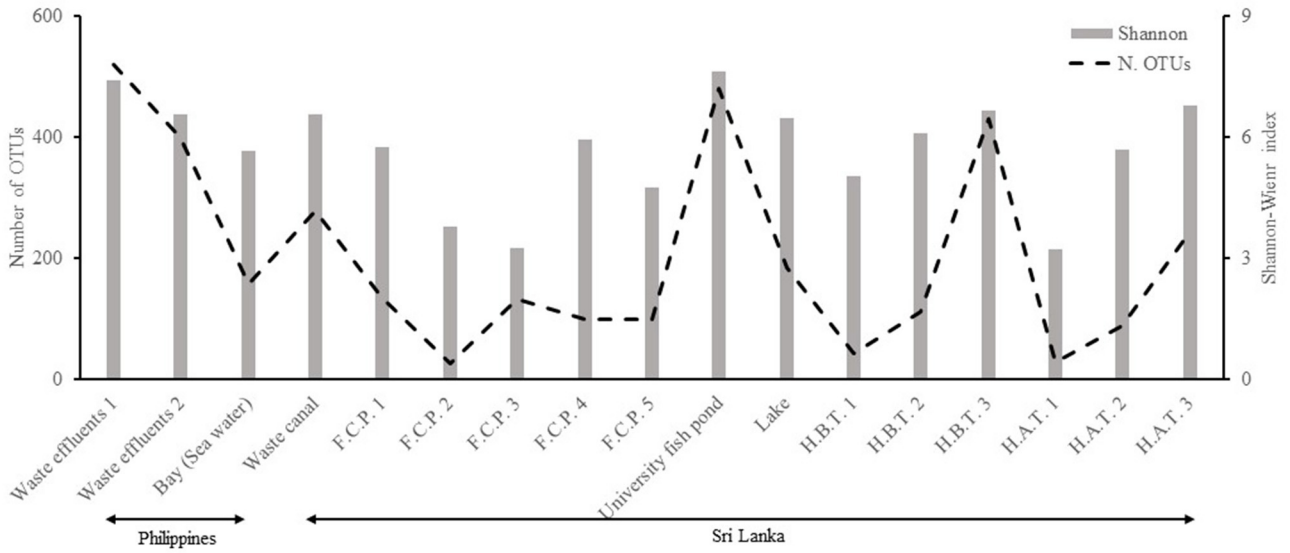


FIGURE 2 | Bacterial richness (Shannon–Weaver index and Observed number of operational taxonomic units—OTUs) of water samples. FCP, Commercial Fish Culture Pond; HAT, Hospital Wastewater (After Treatment); HBT, Hospital Wastewater (Before Treatment). The information of sample type is available in **Table 1**.

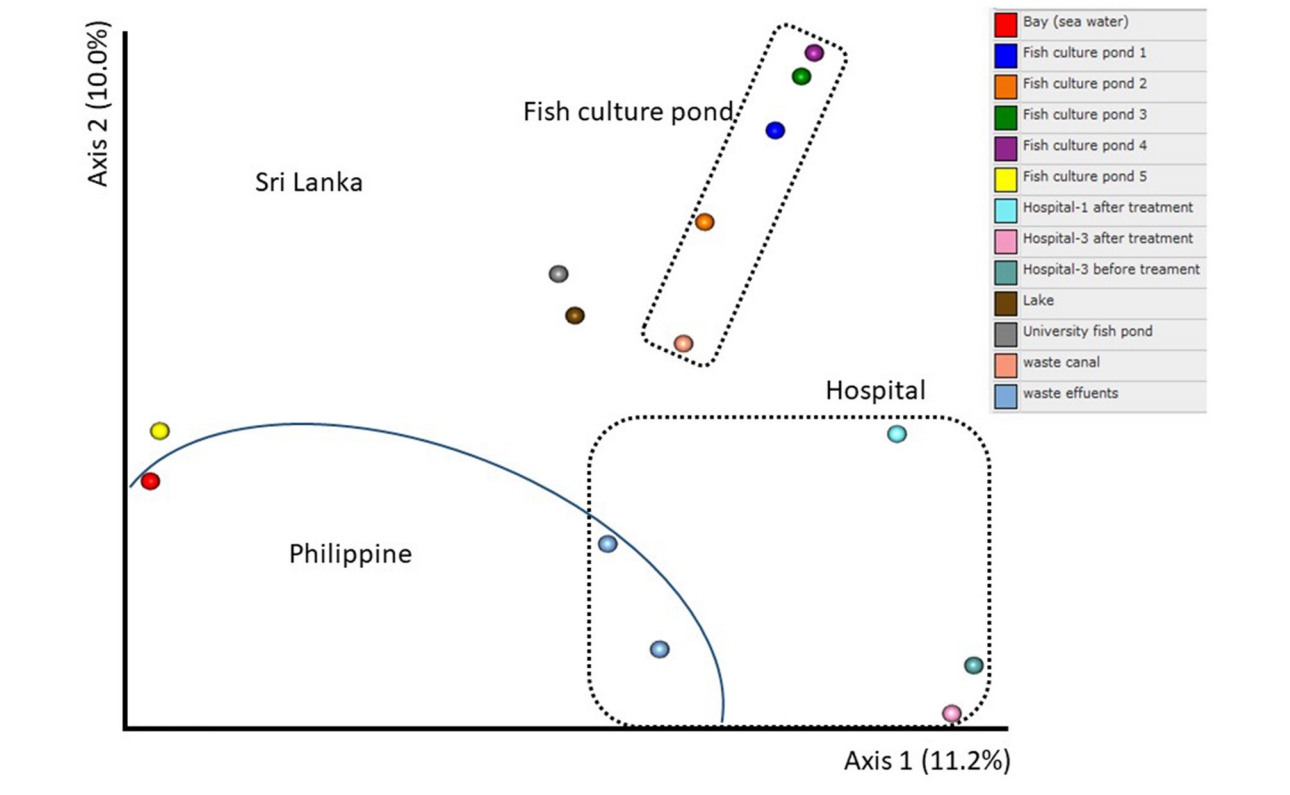


FIGURE 3 | Principal coordinate analysis (PCoA) of 16S rRNA genes in water samples according to the Bray–Curtis method. The information of sample type is available in **Table 1**.

consequently altering the local habitat atmosphere (Saarenheimo et al., 2017) by driving the increase in carbon availability (Garnier et al., 1992). The characterization of the bacterial community in water as a response to change the environment would provide a valuable assessment of the aquatic microbial ecology and their risks (Wang et al., 2016). The distribution and composition of the

bacterial community in water were greatly affected by time and space with special regard to seasonal effects (Zhang et al., 2019).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors. And I detected the following words and expressions: Global list: DNA Data Bank of Japan, Illumina, 16s rRNA, Library construction.

AUTHOR CONTRIBUTIONS

AC: data analysis. RW: text review. KJ: data supplier. BG: literature review. TO: data discussion and conclusions. KG: experimental design. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially supported by the Grant-in-Aid for Scientific Research (B) [Grant no. 17H01865].

REFERENCES

Ahn, Y., and Choi, J. (2016). Bacterial communities and antibiotic resistance communities in a full-scale hospital wastewater treatment plant by high-throughput pyrosequencing. *Water* 8:580. doi: 10.3390/w8120580

Akiba, M., Senba, H., Otagiri, H., Prabhasankar, V. P., Taniyasu, S., Yamashita, N., et al. (2015). Impact of wastewater from different sources on the prevalence of antimicrobial-resistant *Escherichia coli* in sewage treatment plants in South India. *Ecotoxicol. Environ. Saf.* 115, 203–208. doi: 10.1016/j.ecoenv.2015.02.018

Ameer, M. L. (2017). Water-Borne diseases and the their challenges in the coastal of Ampara District in Sri Lanka. *World News Nat. Sci.* 9, 7–18.

Arulnesan, C. A., Kamsan, S., Rajeshkannan, N., Sivapalan, K., and Muruganthan, K. (2015). *Microbial Contamination of Public Drinking Water Sources and Associated Factors in Kopay Medical Officer of Health area Jaffna*. Sri Lanka.

Ashbolt, N. J. (2015). Microbial contamination of drinking water and human health from community water systems. *Curr. Environ. Heal. Rep.* 2, 95–106. doi: 10.1007/s40572-014-0037-5

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., et al. (2018). *QIIME 2: Reproducible, Interactive, Scalable, and Extensible Microbiome Data Science*, PeerJ Preprints.

Chen, J., Wang, P., Wang, C., Wang, X., Miao, L., Liu, S., et al. (2018). Bacterial communities in riparian sediments: a large-scale longitudinal distribution pattern and response to dam construction. *Front. Microbiol.* 9:999. doi: 10.3389/fmicb.2018.00999

Ganzert, L., Bajerski, F., and Wagner, D. (2014). Bacterial community composition and diversity of five different permafrost-affected soils of Northeast Greenland. *FEMS Microbiol. Ecol.* 89, 426–441. doi: 10.1111/1574-6941.12352

Garnier, J., Servais, P., and Billen, G. (1992). Bacterioplankton in the Seine River (France): impact of the Parisian urban effluent. *Can. J. Microbiol.* 38, 56–64. doi: 10.1139/m92-009

Guruge, K. S., Tamamura, Y. A., Goswami, P., Tanoue, R., Jinadasa, K. B. S. N., Nomiyama, K., et al. (2021). The association between antimicrobials and the antimicrobial-resistant phenotypes and resistance genes of *Escherichia coli* isolated from hospital wastewaters and adjacent surface waters in Sri Lanka. *Chemosp.* 279:130591. doi: 10.1016/j.chemosphere.2021.130591

Gwimbi, P., George, M., and Ramphalile, M. (2019). Bacterial contamination of drinking water sources in rural villages of Mohale Basin, Lesotho: exposures through neighbourhood sanitation and hygiene practices. *Environ. Health Prev. Med.* 24:33. doi: 10.1186/s12199-019-0790-z

Ji, B., Qin, H., Guo, S., Chen, W., Zhang, X., and Liang, J. (2018). Bacterial communities of four adjacent fresh lakes at different trophic status. *Ecotoxicol. Environ. Saf.* 157, 388–394. doi: 10.1016/j.ecoenv.2018.03.086

Jung, A.-V., Le Cann, P., Roig, B., Thomas, O., Baurès, E., and Thomas, M. F. (2014). Microbial contamination detection in water resources: interest of current optical methods, trends and needs in the context of climate change. *Int. J. Environ. Res. Public Health* 11, 4292–4310. doi: 10.3390/ijerph110404292

Kageyama, K., Komatsu, T., and Suga, H. (2003). Refined PCR protocol for detection of plant pathogens in soil. *J. Gen. Plant Pathol.* 69, 153–160. doi: 10.1007/s10327-002-0037-4

Kalyuhznaya, M. G., Martens-Habbena, W., Wang, T., Hackett, M., Stolyar, S. M., Stahl, D. A., et al. (2009). Methylophilaceae link methanol oxidation to denitrification in freshwater lake sediment as suggested by stable isotope probing and pure culture analysis. *Environ. Microbiol. Rep.* 1, 385–392. doi: 10.1111/j.1758-2229.2009.00046.x

Khatri, N., and Tyagi, S. (2015). Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Front. Life Sci.* 8, 23–39. doi: 10.1080/21553769.2014.933716

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808

Kumar, M., Chaminda, G. G. T., and Honda, R. (2020). Seasonality impels the antibiotic resistance in Kelani River of the emerging economy of Sri Lanka. *NPJ Clean Water* 3:12. doi: 10.1038/s41545-020-0058-6

Li, X., Meng, D., Li, J., Yin, H., Liu, H., Liu, X., et al. (2017). Response of soil microbial communities and microbial interactions to long-term heavy metal contamination. *Environ. Pollut.* 231, 908–917. doi: 10.1016/j.envpol.2017.08.057

Mahagamage, M., Pathirage, M., and Manage, P. M. (2016). *Microbiological Quality in Ground Water of the Kelani River Basin*, Sri Lanka.

Numberger, D., Ganzert, L., Zoccarato, L., Mühldorfer, K., Sauer, S., Grossart, H. P., et al. (2019). Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. *Sci. Rep.* 9:9673. doi: 10.1038/s41598-019-46015-z

Onyango, A. E., Okoth, M. W., Kunyanga, C. N., and Aliwa, B. O. (2018). Microbiological quality and contamination level of water sources in Isiolo County in Kenya. *J. Environ. Public Health* 2018:2139867. doi: 10.1155/2018/2139867

Oshiki, M., Onuki, M., Satoh, H., and Mino, T. (2013). Microbial community composition of polyhydroxyalkanoate-accumulating organisms in full-scale wastewater treatment plants operated in fully aerobic mode. *Microbes Environ.* 28, 96–104. doi: 10.1264/jsme2.ME12141

Ou, D., Li, H., Li, W., Wu, X., and Wang, Y. (2018). Salt-tolerance aerobic granular sludge: formation and microbial community characteristics. *Bioresour. Technol.* 249, 132–138. doi: 10.1016/j.biortech.2017.07.154

Pandey, P. K., Kass, P. H., Soupir, M. L., Biswas, S., and Singh, V. P. (2014). Contamination of water resources by pathogenic bacteria. *AMB Express* 4:51. doi: 10.1186/s13568-014-0051-x

Qin, H., Ji, B., Zhang, S., and Kong, Z. (2018). Study on the bacterial and archaeal community structure and diversity of activated sludge from three wastewater treatment plants. *Mar. Pollut. Bull.* 135, 801–807. doi: 10.1016/j.marpolbul.2018.08.010

Saarenheimo, J., Aalto, S. L., Rissanen, A. J., and Tiirola, M. (2017). Microbial community response on wastewater discharge in boreal lake sediments. *Front. Microbiol.* 8:750. doi: 10.3389/fmicb.2017.00750

Swan, B. K., Ehrhardt, C. J., Reifel, K. M., Moreno, L. I., and Valentine, D. L. (2010). Archaeal and bacterial communities respond differently to environmental gradients in anoxic sediments of a California hypersaline lake, the Salton Sea. *Appl. Environ. Microbiol.* 76, 757–768. doi: 10.1128/AEM.02409-09

Wang, W., Wang, H., Feng, Y., Wang, L., Xiao, X., Xi, Y., et al. (2016). Consistent responses of the microbial community structure to organic farming

along the middle and lower reaches of the Yangtze River. *Sci. Rep.* 6:35046. doi: 10.1038/srep35046

Wei, C., Bao, S., Zhu, X., and Huang, X. (2008). Spatio-temporal variations of the bacterioplankton community composition in Chaohu Lake, China. *Prog. Nat. Sci.* 18, 1115–1122. doi: 10.1016/j.pnsc.2008.04.005

Wei, G., Li, J., Wang, N., and Gao, Z. (2014). Spatial abundance and diversity of bacterioplankton in a typical stream-forming ecosystem, Huangqian reservoir, China. *J. Microbiol. Biotechnol.* 24, 1308–1318. doi: 10.4014/jmb.1403.03067

Yang, Y., Wang, L., Xiang, F., Zhao, L., and Qiao, Z. (2020). Activated sludge microbial community and treatment performance of wastewater treatment plants in industrial and municipal zones. *Int. J. Environ. Res. Public Health* 17:436. doi: 10.3390/ijerph17020436

Zhang, L., Zhao, T., Wang, Q., Li, L., Shen, T., and Gao, G. (2019). Bacterial community composition in aquatic and sediment samples with spatiotemporal dynamics in large, shallow, eutrophic Lake Chaohu, China. *J. Freshw. Ecol.* 34, 575–589. doi: 10.1080/02705060.2019.1635536

Zinger, L., Gobet, A., and Pommier, T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Mol. Ecol.* 21, 1878–1896. doi: 10.1111/j.1365-294X.2011.05362.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Water

ORIGINAL RESEARCH
published: 02 February 2022
doi: 10.3389/frwa.2021.750673

Factors Associated With *E. coli* Levels in and *Salmonella* Contamination of Agricultural Water Differed Between North and South Florida Waterways

*Claire M. Murphy[1], Laura K. Strawn[1], Travis K. Chapin[2], Rachel McEgan[2], Sweeya Gopidi[3], Loretta Friedrich[2], Lawrence D. Goodridge[4], Daniel L. Weller[5], Keith R. Schneider[3] and Michelle D. Danyluk[2,3]**

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [2] Citrus Research and Education Center, University of Florida, Lake Alfred, FL, United States, [3] Food Science and Human Nutrition Department, University of Florida, Gainesville, FL, United States, [4] Canadian Research Institute for Food Safety, University of Guelph, Guelph, ON, Canada, [5] Department of Environmental and Forest Biology, SUNY College of Environmental Sciences and Forestry, Syracuse, NY, United States

OPEN ACCESS

Edited by:
Stefanie Lutz,
Utrecht University, Netherlands

Reviewed by:
Vasiliki Syngouna,
University of Patras, Greece
Camila Rodrigues,
Auburn University, United States

***Correspondence:**
Michelle D. Danyluk
mddanyluk@ufl.edu

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *31 July 2021*
Accepted: *16 November 2021*
Published: *02 February 2022*

Citation:
Murphy CM, Strawn LK, Chapin TK, McEgan R, Gopidi S, Friedrich L, Goodridge LD, Weller DL, Schneider KR and Danyluk MD (2022) Factors Associated With E. coli Levels in and Salmonella Contamination of Agricultural Water Differed Between North and South Florida Waterways. Front. Water 3:750673. doi: 10.3389/frwa.2021.750673

The microbial quality of agricultural water is often assessed using fecal indicator bacteria (FIB) and physicochemical parameters. The presence, direction, and strength of associations between microbial and physicochemical parameters, and the presence of human pathogens in surface water vary across space (e.g., region) and time. This study was undertaken to understand these associations in two produce-growing regions in Florida, USA, and to examine the pathogen ecology in waterways used for produce production. The relationship between *Salmonella* presence, and microbial and physicochemical water quality; as well as weather and land use factors were evaluated. Water samples were collected from six sites in North Florida ($N = 72$ samples) and eight sites in South Florida ($N = 96$ samples) over 12 sampling months. Land use around each sampling site was characterized, and weather and water quality data were collected at each sampling. *Salmonella*, generic *Escherichia coli*, total coliform, and aerobic plate count bacteria populations were enumerated in each sample. Univariable and multivariable regression models were then developed to characterize associations between microbial water quality (i.e., *E. coli* levels and *Salmonella* presence), and water quality, weather, and land use factors separately for North and South Florida. The *E. coli* and total coliforms mean concentrations ($\log_{10}$ MPN/100 mL) were 1.8 ± 0.6 and $>3.0 \pm 0.4$ in North and 1.3 ± 0.6 and $>3.3 \pm 0.2$ in South Florida waterways, respectively. While *Salmonella* was detected in 23.6% (17/72) of North Florida and 28.1% (27/96) of South Florida samples, the concentration ranged between <0.48 and 1.4 $\log_{10}$ MPN/100 mL in North Florida, and <0.48 and 3.0 $\log_{10}$ MPN/100 mL in South Florida. Regression analyses showed no evidence of a correlation between either $\log_{10}$ total coliforms or *E. coli* levels, and if a sample was *Salmonella*-positive. The factors associated with

Salmonella presence and $\log_{10}$ *E. coli* levels in North Florida differed from those in South Florida; no factors retrained in multivariable regression models were the same for the North and South Florida models. The differences in associations between regions highlight the complexity of understanding pathogen ecology in freshwater environments and suggest substantial differences between intra-state regions in risk factors for *Salmonella* contamination of agricultural water.

Keywords: *Salmonella*, *E. coli*, surface water, fecal indicator bacteria, food safety, produce safety

INTRODUCTION

Agricultural water used in produce production environments has been identified as a probable route of contamination in past produce outbreaks (Greene et al., 2008; Klontz et al., 2010; Park et al., 2012; Rodrigues et al., 2020). When agricultural water comes into contact with the harvestable portion of a plant during production (e.g., during foliar irrigation, frost protection) or farm environment (e.g., through splash from contaminated soil and feces) fresh produce can become contaminated by human pathogens present in the water (Miles et al., 2009; Erickson et al., 2010; Fatica and Schneider, 2011; Ijabadeniyi et al., 2011; Atwill et al., 2015; Jeamsripong et al., 2019; Lee et al., 2019). Understanding pathogen ecology in freshwater environments used for produce production is critical for ensuring safety and assessing the risk of potential contamination events.

The Produce Safety Rule (PSR), part of the 2011 Food Safety Modernization Act (FSMA), defines requirements for pre- and post-harvest agricultural water quality, including the frequency of testing for *Escherichia coli* (as an indicator of probable fecal contamination), and criteria for *E. coli* limits. While under review at the time this manuscript was written (US Food Drug Administration, 2015), the final PSR (i) required that agricultural surface water used during produce production establish a microbial water quality profile (MWQP) using 20 samples collected over 2–4 years on a rolling basis, and (ii) that the geometric mean (GM) and statistical threshold value (STV) of *E. coli* in this sample be $\leq$126 and $\leq$410 CFUs/100 mL, respectively. Water that exceeds these requirements is to be re-tested, not used, or treated to reduce the potential contamination risk of produce (US Food Drug Administration, 2015). While the PSR relies on the use of *E. coli*, an indicator of potential fecal contamination, there is conflicting data within the scientific literature on the efficacy of *E. coli* as a fecal indicator (Ishii and Sadowsky, 2008; Jang et al., 2017), and the association between *E. coli* and the presence of food safety hazards in agricultural waters (McEgan et al., 2013; Luo et al., 2015; Topalcengiz et al., 2017; Truitt et al., 2018). Past studies have shown that meeting the PSR standard may not be indicative of the presence of food safety hazards at the time of water use.

Physicochemical water quality (e.g., turbidity, pH), weather (e.g., rainfall, relative humidity), and land-use factors (e.g., proximity to forest and wetland, elevation) are well-established in the literature as being associated with microbial water quality (Strawn et al., 2013b; Chapin et al., 2014; Weller et al., 2015, 2020a,b; Liu et al., 2018; Truitt et al., 2018; Gu et al., 2019). Multiple studies have discussed using physicochemical water quality monitoring as alternatives or supplements to *E. coli* monitoring; others have shown that models that use these environmental factors as features can accurately predict microbial water quality for agricultural waterways (Topalcengiz et al., 2017; Weller et al., 2021). Prior research (McEgan et al., 2013; Strawn et al., 2013a, 2014; Weller et al., 2015; Gu et al., 2018; Truitt et al., 2018) has also shown that microbial water quality is associated with spatial (e.g., within and between waterways, regions) and temporal (e.g., over a day, year) factors. For example, one study observed markedly different *Salmonella* prevalence rates, 9.4 and 37.5% in agricultural water (250 mL sample), collected from New York and South Florida produce farms, respectively (Strawn et al., 2014). Luo et al. (2015) found that *Salmonella* concentrations (MPN/L) in Florida and Georgia ponds were seasonally driven and were significantly correlated with temperature and rainfall. The researchers in this study also noted that generic *E. coli* levels were significantly associated with the likelihood of *Salmonella* detection (Luo et al., 2015). McEgan et al. (2013) found no consistent correlation between *Salmonella* presence or *E. coli* levels, and multiple environmental factors (e.g., water and air temperature, pH, ORP, turbidity, conductivity) when samples from multiple sites were aggregated into a single dataset for analysis. However, when the correlation was assessed separately for each site, McEgan et al. (2013) found evidence of weak correlations between microbial water quality and environmental factors, including $\log_{10}$ MPN *Salmonella*/100 mL being correlated with the air temperature at one and turbidity at two out of 18 sites. The relationship between microbial water quality and environmental factors appears complex and varies by study and over space (e.g., between waterway/site) and time (e.g., season). Additional studies on pathogen ecology in agricultural waterways are important to better characterize this variability and to understand conditions favorable to pathogen contamination of surface water. This data is key for the development of risk management strategies for agricultural water used in preharvest applications. The objectives of this study were to characterize and compare (1) the associations between microbial water quality, including pathogen presence, and environmental factors (e.g., water quality, weather, land use) in North and South Florida waterways; and (2) *Salmonella* diversity in North and South Florida waterways.

METHODS

Water Sampling

Surface water samples were collected from both North and South Florida. The surface water in North Florida was collected from rivers ($n = 5$) and a lake ($n = 1$). The surface water in South Florida was collected from canals ($n = 8$). The North Florida samples were collected from each site monthly over 12 consecutive months beginning in November 2011. The South Florida samples were collected 12 times between May 2015 and November 2016. All samples were collected as previously described by McEgan et al. (2013). Briefly, a sterile carboy (Nalgene, Rochester, NY) was fitted with 4 kg lead weights, attached to a rope, and lowered 20 cm below the water surface at least 3 m from the shore. Carboys were filled with at least 1 L of water, transported to the lab at 4°C, and processed within 24 h of collection. Sampling always occurred before solar noon.

Physicochemical Water Quality, Weather, and Adjacent Land Use Factors

At each sampling physicochemical water quality, including turbidity, water temperature, pH, conductivity, colorimetric reading, and oxidation-reduction potential (ORP), was assessed. Each physicochemical parameter was measured in triplicate, and the value averaged. Turbidity was measured using a portable colorimeter (DR/850; Hach Company, Loveland, CO, USA) according to the manufacturer's instructions. Water and air temperatures were measured with a portable temperature probe (SH66A; Cooper Instrument Corporation, Middlefield, CT, USA). The ORP and pH were measured with a portable ORP/pH meter (pH 6 Acorn series: Oakton, Vernon Hills, IL, USA). Conductivity was measured with a portable conductivity tester (HI98304 DIST1 4 EC, HANNA Instruments, Woonsocket, RI, USA).

The weather data, including air temperature, relative humidity, and rainfall, were obtained for each sampling from the Florida Automated Weather Network (https://fawn.ifas.ufl.edu/) using the closest weather station to the sampling site. To characterize the land use around each sampling site, land cover data were downloaded from the National Land Cover dataset (https://www.mrlc.gov/national-land-cover-database-nlcd-2016). The proportion of land within 122,366 and 1,098 m of each site under pasture-hay, cropland, forest-wetland, and developed (>20% impervious) cover was then calculated using the code developed by D. Weller (https://github.com/wellerd2/Calculating-land-use-land-cover-and-landscape-structure-parameters) as previously described (Liao et al., 2021). Buffer distances (122,366 and 1,098 m) were selected based on the recommendations from the Leafy Green Marketing Agreement on how far pre-harvest agricultural water sources should be from land uses that may contaminate the water with human pathogens (Table 7 Crop Land and Water Sources Adjacent Land Uses in California Leafy Greens Marketing Agreement, 2020). For example, the recommended distance from a crop land to a concentrated animal feeding operation (CAFO) with >1,000 animals is 1,200 feet (∼366 m); this buffer (366 m), and buffers 1/3 smaller (122 m) and 1/3 larger (1,098 m) were used here. The elevation was obtained from the United States Geological Survey (https://apps.nationalmap.gov/elevation/#/%23bottom) for each water collection site.

Enumeration of Aerobic Plate Count, Total Coliform, and *E. coli* Levels

The aerobic plate count (APC) (CFU/100 mL), total coliform (MPN/100 mL), and *E. coli* (MPN/100 mL) levels were enumerated in each sample as previously described (McEgan et al., 2013). Briefly, for APC, water samples were serially diluted in 0.1% peptone water (Difco, Sparks, MD), and 100 μL aliquots were spread plated in duplicate on a tryptic soy agar (TSA) (Difco, Sparks, MD, USA). The TSA plates were incubated at 35 ± 2°C for 24 h. Colonies were enumerated by hand and CFU/100 mL calculated. Colisure presence/absence snap packs (IDEXX Laboratories, Inc., Westbrook, ME, USA) were used to determine the coliform and *E. coli* most probable numbers (MPN) in a five-by-three MPN configuration (10-, 1-, and 0.1 mL dilutions). The tubes were incubated at 35 ± 2°C for 24 h. The yellow color indicated coliforms, and *E. coli* was identified by observing fluorescence using a 6-watt fluorescent, 365 nm long-wave UV lamp with bulb from IDEXX Laboratories, Inc., Westbrook, ME. The MPN/100 mL was determined from the table in Standard Methods for the Examination of Water and Wastewater, 18th ed (American Public Health Association, 1992).

Salmonella Enumeration and Characterization

The methods for the MPN estimation of *Salmonella* in each water sample were based on the US Food and Drug Administration (FDA) Bacteriological Analytical Manual (BAM) (Andrews et al., 2011). Briefly, a three-by-three MPN method using dilutions (i) 10 mL of water sample in 10 mL double-strength lactose broth, (ii) 1 mL of water sample in 9 mL single-strength lactose broth, and (iii) 0.1 mL of water sample in 9 mL single-strength lactose broth were done. The MPN tubes were incubated at 35 ± 2°C for 24 h. Selective enrichment was performed by transferring 1.0- and 0.1 mL aliquots of each tube to tetrathionate (TT) broth (Difco) and Rappaport-Vassiliadis (RV) broth (Difco), respectively. The TT and RV broths were incubated at 35 ± 2°C and 41 ± 2°C for 24 and 48 h, respectively. A 10-μL aliquot from each TT and RV broth were streaked onto xylose lysine Tergitol 4 (XLT-4) (Difco) and CHROMagar *Salmonella* Plus (DRG International, Inc., Springfield, NJ, USA), and incubated at 35 ± 2°C for 24 h. Presumptive *Salmonella* colonies were streaked on lysine iron agar slants (LIA) (Difco) and triple sugar iron agar slants (TSI) (Difco). The slants were incubated at 35 ± 2°C for 24 h.

Further confirmation was performed using PCR for the *invA* and *oriC* genes, as previously described (Malorny et al., 2003). PCR confirmed *Salmonella* isolates were preserved at −80°C in 15% glycerol. One *Salmonella* isolate per sample enrichment scheme (e.g., TT-XLT-4, RV-XLT-4) was sent to the National Veterinary Services Laboratory (Ames, Iowa, US) for serotyping.

Samples that were below the limit of detection (LOD) for *Salmonella* (LOD; <0.48 $\log_{10}$ MPN/100 mL) were considered *Salmonella*-negative. Samples yielding an MPN value, above the LOD (with an upper limit of detection of 3.3 $\log_{10}$ MPN/100 mL), were considered positive for *Salmonella* (volume of water tested was 33.3 mL).

Statistical Analysis

All analyses were performed in R version 3.3.5 (R Foundation for Statistical Computing, Vienna, Austria). Initial descriptive analysis was performed, and summary statistics were calculated separately for all continuous factors (e.g., microbial levels, weather conditions) in North and South Florida. Using the vegan package, Simpson's Index of Diversity was calculated to quantify and compare the *Salmonella* serotype diversity in the North and South Florida water samples. Multiple samples had total coliform levels above the upper LOD, and the upper LOD times 1.5 were used for the total coliform value for these samples in the regression and tree analyses.

Bayesian mixed models were implemented to characterize the differences in microbial concentration, and the presence-absence of *Salmonella* between regions, seasons, and water types. The outcomes considered are listed in **Tables 1**, **2**. Due to the number of samples below the detection limit for *Salmonella* and above the limit for total coliforms, hurdle models were implemented. Briefly, logistic regression was used to characterize the associations if a sample was *Salmonella* positive or negative, or if the total coliform levels were above or below the limit of detection. Then for those samples where *Salmonella* or total coliforms were enumerable, a separate log-linear model was fit. For models where the outcome was binary, a Bernoulli distribution with a logit link function was used. All models included a random effect of site and fixed effect of the season to account for pseudo-replication and temporal autocorrelation. Separate models were fit for each outcome with either water type (canal, lake, or river) or region (North or South Florida) as the covariates. Separate models were also used to characterize the relationship between *Salmonella* contamination and $\log_{10}$ *E. coli* levels (both as a continuous concentration variable, and as two binary variables indicating if *E. coli* levels in the sample were above or below the geometric mean and STV cut-offs prescribed in the PSR). Models were fit using the brms package for Bayesian Regression Models using 'Stan', uninformative priors, 3 chains, and thinning set to 10 (Bürkner, 2017a,b). While the number of iterations per chain was set to 5,000 (burn-in of 2,500) for most models. The maximum a posteriori (MAP) and 95% credibility interval (CI) for the effect estimates were calculated using the bayestestR package (Makowski et al., 2019a,b). The method of interpreting the MAP and 95% CI estimates is described in the footnotes for **Tables 1**, **2** as the interpretation of the probability of direction (PD), practical significance (PS), and regional of practical equivalence (ROPE) indices, which were quantified and used to determine if the (i) given factor had a substantial effect on the outcome, and (ii) if a positive or negative effect exists regardless of if that effect is negligible or non-negligible.

TABLE 1 | Results of the Bayesian mixed models used to characterize differences in *E. coli* and total coliform levels between regions, water types; REF, reference-level for categorical factors.

Outcome	Covariate	MAP[a]	95% CI[b]	PD[c]	PS[d]	ROPE[e]
$\log_{10}$ *E. coli* levels (MPN/100 mL)						
	Region (Ref = South FL)	0.50	0.12, 0.90	0.99*	0.98*	<0.01*
	Season (Ref = Fall)					
	Spring	−0.33	−0.54, −0.09	1.00*	0.98*	<0.01*
	Summer	0.04	−0.17, 0.30	0.71	0.39	0.56
	Winter	0.13	−0.11, 0.34	0.80	0.52	0.46
	Water type (Ref = Canal)					
	Lake	0.90	0.19, 1.70	0.99*	0.99*	<0.01*
	River	0.41	0.02, 0.80	0.98*	0.94*	0.04
E. coli levels below with PSR geometric mean threshold (<126 CFU/100 mL)						
	Region (Ref = South FL)	5.43	1.19, 52.52	0.99*	0.98*	<0.01*
	Season (Ref = Fall)					
	Spring	0.44	0.10, 1.48	0.91	0.88*	0.05
	Summer	1.33	0.32, 3.61	0.62	0.56	0.13
	Winter	0.78	0.19, 2.61	0.60	0.55	0.12
	Water type (Ref = Canal)					
	Lake	17.04	0.93, 1,665.34	0.97*	0.97*	0.01*
	River	3.67	0.74, 47.93	0.97*	0.96*	0.01*
E. coli levels below with PSR STV threshold (<410 CFU/100 mL)						
	Region (Ref = South FL)	4.22	0.56, 47.22	0.95*	0.95*	0.02*
	Season (Ref = Fall)					
	Spring	1.03	0.12, 0.91	0.53	0.48	0.07
	Summer	3.10	0.55, 26.92	0.90	0.88*	0.03*
	Winter	0.96	0.08, 11.09	0.52	0.48	0.08
	Water type (Ref = Canal)					
	Lake	15.52	1.89, 1,269.61	0.99*	0.99*	<0.01*
	River	1.59	0.24, 2.88	0.86	0.83*	0.05
Total coliform levels (above vs. below the upper LOD)[f]						
	Region (Ref = South FL)	0.22	0.01, 1.89	0.91	0.89*	0.03
	Season (Ref = Fall)					
	Spring	0.51	0.17, 1.59	0.90	0.86*	0.07
	Summer	2.27	0.60, 7.82	0.89	0.85*	0.07
	Winter	0.28	0.09, 0.85	0.99*	0.99*	<0.01*
	Water type (Ref = Canal)					
	Lake	0.58	0.00, 72.42	0.60	0.58	0.04
	River	0.17	0.01, 2.45	0.93	0.92*	0.03*
$\log_{10}$ total coliform levels (MPN/100 mL)[f]						
	Region (Ref = South FL)	−0.41	−0.61, −0.13	1.00*	0.98*	<0.01*
	Season (Ref = Fall)					
	Spring	−0.08	−0.35, 0.22	0.66	0.38	0.50
	Summer	0.01	−0.32, 0.36	0.51	0.29	0.45
	Winter	0.19	−0.06, 0.49	0.92	0.75*	0.24

(Continued)

TABLE 1 | Continued

Outcome	Covariate	MAP[a]	95% CI[b]	PD[c]	PS[d]	ROPE[e]
	Water type (Ref = Canal)					
	Lake	−0.51	−1.07, 0.03	0.97*	0.94*	0.04
	River	−0.38	−0.63, −0.09	1.00*	0.97*	<0.01*

[a]*MAP, Maximum a posteriori estimate, or the mode of the posterior distribution for the effect estimate of the current parameter. It can be interpreted as the unit change in a continuous outcome associated with a one-unit change in a continuous explanatory factor or changing from the reference level to a non-reference level for a categorical explanatory factor. For binary outcomes, the effect should be interpreted as an odds ratio (e.g., odds of total coliform levels being above the upper limit of detection).*

[b]*95% Credibility Interval (CI), which indicates the central portion of the posterior distribution of possible effect estimates and is interpreted as: "Given the observed data, the effect estimate has a 95% probability of falling between x and y" (Makowski et al., 2019b). As such, having a 95% CI that includes 0 (for continuous outcomes) or 1 (for binary outcomes) is not necessarily indicative of the absence of an association; the 95% CI should be used in conjunction with the probability of direction (PD), practical significance (PS), and regional of practical equivalence (ROPE) to make that determination.*

[c]*Probability of direction, an index of if a positive or negative effect exists regardless of if that effect is negligible or non-negligible. The PD correlates strongly with frequentist p-values with PD values near 1.0 indicating greater certainty that the effect of the factor is truly positive or negative (i.e., indicates confidence in the direction of the association (Makowski et al., 2019a,b). Specifically, PD values of 0.95, 0.975, 0.995, and 0.9995 correspond to two-sided frequentist p-values of 0.10, 0.05, 0.01, and 0.001, respectively (Makowski et al., 2019a,b). Values above 0.95 are marked with*.*

[d]*PS, practical significance, which indicates the probability that the parameter's effect is above a given threshold representing a negligible effect in the median's direction; this is a unidirectional equivalence test that indicates if the effect is both non-negligible and in a given direction (Makowski et al., 2019a,b). Values should be larger than 0.5 to indicate practical significance; a cut-off of 0.75 was used here to be conservative and is marked with.*

[e]*ROPE, regional of practical equivalence, which indicates if the parameter is outside of a range of practically negligible effect (i.e., it indicates the magnitude of effect), and is calculated by determining the percent overlap between the 95% credibility interval and the range of practically no effect. The closer the ROPE percentage is to 0, the more confident we can be that the given factor has a substantial effect on FIB levels or the probability of FST detection. Specifically, we use the following cutoffs for ROPE interpretation: >99% negligible effect, >97.5% probably negligible effect, between 2.5 and 97.5% uncertain effect, <2.5% non-negligible effect, <1% significant effect (Makowski et al., 2019b). Values <2.5% are marked with*.*

[f]*Due to the large number of samples with total coliforms levels above the limit of detection (LOD), a hurdle model approach was used and the model. As such, logistic regression was fit for if the coliform concentration was above or below (below = reference-level) the upper LOD> Then for those samples below the upper LOD, a separate log_{10}-linear model was fit.*

Separately from the Bayesian regression, conditional inference trees were implemented using the partykit package and the defaults recommended by the package authors. Trees were used to determine if there were differences between regions in environmental factors associated with log_{10} *E. coli,* total coliform, *Salmonella* levels, and *Salmonella* presence-absence. For the model where the outcome was log_{10} *Salmonella* levels, only samples positive for *Salmonella* were used. For the total coliform model, coliform values for samples where coliform levels were above the upper LOD were set to 1.5*LOD. Conditional trees were used since they are robust to collinearity and correlation between explanatory factors, can handle missing data, can handle hierarchical relationships (and account for all possible interactions), and can be easily interpreted (Weller et al., 2020a,c). In interpreting the results, it is important to note that region and water type are collinear, with all South Florida

TABLE 2 | Results of the Bayesian mixed models used to characterize differences in *Salmonella* between regions and water types.

Outcome	Covariate	MAP[a]	95% CI[b]	PD[c]	PS[d]	ROPE[e]
Salmonella presence-absence[f]						
	E. coli levels in relation to PSR standard					
	Geometric mean cut-off (Ref = Below)	0.99	0.40, 2.33	0.51	0.43	0.17
	STV cut-off (Ref = Below)	2.54	0.66, 8.75	0.89	0.87*	0.06
	Log_{10} *E. coli* (MPN/100 mL)	1.12	0.65, 2.14	0.78	0.66	0.21
	Region (Ref = South FL)	0.76	0.31, 1.67	0.71	0.64	0.17
	Season (Ref = Fall)					
	Spring	1.90	0.20, 1.46	0.87	0.83*	0.09
	Summer	5.69	0.61, 3.84	0.83	0.77*	0.11
	Winter	1.70	0.20, 1.62	0.85	0.80*	0.10
	Water type (Ref = Canal)					
	Lake	1.54	0.26, 6.08	0.63	0.58	0.11
	River	0.70	0.30, 1.59	0.82	0.75*	0.14
Log_{10} *Salmonella* (MPN/100 mL)[f]						
	E. coli levels in relation to PSR standard					
	Geometric mean cut-off (Ref = Below)	−0.01	−0.39, 0.40	0.52	0.33	0.41
	STV cut-off (Ref = Below)	0.11	−0.35, 0.59	0.64	0.48	0.31
	Log_{10} *E. coli* (MPN/100 mL)	0.05	−0.17, 0.28	0.69	0.35	0.59
	Region (Ref = South FL)	−0.20	−0.58, 0.14	0.86	0.73	0.23
	Season (Ref = Fall)					
	Spring	−0.04	−0.43, 0.42	0.52	0.34	0.36
	Summer	0.31	−0.07, 0.64	0.94	0.85*	0.13
	Winter	−0.07	−0.52, 0.31	0.69	0.49	0.37
	Water type (Ref = Canal)					
	Lake	−0.06	−0.74, 0.54	0.60	0.46	0.26
	River	−0.23	−0.62, 0.24	0.88	0.75*	0.20

[a]*MAP, Maximum a posteriori estimate, or the mode of the posterior distribution for the effect estimate of the current parameter. It can be interpreted as the unit change in a continuous outcome associated with a one-unit change in a continuous explanatory factor or changing from the reference level to a non-reference level for a categorical explanatory factor. For binary outcomes, the effect should be interpreted as an odds ratio (e.g., odds of Salmonella being detected as opposed to not).*

[b]*95% CI, which indicates the central portion of the posterior distribution of possible effect estimates and is interpreted as: "Given the observed data, the effect estimate has a 95% probability of falling between x and y" (Makowski et al., 2019b). As such, having a 95% CI that includes 0 (for continuous outcomes) or 1 (for binary outcomes) is not necessarily indicative of the absence of an association; the 95% CI should be used in conjunction with PD, PS, and ROPE to make that determination.*

[c]*Probability of direction, an index of if a positive or negative effect exists regardless of if that effect is negligible or non-negligible. The PD correlates strongly with frequentist p-values with PD values near 1.0 indicating greater certainty that the effect of the factor is truly positive or negative (i.e., indicates confidence in the direction of the association (Makowski et al., 2019a,b). Specifically, PD values of 0.95, 0.975, 0.995, and 0.9995 correspond to two-sided frequentist p-values of 0.10, 0.05, 0.01, and 0.001, respectively (Makowski et al., 2019a,b). Values above 0.95 are marked with*.*

[d]*PS, practical significance, which indicates the probability that the parameter's effect is above a given threshold representing a negligible effect in the median's direction; this is a unidirectional equivalence test that indicates if the effect is both non-negligible and in a given direction (Makowski et al., 2019a,b). Values should be larger than 0.5 to indicate practical significance; a cut-off of 0.75 was used here to be conservative and is marked with.*

[e]*ROPE, regional of practical equivalence, which indicates if the parameter is outside of a range of practically negligible effect (i.e., it indicates the magnitude of effect), and is calculated by determining the percent overlap between the 95% credibility interval and the range of practically no effect. The closer the ROPE percentage is to 0, the more confident we can be that the given factor has a substantial effect on FIB levels or the probability of FST detection. Specifically, we use the following cutoffs for ROPE interpretation: >99% negligible effect, >97.5% probably negligible effect, between 2.5 and 97.5% uncertain effect, <2.5% non-negligible effect, <1% significant effect (Makowski et al., 2019b). Values <2.5% are marked with*.*

[f]*Due to the large number of Salmonella-negative samples a hurdle model approach was used. As such, Salmonella presence-absence indicates a model where the outcome is the likelihood of detecting (as opposed to not detecting, which is the reference-level) Salmonella (i.e., MAP and 95% CI should be interpreted as odds ratios). The models where the outcome was log_{10} Salmonella concentration, were built only using data from Salmonella-positive samples.*

sites being canals, and North Florida sites being either lakes or rivers.

RESULTS AND DISCUSSION

The goals of the study were to characterize and compare (1) the associations between microbial water quality, including pathogen presence, and environmental factors (e.g., water quality, weather, land use) in North and South Florida waterways; and (2) *Salmonella* diversity in North and South Florida waterways. Samples were collected at six North Florida and eight South Florida sites, each representing a separate waterway. All six sites in North Florida were natural waterways (e.g., rivers, lakes) while the eight South Florida sites were all canals. Each site was visited 12 times during the study, yielding 168 water samples (72 North and 96 South Florida samples). Physicochemical water quality, weather, and land use data for each sample and site are summarized in **Supplementary Tables 1, 2**.

Microbial Water Quality in Two Florida Growing Regions

Of the 168 samples, total coliform levels in 112 (40 from North and 72 from South) were above the upper limit of detection ($\log_{10}$ > 3.3 MPN/100 mL); no samples fell below the lower limit of detection. For the 112 counts that fell above the limit of detection, the value of 3.3 $\log_{10}$ MPN/100 mL was used. The mean and median $\log_{10}$ MPN/100 mL of total coliforms was >3.0 (Range = 1.7, >3.3) and >3.2 (IQR = >2.7, >3.3) in North Florida, and >3.3 (Range = 2.4, >3.3) and >3.3 (IQR = >3.3, >3.3) in South Florida (**Table 3**), respectively. Unlike total coliforms, no sample had *E. coli* levels below the lower limit or above the upper limit of detection. For North Florida samples the mean was 1.8 $\log_{10}$ MPN/100 mL (Range = 0.6, 3.2) and the median was 1.7 $\log_{10}$ MPN/100 mL (IQR = 1.4, 2.2), while in South Florida samples the mean was 1.3 (Range = 0.0, 2.8) and the median was 1.3 (IQR = 0.9, 1.6; **Table 3**). While only 9% of the samples collected in South Florida exceed the PSR GM standard for *E. coli* (126 CFU/100 mL), 32% of samples collected in North Florida exceeded this cut-off. Similarly, 3% of South Florida samples and 11% of North Florida samples had *E. coli* levels that exceeded the PSR STV standard (410 MPN/100 mL). Multiple sources show that the difference between CFU and MPN is not significant to change the interpretation of the findings or conclusions (Cowburn et al., 1994; Hargett and Goyn, 2004; Gronewold and Wolpert, 2008; Fricker et al., 2010).

Salmonella was detected in 26% (44/168) of water samples (**Table 4**; 124 samples were below the limit of detection, <0.48 MPN/100 mL). More *Salmonella* was detected in South Florida (27/96; 28%), than in North Florida (17/72; 24%). For the 124 counts that fell below the limit of detection, the value of.48 $\log_{10}$ MPN/100 mL was used. The mean and median $\log_{10}$ MPN/100 mL of *Salmonella* was <0.5 (Range ≤ 0.5, 1.4) and <0.5 (IQR ≤ 0.5, <0.5) in North Florida, and <0.6 (Range ≤ 0.5, 3.0) and <0.5 (IQR ≤ 0.5, <0.5) in South Florida (**Table 3**). **Figure 1** describes the distribution of total coliforms, *E. coli,* and *Salmonella* by region, season, and water type. The *Salmonella* prevalence fell within the wide range reported by past Florida studies (McEgan et al., 2013; Luo et al., 2015; Topalcengiz et al., 2017). A Central Florida study reported a *Salmonella* prevalence of 4.8% (26/540) in 250 mL pond samples (Topalcengiz et al., 2017), while an independent Central Florida study detected *Salmonella* in all 202 10-L samples collected from multiple surface water types (e.g., ponds, canals) (McEgan et al., 2013). The *Salmonella* prevalence reported here is also within the range reported by studies conducted in other states, including North Carolina [e.g., 54.7% (47/86) of 25 mL water samples; (Patchanee et al., 2010)], and Georgia [e.g., 11.9% (34/285) of 222 mL water samples (Antaki et al., 2016); 79.2% (57/72) of 111 mL water

TABLE 3 | Summary statistics for total coliforms, generic *E. coli*, anerobic plate count, and *Salmonella* for North and South Florida waterways.

Factor	Number of observations	Min.	Max.	Mean	Median	SD[a]	Quartiles	
							1st	3rd
Total coliforms ($\log_{10}$ MPN/100 mL)[b]	168	1.7	>3.3	>3.1	>3.2	0.3	>3.2	>3.3
North Florida	72	1.7	>3.3	>3.0	>3.2	0.4	>2.7	>3.3
South Florida	96	2.4	>3.3	>3.3	>3.3	0.2	>3.3	>3.3
Generic *E. coli* ($\log_{10}$ MPN/100 mL)	168	0.0	3.4	1.5	1.5	0.7	1.0	1.9
North Florida	72	0.6	3.2	1.8	1.7	0.6	1.4	2.2
South Florida	96	0.0	2.8	1.3	1.3	0.6	0.9	1.6
Anerobic plate count ($\log_{10}$ CFU/100 mL)	168	3.8	6.4	5.0	4.9	0.6	4.6	5.4
North Florida	72	3.8	6.4	5.0	4.9	0.7	4.5	5.7
South Florida	96	3.8	6.1	4.9	4.9	0.5	4.6	5.3
Salmonella ($\log_{10}$ MPN/100 mL)[c]	168	<0.5	3.0	<0.6	<0.5	0.3	<0.5	<0.5
North Florida	72	<0.5	1.4	<0.5	<0.5	0.1	<0.5	<0.5
South Florida	96	<0.5	3.0	<0.6	<0.5	0.4	<0.5	<0.5

[a] *SD, Standard deviation.*

[b] *Upper limit of detection of >3.3 $\log_{10}$ MPN/100 mL. One hundred twelve counts were above the limit of detection and were given the value of 3.3 $\log_{10}$ MPN/100 mL for calculations.*

[c] *Lower limit of detection of <0.48 $\log_{10}$ MPN/100 mL. One hundred twenty-four counts were below the limit of detection and were given the value of 0.48 $\log_{10}$ MPN/100 mL for calculations.*

samples (Haley et al., 2009)]. One reason likely responsible for the wide range in *Salmonella* prevalence reported by previous studies is the methodological differences between these studies. The study that reported the highest *Salmonella* prevalence (McEgan et al., 2013) collected 10 L samples, 5 to 40 times larger than the samples collected in the other studies discussed here (Haley et al., 2009; Patchanee et al., 2010; Antaki et al., 2016; Topalcengiz et al., 2017), and were 30 times larger than the samples collected in the present study. The likelihood of detecting *Salmonella* is higher for larger volumes of water, as increasing the volume tested decreases the lower limit of detection. Methodological differences between studies confound the comparison of results and reduce the ability to determine if the observed prevalence is consistent within and between studies. Development of standardized practices, including standard volumes for studies focused on agricultural water used for produce production may be appropriate. Alternatively, it may be possible to develop a statistical model, that accounts for sample volume, to enable the comparison of results between studies.

TABLE 4 | Effect of factors season and month on the frequency of *Salmonella*-positive[a] water samples in North and South Florida waterways.

Factor	North florida waterway *Salmonella* frequency (percent)	South florida waterway *Salmonella* frequency (percent)
Total	23.6 (17/72)	28.1 (27/96)
Month		
January	33.3 (2/6)	37.5 (3/8)
February	16.7 (1/6)	0.0 (0/8)
March	0.0 (0/6)	25.0 (2/8)
April	0.0 (0/6)	37.5 (3/8)
May	0.0 (0/6)	37.5 (3/8)
June	50.0 (3/6)	25.0 (2/8)
July	50.0 (3/6)	62.5 (5/8)
August	50.0 (3/6)	0.0 (0/8)
September	0.0 (0/6)	62.5 (5/8)
October	0.0 (0/6)	25.0 (2/8)
November	66.7 (4/6)	12.5 (1/8)
December	16.7 (1/6)	12.5 (1/8)
Season		
Fall	22.2 (4/18)	33.3 (8/24)
Winter	22.2 (4/18)	16.7 (4/24)
Spring	0.0 (0/18)	33.3 (8/24)
Summer	50.0 (9/18)	29.2 (7/24)

[a]*Salmonella positive is defined as a sample at or above the limit of detection where a Salmonella negative is below the limit of detection.*

Microbial Water Quality Varied Between Regions and Water Types

The total coliform and *Salmonella* levels reported are higher in South Florida than North Florida; the opposite is reported for *E. coli* (**Figure 1**). Total coliform and *Salmonella* levels reported are highest in canals, with total coliform levels higher in rivers, than lakes, and *Salmonella* levels were higher in lakes than rivers; for *E. coli* levels appeared lowest in canals, followed by rivers and lakes (**Figure 1**).

In the Bayesian mixed models, the $\log_{10}$ *E. coli* levels (measured as both a continuous factor, and as a binary factor representing if samples were above or below the PSR cut-offs of 126 and 410 CFU/100 mL) and $\log_{10}$ total coliform levels (measured as both a continuous factor, and a binary factor representing if samples were above or below the upper LOD) differed substantially between North and South Florida, and between water types (**Table 1**). The *E. coli* levels were ∼0.50 $\log_{10}$ MPN/100 mL higher in North Florida than in South Florida (MAP = 0.50; 95% CI = 0.12,0.90). The results of the conditional inference trees were generally consistent with the regression models. However, given collinearity between the region and water type (i.e., that canals were only sampled in

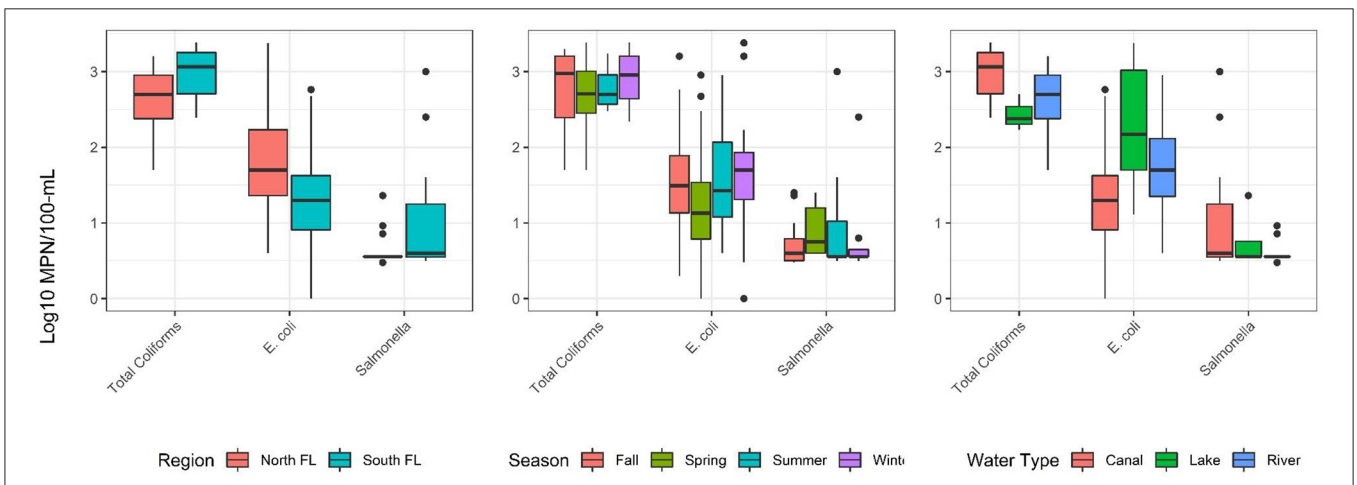


FIGURE 1 | Distribution of $\log_{10}$ total coliform, *E. coli,* and *Salmonella* levels (MPN/100 mL) in each region[a], season[b], and water type[c]. [a]*N* = 72 and 96 samples for North and South FL, respectively. [b]*N* = 42 samples for fall, spring, summer, and winter. [c]*N* = 96, 12, and 60 samples for canal, lake, and river, respectively.

South Florida, and lakes and rivers in North Florida), it is difficult to determine if differences are driven by water type or region. To assist in overcoming this difficulty (i.e., collinearity), and probe if differences are driven more by water type or region, future studies should include a variety of water types within each region. The different surface water sampling locations reflect the different, and complex, watersheds in the state, resulting from disjointed drainage systems and atypical elevation gradients. In North Florida a karst topography results in numerous rivers, streams, lakes, and springs; in South Florida, canal systems were developed to divert the water that historically flowed as a sheet of water to allow for agricultural production (Purdum et al., 2002). The findings, in light of these geographical differences, highlight the heterogeneity inherent to freshwater environments, and the need for improved understanding of pathogen ecology for specific, intrastate produce growing regions.

Collinearity between the region and water type is not a factor for *Salmonella* result interpretation since neither *Salmonella* levels nor the odds of *Salmonella* detection differed substantially between regions. The *Salmonella* levels and the odds of *Salmonella* detection did differ between water types (e.g., canals, rivers, lakes) (**Table 2**). Specifically, the odds of *Salmonella* detection (OR = 0.70; 95% CI = 0.30, 1.59), and *Salmonella* levels (Effect Estimate = −0.23; 95% CI = −0.62, 0.24) were lower for river samples, compared with canal samples (**Table 2**) based on the practical significance (PS) index being ≥0.75, indicating the observed effect is both non-negligible and, in the direction indicated by maximum a posteriori estimate (MAP) (**Table 2**). The *Salmonella* tree (**Figure 2**) found evidence of a significant regional difference. However, this difference was dependent on environmental conditions (i.e., when temperatures were high, *Salmonella* levels were higher, regardless of region, but at lower water temperatures, *Salmonella* levels were higher in South Florida than in North Florida). These findings indicate that microbial water quality varied both between regions and between water types. This finding was not surprising, as past studies that compared microbial quality between growing regions have also found evidence of regional differences (Strawn et al., 2013a, 2014; Chapin et al., 2014; Weller et al., 2020b). Weller et al. (2020b) sampled Arizona canals and New York streams and reported higher *E. coli* levels and a higher prevalence of *Listeria monocytogenes*, pathogenic *E. coli* markers, and *Salmonella* in New York streams. Additionally, Strawn et al. (2014) found a higher prevalence of *Salmonella* positive overall environmental samples and water samples in south Florida (35 and 38%, respectively), compared with New York (5 and 9%, respectively). The findings reported here, coupled with previous studies (Strawn et al., 2013a, 2014; Chapin et al., 2014; Weller et al., 2020b) highlight the differences in microbial water quality between water sources and growing regions, and underscores the challenges of developing a one-size-fits-all approach for managing microbial hazards in agricultural water. The findings reported here, identify the differences in microbial water quality, and subsequent resulting challenges with recommendations for managing microbial hazards, even within a single state. These findings suggest that risk management approaches may need to be tailored to specific water types within localized (e.g., intra-state) regions.

Salmonella Serovars Were Diverse and Differed Between Regions

Serotyping was performed on one representative *Salmonella* isolate per isolation scheme (up to 4 isolates per sample) and yielded 45 *Salmonella* isolates from the 44 positive samples (**Supplementary Table 3**). One of the 44 positive samples yielded two serovars: *S. enterica* subspecies *enterica* Invernes and Muenchen (in North Florida waterways) (**Supplementary Table 3**). The remaining 43 *Salmonella*-positive samples represented one serovar including Anatum (1), Florida (3), Hartford (1), Inverness (4), Muenchen (4), Saintpaul (3) and IV 40:z4,z24 (2) in North Florida; and Agona (1), Baildon (1), Braenderup (2), Enteritidis (2), Javiana (1), Litchfield (1), Muenchen (1), Rough (1), Rubislaw (2), Tennessee (4), Typhimurium (1), III 60:r:z (1), and IV 53:z4.z23 (9) in South Florida waterways (**Supplementary Table 3**). Overall serotype richness was higher in South Florida (13 serotypes) than North Florida (7 serotypes), as was Simpson's Index of Diversity (0.67 in South Florida;0.58 in North Florida). *Salmonella* serotype IV 53:z4.z23 was isolated most frequently from water samples in South Florida waterways, consisting of approximately 33% (9/27). While no one serotype was predominant among the *Salmonella* isolates from North Florida waterways, four serotypes (Florida, Inverness, Muenchen, and Saintpaul) represented 78% (14/18). A previous study in central Florida identified 33 *Salmonella enterica* serotypes from 165 surface water samples with the most frequent serotypes being Muenchen, Rubislaw, Anatum, Gaminara, and IV_50:z4,z23:- (McEgan et al., 2014). When Strawn et al. (2014) looked at the *Salmonella* diversity between two growing regions (South Florida and New York), it was identified that a high PFGE type diversity (Simpson's diversity index,0.90,0.02) was observed among *Salmonella* isolates across both regions and only three Pulsed-field gel electrophoresis (PFGE) types were shared between the two regions. Similarly, prior research has shown that specific *Salmonella* serovars may be associated with certain regions, such as *Salmonella* Newport repeatedly being isolated from the eastern shore of Virginia (Greene et al., 2008; Truitt et al., 2018) while several *Salmonella* strains, all with the same PFGE type, have been repeatedly isolated from the surface water in the same region in California (Gorski et al., 2013). These previous findings demonstrate that the diversity of *Salmonella* varies by space and sub-regions. These findings indicate that not only did *Salmonella* levels (under specific weather conditions) differ significantly between Florida regions, but that the composition and diversity of the *Salmonella* populations also differed substantially.

Weather Was an Important Driver Across All Three Microbial Targets

Bayesian regression indicates that microbial targets considered showed evidence of seasonal patterns (**Figure 1**, **Tables 1**, **2**). The *E. coli* levels were 0.33 (95% CI = −0.54, −0.09) $\log_{10}$ MPN/100 mL lower in spring than in fall; the differences between

summer and fall, and winter and fall were negligible and of indeterminate direction based on PD and ROPE. Odds of total coliform levels being above the upper LOD was higher in summer (OR = 2.27; 95% CI = 0.60, 7.82), and lower in spring (OR = 0.51; 95% CI = 0.17, 1.59) and winter (OR = 0.28; 95% CI = 0.09, 0.85) based on the PS (probability of significance) index being ≥0.75, indicating the observed effect is both non-negligible and in the direction indicated by MAP. Based on the PS from the Bayesian regression, the likelihood of *Salmonella* detection was higher in spring, summer, and winter compared with fall, but that *Salmonella* levels (in *Salmonella* positive samples) were only higher in summer compared with fall (**Table 2**). Overall, the identification of seasonal patterns in water quality is consistent with the literature (Carter et al., 1987; Haley et al., 2009; Gorski et al., 2011; Cooley et al., 2014). For example, Haley et al. (2009) found that *Salmonella* concentrations in Georgia surface waters were significantly higher in the summer months compared with other seasons ($P < 0.05$). While other Florida studies either did not sample in Summer due to fewer crops or found no association with summer, the results from this and previous studies suggest an elevated risk during the summer months and therefore, future work will need to test this hypothesis.

Conditional inference tree analysis indicates that water temperature may drive seasonal trends in total coliform and *Salmonella* concentrations, and *Salmonella* detection (**Figures 2**, **3**). Across all three trees, warmer water temperatures were associated with elevated levels or an increased likelihood of detecting the target (**Figures 2**, **3**). This is consistent with past studies reporting seasonal trends in microbial water quality and linked elevated temperatures with an increased likelihood of detecting foodborne pathogens (Polo et al., 1999; Martinez-Urtaza et al., 2004; Haley et al., 2009; Huang et al., 2014; Antaki et al., 2016; Liu et al., 2018). Haley et al. (2009) and Antaki et al. (2016) both note higher *Salmonella* concentrations when water temperatures are warmer in Georgia, USA. Higher *Salmonella* detection in water samples collected seasonally from 34 locations along the Puzih River in Taiwan rates in the summer coincided with higher air and water temperatures (Huang et al., 2014). Since conditional trees can visualize hierarchical relationships, it is of interest that in the present study, the primary split for indicator organisms like total coliforms and generic *E. coli* were based on region and water type (**Figures 3**, **4**), while the primary split for *Salmonella* was based on weather (**Figure 2**). Across all three microbial targets, splits for land use are below those for region and water type, which indicates that the land-use relationships identified here were based on region and or water type-specific (**Figures 2**–**4**). The tree analysis indicates that *E. coli* levels were highest in North Florida (lakes and rivers), and lowest in South Florida (canal) sites, with more than ∼3% of the land (in a 366 m buffer) under developed cover. Conversely, when water temperatures were higher, *Salmonella* levels were higher regardless of region, and only at lower water temperatures did *Salmonella* levels begin to differ between regions (with levels in samples from South Florida being higher, compared with levels

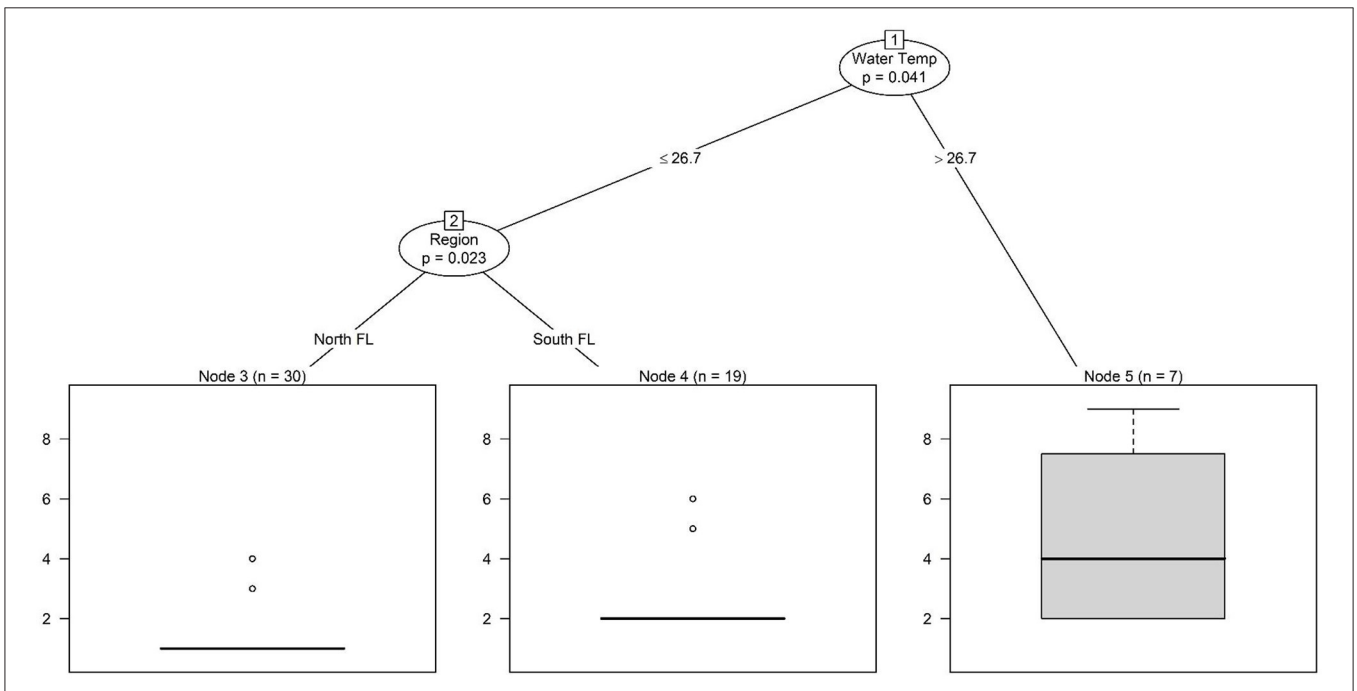


FIGURE 2 | Conditional inference tree showing hierarchical associations between environmental factors and $\log_{10}$ *Salmonella* levels (MPN/100 mL) in samples that were *Salmonella* positive. A separate tree was fit for if a sample was *Salmonella* positive or negative, the only split in this tree was based on if air temperature at time of sample collection was ≤28.2°C (expected *Salmonella* prevalence of ∼22%) and >28.2 C (expected *Salmonella* prevalence ∼50%).

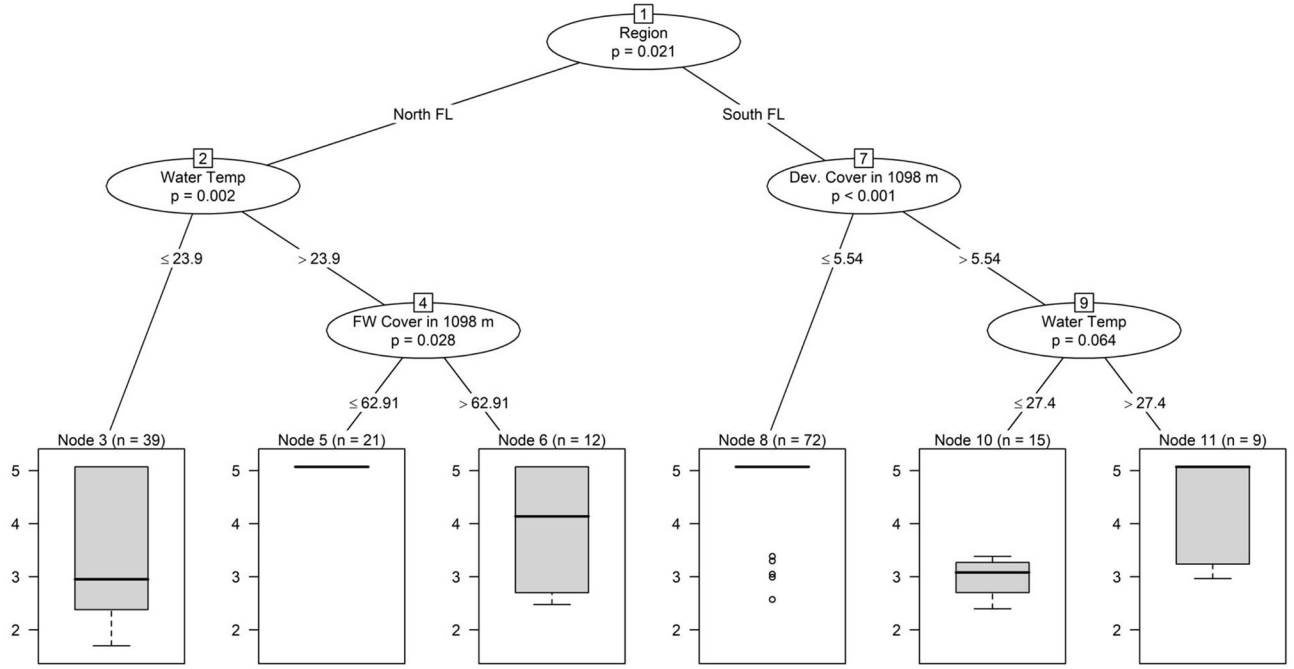


FIGURE 3 | Conditional inference tree showing hierarchical associations between environmental factors and $\log_{10}$ total coliform levels (MPN/100 mL); for samples above the upper limit of detection (LOD) a value of 1.5*LOD was used.

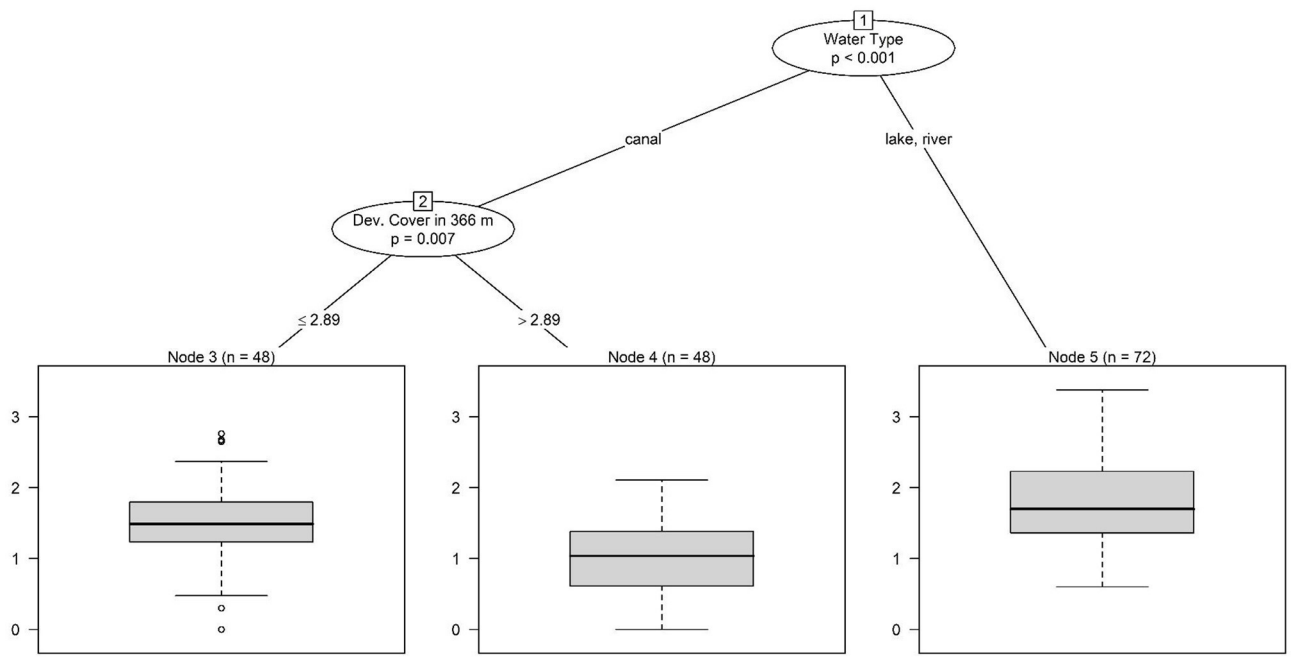


FIGURE 4 | Conditional inference tree showing hierarchical associations between environmental factors and $\log_{10}$ *E. coli* levels (MPN/100 mL); note since lakes and rivers were samples in North Florida and canals in South Florida, region, and water type are interchangeable as the first split.

from North Florida; **Figure 2**). According to the tree analysis, the likelihood of *Salmonella* detection was only dependent on on-air temperature, with no differences between regions (air and water temperature are correlated). The relationships identified in the total coliform tree are more complex, with the highest levels being observed in South Florida (canals) when samples were collected with developed cover (in a 1,098 m buffer) above ∼6% and the water temperature was above ∼27°C; and in

North Florida (river, lake) when samples were collected from water temperature above ∼24°C and forest-wetland cover (in a 1,098 m buffer) was above ∼63% (**Figure 3**). Overall, the fact that the first split in the *Salmonella* tree was based on weather indicates a stronger effect of weather than of region or water type on microbial water quality. The opposite conclusion can be made about land use in the present study, since (i) land use variables were all lower in the trees and thus dependent on specific weather and either region/water type conditions being met. These findings suggest that, for Florida, weather conditions may be useful for monitoring when food safety hazards are more likely to be present in agricultural waterways. Additional research is needed to confirm this finding as previous studies, have found varying pathogen-temperature relationships, and that those relationships are complex based on spatiotemporal factors (e.g., year or site of sample collection) (McEgan et al., 2013; Topalcengiz et al., 2017; Weller et al., 2020b). Weller et al. (2020b) compared pathogen levels between Arizona and New York found a positive relationship between temperature and likelihood of detecting *Salmonella* in Arizona, but a complex, polynomial relationship in New York.

Salmonella Levels Were Not Associated With *E. coli* Levels in the Present Study

The PSR proposed microbial standards for pre-harvest agricultural water and are under review at the time of writing this manuscript. Currently, the PSR standards require (i) that agricultural surface water used during production establish a microbial water quality profile (MWQP) using 20 samples collected over 2 to 4 years on a rolling basis, and (ii) that geometric mean (GM) and statistical threshold value (STV) of *E. coli* in this sample be ≤126 and ≤410 CFUs/100 mL, respectively (US Food Drug Administration, 2015). An MWQP for each site could not be created in the present study as we did not collect 20 samples over 2–4 years, we were able to compare the likelihood of *Salmonella* detection and *Salmonella* concentration in *Salmonella*-positive samples to *E. coli* levels in the same samples. *E. coli* levels in individual samples were more likely to exceed both PSR mean (OR = 5.43; 95% CI = 1.19, 52.52) and STV (OR = 4.22; 95% CI = 0.56, 47.22) cut-offs in North Florida than South Florida (**Table 1**). *Salmonella* levels in *Salmonella* positive samples were not associated with $\log_{10}$ *E. coli* levels (as a continuous factor) or if the *E. coli* levels exceeded (or failed to exceed) the PSR cut-offs (**Table 1**). Odds of *Salmonella* detection was not associated with $\log_{10}$ *E. coli* levels or if the levels exceeded the PSR mean cut-off; odds of *Salmonella* detection was 2.54 higher (95% CI = 0.66, 8.75) in samples that exceeded the PSR STV cut-off than in samples that met the cut-off. The association between *E. coli* levels and foodborne pathogen presence in agricultural water is consistent with some studies reporting an association and others failing to detect an association (Harwood et al., 2005; McEgan et al., 2013; Pachepsky et al., 2016; Truitt et al., 2018). McEgan et al. (2013) found that the presence and strength of the *E. coli* and *Salmonella* relationship differed between sites in the same region of Central Florida. *E. coli* was an adequate predictor of the presence of *Salmonella* in 150 mL samples in West Central Florida ponds; when *E. coli* populations were higher, *Salmonella* presence was more likely, but the relationship between populations differed between ponds (Havelaar et al., 2017; Topalcengiz et al., 2017). The results presented here support the conclusion from earlier studies that *E. coli* levels are unreliable as an indicator for the presence and concentration of microbial hazards in agricultural water. As *E. coli* is an indicator of fecal contamination and not an index for pathogen presence, this aligns with traditional convention.

CONCLUSION

The goals of the study were to characterize and compare (1) the associations between microbial water quality, including pathogen presence and environmental factors (e.g., water quality, weather, land use) in North and South Florida waterways; and (2) *Salmonella* diversity in North and South Florida waterways. While drivers of microbial water quality can differ between intrastate growing regions (e.g., North versus South Florida); this conclusion must be interpreted cautiously as reported differences may also be due to the fact that the predominant water sources used in North (i.e., river, lake) and South (i.e., canals) Florida differ. Despite this limitation, this study highlights the heterogeneity inherent to freshwater environments, and the need for the improved understanding of pathogen ecology for specific, intrastate produce growing regions. Future studies are needed to untangle the relative contribution of the intrastate growing region and water type to the type of differences reported here. This understanding will help with the development of evidence-based risk management strategies for producing safety risks associated with pre-harvest surface water use. This study also highlights the need for alternative approaches for assessing the presence of potential food safety hazards in agricultural water.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CM, DW, and LS: data analysis and manuscript writing. TC: sample collection, laboratory analysis, and manuscript writing. RM and LG: experimental design. SG: sample collection and laboratory analysis. LF: laboratory analysis. KS: experimental design and manuscript writing. MD: experimental design, data analysis, and manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Specialty Crops Research Initiative projects 2008-51180-04846, 2011-51181-30767, and

2019-51181-30016 from the US Department of Agriculture National Institute of Food and Agriculture.

ACKNOWLEDGMENTS

We are grateful for the technical assistance of Ashwathy Sreedharan, Alina Balaguero, Susanna Richardson, and Lorraine Rodriguez-Rivera, for the North Florida sample collection and laboratory analysis, and to Gwen Lundy, Luis Martinez, and Karen Plant for the South Florida sample collection and laboratory analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.750673/full#supplementary-material

REFERENCES

American Public Health Association (1992). *Water Environment Federation: Standard Methods for the Examination of Water and Wastewater, 18th Edn*, eds A. E. Clesceri, L. S. Eaton, AD. Washington, DC: APHA.

Andrews, W. H., Acobson, A., and Hammack, T. (2011). *Bacteriological Analytical Manual (BAM). Chapter 5 Salmonella.* Available online at: https://www.fda.gov/food/laboratory-methods-food/bacteriological-analytical-manual-bam-chapter-5-salmonella (accessed April 20, 2020).

Antaki, E. M., Vellidis, G., Harris, C., Aminabadi, P., Levy, K., and Jay-Russell, M. T. (2016). Low concentration of *Salmonella enterica* and generic *Escherichia coli* in farm ponds and irrigation distribution systems used for mixed produce production in southern Georgia. *Foodborne Pathog Dis.* 13, 551–558. doi: 10.1089/fpd.2016.2117

Atwill, E. R., Chase, J. A., Oryang, D., Bond, R. F., Koike, S. T., Cahn, M. D., et al. (2015). Transfer of *Escherichia coli* O157:H7 from simulated wildlife scat onto romaine lettuce during foliar irrigation. *J. Food Protect.* 78, 240–247. doi: 10.4315/0362-028X.JFP-14-277

Bürkner, P. C. (2017a). Advanced bayesian multilevel modeling with the R package brms. *arXiv [Preprint]* arXiv:1705.11123. doi: 10.32614/RJ-2018-017

Bürkner, P. C. (2017b). brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

California Leafy Greens Marketing Agreement (2020). *California LGMA Commodity Specific Food Safety Guidelines.* California Leafy Greens Marketing Agreement. Available online at: https://lgmatech.com/wp-content/uploads/2020/08/CA-LGMA-Metrics-August-2020_Final_Clean_9-18-20.pdf

Carter, A., Pacha, R. E., Clark, G. W., and Williams, E. (1987). Seasonal occurrence of *Campylobacter* spp. in surface waters and their correlation with standard indicator bacteria. *Appl. Environ. Microbiol.* 53, 523–526. doi: 10.1128/aem.53.3.523-526.1987

Chapin, T. K., Nightingale, K. K., Worobo, R. W., Wiedmann, M., and Strawn, L. K. (2014). Geographical and meteorological factors associated with isolation of *Listeria* species in New York State produce production and natural environments. *J. Food Protect.* 77, 1919–1928. doi: 10.4315/0362-028X.JFP-14-132

Cooley, M. B., Quiñones, B., Oryang, D., Mandrell, R. E., and Gorski, L. (2014). Prevalence of shiga toxin producing *Escherichia coli*, *Salmonella enterica*, and *Listeria monocytogenes* at public access watershed sites in a California Central Coast agricultural region. *Front. Cell. Infect. Microbial.* 4:30. doi: 10.3389/fcimb.2014.00030

Cowburn, J., Goodall, T., Fricker, E., Walter, K., and Fricker, C. (1994). A preliminary study of the use of colilert for water quality monitoring. *Lett. Appl. Microbiol.* 19, 50–52. doi: 10.1111/j.1472-765X.1994.tb00902.x

Erickson, M. C., Webb, C. C., Diaz-Perez, J. C., Phatak, S. C., Silvoy, J. J., Davey, L., et al. (2010). Surface and internalized *Escherichia coli* O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water. *J. Food Protect.* 73, 1023–1029. doi: 10.4315/0362-028X-73.6.1023

Fatica, M. K., and Schneider, K. R. (2011). *Salmonella* and produce: survival in the plant environment and implications in food safety. *Virulence* 2, 573–579. doi: 10.4161/viru.2.6.17880

Fricker, C., Warden, P., Desarno, M., and Eldred, B. (2010). *Significance of Methods and Sample Volumes for E. coli and Total Coliform Measurements.* Available online at: http://www.waterrf.org/publicreportlibrary/4024.pdf (accessed April 25, 2020).

Gorski, L., Jay-Russell, M. T., Liang, A. S., Walker, S., Bengson, Y., Govoni, J., et al. (2013). Diversity of pulsed-field gel electrophoresis pulsotypes, serovars, and antibiotic resistance among *Salmonella* isolates from wild amphibians and reptiles in the California central coast. *Foodborne Pathog. Dis.* 10, 540–548. doi: 10.1089/fpd.2012.1372

Gorski, L., Parker, C. T., Liang, A., Cooley, M. B., Jay-Russell, M. T., Gordus, A. G., et al. (2011). Prevalence, distribution, and diversity of *Salmonella enterica* in a major produce region of California. *Appl. Environ. Microbiol.* 77, 2734–2748. doi: 10.1128/AEM.02321-10

Greene, S., Daly, E., Talbot, E., Demma, L., Holzbauer, S., Patel, N., et al. (2008). Recurrent multistate outbreak of *Salmonella* newport associated with tomatoes from contaminated fields, 2005. *Epidemiology. Infect.* 136, 157–165. doi: 10.1017/S095026880700859X

Gronewold, A. D., and Wolpert, R. L. (2008). Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. *Water Res.* 42, 3327–3334. doi: 10.1016/j.watres.2008.04.011

Gu, G., Strawn, L. K., Oryang, D. O., Zheng, J., Reed, E. A., Ottesen, A. R., et al. (2018). Agricultural practices influence *Salmonella* contamination and survival in pre-harvest tomato production. *Front. Microbiol.* 9:2451. doi: 10.3389/fmicb.2018.02451

Gu, G., Strawn, L. K., Zheng, J., Reed, E. A., and Rideout, S. L. (2019). Diversity and dynamics of *Salmonella enterica* in water sources, poultry litters, and field soils amended with poultry litter in a major agricultural area of Virginia. *Front. Microbiol.* 10:2868. doi: 10.3389/fmicb.2019.02868

Haley, B. J., Cole, D. J., and Lipp, E. K. (2009). Distribution, diversity, and seasonality of waterborne *Salmonellae* in a rural watershed. *Appl. Environ. Microbiol.* 75, 1248–1255. doi: 10.1128/AEM.01648-08

Hargett, E., and Goyn, L. (2004). *Modified mTEC agar, Colilert®, and M-FC Agar–Field Trial Comparison of Bacteria Enumeration Methods in Surface Waters of Eastern Wyoming.* Available online at: http://deq.wyoming.gov/media/attachments/Water%20Quality/Surface%20Water%20Monitoring/Publications/wqd-wpp-monitoring_comp_study_e.coli_final2.pdf (accessed April 25, 2020).

Harwood, V. J., Levine, A. D., Scott, T. M., Chivukula, V., Lukasik, J., Farrah, S. R., et al. (2005). Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol.* 71, 3163–3170. doi: 10.1128/AEM.71.6.3163-3170.2005

Havelaar, A. H., Vazquez, K. M., Topalcengiz, Z., Muñoz-Carpena, R., and Danyluk, M. D. (2017). Evaluating the US food safety modernization act produce safety rule standard for microbial quality of agricultural water for growing produce. *J. Food Prot.* 80, 1832–1841. doi: 10.4315/0362-028X.JFP-17-122

Huang, K.-H., Hsu, B.-M., Chou, M.-Y., Tsai, H.-L., Kao, P.-M., Wang, H.-J., et al. (2014). Application of molecular biological techniques to analyze *Salmonella* seasonal distribution in stream water. *FEMS Microbiol. Lett.* 352, 87–96. doi: 10.1111/1574-6968.12381

Ijabadeniyi, O. A., Debusho, L. K., Vanderlinde, M., and Buys, E. M. (2011). Irrigation water as a potential preharvest source of bacterial contamination of vegetables. *J. Food Saf.* 31, 452–461. doi: 10.1111/j.1745-4565.2011.00321.x

Ishii, S., and Sadowsky, M. J. (2008). *Escherichia coli* in the environment: implications for water quality and human health. *Microb. Environ.* 23, 101–108. doi: 10.1264/jsme2.23.101

Jang, J., Hur, H. G., Sadowsky, M. J., Byappanahalli, M., Yan, T., and Ishii, S. (2017). Environmental *Escherichia coli*: ecology and public health implications—a review. *J. Appl. Microbiol.* 123, 570–581. doi: 10.1111/jam.13468

Jeamsripong, S., Chase, J. A., Jay-Russell, M. T., Buchanan, R. L., and Atwill, E. R. (2019). Experimental in-field transfer and survival of *Escherichia coli* from

animal feces to romaine lettuce in Salinas Valley, California. *Microorganisms* 7:408. doi: 10.3390/microorganisms7100408

Klontz, K. C., Klontz, J. C., Mody, R. K., and Hoekstra, R. M. (2010). Analysis of tomato and jalapeño and serrano pepper imports into the United States from Mexico before and during a national outbreakof *Salmonella* serotype saintpaul infections in 2008. *J. Food Protect.* 73, 1967–1974. doi: 10.4315/0362-028X-73.11.1967

Lee, D., Tertuliano, M., Harris, C., Vellidis, G., Levy, K., and Coolong, T. (2019). *Salmonella* survival in soil and transfer onto produce via splash events. *J. Food Protect.* 82, 2023–2037. doi: 10.4315/0362-028X.JFP-19-066

Liao, J., Guo, X., Weller, D. L., Pollak, S., Buckley, D. H., Wiedmann, M., et al. (2021). Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of selection and population ecology on pangenome evolution. *Nat. Microbiol.* 6, 1021–1030. doi: 10.1038/s41564-021-00935-7

Liu, H., Whitehouse, C. A., and Li, B. (2018). Presence and persistence of *Salmonella* in water: the impact on microbial quality of water and food safety. *Front. Public Health* 6:159. doi: 10.3389/fpubh.2018.00159

Luo, Z., Gu, G., Ginn, A., Giurcanu, M. C., Adams, P., Vellidis, G., et al. (2015). Distribution and characterization of *Salmonella enterica* isolates from irrigation ponds in the southeastern United States. *Appl. Environ. Microbiol.* 81, 4376–4387. doi: 10.1128/AEM.04086-14

Makowski, D., Ben-Shachar, M. S., Chen, S., and Lüdecke, D. (2019a). Indices of effect existence and significance in the bayesian framework. *Front. Psychol.* 10:2767. doi: 10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., and Lüdecke, D. (2019b). bayestestR: describing effects and their uncertainty, existence and significance within the bayesian framework. *J. Open Source Softw.* 4:1541. doi: 10.21105/joss.01541

Malorny, B., Hoorfar, J., Bunge, C., and Helmuth, R. (2003). Multicenter validation of the analytical accuracy of *Salmonella* PCR: towards an international standard. *Appl. Environ. Microbiol.* 69, 290–296. doi: 10.1128/AEM.69.1.290-296.2003

Martinez-Urtaza, J., Saco, M., De Novoa, J., Perez-Piñeiro, P., Peiteado, J., Lozano-Leon, A., et al. (2004). Influence of environmental factors and human activity on the presence of *Salmonella* serovars in a marine environment. *Appl. Environ. Microbiol.* 70, 2089–2097. doi: 10.1128/AEM.70.4.2089-2097.2004

McEgan, R., Chandler, J. C., Goodridge, L. D., and Danyluk, M. D. (2014). Diversity of *Salmonella* isolates from central Florida surface waters. *Appl. Environ. Microbiol.* 80, 6819–6827. doi: 10.1128/AEM.02191-14

McEgan, R., Mootian, G., Goodridge, L. D., Schaffner, D. W., and Danyluk, M. D. (2013). Predicting *Salmonella* populations from biological, chemical, and physical indicators in Florida surface waters. *Appl. Environ. Microbiol.* 79, 4094–4105. doi: 10.1128/AEM.00777-13

Miles, J. M., Sumner, S. S., Boyer, R. R., Williams, R. C., Latimer, J. G., and Mckinney, J. M. (2009). Internalization of *Salmonella enterica* serovar montevideo into greenhouse tomato plants through contaminated irrigation water or seed stock. *J. Food Protect.* 72, 849–852. doi: 10.4315/0362-028X-72.4.849

Pachepsky, Y., Shelton, D., Dorner, S., and Whelan, G. (2016). Can *E. coli* or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters? *Crit. Rev. Microbiol.* 42, 384–393. doi: 10.3109/1040841X.2014.954524

Park, S., Szonyi, B., Gautam, R., Nightingale, K., Anciso, J., and Ivanek, R. (2012). Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. *J. Food Protect.* 75, 2055–2081. doi: 10.4315/0362-028X.JFP-12-160

Patchanee, P., Molla, B., White, N., Line, D. E., and Gebreyes, W. A. (2010). Tracking *Salmonella* contamination in various watersheds and phenotypic and genotypic diversity. *Foodborne Pathog. Dis.* 7, 1113–1120. doi: 10.1089/fpd.2010.0572

Polo, F., Figueras, M., Inza, I., Sala, J., Fleisher, J. M., and Guarro, J. (1999). Prevalence of *Salmonella* serotypes in environmental waters and their relationships with indicator organisms. *Antonie Van Leeuwenhoek* 75, 285–292. doi: 10.1023/A:1001817305779

Purdum, E. D., Krafft, P. A., Anderson, J., Bartos, B., McPherson, S., Penson, G., et al. (2002). *Florida Water: A Water Resources Manual From Florida's Water Managemnet District.* Available online at: https://www.sfwmd.gov/sites/default/files/documents/florida_waters.pdf (accessed November 9, 2021).

Rodrigues, C., Da Silva, A. L. B. R., and Dunn, L. L. (2020). Factors impacting the prevalence of foodborne pathogens in agricultural water sources in the southeastern United States. *Water* 12:51. doi: 10.3390/w12010051

Strawn, L. K., Danyluk, M. D., Worobo, R. W., and Wiedmann, M. (2014). Distributions of *Salmonella* subtypes differ between two US produce-growing regions. *Appl. Environ. Microbiol.* 80, 3982–3991. doi: 10.1128/AEM.00348-14

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Gröhn, Y. T., Worobo, R. W., et al. (2013a). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol.* 79, 588–600. doi: 10.1128/AEM.02491-12

Strawn, L. K., Gröhn, Y. T., Warchocki, S., Worobo, R. W., Bihn, E. A., and Wiedmann, M. (2013b). Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Appl. Environ. Microbiol.* 79, 7618–7627. doi: 10.1128/AEM.02831-13

Topalcengiz, Z., Strawn, L. K., and Danyluk, M. D. (2017). Microbial quality of agricultural water in Central Florida. *PLoS ONE* 12:e0174889. doi: 10.1371/journal.pone.0174889

Truitt, L. N., Vazquez, K. M., Pfuntner, R. C., Rideout, S. L., Havelaar, A. H., and Strawn, L. K. (2018). Microbial quality of agricultural water used in produce preharvest production on the eastern shore of Virginia. *J. Food Protect.* 81, 1661–1672. doi: 10.4315/0362-028X.JFP-18-185

US Food Drug Administration (2015). Standards for the growing, harvesting, packing, and holding of produce for human consumption. *Fed. Regist.* 80, 74353–74642. Available online at: https://www.govinfo.gov/content/pkg/FR-2015-11-27/pdf/2015-28159.pdf (accessed September 11, 2020).

Weller, D., Belias, A., Green, H., Roof, S., and Wiedmann, M. (2020a). Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Front. Sust. Food Syst.* 3:124. doi: 10.3389/fsufs.2019.00124

Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., et al. (2020b). Complex interactions between weather, and microbial and physicochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water. *Front. Microbiol.* 11:134. doi: 10.3389/fmicb.2020.00134

Weller, D., Wiedmann, M., and Strawn, L. K. (2015). Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Appl. Environ. Microbiol.* 81, 6059–6069. doi: 10.1128/AEM.01286-15

Weller, D. L., Love, T. M., Belias, A., and Wiedmann, M. (2020c). Predictive models may complement or provide an alternative to existing strategies for assessing the enteric pathogen contamination status of northeastern streams used to provide water for produce production. *Front. Sust. Food Syst.* 4:151. doi: 10.3389/fsufs.2020.561517

Weller, D. L., Love, T. M., and Wiedmann, M. (2021). Interpretability versus accuracy: a comparison of machine learning models built using different algorithms, performance measures, and features to predict *E. coli* levels in agricultural water. *Front. Artif. Intell.* 4:628441. doi: 10.3389/frai.2021.628441

Author Disclaimer: Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the US Department of Agriculture.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Water

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/frwa.2021.741676

Land Use, Weather, and Water Quality Factors Associated With Fecal Contamination of Northeastern Streams That Span an Urban-Rural Gradient

*Daniel L. Weller[1,2], Claire M. Murphy[3], Stephanie Johnson[4], Hyatt Green[1], Edward M. Michalenko[4], Tanzy M. T. Love[2] and Laura K. Strawn[3]**

[1] Department of Environmental and Forest Biology, SUNY College of Environmental Sciences and Forestry, Syracuse, NY, United States, [2] Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States, [3] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [4] Onondaga Environmental Institute, Syracuse, NY, United States

OPEN ACCESS

Edited by:
Pengfei Zhang,
City College of New York (CUNY), United States

Reviewed by:
Cheng Yan,
China University of Geosciences Wuhan, China
Gregory O'Mullan,
Queens College (CUNY), United States
Doris Yoong Wen Di,
University of Hawaii at Manoa, United States

***Correspondence:**
Laura K. Strawn
Lstrawn@vt.edu

Specialty section:
This article was submitted to Environmental Water Quality, a section of the journal Frontiers in Water

Received: *15 July 2021*
Accepted: *15 November 2021*
Published: *24 February 2022*

Citation:
Weller DL, Murphy CM, Johnson S, Green H, Michalenko EM, Love TMT and Strawn LK (2022) Land Use, Weather, and Water Quality Factors Associated With Fecal Contamination of Northeastern Streams That Span an Urban-Rural Gradient. Front. Water 3:741676. doi: 10.3389/frwa.2021.741676

Fecal contamination of surface water has been associated with multiple enteric disease outbreaks and food recalls. Thus, it is important to understand factors associated with fecal contamination of agricultural water sources. Since fecal indicator bacteria (FIB) were used to monitor surface water for potential fecal contamination, the purpose of the present study was to characterize associations between environmental factors, and (i) FIB (*E. coli, Enterococcus,* and coliform) levels, and (ii) host-specific fecal marker detection. This study used data collected from 224 sites along 3 waterways, which spanned an urban-rural gradient around Syracuse, New York. Between 2008 and 2017, 2,816 water samples were collected, and *E. coli, Enterococcus,* and/or coliform concentrations were enumerated. Thirty-one samples were also tested for human and ruminant microbial source-tracking markers. Water quality (e.g., turbidity, nitrate) and weather data were also collected for each site. Univariable Bayesian regression was used to characterize the relationship between each microbial target and land use, water quality, and weather factor. For each model, probability of direction and region of practical equivalence overlap (ROPE) were calculated to characterize the association's direction and strength, respectively. While levels of different FIB were not correlated with each other, FIB levels were associated with environmental conditions. Specifically, FIB levels were also positively associated with temperature, nutrient and sediment levels. Log10 *E. coli* levels increased by 0.20 (CI = 0.11, 0.31) and log10 *Enterococcus* levels increased by 0.68 (CI = 0.08, 1.24) for each log10 increase in salinity and nitrate, respectively. These findings may indicate that similar processes drove microbial, sediment, and nutrient contamination of the sampled watersheds. While fecal contamination was strongly associated with land use, the direction of association varied between FIBs and the buffer distance used to calculate land use metrics. *E. coli* levels and human marker detection were positively associated with percent pasture cover within 122, 366, and 1,098 m of the sampling site, while *Enterococcus* and coliform levels were only associated with pasture

cover within 1,098 m (not 122 or 366 m). Ruminant markers were positively associated with pasture cover within 122 m, but not 366 or 1,098 m. These findings highlight the importance of considering (i) adjacent land use (and associated non-point sources of contamination) when developing strategies for managing fecal hazards associated in agricultural and recreational water, and (ii) spatial scale (e.g., 122 vs. 1,098 m) when developing these strategies.

Keywords: fecal indicator bacteria, *Escherichia coli*, *Enterococcus*, fecal source tracking, Bayesian regression, water quality

INTRODUCTION

Fecal contamination of surface water represents a public health hazard, and threatens the economic and recreational value of waterbodies (Rabinovici et al., 2004; Dwight et al., 2005; Given et al., 2006; DeFlorio-Barker et al., 2017; U.S. Environmental Protection Agency SIM., 2018; Calderón-Arrieta et al., 2019). Indeed, fecal contamination of agricultural or recreational water has been identified as the probable cause of multiple enteric disease outbreaks (Ackers et al., 1998; Wachtel et al., 2002; Johnson, 2006; Greene et al., 2008; Barton Behravesh et al., 2011; Food Drug Administration, 2019, 2020). For example, a 2008 multistate *Salmonella* outbreak in the United States was traced back to the use of contaminated water to irrigate hot peppers in 2008. While this outbreak caused 1,200 salmonellosis cases, it also cost tomatoes growers $25 million in the US, as tomatoes were originally misidentified as the food vehicle, before hot peppers were eventually identified (Barton Behravesh et al., 2011; Ribera et al., 2012). Similarly, researchers estimate that enteric illness attributable to recreational water exposures costs $1,220 per 1,000 recreators, or ∼$2.2–$3.7 billion annually in the US (DeFlorio-Barker et al., 2017; U.S. Environmental Protection Agency SIM., 2018). Due to the substantial economic burden and threats of disease associated with fecal contamination of surface water sources, there is considerable interest in mitigating fecal inflows into waterbodies. However, effective mitigation of fecal inflows, requires identifying point and non-point sources of fecal contamination and understanding spatiotemporal variation in these inflows, and in surface water quality. Multiple studies have been conducted to link impaired surface water quality, as indicated by fecal indicator bacteria (FIB) levels, with specific point and non-point sources, and to understand the processes that drive contaminant movement from sources to surface water bodies. For instance, a study that intensively surveyed water quality within an Iowa watershed found that levels of *E. coli*, a FIB, were positively associated with agricultural land use near the sampling site (Pandey et al., 2012). Conversely, a study that monitored FIB levels in Pennsylvania streams found that FIB levels were positively and strongly associated with urban proximal land use, but only weakly associated with agricultural land use at the watershed scale (Duris et al., 2013). Multiple studies have also shown that stream sediments can act as in-channel stores for FIB, and that disturbance of these sediments (e.g., during storm events) can re-introduce the bacteria into the water column (Nagels et al., 2002; Muirhead et al., 2004). However, resuspension may be affected by waterway-specific characteristics, including flow rate and sediment size (Zhou et al., 2017; Fluke et al., 2019). Similarly, potential point and non-point sources of fecal contamination will vary between and within waterways. As a result, it is difficult to adapt the findings of these and other studies to develop effective, practicable strategies for mitigating fecal contamination for individual waterways since (i) potential sources can differ substantially between waterways, or even between reaches in the same waterway, (ii) fecal contamination is driven by complex processes that are affected by the environmental heterogeneity inherent to freshwater environments (Weller et al., 2020a). Many of the streams that pass-through Syracuse, New York (NY) and flow into Onondaga Lake illustrate the difficulties associated with identifying strategies for optimally mitigating fecal contamination. Three of these streams, Onondaga Creek, Ley Creek and Harbor Brook (**Figure 1**), are listed as impaired by the NY State Department of Environmental Conservation (https://www.dec.ny.gov/chemical/31290.html) due to excess microbial, nutrient, ammonia and/or turbidity levels, and have been the focused of remediation efforts for more than 30 years (Ganley et al., 1982). Onondaga Creek was identified in the mid-1990s as being responsible for most of the fecal coliform loads and combined sewer overflow (CSO) flows entering Onondaga Lake (Steven, 1996), and the dominant source for sediments entering the lake (Prestigiacomo et al., 2007). Despite long-term monitoring efforts and attempts at remediation in these waterways (e.g., by reducing CSOs; Ganley et al., 1982; Effler et al., 2009; Walker et al., 2013), impairment remains a problem, and as a result, recreational uses, including swimming and fishing, are limited or prohibited in these waterways. For example, monitoring efforts between 2000 and 2007 showed that fecal coliform concentrations in Onondaga Creek exceeded NY State water quality standards on 16% (34/215) and 75% (162/215) of dry weather days at a rural and urban sampling site, respectively (https://static.ongov.net/WEP/AMP/DATA_DOWNLOAD_AREA/NYSDEC%20Data%20Request/Microbial%20Trackdown%20Study/MTS%20Reports/MTS%20Phase%203%20Final%20Report/MTS%20Phase%203%20Final%20Report_4-11-19.pdf). Such findings indicate that combined CSO discharge is not solely responsible for bacteria release to Onondaga Creek, or other Syracuse-area waterways. Data from other monitoring efforts in the area supported this conclusion and found that fluctuations in fecal coliform bacteria levels could not be explained by precipitation-driven discharges alone, and

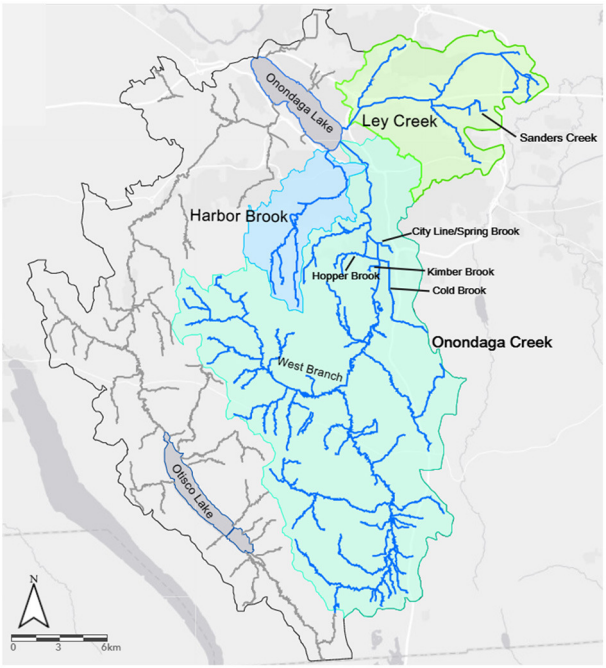


FIGURE 1 | Location and watersheds for the three streams (Harbor Brook, Ley Creek, and Onondaga Creek), and the tributaries of these streams, that were sampled in the present study. Note: Cold Brook and City Line/Spring Brook were treated as a single waterway.

that there appeared to be a strong spatial signal in the data with bacterial concentrations being substantially higher within, compared to outside, the Syracuse city limits (https://www.oei2.org/microbial-trackdown-study/). Indeed, dry weather release was identified as a significant factor, and it was hypothesized that dry weather discharge was the product of an aging sewer system due to broken pipes, poor cross linkages, or illicit connections. Such a hypothesis is consistent with the literature (Ahmed et al., 2005; Sowah et al., 2014, 2017; Weller et al., 2020b), as multiple studies have linked septic system density, sewer system age, and distance to discharge sites/septic systems with ground and surface water impairment. Therefore, this study aimed to (i) identify potential sources of fecal contamination in urban and rural reaches of Onondaga Creek, Ley Creek, and Harbor Brook (Syracuse, NY area waterways), and (ii) characterize associations between microbial indicators of fecal contamination and environmental factors (e.g., weather, nutrient levels), and (iii) provide recommendations on how to prioritize mitigation efforts to maximize restoration efforts.

MATERIALS AND METHODS

Study Design and Water Collection Sites

The data reported here were generated as part of a three-phase collection. During the first phase of sampling (No of Samples = 189; June 11 to October 8, 2014), sampling occurred irrespective of ambient or antecedent weather conditions; phases two and three were conducted to reflect dry weather (i.e., sampling was only conducted if <0.20 cm of rain fall during the preceding 48 h). Three waterways were the main focus of sampling (**Figure 1**): Ley Creek (No. of samples tested for ≥1 microbial targets (N) = 291; Number of sites where these samples were collected (S) = 58], Onondaga Creek (N = 1,700; S = 109), and Harbor Brook (N = 612; S = 38; **Figure 2**). Additionally, 1 tributary of Ley Creek [Sanders Creek (N = 1; S = 1)], and 4 tributaries of Onondaga Creek [Cold Brook (N = 34; S = 6), Hopper Brook (N = 48; S = 8), Kimber Brook (N = 4; S = 1), and West Branch (N = 126; S = 3); **Figure 1**] were sampled.

All sampled waterways follow a gradient from rural areas dominated by agriculture and forested to outlets on Onondaga Lake, which are located in the urban core of Syracuse, NY. The watersheds for the three waterways are 35 km^2 (Harbor Brook), 76 km^2 (Ley Creek), and 285 km^2 (Onondaga Creek), representing 5, 10, and 49% of the Onondaga Lake drainage area, respectively (**Figure 1**). Sample site locations along each waterway were selected to ensure accessibility for samplers, and to ensure both rural and urban sites were represented. A subset of samples (N = 31) were collected for microbial source-tracking marker (MST) analysis between August 2015 and August 2017 from Cold Brook (N = 8), Harbor Brook (N = 10), Hopper Brook (N = 10), and Onondaga Creek (N = 3). The sites used to collect MST data were selected to reflect sites where (i) more than one fecal source appeared to be present, including human and non-human (e.g., agricultural inputs, wetlands or ponds with large waterfowl populations), (ii) where past source-tracking efforts have been unable to identify sources of fecal contamination, and (iii) where fecal coliforms were consistently high during routine monitoring efforts. Samples from these sites were only tested for MST presence if fecal coliforms in a concomitantly collected sample were ≥ 200 cfu/100 mL.

To characterize the impact of land use in the area immediately around each sample collection site the proportion of land within 122, 366, and 1,098 m of each site under pasture-hay, cropland, forest-wetland, and developed (>20% impervious) cover was determined [for an example of the code used see: https://github.com/wellerd2/Calculating-land-use-land-cover-and-landscape-structure-parameters] (Rehmann and Soupir, 2009). Briefly, land use data was extracted from the National Land Cover database (https://www.mrlc.gov/national-land-cover-database-nlcd-2016) for the relevant year, and buffer distance. Buffer distances were selected based on Leafy Green Marketing Agreement (https://lgmatech.com/wp-content/uploads/2020/08/CA-LGMA-Metrics-August-2020_Final_Clean_9-18-20.pdf) recommended buffers from the location of any adjacent land uses that are likely to present a food safety risk (Table 7 from Crop Land and Water Sources Adjacent Land Uses). For example, the recommended distance from cropland to a concentrated animal feeding operation (CAFO) with >1,000 head is ∼1,200 feet (366 m). We selected this buffer, and buffers 1/3 smaller (122 m) and larger (1,098 m) to obtain land use data. Additionally, to compare water quality between rural vs. urban areas, sites were characterized as urban if they were within the boundaries of the city of Syracuse (**Figure 2**). It is also important to note that four of the five sites to the north of the city limits, while in a predominantly developed landscape are in

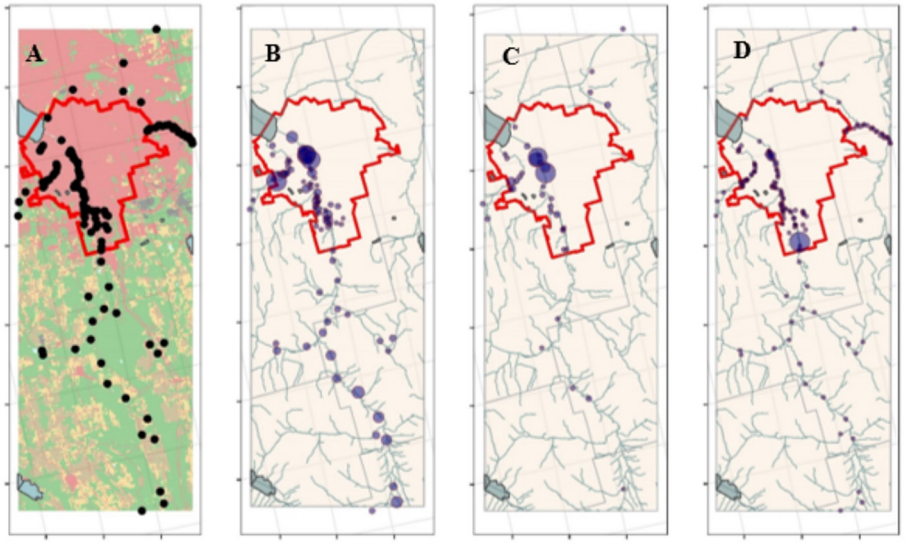


FIGURE 2 | (A) Developed (red), forest-wetland (green), pasture (yellow), crop (brown), and water (blue) cover in the study area; dots show sampling sites. In **(B–D)**, the size of the dots is proportional to *E. coli, Enterococcus*, and fecal coliform levels, respectively, in each sample. Since this was a longitudinal study, locations were jittered to facilitate visualization of levels in samples collected at the same site but different times. The red lines show the Syracuse, NY city limits; sites inside these limits were considered urban; all other sites were considered rural.

patches of non-developed land cover. However, recognizing that dichotomizing into rural vs. urban oversimplifies that complex landscape around Syracuse, separate analyses using the amount of developed cover around the site were also implemented.

Sample Collection and Analysis

Each site was sampled up to twice a month between May and November, with samples being collected annually from 2008 to 2017 (except 2010 or 2011). Sampling within each waterway was performed in a downstream to upstream manner to avoid sampling the same slug of water. All samples were collected at the same time, varying volumes outlined were collected in different bottles for the various assays including microbiological, physiochemical, and nutrient/sediment analyses.

For microbiological analysis, direct grab samples were collected from the centerline of the waterway by submerging a sterilized 150-mL plastic bottle just below the water surface. Approximately 125 mL of each water sample was transferred to a sterilized and pre-preserved (with sodium thiosulfate) 150-mL coliform bottle. Approximately 10 mL of each water sample was transferred to a 25-mL glass vial for chloride analysis (see below). When samples were collected for MST analysis, a separate 1-L grab sample was also obtained using the direct grab method. Total coliform, fecal coliform, *E. coli* and/or *Enterococcus* levels were enumerated using the 125-mL sample <8 h after collection. FIB and total coliform levels were enumerated by the Onondaga County Health Department (a certified laboratory) using NY State Department of Health standard methods (**Supplementary Table 1**). Separately, samples collected for MST detection were processed <24 h after collection by the Wadsworth Laboratory (NY State Department of Health, Albany, NY; New York State Department of Environmental Conservation, 2017). Briefly, after undergoing anerobic enrichment to select for members of *Bacteroidales* (Green et al., 2012, 2014, 2019), separate PCR-screens were performed, for host-specific avian (Lu et al., 2008; Green et al., 2012), canid (New York State Department of Environmental Conservation, 2017), human (Bernhard and Field, 2000a; Shanks et al., 2007, 2009; Green et al., 2014), and ruminant fecal markers (Bernhard and Field, 2000b; **Supplementary Table 1**).

At each sampling event, data were also collected on physicochemical water quality parameters and weather factors. Data on water temperature, pH, dissolved oxygen, specific conductance, and turbidity were collected *in situ* using a YSI 650 MDS handheld device (YSI Inc., Yellow Springs, OH, USA) equipped with a 6600 or 6820-V2 multi-parameter water quality probe. Chloride was measured *in situ* using a Hach™ pocket colorimeter. To enumerate nutrient and sediment levels, a separate 7.6 L sample was collected from the same location, and at the same time as the sample used to enumerate FIB levels. The 7.6 L sample was divided into separate 1 and 0.5 L aliquots for nutrient and sediments analysis, respectively (**Supplementary Table 1**). It is important to note that not all parameters were measured in all samples (e.g., due to equipment malfunctions, availability of staff, time conflicts, feasibility of obtaining samples; S2). Weather data (air temperature, rainfall, relative humidity, and wind speed) for each sampling event were obtained from the Network for Environmental and Weather Applications (NEWA). Data was obtained using the weather station

closest to each sampling site location (http://newa.cornell.edu/).

Statistical Analysis

All analyses were performed using R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). For all analyses, all data were log_{10}-transformed. While 2,816 samples were collected, not all analytes were enumerated in all samples; as a result, 5-point summaries and proportion of missingness were calculated for each variable (**Supplementary Tables 2**, **5**, **6**).

Bayesian mixed models were developed to (i) characterize spatial and temporal patterns in log_{10} FIB levels, and human and ruminant MST markers presence, and (ii) identify environmental factors associated with log10 FIB levels or probability of MST detection. Log-linear Bayesian mixed models were developed when the outcome was log10 *E. coli, Enterococcus*, fecal coliform, or total coliform levels, while Bayesian linear mixed models were developed when the outcome was probability of human or ruminant MST detection. To characterize spatial and temporal patterns, models were developed with fixed effects of elevation, latitude, longitude, rurality (i.e., if the sample was collected inside or outside Syracuse, NY city-limits), and waterway, and of year and month, respectively. For models with a spatial fixed effect, month was included as a random effect, while waterway was included as a random effect for models with temporal fixed effects.

To identify relationships between fecal indicators and environmental factors, Bayesian mixed models were developed with random effects of month and waterway. These models included a single fixed effect for land use (proportion of land within 122, 366, or 1,098 m of the sampling site under crop, pasture-hay, forest-wetland or developed cover), weather (air temperature at sampling and 0–3 d before sampling, rainfall 0–1, 1–2, and 2–3 d before sampling, and wind speed 0–1 d before sampling), or water quality parameters (log10 chloride, conductivity, total dissolved solids, nitrate, total organic carbon, total phosphorous, salinity, total suspended solids, and turbidity levels as well as pH). Since not all analytes were measured in all samples, models were implemented only if there were at least 20 pairwise observations for both the microbial target and environmental factor of interest (**Supplementary Tables 5**, **6**). Due to the large number of samples tested for fecal coliforms ($N = 2{,}658$) compared to the other FIBs, fecal coliform models could be implemented for all covariates considered here. Conversely, *Enterococcus* models could not be implemented with chloride, total dissolved solids, total organic carbon, or total phosphorous as the explanatory variables.

Models were fit using the brms package, uninformative priors, 3 chains, and thinning set to 10 (Bürkner, 2017, 2018). While the number of iterations per chain was set to 5,000 (burn-in of 2,500) for most models, there were convergence issues for a subset of models (Bürkner, 2017, 2018). For those models with convergence issues, the number of iterations per chain and burn-in iterations were increased per the package author's recommendation (Bürkner, 2017, 2018). The median, mean, maximum a posteriori (MAP), and 89% credibility interval for the effect estimates were calculated using the bayestestR package (Makowski et al., 2019a). Bayesian credibility intervals are not interpreted in the same way as frequentist confidence intervals. Instead the 89% credibility interval indicates the highest probability portion of the posterior distribution of possible effect estimates and should be interpreted as: "Given the observed data, the effect estimate has a 89% probability of falling between x and y" (Makowski et al., 2019a).

Unlike frequentist regression models, Bayesian models are not based on assessing statistical significance; instead, Bayesian models offer a probabilistic (more flexible) view of the parameters and their corresponding uncertainty. As such, Bayesian analyses are more robust to small sample size, do not rely on a researcher-specified *p*-value threshold, and are less prone to Type I errors compared to frequentist approaches. Therefore, rather than concluding that an association is present when the *P*-value is below a certain threshold, conclusions can be drawn by determining if the parameter is outside of a range of practically negligible effect. This measure, called the ROPE percentage, indicates the magnitude of effect, and is calculated by determining the percent overlap between the 89% credibility interval and the range of practically no effect. Thus, the closer the ROPE percentage is to 0, the more confident we can be that the given factor has a substantial effect on FIB levels or probability of MST detection. Specifically, we use the following cutoffs for ROPE interpretation: >99% ∼ negligible effect, >97.5% ∼ probably negligible effect, between 2.5 and 97.5% ∼ uncertain effect, <2.5% ∼ non-negligible effect, <1% ∼ significant effect (Makowski et al., 2019a). Associations in Bayesian regression can also be assessed using the probability of direction (PD), which is an index of if a positive or negative effect exists regardless of if that effect is negligible or non-negligible. The PD correlates strongly with frequentist *P*-values with PD values near 1.0 indicating greater certainty that the effect of the factor is truly positive or negative (i.e., indicates confidence in the direction of the association (Makowski et al., 2019a,b). Specifically, PD values of 0.95, 0.975, 0.995, and 0.9995 correspond to two-sided frequentist *P*-values of 0.10, 0.05, 0.01, and 0.001, respectively (Makowski et al., 2019a,b). The ROPE and PD measures are independent, and a factor can have a high PD, but low ROPE value, indicating high certainty in the direction of a negligible effect. The inverse would indicate that the parameter has a non-negligible (or substantial) effect, but the direction of that effect is unclear. Overall, PD and ROPE can be used in conjunction to make a statement of the form, "the effect of the given factor has a probability of PD of being negative/positive, and can be considered as significant/nonsignificant (Makowski et al., 2019a)." Measures of (PS) were also calculated, and indicate the probability that the parameter's effect is above a given threshold, representing a negligible effect in the median's direction; this is a unidirectional equivalence test that indicates if the effect is both non-negligible, and in a given direction (Makowski et al., 2019a,b). Values should be larger than 0.5 to indicate practical significance; a cut-off of 0.75 was used here to be conservative. Based on the recommended cut-offs and interpretations for PD, PS, and ROPE, we determined that with PS >0.50, PD >0.75, and ROPE <0.25 warranted reporting here.

RESULTS AND DISCUSSION

For the 2,816 samples collected, microbial and physicochemical water quality parameters were summarized in **Supplementary Table 2**. In total, *E. coli*, *Enterococcus*, total coliform, and fecal coliform levels were enumerated in 281, 288, 96, and 2,658 samples, respectively. The average log10 concentration (CFU/100-mL) of *E. coli*, *Enterococcus,* total coliform, and fecal coliform levels were 2.2 [Standard Deviation (SD) = 0.9; Range = −0.3, 4.2), 2.0 (SD = 0.9; Range = −0.3, 5.0), 3.0 (SD = 1.0; Range = −0.3, 4.2), and 2.5 (SD = 0.8; Range = −0.4, 7.2), respectively (**Supplementary Table 2**). Similar summaries of physicochemical water quality in the present study can be found in **Supplementary Table 2**. For example, average total suspended solids was 7.0 mg/L (SD = 110.5; Range = 1.0, 2,077.0; **Supplementary Table 2**).

Lack of Association Between MST Markers and FIB May Indicate Multiple Sources of Contamination

Fecal coliform levels were positively associated with *Enterococcus* and total coliform levels (**Table 1**, **Supplementary Table 3**). While *E. coli* levels were associated with levels of other FIB, land use and spatial factors, such as rurality, that were positively associated with fecal coliform and *Enterococcus* levels tended to be negatively associated with *E. coli* levels, and vice versa (**Tables 1**–**3**). Of the 31 samples tested for host-specific microbial source tracking markers (MST), 22 (71%) and 11 (35%) samples were positive for human and ruminant markers. Detection of human markers were positively associated with detection of ruminant markers (PD = 0.91; PS = 0.80); however, the magnitude of this effect is uncertain (ROPE = 0.16; **Supplementary Table 3**). Additionally, neither marker was associated with levels of any of the other microbial targets considered here (PS <0.50, PD <0.95, and ROPE >0.025). This may indicate that (i) human and ruminant fecal contamination were coming from the same sources, (ii) that sources of FIB were different from sources of human and ruminant contamination, and/or (iii) that other host-specific sources of contamination not tested here were driving fecal contamination in the present study. For example, past studies have found that FIBs, such as *E. coli* can naturalize and survive in non-host environments (e.g., algal mats, water, soil; (Hendricks, 1967; Byappanahalli et al., 2003; Ksoll et al., 2007; Nautiyal et al., 2010; Goto and Yan, 2011; NandaKafle et al., 2018)). As a result, FIB levels may not reflect recent fecal contamination events, resulting in a lack of association between host-specific markers and FIB levels. It is also interesting to note that the findings reported here are inconsistent with a previous study conducted on Onondaga Creek (NY, US), which found a significant positive association between human marker detection and *Enterococcus* levels, and a significant inverse association between detection of human and ruminant markers (Green et al., 2019). This discrepancy may be due to the fact that this former study collected samples from Onondaga Creek in 2015, while the present study represents a larger number of waterways and a longer sampling period (2008–2017). Water quality is known to vary over space and time, and to vary by the spatial scale of analysis (e.g., Goyal et al., 1977; Pandey et al., 2012; Partyka et al., 2018a; Fluke et al., 2019; Weller and Jordan, 2020; Weller et al., 2020b). It may also indicate that processes driving fecal contamination differ for each of the watersheds sampled here, complicating detection of watershed-specific patterns of contamination (which were the focus of the previous study on Onondaga Creek, NY, US). However, since only three samples from Onondaga Creek and 31 samples total were tested for MST markers, examining watershed-specific patterns of association were outside the scope of the study reported here, and should be investigated in future studies on Onondaga Lake, and other watersheds.

Presence and Strength of Association Between FIB Levels and Adjacent Land Use Was Dependent on the Size of the Buffer Used to Calculate Land Use Parameters

The distribution of FIB was non-uniform with *Enterococcus* and fecal coliform levels being higher and *E. coli* levels being lower, on average, within the city of Syracuse, NY, compared to outside it (i.e., urban vs. rural samples; **Figure 1**; **Table 2**); neither the probability of detecting human or ruminant MST markers were associated with rurality (**Supplementary Table 3**). In fact, while *E. coli* levels were, on average, 0.35 log10 CFU/100-mL (89% Credibility Interval [CI] = −0.61, 0.00; PF = 0.94; ROPE = 0.10) lower, and *Enterococcus* and fecal coliform levels were, on average, 0.32 log CFU/100-mL (CI = 0.13, 0.52; PD = 0.99; ROPE < 0.01) and 0.51 log CFU/100-mL (CI = 0.44, 0.59; PD = 1.00; ROPE < 0.01) higher in samples collected from urban, compared to rural sites. Since latitude increased and elevation decreased along the rural-urban gradient, it is not surprising that fecal coliforms and *Enterococcus* were positively associated with latitude and negatively associated with elevation (Syracuse, NY is farther north and at a lower elevation, than surrounding rural areas), while the reverse patterns were observed for *E. coli*, and human and ruminant MST markers (**Table 2**). Overall, these patterns are also consistent with the land use associations observed here. Specifically, the only association between land use and *E. coli* levels was a positive association with pasture cover (**Table 3**), which mirrors the positive association between probability of detecting human and ruminant MST markers and pasture cover (**Supplementary Table 4**). Conversely, *Enterococcus* and fecal coliform levels were negatively associated with agricultural land uses and forest-wetland cover, but positively associated with developed cover (**Table 3**). Interestingly, while the presence and magnitude of the association between *E. coli* levels and pasture cover appeared robust to buffer size, the associations between *Enterococcus* and fecal coliform levels and pasture cover, and between ruminant detection and pasture cover were only evident for 1,098 and 122 m buffers, respectively (**Table 3**, **Supplementary Table 4**). The relationship between total coliform levels and cropland, human marker detection, and cropland, and ruminant marker detection and developed

TABLE 1 | Results of Bayesian mixed models where there was evidence of a positive or negative association [i.e., practical significance (PS) > 0.50, probability of direction (PD) > 0.75, and overlap with region of practical equivalence (ROPE) < 0.25] between log_{10} FIB levels, and either temporal, water quality, or weather features.

Factor	Effect estimate	89% Credibility interval		PD	PS	ROPE
		Lower	Upper			
E. coli						
Average air temp. 0–3 d (°C)	0.19	0.06	0.31	0.98	0.88	0.08
Conductivity (umHos/cm)	0.21	0.11	0.31	1.00	0.96	<0.01
Enterococcus (log10 CFU/100-mL)	0.39	0.15	0.64	1.00	0.98	<0.01
Fecal coliforms (log10 CFU/100-mL)	0.75	0.70	0.79	1.00	1.00	<0.01
Month (November = reference)[a]						
April	−0.03	−0.43	0.35	0.87	0.37	0.37
July	0.62	0.26	1.00	1.00	0.98	<0.01
August	0.25	−0.10	0.59	0.87	0.76	0.21
September	0.41	0.06	0.80	0.96	0.91	0.03
October	−0.19	−0.57	0.18	0.78	0.63	0.27
Nitrate (mg/L)	0.16	0.06	0.26	0.99	0.85	0.11
Salinity (ppt)	0.20	0.11	0.31	1.00	0.95	<0.01
Total coliforms (log10 CFU/100–mL)	0.86	0.69	1.03	1.00	1.00	<0.01
Total rainfall 0–1 d (mm)	0.14	0.04	0.27	0.96	0.71	0.24
Turbidity (NTU)	0.26	0.19	0.33	1.00	1.00	<0.01
Enterococcus						
Air temp. at sampling (°C)	0.19	0.06	0.31	0.99	0.85	0.10
Average air temp. 0–3 d (°C)	−0.20	−0.46	0.10	0.88	0.74	0.24
Fecal coliforms (log10 CFU/100-mL)	0.75	0.69	0.81	1.00	1.00	<0.01
Month (November = reference) [a]						
July	0.68	0.41	0.98	1.00	1.00	<0.01
August	0.46	0.19	0.74	1.00	0.98	<0.01
September	0.49	0.21	0.75	1.00	0.98	<0.01
October	0.85	0.40	1.22	1.00	1.00	<0.01
Nitrate (mg/L)	0.68	0.08	1.24	0.96	0.95	0.01
pH	−0.30	−0.42	−0.16	1.00	0.99	<0.01
Total rainfall 2–3 d (mm)	0.23	0.09	0.37	1.00	0.92	0.02
Total suspended solids (mg/L)	1.03	−0.48	2.39	0.89	0.87	0.04
Turbidity (NTU)	0.20	0.08	0.33	0.99	0.89	0.05
Wind speed 0–1 d (average km/h)	0.30	0.18	0.41	1.00	1.00	<0.01
Year	0.22	0.16	0.28	1.00	1.00	<0.01
Fecal coliforms						
Conductivity (umHos/cm)	0.18	0.15	0.21	1.00	1.00	<0.01
Month (November = reference)[a]						
March	−0.56	−1.10	−0.03	0.96	0.92	0.04
April	−0.51	−0.74	−0.29	1.00	1.00	<0.01
May	−0.58	−0.78	−0.37	1.00	1.00	<0.01
June	−0.06	−0.23	0.09	0.72	0.33	0.67
July	0.08	−0.07	0.20	0.80	0.39	0.64
August	0.03	−0.11	0.15	0.61	0.18	0.83
September	0.02	−0.12	0.15	0.62	0.17	0.85
October	−0.23	−0.37	−0.09	0.99	0.92	0.03
Nitrate (mg/L)	0.33	0.26	0.41	1.00	1.00	<0.01
Salinity (ppt)	0.23	0.20	0.27	1.00	1.00	<0.01
Total coliforms (log10 CFU/100–mL)	1.04	0.88	1.18	1.00	1.00	<0.01
Total dissolved solids (mg/L)	0.39	0.29	0.49	1.00	1.00	<0.01
Total organic carbon (mg/L)	1.07	0.71	1.48	1.00	1.00	<0.01

(Continued)

TABLE 1 | Continued

Factor	Effect estimate	89% Credibility interval		PD	PS	ROPE
		Lower	Upper			
Total phosphorous (mg/L)	0.48	0.42	0.54	1.00	1.00	<0.01
Total suspended solids (mg/L)	0.17	0.10	0.23	1.00	0.93	0.01
Turbidity (NTU)	0.25	0.22	0.29	1.00	1.00	<0.01

[a]*Note for multi-level categorical variables, if any level met the PS, PD, and ROPE thresholds [i.e., PS > 0.50, PD > 0.75, and ROPE < 0.25], results for all levels of the variable are reported. Variables not shown in the table, were not significant, and were not shown.*

and forest-wetland cover also seemed dependent on buffer size. Thus, while there were associations between specific land uses and each fecal indicator, these associations appeared dependent on the scale of analysis. This is not unexpected, as a previous study that examined nutrient pollution in Maryland waterways also found that spatial scale of analysis affected model results (Weller and Jordan, 2020). Similarly, a study in Arkansas, US that characterized associations between *E. coli* levels and riparian land use found that the size and area of the riparian buffer affected the change-point (i.e., the point within the land use parameter that demarcates a significant change in *E.coli* levels; (Scott et al., 2017)). Since the buffers used here were selected to replicate the recommended distance between large livestock operations (>1,000 head) and agricultural water sources in the Leafy Green Marketing Agreement (i.e., 366 m), this finding has direct implications for produce safety, and highlights the importance of considering spatial scale (e.g., using 122, 366, or 1,098 m buffers), when developing strategies and/or guidance for identifying potential fecal contamination risks.

Regardless of the scale for analysis, the study reported here identified strong associations between adjacent land use and fecal contamination, which is consistent with past studies conducted in the Northeastern US, and other areas (Shiels and Guebert, 2010; Wilkes et al., 2011; Duris et al., 2013; Scott et al., 2017; Wu, 2019; Weller et al., 2020b). For instance, we found that *E. coli* levels increased by 0.30 (89% CI = 0.17, 0.39; PD = 1.00; ROPE <0.01) log10 CFU/100-mL for each percent increase in pasture cover within 1,098 m of the sampling site (**Table 1**); this positive linear association is consistent with the findings of the aforementioned Arkansas (US) study (Scott et al., 2017). Conversely, *Enterococcus* and fecal coliform levels both decreased as pasture cover increased, but increased as developed cover increased (**Table 1**). Given the pattern of land use associations observed here, our findings suggest that the sources of *Enterococcus* and fecal coliform contamination differ from sources of *E. coli* or MST markers, with the former coming from developed areas, and the latter from agricultural areas. While we would expect human markers to be associated with urban areas and developed land uses, the association between human markers and agricultural land uses and rural areas may indicate the presence of human fecal contamination sources in rural hinterlands. For example, areas within Syracuse are on municipal water and sewer, while the rural areas to the south largely utilize septic; as such one source of fecal contamination in rural areas may be failing septic. While such a hypothesis is supported by past studies conducted in Queensland, Australia (Ahmed et al., 2005), Georgia, US (Sowah et al., 2014, 2017), Michigan, US (Verhougstraete et al., 2015), and New York, US (Rao et al., 2015), data on the age, location, and density of rural fecal contamination sources, such as septic systems, were not available and this could not be investigated in the present study. This highlights a specific research gap, the impact and source of rural fecal contamination, that future studies in the Onondaga Creek Watershed should address.

Despite differences in land use associations for the various fecal indicators, levels of all indicators were higher in Onondaga Creek, compared to the other waterways sampled (**Table 2**, **Supplementary Table 3**). For instance, *E. coli* levels in Kimber Brook were almost 2 log10 (89% CI = −3.29, −0.31; PD = 0.97; ROPE<0.01) lower, on average, than levels in Onondaga Creek. These findings support previous studies' conclusions that Onondaga Creek was responsible for most of the bacterial and sediment loads entering Onondaga Lake (Steven, 1996; Prestigiacomo et al., 2007). Given the consistency across fecal indicators, including the host-specific markers, this suggests that, independent of land use near each sampling site, fecal contamination patterns varied between the waterways sampled here with Onondaga Creek being more impaired than the other waterways. Thus, while this finding is not surprising, it is important for guiding efforts to mitigate fecal contamination of surface waters that flow into Onondaga Lake. Specifically, since waterways sampled here drain into Onondaga Lake, this finding suggests that mitigation efforts should target Onondaga Creek, as opposed to Ley Creek, Harbor Brook, or Hopper Brook.

Nutrient and Sediment Pollution Were Associated With FIB Contamination, Which May Indicate Shared Contamination Sources or Similar Processes for Transporting Contaminants From Source to Waterway

All FIB tested for in the present study were positively associated with nutrient (e.g., nitrate, phosphorous), chemical (e.g., salinity, conductivity), or sediment (e.g., turbidity, total suspended solid) contamination. For example, the concentration of *E. coli*, *Enterococcus,* and fecal coliforms increased by 0.16 (CI = 0.06, 0.26; PD = 0.99; ROPE = 0.11), 0.68 (CI = 0.08, 1.24; PD = 0.96; ROPE = 0.01), and 0.33 (CI = 0.26, 0.41; PD = 1.00; ROPE < 0.01) log10 CFU/100-mLs, respectively, for each

TABLE 2 | Results of Bayesian mixed models where there was evidence of a positive or negative association [i.e., practical significance (PS) > 0.50, probability of direction (PD) > 0.75, and overlap with region of practical equivalence (ROPE) < 0.25] between log10 FIB levels, and spatial features.

Factor	Effect estimate	89% Credibility interval		PD	PS	ROPE
		Lower	Upper			
E. coli						
Latitude (°)	−0.22	−0.34	−0.11	1.00	0.93	<0.01
Urban site (Rural = Reference)	−0.35	−0.61	0.00	0.94	0.85	0.10
Waterway (Onondaga Creek = reference)[a]						
Cold Brook	−1.03	−1.73	−0.28	0.99	0.98	<0.01
Harbor Brook	−0.30	−0.64	0.05	0.91	0.83	0.13
Hopper Brook	−0.44	−1.02	0.22	0.87	0.80	0.12
Kimber Brook	−1.80	−3.29	−0.31	0.97	0.96	<0.01
West Branch	−0.22	−0.57	0.15	0.83	0.70	0.24
Enterococcus						
Elevation	−0.35	−0.53	−0.15	1.00	0.98	<0.01
Latitude (°)	0.16	0.06	0.26	0.99	0.81	0.15
Urban site (Rural = reference)	0.32	0.13	0.52	0.99	0.95	<0.01
Waterway (Onondaga Creek = reference)[a]						
Harbor Brook	−0.28	−0.50	−0.06	0.99	0.91	0.04
Ley Creek	−0.35	−0.58	−0.07	0.99	0.93	0.03
West Branch	−0.60	−1.05	−0.13	0.98	0.96	<0.01
Fecal coliforms						
Elevation	−0.20	−0.24	−0.17	1.00	1.00	<0.01
Latitude (°)	0.14	0.11	0.17	1.00	0.98	<0.01
Urban site (Rural = Reference)	0.51	0.44	0.59	1.00	1.00	<0.01
Waterway (Onondaga Creek = reference)[a]						
Cold Brook	−0.11	−0.39	0.14	0.77	0.53	0.42
Harbor Brook	−0.10	−0.18	−0.03	0.98	0.51	0.48
Hopper Brook	0.48	0.28	0.75	1.00	1.00	<0.01
Kimber Brook	−0.49	−1.20	0.28	0.86	0.81	0.10
Ley Creek	−0.23	−0.31	−0.10	1.00	0.98	<0.01
Sanders Creek	−0.14	−1.67	1.37	0.56	0.51	0.10
West Branch	−0.56	−0.70	−0.39	1.00	1.00	<0.01

[a]*Note for multi–level categorical variables, if any level met the PS, PD, and ROPE thresholds [i.e., PS > 0.50, PD > 0.75, and ROPE < 0.25], results for all levels of the variable are reported. Variables not shown in the table, were not significant, and were not shown.*

log10 increase in nitrate levels (**Table 1**, **Supplementary Table 3**). Similar patterns were observed for conductivity, salinity, total organic carbon, total phosphorous, total dissolved solids, total suspended solids, and turbidity (**Table 1**). In fact, *Enterococcus* and fecal coliform levels increased by ∼1 log for each log10 increase in total suspended solid levels (89% CI = −0.48, 2.39; PD = 0.89; ROPE = 0.03) and for each log10 increase in total organic carbon levels (89% CI = 0.71, 1.48; PD = 1.00; ROPE < 0.01), respectively (**Table 1**). The discovery of positive associations between physicochemical and microbial water quality parameters is consistent with the previous studies (Soupir et al., 2010; Viau et al., 2011; Wilkes et al., 2011; Rao et al., 2015; Verhougstraete et al., 2015). For example, a survey of 64 Michigan, US rivers found that nutrient levels were able to account for 48% of variance in *E. coli* levels (Verhougstraete et al., 2015). Additionally, a study that identified factors associated with detection of parasites of fecal origin in an Ontario, Canada watershed, found that *Giardia* detection was associated with both conductivity and total phosphorous levels (Wilkes et al., 2011). Viau et al. surveyed the microbial and physicochemical quality of Hawaiian streams, and found evidence of positive associations between multiple microbial and physicochemical parameters, including *E. coli* and total phosphorous levels, and between enterococci levels and turbidity (Viau et al., 2011). In fact, a positive association has been repeatedly identified between microbial water quality and turbidity levels in studies conducted in different water types (e.g., reservoirs, canals, ponds), regions (e.g., the Northeastern and Southeastern US, Ecuador), and where different study designs were used (e.g., microbial targets, sampling strategies; (Viau et al., 2011; Francy et al., 2013; Rao et al., 2015; Topalcengiz et al., 2017; Partyka et al., 2018b; Weller et al., 2020a)). Given the reproducibility of the turbidity-microbial water quality relationship, multiple studies have suggested using turbidity as a supplemental indicator for

TABLE 3 | Results of Bayesian mixed models where there was evidence of a positive or negative association [i.e., practical significance (PS) > 0.50, probability of direction (PD) > 0.75, and overlap with region of practical equivalence (ROPE) < 0.25] between log10 FIB levels, and the percent of buffer around each sampling site under crop, developed, forest–wetland or pasture cover.

Factor	Buffer distance (m)	Effect estimate	89% Credibility interval		PD	PS	ROPE
			Lower	Upper			
E. coli							
Pasture (%)							
	122	0.25	0.17	0.34	1.00	1.00	<0.01
	366	0.25	0.16	0.35	1.00	0.99	<0.01
	1,098	0.30	0.17	0.39	1.00	1.00	<0.01
Enterococcus							
Cropland (%)							
	122	−0.35	−0.64	−0.12	0.98	0.94	<0.01
	366	−0.36	−0.55	−0.20	1.00	0.99	<0.01
	1,098	−0.24	−0.36	−0.10	1.00	0.96	<0.01
Developed (%)							
	122	0.33	0.23	0.44	1.00	1.00	<0.01
	366	0.32	0.23	0.42	1.00	1.00	<0.01
	1,098	0.34	0.23	0.43	1.00	1.00	<0.01
Forest–Wetland (%)							
	122	−0.22	−0.33	−0.12	1.00	0.97	<0.01
	366	−0.26	−0.34	−0.14	1.00	0.99	<0.01
	1,098	−0.28	−0.38	−0.17	1.00	1.00	<0.01
Pasture (%)							
	1,098	−0.20	−0.29	−0.11	1.00	0.96	<0.01
Fecal coliforms							
Developed (%)							
	122	0.23	0.19	0.27	1.00	1.00	<0.01
	366	0.27	0.23	0.30	1.00	1.00	<0.01
	1098	0.27	0.23	0.30	1.00	1.00	<0.01
Forest–Wetland (%)							
	122	−0.19	−0.22	−0.15	1.00	1.00	0.00
	366	−0.28	−0.32	−0.25	1.00	1.00	<0.01
	1,098	−0.27	−0.31	−0.24	1.00	1.00	<0.01
Pasture (%)							
	1,098	−0.15	−0.19	−0.11	1.00	0.97	<0.01

identifying when a waterway may be contaminated by FIB. Overall, the positive association between microbial water quality (e.g., FIB levels), and nutrient (e.g., nitrate) and sediment (e.g., turbidity) levels in this and other studies, may indicate that similar processes are driving microbial, sediment and nutrient contamination. Such a conclusion is supported by literature on sources, fate, and transport of microbial, nutrient, and sediment contaminants. For instance, it is well-established in the scientific literature that rain events facilitate the release and transport of microbes, nutrients, and sediment from sources to surface waterways (Nagels et al., 2002; Muirhead et al., 2004; Zhou et al., 2017; Fluke et al., 2019). Indeed, rainfall antecedent of sampling events, and wind speed (a proxy for storm events) were associated with elevated FIB levels in this and previous studies (Francy et al., 2013). Additionally, the reported study also highlights that even though sampling in phases two and three of the present study only occurred during dry weather, bacterial pollution was still persistent in sampled waterways. Moreover, bacteria are often bound to particles, with one study reporting that 49% of all *Enterococci* detected were attached to soil or manure particles (Soupir et al., 2010), and streambed sediments have been repeatedly identified as an in-channel store for FIBs, whose resuspension can result in elevated FIB levels in the water column (Goyal et al., 1977; Muirhead et al., 2004; Rehmann and Soupir, 2009; Kim et al., 2010; Zhou et al., 2017). Thus, differential persistence of FIBs in streambed sediments, may account for the lack of correlation between FIB groups in the present study. From an applied perspective, these findings suggest that strategies to mitigate fecal contamination may concurrently reduce nutrient and sediment contamination; this is of specific interest for land managers in the Syracuse (NY, US) area, given the aforementioned interest in improving the quality

of water entering Onondaga Lake; as well as, in regions with similarly impaired waterways.

CONCLUSION

This study aimed to (i) identify potential sources of fecal contamination in urban and rural reaches of Onondaga Creek, Ley Creek, and Harbor Brook (Syracuse, NY area waterways), (ii) characterize associations between microbial indicators of fecal contamination and environmental factors (e.g., weather, nutrient levels), and (iii) provide recommendations on how to prioritize mitigation efforts to maximize restoration efforts. Briefly, the present study failed to find evidence of an association between FIB levels and host-specific MST markers. Additionally, land use and spatial factors, such as rurality, that were positively associated with fecal coliform and *Enterococcus* levels tended to be negatively associated with *E. coli* levels, and vice versa. These two findings may suggest multiples sources of fecal contamination in the sampled streams. Conversely, the study reported here also found evidence of strong associations between FIB levels, and nutrient and sediment levels in the sampled waterways, which may indicate shared contamination sources or similar processes for transporting contaminants from source to waterway. Given the aforementioned interest in improving the quality of water entering Onondaga Lake, this finding is of particular interest from an applied perspective, as it suggests efforts to reduce fecal contamination may reduce nutrient and sediment contamination. Lastly, this study found that FIB levels were not uniform over space, with levels being (i) higher in Onondaga Creek, compared to all other waterways, and (ii) strongly associated with adjacent land use. These findings highlight the importance of considering adjacent land use and stream-to-stream differences when developing strategies for managing fecal hazards in agricultural and recreational water when developing watershed management plans. Specifically, our finding suggests that mitigation efforts should focus on Onondaga Creek, as opposed to Ley Creek, Harbor Brook, or Hopper Brook. By focusing on the most impaired waterway entering Onondaga Lake, rather than all waterways, this type of targeted strategy may provide a way to maximize funds and water quality simultaneously.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SJ and EM conceived of the study, designed the data collection protocols, and led data collection efforts. DW and LS developed the statistical analysis plan. DW analyzed the data with support from TL. CM developed the data summaries. All authors contributed to manuscript preparation and revision.

FUNDING

This work was supported by Onondaga Environmental Institute (OEI). Manuscript preparation and data analyses were supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (NIH) under award number T32ES007271, and Specialty Crops Research Initiative project 2019-51181-30016 from the United States Department of Agricultural (USDA) National Institute of Food and Agriculture.

ACKNOWLEDGMENTS

We are grateful for the laboratory assistance of all technicians at the Onondaga County Health Department (Syracuse, NY) and Wadsworth Laboratory, NY State Department of Health (Albany, NY). Additionally, we are thankful for staff and experts at the Onondaga Environmental Institute.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.741676/full#supplementary-material

REFERENCES

Ackers, M., Mahon, B., Leahy, E., Goode, B., Damrow, T., Hayes, P., et al. (1998). An outbreak of *Escherichia coli* O157, H7. infections associated with leaf lettuce consumption. *J Infect Dis* 177:1588–1593. doi: 10.1086/515323

Ahmed, W., Neller, R., and Katouli, M. (2005). Evidence of septic system failure determined by a bacterial biochemical fingerprinting method. *J. Appl. Microbiol.* 98, 910–920. doi: 10.1111/j.1365-2672.2004.02522.x

Barton Behravesh, C., Mody, R. K., Jungk, J., Gaul, L., Redd, J. T., Chen, S., et al. (2011). 2008 outbreak of Salmonella Saintpaul infections associated with raw produce. *N. Engl. J. Med.* 364, 918–927. doi: 10.1056/NEJMoa1005741

Bernhard, A. E., and Field, K. G. (2000a). Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl. Environ. Microbiol.* 66, 1587–1594. doi: 10.1128/AEM.66.4.1587-1594.2000

Bernhard, A. E., and Field, K. G. (2000b). A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 66, 4571–4574. doi: 10.1128/AEM.66.10.4571-4574.2000

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-017

Byappanahalli, M. N., Shively, D. A., Nevers, M. B., Sadowsky, M. J., and Whitman, R. L. (2003). Growth and survival of *Escherichia coli* and enterococci populations in the macro-alga Cladophora (Chlorophyta). *FEMS Microbiol. Ecol.* 46, 203–211. doi: 10.1016/S0168-6496(03)00214-9

Calderón-Arrieta, D., Caudill, S. B., and Mixon, F. G. (2019). Valuing recreational water clarity and quality: evidence from hedonic pricing models of lakeshore properties. *Appl. Econ. Lett.* 26, 237–244. doi: 10.1080/13504851.2018.1458187

DeFlorio-Barker, S., Wade, T. J., Jones, R. M., Friedman, L. S., Wing, C., and Dorevitch, S. (2017). Estimated costs of sporadic gastrointestinal illness associated with surface water recreation: a combined analysis of data from NEEAR and CHEERS studies. *Environ. Health Perspect.* 125, 215–222. doi: 10.1289/EHP130

Duris, J. W., Reif, A. G., Krouse, D. A., and Isaacs, N. M. (2013). Factors related to occurrence and distribution of selected bacterial and protozoan pathogens in Pennsylvania streams. *Water Res.* 47, 300–314. doi: 10.1016/j.watres.2012.10.006

Dwight, R. H., Fernandez, L. M., Baker, D. B., Semenza, J. C., and Olson, B. H. (2005). Estimating the economic burden from illnesses associated with recreational coastal water pollution - a case study in Orange County, California. *J. Environ. Manage.* 76, 95–103. doi: 10.1016/j.jenvman.2004.11.017

Effler, S. W., O'Donnell, S. M., Prestigiacomo, A. R., O'Donnell, D. M., Matthews, D. A., Owens, E. M., et al. (2009). Tributary plunging in an urban lake (Onondaga Lake): drivers, signatures, and implications. *JAWRA J. Am. Water Resour. Assoc.* 45, 1127–1141. doi: 10.1111/j.1752-1688.2009.00348.x

Fluke, J., González-Pinzón, R., and Thomson, B. (2019). Riverbed sediments control the spatiotemporal variability of *E. coli* in a highly managed, Arid River. *Front Water* 1:4. doi: 10.3389/frwa.2019.00004

Food and Drug Administration (2019). *Investigation Summary: Factors Potentially Contributing to the Contamination of Romaine Lettuce Implicated in the Fall 2018. Multi-State Outbreak of E.* coli *O 157, H7.* Silver Spring, MD: Food and Drug Administration.

Food and Drug Administration (2020). *Outbreak Investigation of E. coli: Romaine from Salinas, California*. Washington, DC: Food and Drug Administration.

Francy, D. S., Stelzer, E. A., Duris, J. W., Brady, A. M. G. G., Harrison, J. H., Johnson, H. E., et al. (2013). Predictive models for Escherichia coli concentrations at inland lake beaches and relationship of model variables to pathogen detection. *Appl. Environ. Microbiol.* 79, 1676–1688. doi: 10.1128/AEM.02995-12

Ganley, R. C., Kirsch, G. N., Oliver, A. J., Karanik, J. M., and Ott, R. R. (1982). Cost-effective program for combined sewer overflow abatement. *J. Water Pollut. Control Fed.* 54, 1215–1223.

Given, S., Pendleton, L. H., and Boehm, A. B. (2006). Regional public health cost estimates of contaminated coastal waters: a case study of gastroenteritis at southern California beaches. *Environ. Sci. Technol.* 40, 4851–4858. doi: 10.1021/es060679s

Goto, D. K., and Yan, T. (2011). Genotypic diversity of *Escherichia coli* in the water and soil of tropical watersheds in Hawaii. *Appl. Environ. Microbiol.* 77, 3988–3997. doi: 10.1128/AEM.02140-10

Goyal, S. M., Gerba, C. P., Melnick, J. L., Bonsdorff, C.-., H. von, Torvela, N., Heikinheimo, A., et al. (1977). Occurrence and distribution of bacterial indicators and pathogens in canal communities along the Texas coast. *Appl. Environ. Microbiol.* 34, 139–149. doi: 10.1128/aem.34.2.139-149.1977

Green, H., Weller, D. D. L., Johnson, S., and Michalenko, E. (2019). Microbial source-tracking reveals origins of fecal contamination in a recovering watershed. *Water* 11:2162. doi: 10.3390/w11102162

Green, H. C., Dick, L. K., Gilpin, B., Samadpour, M., and Field, K. G. (2012). Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Appl. Environ. Microbiol.* 78, 503–510. doi: 10.1128/AEM.05734-11

Green, H. C., Haugland, R. A., Varma, M., Millen, H. T., Borchardt, M. A., Field, K. G., et al. (2014). Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl. Environ. Microbiol.* 80, 3086–3094. doi: 10.1128/AEM.04137-13

Greene, S. K., Daly, E. R., Talbot, E. A., Demma, L. J., Holzbauer, S., Patel, N. J., et al. (2008). Recurrent multistate outbreak of Salmonella Newport associated with tomatoes from contaminated fields, 2005. *Epidemiol. Infect.* 136, 157–165. doi: 10.1017/S095026880700859X

Hendricks, C. (1967). Multiplication and growth of selected enteric bacteria in clear mountain stream water. *Water Res.* 1, 567–576. doi: 10.1016/0043-1354(67)90039-5

Johnson, S. (2006). *The Ghost Map: The Story of London's Most Terrifying Epidemic–and How It Changed Science, Cities, and the Modern World*. New York, NY: Penguin Books.

Kim, J.-., W., Pachepsky, Y. A., Shelton, D. R., and Coppock, C. (2010). Effect of streambed bacteria release on *E. coli* concentrations: monitoring and modeling with the modified SWAT. *Ecol. Modell.* 221, 1592–1604. doi: 10.1016/j.ecolmodel.2010.03.005

Ksoll, W. B., Ishii, S., Sadowsky, M. J., and Hicks, R. E. (2007). Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. *Appl. Environ. Microbiol.* 73, 3771–3778. doi: 10.1128/AEM.02654-06

Lu, J., Santo Domingo, J. W., Lamendella, R., Edge, T., and Hill, S. (2008). Phylogenetic diversity and molecular detection of bacteria in gull feces. *Appl. Environ. Microbiol.* 74, 3969–3976. doi: 10.1128/AEM.00019-08

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., and Lüdecke, D. (2019b). Indices of effect existence and significance in the bayesian framework. *Front. Psychol.* 10:2767. doi: 10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., and Lüdecke, D. (2019a). bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework software. *J. Open Source Softw.* 4:1541. doi: 10.21105/joss.01541

Muirhead, R. W., Davies-Colley, R. J., Donnison, A. M., and Nagels, J. W. (2004). Faecal bacteria yields in artificial flood events: quantifying in-stream stores. *Water Res.* 38, 1215–1224. doi: 10.1016/j.watres.2003.12.010

Nagels, J. W., Davies-Colley, R. J., Donnison, A. M., and Muirhead, R. W. (2002). Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. *Water Sci. Technol.* 45, 45–52. doi: 10.2166/wst.2002.0408

NandaKafle, G., Christie, A. A., Vilain, S., and Brözel, V. S. (2018). Growth and extended survival of *Escherichia coli* O157, H7. in soil organic matter. *Front. Microbiol.* 9:762. doi: 10.3389/fmicb.2018.00762

Nautiyal, C. S., Rehman, A., and Chauhan, P. S. (2010). Environmental *Escherichia coli* occur as natural plant growth-promoting soil bacterium. *Arch. Microbiol.* 192, 185–193. doi: 10.1007/s00203-010-0544-1

New York State Department of Environmental Conservation (2017). *LAKE GEORGE BEACH Pollution Source Investigation*. Albany, NY: New York State Department of Environmental Conservation.

Pandey, P. K., Soupir, M. L., Haddad, M., and Rothwell, J. J. (2012). Assessing the impacts of watershed indexes and precipitation on spatial in-stream *E. coli* concentrations. *Ecol. Indic.* 23, 641–652. doi: 10.1016/j.ecolind.2012.05.023

Partyka, M. L., Bond, R. F., Chase, J. A., and Atwill, E. R. (2018a). Spatial and temporal variability of bacterial indicators and pathogens in six California reservoirs during extreme drought. *Water Res.* 129, 436–446. doi: 10.1016/j.watres.2017.11.038

Partyka, M. L., Bond, R. F., Chase, J. A., and Atwill, E. R. (2018b). Spatiotemporal variability in microbial quality of western US agricultural water supplies: a multistate study. *J. Environ. Qual.* 47, 939–948. doi: 10.2134/jeq2017.12.0501

Prestigiacomo, A. R., Effler, S. W., O'Donnell, D., Hassett, J. M., Michalenko, E. M., Lee, Z., et al. (2007). Turbidity and suspended solids levels and loads in a sediment enriched stream: implications for impacted lotic and lentic ecosystems. *Lake Reserv. Manag.* 23, 231–244. doi: 10.1080/07438140709354012

Rabinovici, S. J. M., Bernknopf, R. L., Wein, A. M., Coursey, D. L., and Whitman, R. L. (2004). Economic and health risk trade-offs of swim closures at a Lake Michigan Beach. *Environ. Sci. Technol.* 38, 2737–2745. doi: 10.1021/es034905z

Rao, G., Eisenberg, J., Kleinbaum, D., Cevallos, W., Trueba, G., Levy, K., et al. (2015). Spatial variability of *Escherichia coli* in rivers of northern coastal ecuador. *Water* 7, 818–832. doi: 10.3390/w7020818

Rehmann, C. R., and Soupir, M. L. (2009). Importance of interactions between the water column and the sediment for microbial concentrations in streams. *Water Res.* 43, 4579–4589. doi: 10.1016/j.watres.2009.06.049

Ribera, L. A., Palma, M. A., Paggi, M., Knutson, R., Masabni, J. G., and Anciso, J. (2012). Economic analysis of food safety compliance costs and foodborne illness outbreaks in the United States. *Horttechnology* 22, 150–156. doi: 10.21273/HORTTECH.22.2.150

Scott, E. E., Leh, M. D. K., and Haggard, B. E. (2017). Spatiotemporal variation of bacterial water quality and the relationship with pasture land cover. *J. Water Health* 15, 839–848. doi: 10.2166/wh.2017.101

Shanks, O. C., Kelty, C. A., Sivaganesan, M., Varma, M., and Haugland, R. A. (2009). Quantitative PCR for genetic markers of human fecal pollution. *Appl. Environ. Microbiol.* 75, 5507–5513. doi: 10.1128/AEM.00305-09

Shanks, O. C., Santo Domingo, J. W., Lu, J., Kelty, C. A., and Graham, J. E. (2007). Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Appl. Environ. Microbiol.* 73, 2416–2422. doi: 10.1128/AEM.02474-06

Shiels, D. R., and Guebert, M. (2010). Implementing landscape indices to predict stream water quality in an agricultural setting: An assessment of the Lake and River Enhancement (LARE) protocol in the Mississinewa River watershed, East-Central Indiana. *Ecol. Indic.* 10, 1102–1110. doi: 10.1016/j.ecolind.2010.03.007

Soupir, M. L., Mostaghimi, S., and Dillaha, T. (2010). Attachment of *Escherichia coli* and enterococci to particles in runoff. *J. Environ. Qual.* 39, 1019–1027. doi: 10.2134/jeq2009.0296

Sowah, R., Zhang, H., Radcliffe, D., Bauske, E., and Habteselassie, M. Y. Y. (2014). Evaluating the influence of septic systems and watershed characteristics on stream faecal pollution in suburban watersheds in Georgia, USA. *J. Appl. Microbiol.* 117, 1500–1512. doi: 10.1111/jam.12614

Sowah, R. A., Habteselassie, M. Y., Radcliffe, D. E., Bauske, E., and Risse, M. (2017). Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. *Water Res.* 108, 330–338. doi: 10.1016/j.watres.2016.11.007

Steven, W. E. (1996). *Limnological and Engineering Analysis of a Polluted Urban Lake: Prelude to Environmental Management of Onondaga Lake*. New York, NY: Springer.

Topalcengiz, Z., Strawn, L. K., Danyluk, M. M. D., Webster, L., and Gorski, L. (2017). Microbial quality of agricultural water in Central Florida. *PLoS ONE* 12:e0174889. doi: 10.1371/journal.pone.0174889

U.S. Environmental Protection Agency SIM. (2018). *Microbial Source Tracking: How Did That get in There?* EPA Sci Matters Newsl. Washington, DC: USEPA.

Verhougstraete, M. P., Martin, S. L., Kendall, A. D., Hyndman, D. W., and Rose, J. B. (2015). Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10419–10424. doi: 10.1073/pnas.1415836112

Viau, E. J., Goodwin, K. D., Yamahara, K. M., Layton, B. A., Sassoubre, L. M., Burns, S. L., et al. (2011). Bacterial pathogens in Hawaiian coastal streams—Associations with fecal indicators, land cover, and water quality. *Water Res.* 45, 3279–3290. doi: 10.1016/j.watres.2011.03.033

Wachtel, M. R., Whitehand, L. C., and Mandrell, R. E. (2002). Prevalence of *Escherichia coli* associated with a cabbage crop inadvertently irrigated with partially treated sewage wastewater. *J. Food Prot.* 65, 471–475. doi: 10.4315/0362-028X-65.3.471

Walker, W. W., Walker, J. D., and Eng, M. (2013). *Analysis of Fecal Coliform Data from Onondaga Lake Tributaries prepared for Upstate Freshwater Institute and Onondaga County Department of Water Environment Protection*. Syracuse, NY: Wiley.

Weller, D., Belias, A., Green, H., Roof, S., and Wiedmann, M. (2020b). Landscape, water quality, and weather factors associated with an increased likelihood of foodborne pathogen contamination of New York streams used to source water for produce production. *Front. Sustain. Food Syst.* 3:124. doi: 10.3389/fsufs.2019.00124

Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., et al. (2020a). Complex interactions between weather, and microbial and physicochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water. *Front. Microbiol.* 11:134. doi: 10.3389/fmicb.2020.00134

Weller, D. E., and Jordan, T. E. (2020). Inexpensive spot sampling provides unexpectedly effective indicators of watershed nitrogen status. *Ecosphere* 11:e03224. doi: 10.1002/ecs2.3224

Wilkes, G., Edge, T. A., Gannon, V. P. J., Jokinen, C., Lyautey, E., Neumann, N. F., et al. (2011). Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res.* 45, 5807–5825. doi: 10.1016/j.watres.2011.06.021

Wu, J. (2019). Linking landscape patterns to sources of water contamination: Implications for tracking fecal contaminants with geospatial and Bayesian approaches. *Sci. Total Environ.* 650, 1149–1157. doi: 10.1016/j.scitotenv.2018.09.087

Zhou, K., Sassi, H. P., Morrison, C. M., Duan, J. G., and Gerba, C. P. (2017). Resuspension of *Escherichia coli* and MS2 bacteriophage from bed sediment in irrigation canals. *J. Irrig. Drain Eng.* 143:04017005. doi: 10.1061/(ASCE)IR.1943-4774.0001169

Author Disclaimer: Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the NIH and USDA.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership