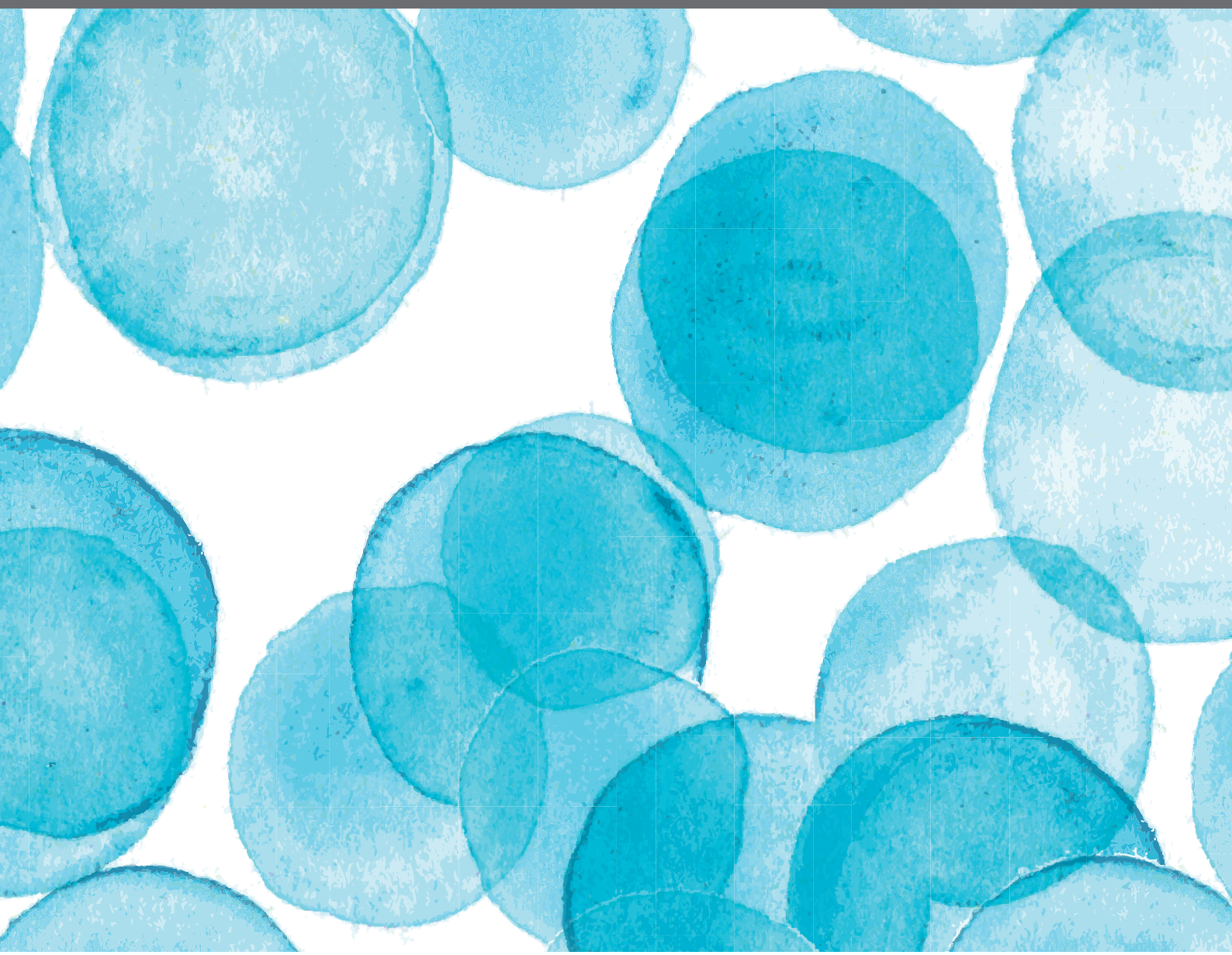
MICROBIAL SOURCE TRACKING

EDITED BY: Michèle Gourmelon, Anicet R. Blanch and Georg H. Reischer
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88974-067-3
DOI 10.3389/978-2-88974-067-3

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

MICROBIAL SOURCE TRACKING

Topic Editors:
Michèle Gourmelon, IFREMER, ODE-DYNECO, Laboratoire d'Ecologie Pélagique, Plouzané, France
Anicet R. Blanch, University of Barcelona, Spain
Georg H. Reischer, Vienna University of Technology, Austria

Citation: Gourmelon, M., Blanch, A. R., Reischer, G. H., eds. (2022). Microbial Source Tracking. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-067-3

Table of Contents

05 ***Editorial: Microbial Source Tracking***
Michèle Gourmelon, Anicet R. Blanch and Georg H. Reischer

08 ***Microbial Source Tracking of Fecal Indicating Bacteria in Coral Reef Waters, Recreational Waters, and Groundwater of Saipan by Real-Time Quantitative PCR***
Christopher Sinigalliano, Kiho Kim, Maribeth Gidley, Kathy Yuknavage, Karen Knee, Dean Palacios, Charito Bautista, Anthony Bonacolta, Hyo Won Lee and Larry Maurin

28 ***Tracking Major Sources of Water Contamination Using Machine Learning***
Jianyong Wu, Conghe Song, Eric A. Dubinsky and Jill R. Stewart

41 ***Human Fecal Pollution Monitoring and Microbial Risk Assessment for Water Reuse Potential in a Coastal Industrial–Residential Mixed-Use Watershed***
Akechai Kongprajug, Thammanitchpol Denpetkul, Natcha Chyerochana, Skorn Mongkolsuk and Kwanrawee Sirikanchana

55 ***Bacteriophages Are Good Estimators of Human Viruses Present in Water***
Elisenda Ballesté, Anicet R. Blanch, Javier Mendez, Laura Sala-Comorera, Leena Maunula, Silvia Monteiro, Andreas H. Farnleitner, Andreas Tiehm, Joan Jofre and Cristina García-Aljaro

64 ***Human Pathogenic Bacteria Detected in Rainwater: Risk Assessment and Correlation to Microbial Source Tracking Markers and Traditional Indicators***
Julia K. Denissen, Brandon Reyneke, Monique Waso, Sehaam Khan and Wesaal Khan

79 ***Evaluation of Low-Cost Phage-Based Microbial Source Tracking Tools for Elucidating Human Fecal Contamination Pathways in Kolkata, India***
Renuka Kapoor, James Ebdon, Ashutosh Wadhwa, Goutam Chowdhury, Yuke Wang, Suraja J. Raj, Casey Siesel, Sarah E. Durry, Wolfgang Mairinger, Asish K. Mukhopadhyay, Suman Kanungo, Shanta Dutta and Christine L. Moe

93 ***The Use of Ribosomal RNA as a Microbial Source Tracking Target Highlights the Assay Host-Specificity Requirement in Water Quality Assessments***
Annastiina Rytkönen, Ananda Tiwari, Anna-Maria Hokajärvi, Sari Uusheimo, Asko Vepsäläinen, Tiina Tulonen and Tarja Pitkänen

109 ***Characterization of Stormwater Runoff Based on Microbial Source Tracking Methods***
Silvia Monteiro, Gaspar Queiroz, Filipa Ferreira and Ricardo Santos

119 ***Antibiotic Resistance and Sewage-Associated Marker Genes in Untreated Sewage and a River Characterized During Baseflow and Stormflow***
Warish Ahmed, Pradip Gyawali, Kerry A. Hamilton, Sayalee Joshi, David Aster, Erica Donner, Stuart L. Simpson and Erin M. Symonds

132 ***Synergistic Application of Molecular Markers and Community-Based Microbial Source Tracking Methods for Identification of Fecal Pollution in River Water During Dry and Wet Seasons***
Hongxia Liang, Zhisheng Yu, Bobo Wang, Fabrice Ndayisenga, Ruyin Liu, Hongxun Zhang and Gang Wu

147 ***Bather Shedding as a Source of Human Fecal Markers to a Recreational Beach***
Dong Li, Laurie C. Van De Werfhorst, Brandon Steets, Jared Ervin, Jill L. S. Murray, Naresh Devarajan and Patricia A. Holden

158 ***Identifying Sources of Faecal Contamination in a Small Urban Stream Catchment: A Multiparametric Approach***
Liam J. Reynolds, Niamh A. Martin, Laura Sala-Comorera, Kevin Callanan, Padraig Doyle, Clare O'Leary, Paul Buggy, Tristan M. Nolan, Gregory M. P. O'Hare, John J. O'Sullivan and Wim G. Meijer

170 ***Genetic Microbial Source Tracking Support QMRA Modeling for a Riverine Wetland Drinking Water Resource***
Julia Derx, Katalin Demeter, Rita Linke, Sílvia Cervero-Aragó, Gerhard Lindner, Gabrielle Stalder, Jack Schijven, Regina Sommer, Julia Walochnik, Alexander K. T. Kirschner, Jürgen Komma, Alfred P. Blaschke and Andreas H. Farnleitner

190 ***Comparative Analysis of Fecal Microbiomes From Wild Waterbirds to Poultry, Cattle, Pigs, and Wastewater Treatment Plants for a Microbial Source Tracking Approach***
Amine M. Boukerb, Cyril Noël, Emmanuelle Quenot, Bernard Cadiou, Julien Chevé, Laure Quintric, Alexandre Cormier, Luc Dantan and Michèle Gourmelon

211 ***Influence of Filter Pore Size on Composition and Relative Abundance of Bacterial Communities and Select Host-Specific MST Markers in Coastal Waters of Southern Lake Michigan***
Muruleedhara N. Byappanahalli, Meredith B. Nevers, Dawn Shively, Cindy H. Nakatsu, Julie L. Kinzelman and Mantha S. Phanikumar

222 ***A Combined Digital PCR and Next Generation DNA-Sequencing Based Approach for Tracking Nearshore Pollutant Dynamics Along the Southwest United States/Mexico Border***
Amity G. Zimmer-Faust, Joshua A. Steele, Xianyi Xiong, Christopher Staley, Madison Griffith, Michael J. Sadowsky, Margarita Diaz and John F. Griffith

237 ***Integrative Survey of 68 Non-overlapping Upstate New York Watersheds Reveals Stream Features Associated With Aquatic Fecal Contamination***
Hyatt Green, Maxwell Wilder, Martin Wiedmann and Daniel Weller

247 ***Monitoring Urban Beach Quality on a Summer Day: Determination of the Origin of Fecal Indicator Bacteria and Antimicrobial Resistance at Prophète Beach, Marseille (France)***
Mylène Toubiana, Christian Salles, Marie-George Tournoud, Patricia Licznar-Fajardo, Isabelle Zorgniotti, Marie-Laure Trémélo, Estelle Jumas-Bilak, Samuel Robert and Patrick Monfort

259 ***Identifying Inorganic Turbidity in Water Samples as Potential Loss Factor During Nucleic Acid Extraction: Implications for Molecular Fecal Pollution Diagnostics and Source Tracking***
Rita B. Linke, Sibel Zeki, René Mayer, Katharina Keiblinger, Domenico Savio, Alexander K. T. Kirschner, Georg H. Reischer, Robert L. Mach, Regina Sommer and Andreas H. Farnleitner

frontiers in Microbiology

EDITORIAL
published: 25 November 2021
doi: 10.3389/fmicb.2021.795564

Editorial: Microbial Source Tracking

Michèle Gourmelon[1]*, *Anicet R. Blanch*[2] and *Georg H. Reischer*[3]

[1] IFREMER, ODE-DYNECO-Pelagos, Laboratoire d'Ecologie Pélagique, Plouzané, France, [2] Department Genetics, Microbiology, and Statistics, University of Barcelona, Barcelona, Spain, [3] Institute of Chemical, Environmental, and Bioscience Engineering, TU Wien, Vienna, Austria

Keywords: microbial source tracking, fecal pollution indicators, microbiota (16S), bacterial communities, molecular marker, water, bacteriophages, modeling

Editorial on the Research Topic

Microbial Source Tracking

INTRODUCTION

Human and animal fecal pollution may adversely affect inland and coastal waters with negative consequences for water supplies, recreational water uses and shellfish production. Fecal pollution of waters is a significant health risk and can lead to economic losses due to shellfish bed closures, bathing prohibitions and serious limitations on water resources. Fecal contamination in water is currently evaluated by the enumeration of traditional fecal indicator bacteria (FIB; i.e., *Escherichia coli* and intestinal enterococci) which are not indicating the fecal pollution source. Microbial source tracking (MST) methods allow the identification of fecal pollution sources that is critical for management and remediation of water quality.

This special issue in *Frontiers in Microbiology* section *Microbiotechnology* offers the collection of 19 original research manuscripts, which contribute to the current knowledge on the microbial source tracking and highlight the latest developments.

MST APPROACHES

A large number of MST methods have been developed targeting bacteria, viruses, chemical compounds and eukaryotic cells. Many of them demonstrate an appropriate host specificity and sufficient sensitivity. Culture-dependent and culture-independent (mainly molecular) methods have been proposed. With the exception of chemical compounds, they are considered in the manuscripts of this special issue.

Multiple varying sources of fecal pollution can influence water quality. Some widespread sources of fecal pollution such as discharges of untreated or treated wastewater, contamination by bathers and wild birds are investigated in beach waters (Li et al.; Sinigalliano et al. and Toubiana et al.), street-level stormwater run-off (Monteiro et al.) and in roof-harvested rainwater (Denissen et al.) in this Research Topic.

MST methods used in the studies collected in this issue are based on the search for bacterial molecular markers, the detection of phages and the application of bacterial community approaches. More specifically, the most frequently used methods are bacterial molecular markers (in particular based on the 16S ribosomal RNA sequence or 16S rDNA gene). Those molecular markers targeting

OPEN ACCESS

Edited by:
Eric Altermann, AgResearch Ltd., New Zealand

Reviewed by:
Kwanrawee Joy Sirikanchana, Chulabhorn Research Institute, Thailand

***Correspondence:**
Michèle Gourmelon
michele.gourmelon@ifremer.fr

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *15 October 2021*
Accepted: *01 November 2021*
Published: *25 November 2021*

Citation:
Gourmelon M, Blanch AR and Reischer GH (2021) Editorial: Microbial Source Tracking. Front. Microbiol. 12:795564. doi: 10.3389/fmicb.2021.795564

human fecal material, or pollution originating from livestock or birds are the methods of choice in at least 15 of the manuscripts in this current Research Topic. In an interesting trend, most studies looked for several sources and combined several methods and targets (e.g., phage detection, bacterial molecular marker quantification, bacterial communities analysis).

The 16S rDNA gene is also the target of bacterial community-based MST approaches by amplicon sequencing of this gene. Boukerb et al. used this approach to identify novel markers for avian and other sources. Liang et al. as well as Zimmer-Faust et al. applied the same method to investigate and model fecal pollution in rural rivers and coastal waters, respectively, combining it with PCR-based detection of host-associated MST markers.

Rytkönen et al. explored the utility of targeting 16S rRNA instead of 16S rDNA as a template in MST studies. Furthermore, manuscripts of this Research Topic have shown that it may be relevant to complement the results from MST methods with chemical parameters indicative of pollution sources. Sinigalliano et al. used isotopic tracking of nitrogen inputs while Reynolds et al. investigated correlations of ammonium, total oxidized nitrogen, and phosphorus levels with MST marker levels.

Highly relevant technical issues of MST investigations were explored by Byappanahalli et al. who investigated the influence of filter pore size on marker- and community-based MST approaches while Linke et al. demonstrated the strong impact inorganic turbidity can have on DNA based water quality analysis. These studies highlight the central importance of quality and process controls in MST investigations.

MST AND HEALTH RISKS—TRACING PATHOGENS AND ANTIMICROBIAL RESISTANCE

If the use of MST markers is currently coupled with the detection of FIB, it is also done in combination with the detection of human pathogens, especially in the context of quantitative microbial risk assessment (QMRA). Denissen et al. investigated the presence of *Mycobacterium tuberculosis*, *Yersinia* spp. and *Listeria monocytogenes* in roof-harvested rainwater. Derx et al. used *Cryptosporidium* and *Giardia* as index pathogens in a catchment wide QMRA approach.

MST methods are also used to trace and investigate the spread of antimicrobial resistant strains and the pollution they originate from in the environment. Ahmed et al. show that MST markers can be useful indicator for the presence of AMR genes while Toubiana et al. combined genetic MST markers with AMR gene detection to assess bathers as sources of contamination.

VIRAL AND PHAGE TARGETS

Ballesté et al. used host-associated bacteriophages as indicators for human pathogenic viruses. Kapoor et al. on the other hand used bacteriophages as a low-cost alternative for differentiate animal waste and human wastewater in Kolkata. Viral indicators together with genetic MST markers were also used as an input for a QMRA investigation by Kongprajug et al.

MST PREDICTION AND MODELING

Wu et al. used various machine learning approaches to predict sources of fecal pollution based on a broad set of parameters from land use to hydrological parameters. In another study, Green et al. used generalized linear mixed models and conditional inference trees to find predictive variable for pollution source in 68 streams in the Finger Lakes region of Upstate New York.

CONCLUSION AND PERSPECTIVES

The selection of MST markers for their combined use with each other or with other markers such as chemical ones, is becoming a useful methodological approach to determine sources of fecal contamination and improve the robustness of prediction models. The incorporation of new computational procedures such as machine learning, neural networks or in the future artificial intelligence is very likely to facilitate and improve these predictive models. Moreover, care may be taken to combine the information provided by MST markers in multilayer data studies with other information from water uses in each geographic region (drinking water consumption, industrial use, agricultural irrigation and livestock, recharge of aquifers, reuse of wastewater, etc.) and hydrological and geological data among others that will allow to have a holistic view of water resources and their exploitability by different agents and uses.

On the other hand, MST studies in recent years are exploring whether the available MST markers can provide significant contributions to the control of waterborne pathogens, the transmission of antibiotic resistance through the water cycle, and to complement microbial quantitative risk assessments. It cannot be ruled out that some of the present MST markers, or others that may appear in the future, may play a key role in these topics.

New contributions from metagenomic and proteomic techniques may, in the not-too-distant future, be new tools for improving the specificity and sensitivity of MST markers or for discovering new ones that provide greater accuracy in operational applications of these MST markers.

AUTHOR CONTRIBUTIONS

MG, AB, and GR prepared the call for this Research Topic, edited the manuscripts that were submitted, and wrote this editorial. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank all authors and reviewers who have contributed to this special issue and the financing agencies involved in this research.

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.

Copyright © 2021 Gourmelon, Blanch and Reischer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fmicb.2020.596650

Microbial Source Tracking of Fecal Indicating Bacteria in Coral Reef Waters, Recreational Waters, and Groundwater of Saipan by Real-Time Quantitative PCR

Christopher Sinigalliano[1], Kiho Kim[2], Maribeth Gidley[1,3], Kathy Yuknavage[4], Karen Knee[2], Dean Palacios[4], Charito Bautista[4], Anthony Bonacolta[1,3], Hyo Won Lee[1,3] and Larry Maurin[4]*

[1] Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL, United States, [2] Department of Environmental Science, American University, Washington, DC, United States, [3] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL, United States, [4] Water Quality Surveillance/Nonpoint Source Program, Bureau of Environmental and Coastal Quality, Commonwealth of the Northern Mariana Islands, Saipan, MP, United States

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Raju Sekar, Xi'an Jiaotong-Liverpool University, China
Megan Devane, Institute of Environmental Science and Research, New Zealand

***Correspondence:**
Christopher Sinigalliano
Christopher.Sinigalliano@noaa.gov

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *20 August 2020*
Accepted: *16 December 2020*
Published: *18 January 2021*

Citation:
Sinigalliano C, Kim K, Gidley M, Yuknavage K, Knee K, Palacios D, Bautista C, Bonacolta A, Lee HW and Maurin L (2021) Microbial Source Tracking of Fecal Indicating Bacteria in Coral Reef Waters, Recreational Waters, and Groundwater of Saipan by Real-Time Quantitative PCR. Front. Microbiol. 11:596650. doi: 10.3389/fmicb.2020.596650

The Commonwealth of the Northern Mariana Islands (CNMI) recently identified the need to improve its capacity for detecting and tracking land-based sources of pollution (LBSP) in coastal waters, particularly microbial contaminants like fecal indicator bacteria (FIB). Reported here is a baseline study of a suite of host-specific FIB microbial source tracking (MST) markers in the coastal shoreline and reef waters around the island of Saipan. Three sampling campaigns were conducted in September 2017, March 2018, and August 2018. Samples were collected from the nearshore surface waters of Saipan, the reef waters of Saipan Lagoon, and groundwater from beaches along the Saipan Lagoon shoreline. Measurements of submarine groundwater discharge (SGD) into nearshore waters and isotopic source tracking of nitrogen inputs were conducted concurrently with MST. Environmental DNA was extracted from the samples and analyzed by quantitative polymerase chain reaction (qPCR) for MST gene markers of fecal Bacteroidales specifically associated with humans, dogs, cows, and pigs, and for an MST gene marker of *Catellicoccus* associated with seabirds. MST assessments were combined with local knowledge, assessments of sanitary infrastructure, and routine watershed surveys. This study identified hotspots of human FIB along the western Saipan Lagoon shoreline in both surface waters and groundwater, plus another hotspot of human FIB at a popular tourist bathing area known as the Grotto. FIB hotspots on the Lagoon shoreline coincided with areas of high SGD and nitrogen isotopic data indicating sewage-derived N inputs. It appears that faulty sanitary infrastructure may be contributing to inputs to Saipan Lagoon, while bather shedding is likely a primary input for the Grotto area. Moderate levels of dog fecal contamination were common and widespread across the island. High levels of seabird fecal contamination were more random, both spatially and temporally, and mostly concentrated along the

less developed northeast region of Saipan. No significant levels of cow or pig fecal marker were detected in coastal water samples. This study provides demonstration and establishment of analytical capacity to resource management in CNMI for MST technology to aid in trouble-shooting water quality issues involving land-based sources of microbial contaminants to CNMI coastal waters.

Keywords: fecal indicating bacteria, microbial source tracking, land-based sources of pollution, water quality, *Enterococcus*, bacteroides, Saipan

INTRODUCTION

The Commonwealth of the Northern Mariana Islands (CNMI) is a territory of the United States that comprises all the islands of the Northern Mariana archipelago chain in the northwest Pacific, except for the island of Guam (which is its own separate US territory). Saipan, the largest of the 14 islands of the CNMI, is the capital and houses the largest population, with the majority of communities situated along Saipan's western coast. Waters along this area are negatively impacted by chronic sediment, nutrient, and microbial pollution that, combined with other stressors such as temperature-induced bleaching, ocean acidification, fishing pressures, reef tourism pressures, etc., can affect many of the island's western and southeastern reefs (Starmer et al., 2008; Yuknavage et al., 2018). Saipan's geology consists of a volcanic basement overlaid with terraced limestone shelves. Despite abundant rainfall, little fresh surface water exists, and the majority of available freshwater occurs in aquifers due to the high permeability of fragmented limestone. As in other small oceanic islands, most of the available fresh water in Saipan is in a freshwater-saltwater coastal-aquifer system where a lens-shaped body of fresh and brackish ground water floats on denser saltwater within the island (Carruth, 2003). Treated aquifer groundwater provides the primary source of fresh tap water for the island population. Saipan has a variety of potential sources of nutrient, chemical, and/or microbial contaminants to both the groundwater aquifer and surface waters, including groundwater infiltration from septic fields, golf courses, some small-scale agriculture, and small industries, plus point sources to the coastal ocean such as the outfalls from two wastewater treatment plants and brackish effluent from reverse osmosis (RO) water treatment surface impoundments (Carruth, 2003; Perreault, 2007). In addition, there is a recognized issue with elderly and/or failing wastewater infrastructure, leading to elevated levels of fecal indicator bacteria in both beach surface water and groundwater (Knapp et al., 2020). Saipan is bordered by a fringing reef off the western coast with an extensive lagoon system that is hydrodynamically separated into three distinct Saipan Lagoon regions: Tanapag, Garapan, and Chalan Kanoa (Kruger et al., 2010). The Garapan Lagoon region in particular borders the populous middle section of the western coastal plain of the island, with substantial urban development, numerous storm drains, and known challenges of outdated and potentially leaking sewage infrastructure (Yuknavage et al., 2018). Since the 1940s, the Saipan Lagoon system has lost about 20% of its seagrass and coral cover (Perez et al., 2018).

Management of the CNMI's coastal resources falls under the jurisdiction of the Bureau of Environmental and Coastal Quality (BECQ). The BECQ is responsible for monitoring, assessing, and protecting water quality within the CNMI, as well as managing land, air, water, and coastal quality. Both commonwealth and US federal laws and regulations mandate this responsibility (Yuknavage et al., 2018). As a whole, the CNMI's marine waters meet the high standards specified by the BECQ for water quality. The majority are designated as "Class AA," which reflects the highest standard of water quality. However, there are still a variety of point and non-point sources of nutrient and microbial contamination that can impair water quality. These pollution sources have the potential to affect both public and ecosystem health, including coral reefs. Environmental managers from Saipan suspect that surface runoff, stormwater discharges, and submarine groundwater discharge (SGD) may serve as significant conduits of microbial and nitrogen pollution. SGD has recently been quantified in this region, using natural tracers radon (Rn) and radium (Ra) to help identify where SGD is occurring (Knapp et al., 2020).

In recent years, the BECQ has issued numerous water quality violations due to elevated enterococci fecal indicating bacteria (FIB) levels in areas of Saipan with heavy stormwater runoff (Yuknavage et al., 2018). Many of these sites are within the highly developed and urbanized Garapan district along the western coast, where drainage issues are in the process of being addressed. Other frequent violations have occurred within Saipan's marinas and in waters surrounding docks, as well as at high-density tourist swimming sites such as the Grotto (Yuknavage et al., 2018). A large number of dogs (known locally as "boonie dogs") also roam the island, especially in developed areas around the central western Saipan Lagoon shore. Dogs can contribute significant loads of fecal bacteria to terrestrial runoff as non-point-source discharges. Dog fecal contamination not only poses a possible public health risk but also potentially confounds water quality assessments based solely on general FIB measurements, as dog feces can contain very high levels of enterococci as compared to some other animals or humans (Solo-Gabriele et al., 2011; Cinquepalmi et al., 2013). A recent population survey of free-roaming dogs in Saipan estimated that more than 21,000 such dogs are on the island (Nimer et al., 2018). Many island residents perceive the large free-roaming dog population as a public health concern, and animal control on the island faces many challenges.

The CNMI has 240.5 miles of ocean shoreline. In 2018, 50.5 of these coastline miles (or 21% of the CNMI's coastline miles) were found to be impaired due to exceedance violations of

enterococci FIB levels (Yuknavage et al., 2018). 32.7 miles of such impaired coastline surrounded the island of Saipan. As in previous years, the BECQ has considered the most common sources of enterococci contamination to be from point sources that include failing sewer lines and other municipal wastewater collection sites, individual onsite wastewater collection systems, and non-point sources. These non-point sources include: (1) sediment-laden storm water runoff with naturally occurring enterococci from urban runoff, erosion from construction sites and new developments, etc.; (2) illicit wastewater discharges from animal pens and outhouses; (3) waste from free-range feral and domestic livestock; and (4) a lack of adequate public restroom facilities at popular, heavily visited tourist sites (Yuknavage et al., 2018).

The BECQ regularly monitors the microbiological water quality of the CNMI's coastal waters through its water quality surveillance program, testing for enterococci fecal bacteria with the EnteroLert system (IDEXX Laboratories). It also regularly tests for *Escherichia coli* with the ColiLert system (IDEXX Laboratories) as per the manufacturer's instructions and guidelines by the US Environmental Protection Agency (EPA). However, these current regulatory approved tests lack any source-tracking capability. Although general FIB monitoring is useful for identifying hotspots and zones of water quality exceedance, it is of less utility in helping resource managers understand why there are such exceedances and where the microbial contamination may be originating (Harwood et al., 2014).

The BECQ already uses a GeneDisc Rapid Microbiology System and a commercially available kit (Pall Corporation) to conduct quantitative polymerase chain reaction (qPCR) enumeration of *E. coli* and enterococci FIB. The system consists of a Pall GeneDisc qPCR thermocycler and Pall Environmental DNA Extractor System used with the GeneDisc Recreation Water *E. coli* and *Enterococcus* spp. assay kit (Pall Corporation). However, these particular commercial qPCR assays are for general FIB enumeration and lack source tracking capability. This project has now adapted the in-house BECQ GeneDisc qPCR thermocycler platform for conducting a range of standard host-specific microbial source tracking (MST) assays.

Many strains and species of gut microbial flora have co-evolved with their animal hosts and contain unique gene sequences that can be diagnostic for particular host-associated FIBs, indicating fecal contamination specific to particular host animals such as humans, dogs, birds, pigs, cows, etc. (Boehm et al., 2013). The ability to enumerate the relative abundance of host-specific FIB is a useful tool for resource managers to investigate hotspots of microbial contamination. It also potentially provides better insight into the possible sources and patterns of transport for land-based sources of pollution (LBSP) that contribute to the microbial contamination of impaired waters and enables a better assessment of the potential environmental and public health risks of such contamination (Harwood et al., 2014). Patterns of excessive human-source fecal indicators can suggest infrastructure problems and sanitary leakage, while canine-source FIB markers may indicate increased contamination inputs from surface runoff and stormwater sources. However, it should be recognized that in some urban communities in some countries, there is an increasing tendency for pet owners to flush pet waste into municipal sanitary systems (Caldwell and Levine, 2009). Despite this, flushing dog fecal waste is not believed to be a common practice in Saipan, and most of the dog population on the island is free-roaming outdoors. In addition, agricultural FIB markers such as from cow and pig sources can indicate livestock fecal inputs, while excessive bird markers may indicate background inputs from wildlife populations.

Microbial source tracking adds another set of tools to the resource management toolbox that, in combination with other methods, can better direct investigative and mitigation efforts. MST is a DNA-based technology that enables the water quality management community to determine whether humans or other animal species are responsible for microbial fecal contamination in an aquatic environment. A variety of methods for molecular MST of fecal indicator bacteria have been developed, tested, and deployed, and applications for MST in water quality management are becoming increasingly common (Boehm et al., 2013; Harwood et al., 2014). Note that there are no MST methods promulgated yet for regulatory criteria, and there are currently no abundance threshold exposure limits promulgated in the US for regulatory purposes for any host-specific MST genetic markers. Rather, MST assessment of the relative abundance of host-specific fecal bacterial genetic markers is currently used more commonly in conjunction with the regulatory general FIB assessments for enterococci and *E. coli* as a troubleshooting approach to investigate chronic microbial water quality problems. This type of approach can be highly effective when integrated into a multi-tool, multi-tiered strategy for water quality assessment, as described in The California Microbial Source Identification Manual: A Tiered Approach to Identifying Fecal Pollution Sources to Beaches (Griffith et al., 2013) by the Southern California Coastal Water Research Project (SCCWRP).

Bacterial water quality criteria in the CNMI are promulgated by the BECQ and are based upon the EPA's water quality recommendations (US Environmental Protection Agency [US EPA], 2012b). The current regulatory microbial water quality criteria in the US and its territories rely on the culture-based enumeration of *E. coli* and/or fecal enterococci (Soller et al., 2010a,b, 2014). The recommended regulatory limits for enterococci in recreational waters, measured by either the membrane filtration plate count method as in EPA method 1600 (US Environmental Protection Agency [US EPA], 2009) or the Chromogenic Substrate method as in the commercial EnteroLert test (IDEXX), is a geomean of 35 CFU/100 mL or 35 MPN/100 mL, respectively. In addition to these regulatory water quality criteria, there is also a recommendation for a single-sample Single Threshold Value (STV) for a "Beach Action Value" or BAV limit of 130 CFU or MPN per 100 mL. This value is based upon the 90th percentile confidence interval (CI) for the STV. Note that this BAV is not a regulatory water quality criteria *per se*, but rather serves as guidance toward decision-making about posting bather warnings or beach closure notifications due to bacterial risk, based on the results of single grab samples rather than a geomean of multiple samples over time. Thus, the BAV is the abundance of viable enterococci in

recreational waters that would trigger posted beach warnings or beach closures. The EPA equates these values to an estimated illness rate of 36 illnesses per 1000 bathers, based on a series of epidemiological studies (US Environmental Protection Agency [US EPA], 2012b, p. 43). For additional precaution, the EPA has suggested an even lower BAV based on an STV of 70 CFU or MPN per 100 mL to be "more conservative," based on the 75th percentile CI rather than the 90th percentile CI. The EPA also has an alternative recommendation for an optional higher protective standard consisting of a geomean of 30 CFU or MPN per 100 mL, and a single-sample BAV limit of 60 CFU or MPN per 100 mL. These alternative values equate to an estimated illness rate of only 32 illnesses per 1000 bathers (US Environmental Protection Agency [US EPA], 2012b, p. 44). In regards to enumeration of total enterococci (including non-viable) by qPCR gene amplification, the EPA has recommended a single-sample BAV limit for general enterococci as measured by the Taqman qPCR Entero1A assay as described in EPA method 1611 (US Environmental Protection Agency [US EPA], 2012b), consisting of 1000 "cell calibrator equivalents" (cce) per 100 mL, for a predicted illness rate of 36 illnesses per 1000 bathers, or an optional alternative of 640 cce per 100 mL for a predicted illness rate of 32 illnesses per 1000 bathers (US Environmental Protection Agency [US EPA], 2012b, p. 44). While terminology in the literature may vary, these cce are essentially equivalent to "target genome copies" (TGC), "genome copies" (GC), or "genome equivalents" (GE). It is up to each US state and territory to consider these various EPA recommendations for bacterial recreational water quality standards and then promulgate their own official regulations, water quality criteria, and state/territory specific BAVs, and to provide justification to the US EPA for these state/territory criteria. In the case of the marine Pacific territories of American Samoa, Guam, and the CNMI, the official BAV for the live enterococci single sample threshold is 130 CFU or MPN per 100 mL of water sample based on the 90% CI of the STV, reflective of 36 illnesses per 1000 bathers. The CNMI has not yet officially promulgated a BAV for qPCR-based enterococci abundance.

To our knowledge, there are currently no regulatory criteria, guidelines, or official recommendations for exposure limits or BAVs of any host-source-specific FIB microbial source tracking markers. However, a quantitative microbial risk assessment (QMRA) study by Boehm et al. (2018) focused on the potential exposure risk of the human-source fecal *Bacteroides* qPCR marker HF183 (the Minor-Grove-Binder Taqman version, as in EPA Method 1696). These researchers concluded that the risk associated with exposure to a fixed HF183 concentration in recreational water decreased with the age of contamination, and that swimmer exposure to sewage after it had aged ∼3 days resulted in a median risk of less than 30 illnesses per 1000 bathers. This QMRA study calculated a risk-based HF183 threshold of 900 copies/100 mL for sewage contamination aged 2.5 days, while a risk-based water quality threshold for HF183 in surface waters that takes into account uncertainty in contamination age was derived to be 4100 copies/100 mL (Boehm et al., 2018). Considering this previous QMRA study by Boehm et al., we suggested to the BECQ that when interpreting the results of this Saipan MST study, they utilize an HF183 marker exposure threshold of 1000 copies/100 ml for human-source fecal contamination less than 3 days old and 4100 copies/1000 mL of HF183 marker for human-source fecal contamination of uncertain age or greater than 3 days old.

As part of the broader MST technology transfer project between the NOAA-Atlantic Oceanographic and Meteorological Laboratory (AOML) and the CNMI-BECQ, an initial study was conducted as a collaborative effort between the BECQ, NOAA-AOML, and American University. This study examined land-based sources of microbial pollution to the coastal waters of Saipan following the principles and protocols of the California Microbial Source Identification Manual (Griffith et al., 2013). The goals were to enumerate host-specific FIB markers, targeting fecal contamination from humans, dogs, seabirds, cows, and pigs in nearshore coastal waters, recreational beach waters, and reef-associated surface waters, as well as in selected groundwater samples that discharge into Saipan Lagoon. These data were supplemented with nitrogen stable isotope analyses of nearshore macroalgae and a survey of SGD on Saipan's western coast to better understand pathways of inputs of land-based pollutants. This information was used for follow-up management recommendations to the BECQ to better inform their resource management efforts regarding microbial and nutrient water quality. An additional goal was to establish an ongoing in-house capability for routine MST analyses in the BECQ laboratory using the qPCR thermocycler instrumentation already in their possession. The process of the MST technology transition to the BECQ is described in more detail in the online **Supplementary Materials** document of this article, and is also reported elsewhere in detail (NOAA-Coral Reef Conservation Program, Final Project Report for CRCP Project # 31184). In addition, the details of the multi-laboratory validation of GeneDisc performance with MST assays and multi-platform comparisons of performance between GeneDisc and other thermocyclers are intended to be published elsewhere. This report focuses on the results of the baseline MST field study in Saipan's coastal waters conducted as part of this technology transfer project.

MATERIALS AND METHODS

Study Region and Field Program

Our study area was the island of Saipan, bounded by north latitude 15.290533°, south latitude 15.091500°, west longitude 145.685494°, and east longitude 145.830316°. Sampling sites included Saipan Lagoon and the Saipan reef tract along the western side of the island, sites at the northern end of the island, particularly the Grotto area (site NEB01), Bird Island (site NEB02), selected beaches along the northeast portion of the island, and the Lao Lao Bay area on the eastern side of Saipan.

Field sampling efforts for the Saipan MST baseline study were conducted on March 12-19, 2018 and July 31-August 8, 2018. A few samples of opportunity were collected on September 13, 2017 (before the baseline MST survey started) as part of a routine BECQ monitoring effort, and were used primarily for demonstration purposes to teach the MST qPCR methods

to BECQ personnel during the first MST technology transfer workshop conducted by NOAA-AOML. During the actual MST baseline surveys in 2018, BECQ researchers collected almost half of the samples during their routine water quality surveillance efforts, while researchers at American University collected shore and reef samples as a courtesy during their own ongoing nitrogen isotope and SGD tracking studies (Knapp et al., 2020), which occurred simultaneously with our MST studies.

Sample sites are shown in **Figure 1**, with samples collected by the BECQ denoted by site numbers starting with NEB (northeast

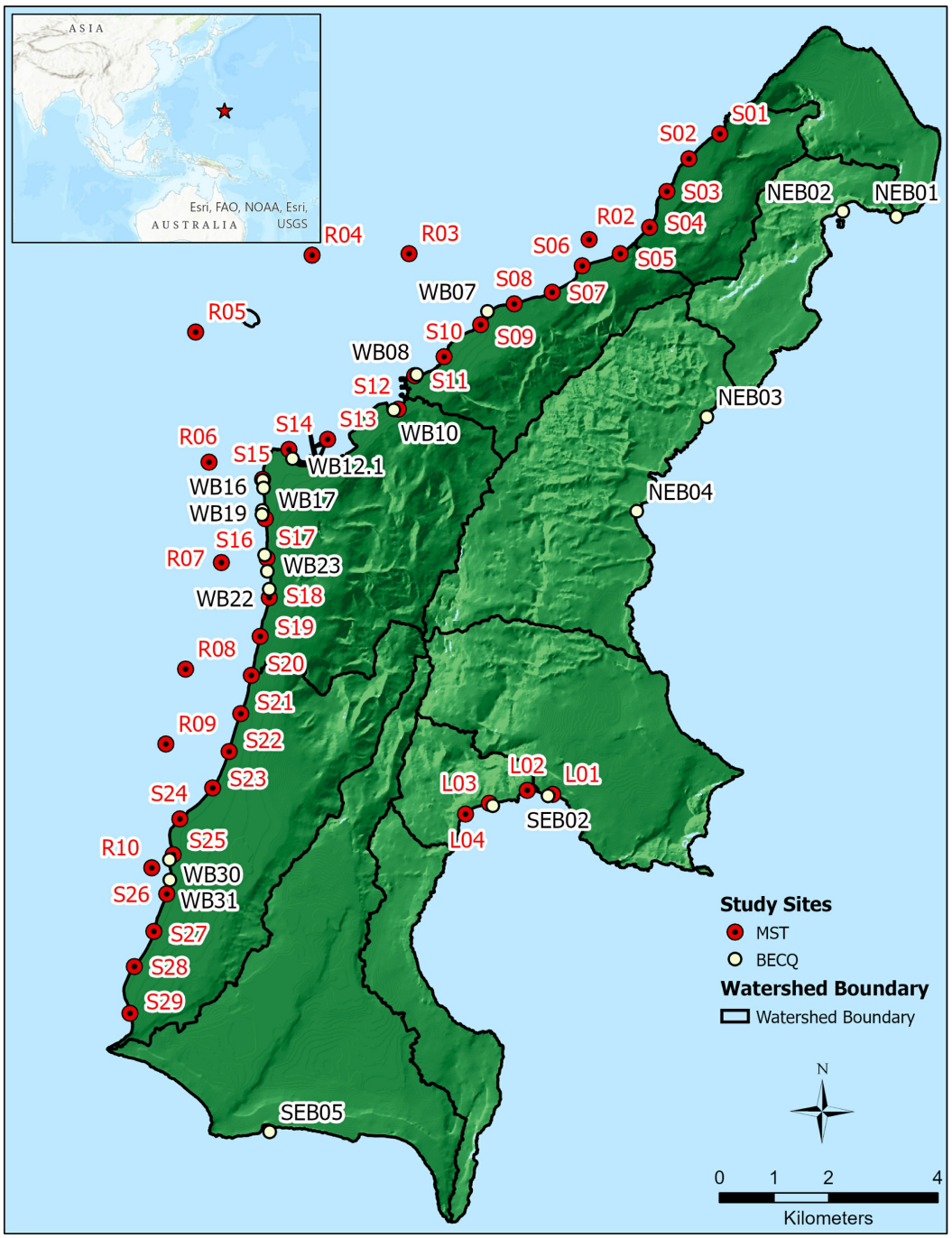


FIGURE 1 | Location of sample sites on the island of Saipan used for MST analyses, plus boundaries of the major watersheds on the island. NEB, SEB, and WB samples were collected by the BECQ from their routine water quality monitoring sites, while S, R, and L samples were collected by American University at additional sites specific for this study. Sites S1–S14 represent the Tanapag lagoon region (designated as "TANA"), sites S15–S25 represent the Garapan lagoon region (designated here as "GARA"), and sites S26–S29 represent the Chalan Kanoa lagoon region (designated as "CHAL").

beach), SEB (southeast beach), and WB (west beach). These sites are part of the BECQ's regular water surveillance monitoring program. Sites with numbers starting with S (shore), R (reef), and L (Lao Lao Bay) were collected by American University. Groundwater samples correspond to approximately 1 m inland of the same location as the same numbered shore site (for example, S18 and GW-S18 are shore water and groundwater respectively, with GW-S18 at one meter inland of S18). GPS coordinates for samples collected by the BECQ are reported in **Supplementary Table 1**, and GPS coordinates for samples collected by American University are reported in **Supplementary Table 2**.

Sample Collection and Handling

Surface water samples were collected approximately 15–30 cm below the surface in sterile 1.5 L Whirl-Pak bags (Nasco). Samples from these nearshore and reef locations were collected as per EPA methods 1611 and 1696, stored on ice, and filtered within 6 h of collection in the BECQ laboratory at Saipan (US Environmental Protection Agency [US EPA], 2012a; US Environmental Protection Agency [US EPA], 2019). Groundwater samples were collected using a 24″ (60 cm) PushPoint field investigation porewater sampler connected to a syringe with tubing (MHE Products, East Tawas, MI, United States). Groundwater samples intended for MST qPCR analyses were placed in sterile 1.5 L Whirl-Pak bags and processed in the BECQ laboratory. More specific details about the methods for the groundwater sampling for this study are presented in a separate publication describing nitrogen stable isotope source tracking and SGD quantification with Rn and Ra isotopes (Knapp et al., 2020).

Viable Enterococci Enumeration via IDEXX EnteroLert

Live enterococci were measured from subsamples of the collected water samples using the standard EPA approved EnteroLert Fluorogenic Substrate method (IDEXX Laboratories), following EPA guidelines, established BECQ protocols, and the manufacturer's instructions. Samples were processed with the IDEXX EnteroLert method by the BECQ Environmental Surveillance Laboratory. However, not all of samples collected by the American University group were tested using EnteroLert. The viable enterococci abundance in units of Most Probable Number (MPN) per 100 mL as estimated by the EnteroLert assay from the manufacturer's statistical tables, is shown in **Figure 2**, and also in the **Supplementary Table 5**.

Sample Filtration and Environmental DNA Extraction

In brief, at the BECQ laboratory, aliquots of water samples were filtered through sterile 0.45 μm pore-size, 47 mm diameter membrane filters using disposable sterile filter funnels (Pall Microfunnel, Pall Corporation). Water samples were filtered to 1 L or until clogging, and the volumes filtered varied from <100 mL to 1 l, depending on turbidity (and in the case of groundwater, on how much sample was originally collected). Most samples ranged from 800 mL to 1 L. In all cases, the actual sample volume that was filtered for each individual sample was recorded and used for quantitative calculations of MST target signal. Actual volumes filtered for each sample are listed in online **Supplementary Table 5**. For the initial samples of opportunity processed in September 2017 as part of the first training demonstrations at BECQ, sterile 0.4 μm pore-size, 47 mm diameter polycarbonate filters were used since these filters came pre-loaded as part of the Pall disposable MicroFunnels (Pall Corporation). For the regular 2018 baseline MST survey samples, cellulose ester filters (Pall GN6, Pall Corporation) were used, so as to be consistent with numerous previous MST studies conducted by NOAA in the marine environment (Sinigalliano et al., 2010, 2019; Campbell et al., 2015; Symonds et al., 2016; Staley et al., 2017). Sample filters were aseptically transferred to Lysing Matrix E bead beat tubes (from the FastDNA Spin Kit for Soil, MP Biomedicals), and preserved by addition of DNAgard Tissue preservative solution (Biomatrica). The MST sample filters were later extracted by gently removing the DNAgard preservative solution, lysing and homogenizing the cells with a SuperFastPrep-2 bead beating instrument (MPBiomedicals), and then purifying total environmental DNA from the lysate with the FastDNA Spin Kit for Soil protocol (MP Biomedicals, Thermo-Fisher) according to the manufacturer's instructions with minor modifications as described in the online **Supplementary Material** of this article.

Modification of MST qPCR Assays for GeneDisc Platform

More specific details of adapting these MST assays to the Pall GeneDisc system are described in the online **Supplementary Material** to this article. The BECQ laboratory was already in possession of a GeneDisc Rapid Microbiology qPCR System (Pall Corporation), which is an automated user-friendly system that normally requires the use of commercial kits with pre-loaded assay reagents. The GeneDisc system did not have any MST specific assays available at the time of this study, and normally the system cannot be customized by the end user. However, the Pall Corporation graciously custom modified the software of the particular GeneDisc instrument at the BECQ lab and specifically provided our study with custom blank "open" MST GeneDisc plates without pre-loaded reagents. This allowed BECQ personnel to run their own thermocycling parameters and conditions, as well as load their own MST reaction cocktails as per the standard MST protocols in the appendices of the California Microbial Source Identification Manual (Griffith et al., 2013).

The specific MST assays used for this Saipan microbial MST baseline study included: (1) human-source *Bacteroides* assay HF183 (EPA Taqman version as per EPA method 1696); (2) dog-source Bacteroidales assay DogBact; (3) cow-source Bacteroidales assay CowM2; (4) pig-source Bacteroidales assay Pig2Bact; and (5) the Gull2 assay specific for *Catellicoccus marimammalium*. *C. marimammalium* are found in the gut of most seagulls, as well as potentially in the gut of other birds (especially seabirds) that may co-habit, scavenge, or nest with seagulls. Depending on the specific geographic location and co-nesting behavior, this may also include species of terns, pelicans, geese, and very often, pigeons (Sinigalliano et al., 2013). Assay protocols followed

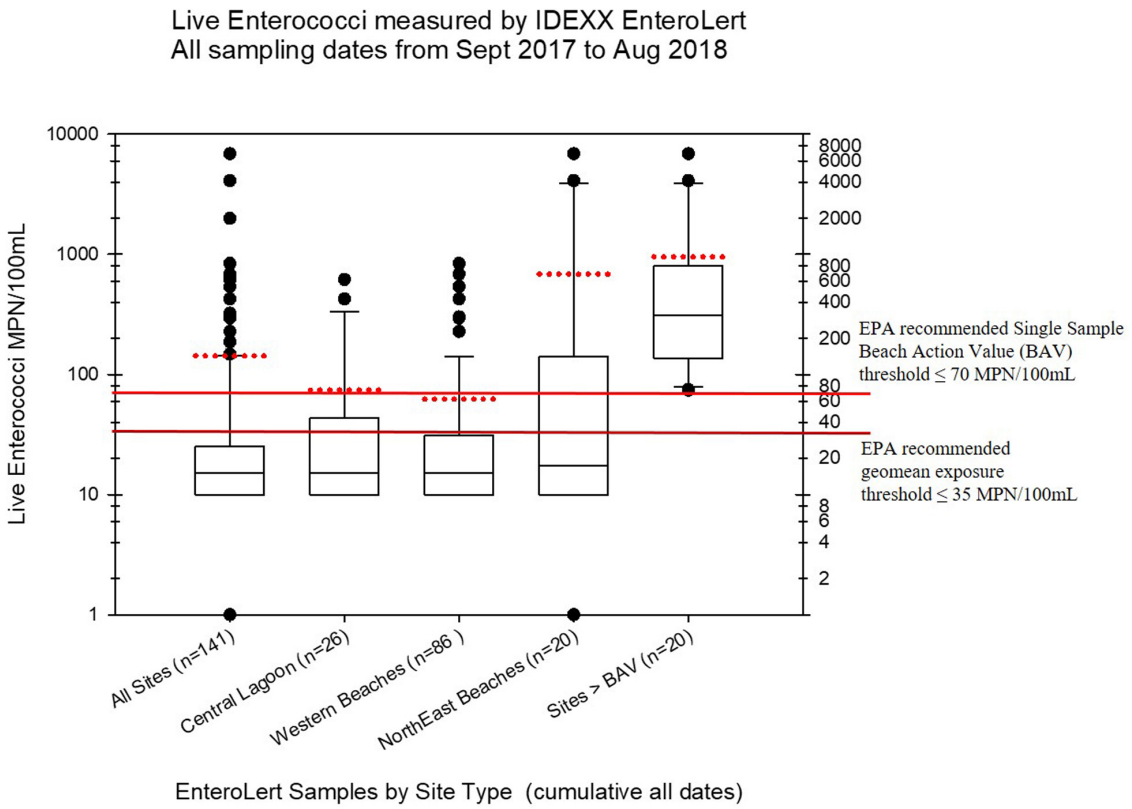


FIGURE 2 | Boxplots summarizing the distribution of live enterococci in the Saipan MST study samples as measured by the regulatory Chromogenic Substrate Most Probable Number Method (i.e., the commercial EnteroLert test by IDEXX Corporation). The black line within the box marks the median, and the red dotted line for each boxplot marks the mean. Note that not all samples were tested by EnteroLert (e.g., only 141 of 182 water samples were tested). The plots show the distribution of enterococci by different regions of Saipan that had hotspots of enterococci exceedances, while the last boxplot shows the distribution of all samples from all sites that exceeded water quality criteria for live enterococci (i.e., exceeding the Beach Action Value with more than 70 MPN/100 mL for a single sample). Fourteen of the 92 samples assessed (15.2%) exceeded regulatory water quality criteria for viable enterococci.

the California Microbial Source Identification Manual (Griffith et al., 2013), with minor modifications for the GeneDisc platform (see the online **Supplementary Methods Information**). The sequences of the oligonucleotides for primers and probes, as well as the synthetic dsDNA standard control fragments, are listed in **Supplementary Table 3**.

Quantitative PCR Quality Assurance and Controls

More specific details are given in the online **Supplementary Material** document. MST qPCR assay quality assurance procedures and controls were as described in EPA Method 1696 for the characterization of human fecal pollution in water by HF183/BacR287 Taqman qPCR (US Environmental Protection Agency [US EPA], 2019). These QA/QC metrics were used as guidance for the QA/QC assessment of all the assays. The standard curve quality control metrics for these qPCR assays as run on the BECQ lab's Pall GeneDisc instrument in Saipan are shown in **Supplementary Table 4**. The lower limit of quantitation (LLOQ) was determined for the qPCR reactions from the standard curves of each batch run. Samples with a Cq value greater than the reaction LLOQ were categorized as "DNQ" or "detected but not quantifiable." The overall reaction sensitivity of the LLOQ for this batch of standard curves was taken to be equivalent to 10 target sequence copies (tsc) per GeneDisc reaction well for each of the MST assays used in this study, which gave a calculated environmental detection LLOQ of 50 tsc/100 mL of water sample. Negative qPCR controls consisted of both "no template controls" (NTC) where no target DNA was added to the reaction wells and method blank (MB) controls where sterile water negative control samples were filtered, extracted, and analyzed in the same manner as the environmental water samples. For sample processing controls (SPC), the variability in sample processing efficiency or inhibition was measured for each environmental sample and method blank sample by using a spike preparation consisting of a fixed concentration of salmon DNA as per EPA method 1696 (US Environmental Protection Agency [US EPA], 2019).

Statistical boxplot analyses of the MST result patterns for host-specific fecal bacterial markers were generated with the SigmaPlot software package, version 14 (Systat Software, Inc.). Final MST positive control concentration standard curves and their associated linear regression statistics were also plotted using the SigmaPlot v.14 software package.

Stable Isotope Analysis

Sampling of benthic seagrass, algae, surface water and groundwater for isotopic analysis are described thoroughly in Knapp et al. (2020). For seagrasses and algae, five replicates for each species were taken, cleaned of foreign materials, air-dried for at least 24 h at 60°C, and ground to a fine powder. The samples were then weighed in either tin or silver capsules (the latter was used if acidification was required to remove inorganic carbon), and then analyzed on an EA-IRMS. Isotope values are reported as $\delta^{15}N$ relative to atmospheric N_2 ($\delta^{15}N = [(R_{sample}/R_{AIR})\text{-}1]^*1000$). Two in-house standards, acetanilide and caffeine (iACET, $\delta^{15}N$ = 1.18‰, caffeine $\delta^{15}N$ = 20.05‰) were used for calibration and the determination of the precision (0.2‰). Although several species of alga were widespread along the coast, none were present at all sites. Thus, we created an integrated primary producer $\delta^{15}N$ value using correlations between the most common species, *Caulerpa*, with three others to predict $\delta^{15}N$ values for *Caulerpa* at sites where it was absent.

Water samples were syringe-filtered (0.2 μm) into 50-mL HDPE screw-top bottles and kept in a cooler until they could be frozen at the end of each sampling day. Then, $\delta^{15}N$ and $\delta^{18}O$ were determined using the denitrifier method (Sigman et al., 2001; Casciotti et al., 2002) at Hong Kong University. An IRMS (Thermo Delta V Advantage) was used for isotopic analysis. Isotope measurements are reported in delta notation (‰) relative to atmospheric N_2 for $\delta^{15}N$ and V-SMOW for $\delta^{18}O$.

Submarine Groundwater Discharge (SGD)

As detailed in Knapp et al. (2020), SGD was estimated using a combination of measurement of the natural groundwater tracers radon (Rn) and radium (Ra). The activities of these isotopes are generally highest in groundwater, intermediate in coastal waters receiving SGD, and lowest in offshore seawater. Briefly, SGD ($m^3\ m^{-1}\ d^{-1}$) was calculated based on Ra flux (F_{Ra}; dpm d^{-1}) and groundwater endmember Ra activity (Ra_{gw}; dpm m^{-3}) at three sites; SGD at other sites was estimated based on the assumption that Ra, Rn, and SGD all followed the same spatial distribution along Saipan's coast. SGD estimates are given as a range (upper and lower) to reflect sampling error and analytical uncertainty propagated through the calculations.

RESULTS

The summary of results of all fecal indicator bacteria measurements for this study are shown in the project data summary table in the online **Supplementary Table 5** of this article. A series of boxplots visualizing the live enterococci measurements by IDEXX EnteroLert Chromogenic Substrate assay are shown in **Figure 2**, while a series of boxplots visualizing regional MST results are shown in **Figures 3–7**. **Figure 8** depicts submarine groundwater discharge and stable nitrogen isotope observations, while specific site MST results are visualized in **Figures 9–10**. These boxplots summarize the overall results of the qPCR-based MST for the host-source fecal indicator bacteria gene marker abundance in Saipan coastal waters for water samples collected from September 2017 through August 2018.

The BECQ reported non-detects for IDEXX EnteroLert as <30 MPN/100 mL based on the IDEXX EnteroLert statistical tables for MPN. For the live enterococci boxplots shown in **Figure 2**, these EnteroLert non-detect values were set at 10 MPN/100 mL for plotting purposes (hence there are no lower whiskers or outliers below the 10 MPN/100 mL level). The lower exposure threshold line shown on these plots represents the 2012 recreational water quality criteria recommendation by the EPA for a geomean of samples (35 MPN/100 mL), for which the EPA has determined these enterococci abundances reflect a predicted illness rate of less than 36 illnesses per 1000 bathers (US Environmental Protection Agency [US EPA], 2012b, p. 43). The upper exposure threshold line represents the EPA recommended BAV for the STV of individual grab samples (70 MPN/100 mL However, it should again be noted that the CNMI BAV as promulgated by the BECQ is actually 130 MPN/100 mL), which is based on the 90% CI reflective of 36 illnesses per 1000 bathers. The US EPA has lowered their suggested BAV to the 75th percentile as it is assumed to be potentially "more conservative" in their suggested protective threshold for posting beach warnings and/or beach closures.

For viable enterococci, 14 of the 92 EnteroLert samples (15.2%) exceeded the EPA recommended BAV threshold of ≤70 MPN/100 mL. The statistical distribution of these exceedances is shown in the last boxplot on the right side of **Figure 2**. The highest and most frequent enterococci exceedances for these particular samples clustered primarily in the region of the northeastern beaches, including the Grotto (NEB01), Bird Island (NEB02), Jeffrey's Beach (NEB03), Old Man by the Sea (NEB04), and Hidden Beach (NEB07). Hidden Beach and Old Man by the Sea, which are close to one another geographically in the same watershed, had the highest levels of observed viable enterococci during the study (6,867 MPN/100 mL and 1,989 MPN/100 mL respectively, occurring on August 8, 2020). Enterococci elevations substantially above the BAV were also observed on multiple occasions for Ladder Beach at the southern end of the island (site SEB05). Along the western coast, elevations of viable enterococci above the BAV were observed at the American Memorial Park Drainage (site WB12.1), Garapan Fishing Dock (WB21), Garapan Drainage (WB23), and Saipan Lagoon shore sites S10 and S21.

Figure 3 shows boxplots summarizing the statistical distribution of each MST marker for all cumulative sites and cumulative sampling events for this study, displaying the full range of variation and average abundances of each marker for all the sites and times combined. In **Figures 3–10**, the calculated environmental LLOQ and DNQ range are indicated in the lower horizontal line marked "LLOQ." The horizontal line marked "exposure threshold 1" indicates our suggestion to the BECQ for an exposure threshold of 1000 copies/100 mL for the HF183 human *Bacteroides* qPCR marker in surface water for recent human fecal contamination (≤3 days old), while the "exposure threshold 2" line indicates our suggested bather exposure threshold for the HF183 marker in surface water for human fecal contamination of uncertain age or >3 days old. Our suggestions

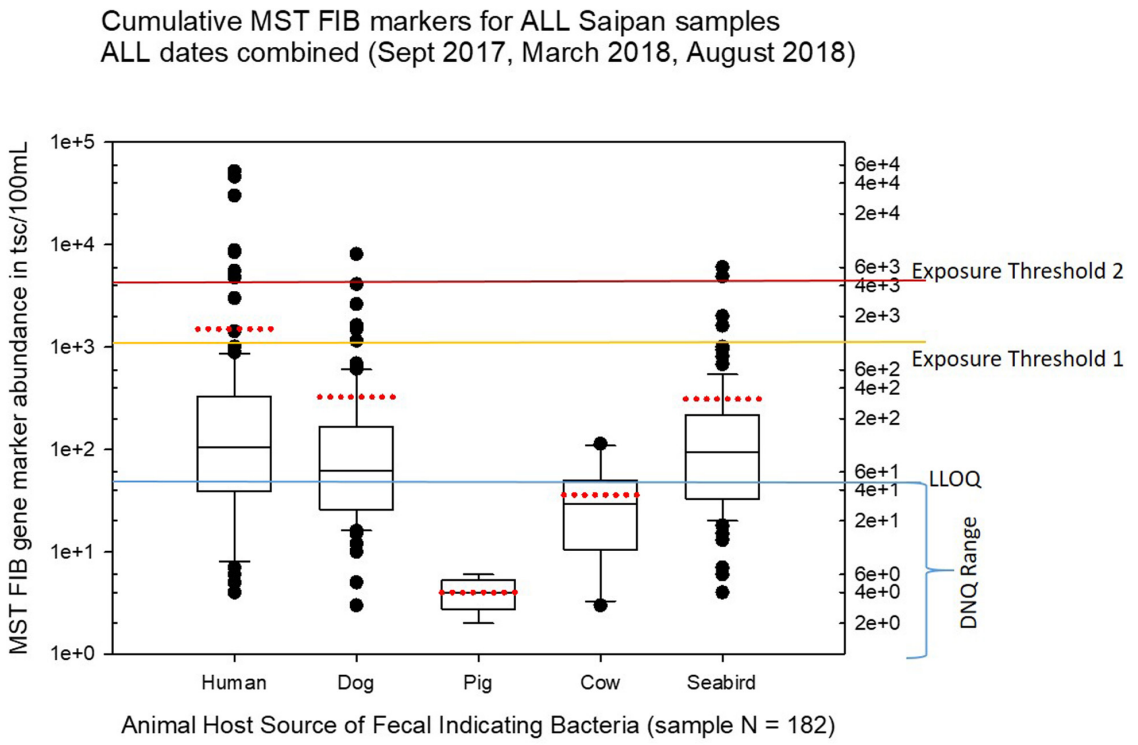


FIGURE 3 | Boxplots summarizing the overall results of MST analyses for host-source fecal indicator bacteria 16S ribosomal gene marker abundance in Saipan coastal waters for all water samples collected during the study from September 2017 through August 2018. The LLOQ line marks the environmental lower limit of quantitation (50 tsc/100 mL). The exposure threshold 1 line marks the suggested HF183 marker exposure limit (1000 tsc/100 mL) for recent human fecal contamination $\leq$ 3 days, while the exposure threshold 2 line marks the suggested HF183 marker exposure limit (4100 tsc/100 mL) for human fecal contamination of uncertain age $>$ 3 days (Boehm et al., 2018). The boundary of each box closest to zero indicates the 25th percentile, the black line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers (error bars) above and below the box indicate the 90th and 10th percentiles, respectively. The red dotted line for each boxplot marks the mean, and the outliers are shown as dots above or below the whiskers.

for these HF183 exposure thresholds used to help interpret the results of this Saipan MST study were adapted from information by a previous Quantitative Microbial Risk Assessment study of the HF183 marker (Boehm et al., 2018).

Our study found that both human and dog fecal bacterial genetic markers were relatively widespread in the Saipan surface water sites tested, while the seagull/seabird marker appeared to be somewhat more localized. Both agricultural markers for cow and pig fecal bacteria were much lower than had been anticipated; they were only rarely observed during the study and never above the background DNQ range (i.e., below the environmental LLOQ). As we utilized synthetic DNA gene blocks containing the assay target sequences for quantitation standards, we report our qPCR results here as "target sequence copies" or tsc. While terminology in the current literature varies, this "tsc" unit used here is essentially equivalent to the more common term "gene copies" (GC) or "target copies" (TC) or just "copies" as used in other papers. Note that tsc does not necessarily represent a "genome equivalent" (GE) or "cell calibrator equivalent" (CCE), as genomes may have multiple target copies of a particular assay target gene per genome depending upon the target gene of the qPCR assay (for example the 16S ribosomal targets such as HF183 can have multiple copies per genome, whereas CowM2 is a single copy target per genome).

For the human-source fecal *Bacteroides* HF183 marker, 59 samples out of 182 total samples (32.4%) demonstrated levels over 100 target sequence copies (tsc) per 100 mL, 11 samples out of 182 were over 1000 tsc/100 mL, and 8 samples out of 182 were over 4100 tsc/100 mL. This indicates a 6.0% exceedance rate for the suggested HF183 exposure threshold for recent human fecal contamination in surface waters and a 4.4% exceedance rate for the suggested exposure threshold for fecal contamination of uncertain age. These exceedances, primarily clustered along the shoreline of the western central Saipan Lagoon region, ranged from sites WB17 to S22, plus site NEB01 (the Grotto).

Seasonal differences in the distribution of MST FIB markers for the three sampling campaigns are shown in the boxplots of **Figures 4–6** for September 2017 (**Figure 4**), March 2018 (**Figure 5**), and August 2018 (**Figure 6**). The September 2017 data set had two observed exceedances of the suggested QMRA exposure limit for the HF183 marker of 1000 tsc/100 mL: these were the Grotto (site NEB01) at 1,012 tsc/100 mL and the Garapan Drainage #1 (site WB17) at 4,786 tsc/100 mL. In March 2018, there was only one exceedance of the QMRA suggested HF183 exposure threshold of 1000 tsc/100 mL at the Grotto (site

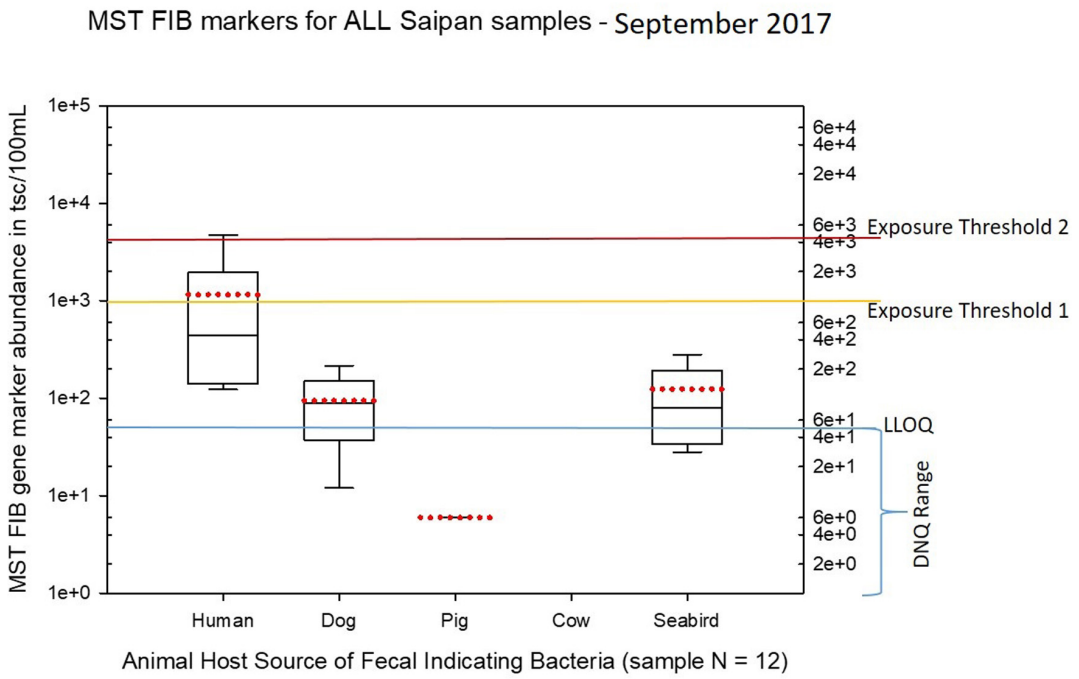


FIGURE 4 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for all Saipan coastal water samples collected in September of 2017. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age.

NEB01) with an observed level of 3,003 tsc/100 mL on March 16, 2018. However, during the August 2018 sampling event, there were four HF183 exceedances of the exposure threshold 1 of 1000 tsc/100 mL for surface water samples, with all exceedance sites clustered in the western central Saipan Lagoon shoreline area: Garapan Fishing Dock (site WB21) at 1,430 tsc/100 mL, site S18 at 8,442 tsc/100 mL, site S20 at 52,410 tsc/100 mL, and site S21 at 8,819 tsc/100 mL. This same western central Saipan Lagoon shoreline area in the Garapan Lagoon region also showed high elevations of HF183 human *Bacteroides* FIB marker in beach groundwater samples. Groundwater site GW-S18 had 46,062 tsc/100 mL on August 3 and 30,293 tsc/100 mL on August 7, while groundwater site GW-S19 had 5,550 tsc/100 mL on August 3 and 4,814 tsc/100 mL on August 7 (see **Supplementary Table 5** for more details). It should be noted that these extreme levels of the HF183 marker over the QMRA suggested threshold of 4100 copies/100 mL were relatively close to one another (American University sample sites S18, S20, and S21) and maintained this elevated level over several days in the region of the western central Saipan Lagoon shoreline. Boxplots showing the statistical pattern of MST markers, particularly for the western central Saipan Lagoon shoreline region, are shown in **Figure 7**. The pattern for submarine groundwater discharge and nitrogen stable isotope tracking is shown in the boxplots for **Figure 8**, and the patterns for MST markers in groundwater specifically for the Garapan Lagoon region is shown in the boxplots of **Figure 9**. In this particular area of the Saipan Lagoon, human fecal marker input was predominant in both frequency and abundance as compared to the other fecal host markers tested.

Integrated primary producer $\delta^{15}N$ values ranged from 4.3‰ (Tanapag region, August) to 12.0‰ (Garapan Lagoon region, March) and were generally higher in samples from the Garapan region of the Lagoon, and among those samples collected in March (**Figure 8**). SGD was also variable, ranging from a low of 0.01 $m^3\ m^{-1}\ d^{-1}$ (Tanapag Lagoon, March) to 10.9 $m^3\ m^{-1}\ d^{-1}$ (Garapan Lagoon, March) and overall, was most prevalent in Garapan Lagoon (**Figure 8**). Overall, the data pointed to an area of consistent SGD-borne, sewage-derived N inputs in Garapan Lagoon (sites S17–S21).

The Grotto (Site NEB01) frequently had elevations of both live enterococci and human fecal HF183 marker, plus some elevations of seabird marker, as shown in **Figures 2**, **10**. For the Grotto water samples in **Figure 10**, several HF183 exceedances of the suggested threshold of 1000 copies/100 mL were observed in multiple samples, but HF183 exceedances were not observed for the exposure threshold of 4100 copies/100 mL.

DISCUSSION

As part of the broader NOAA technology transition effort to aid BECQ in developing an in-house capacity for qPCR-based molecular MST within the Bureau for Environmental and Coastal Quality, NOAA-AOML and CNMI-BECQ researchers conducted a joint MST baseline study to assess the patterns of host-source FIB from LBSP in the coastal waters of Saipan. This study coincided with and was supported by sampling assistance from American University, which was conducting a separately funded

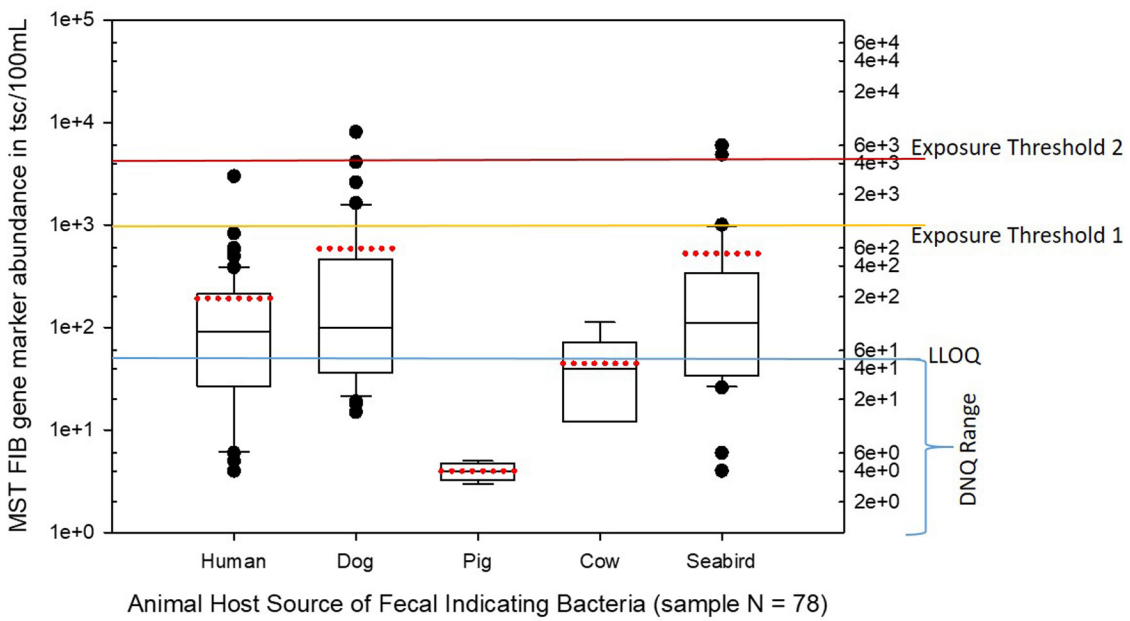


FIGURE 5 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for all Saipan coastal water samples collected in March of 2018. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age.

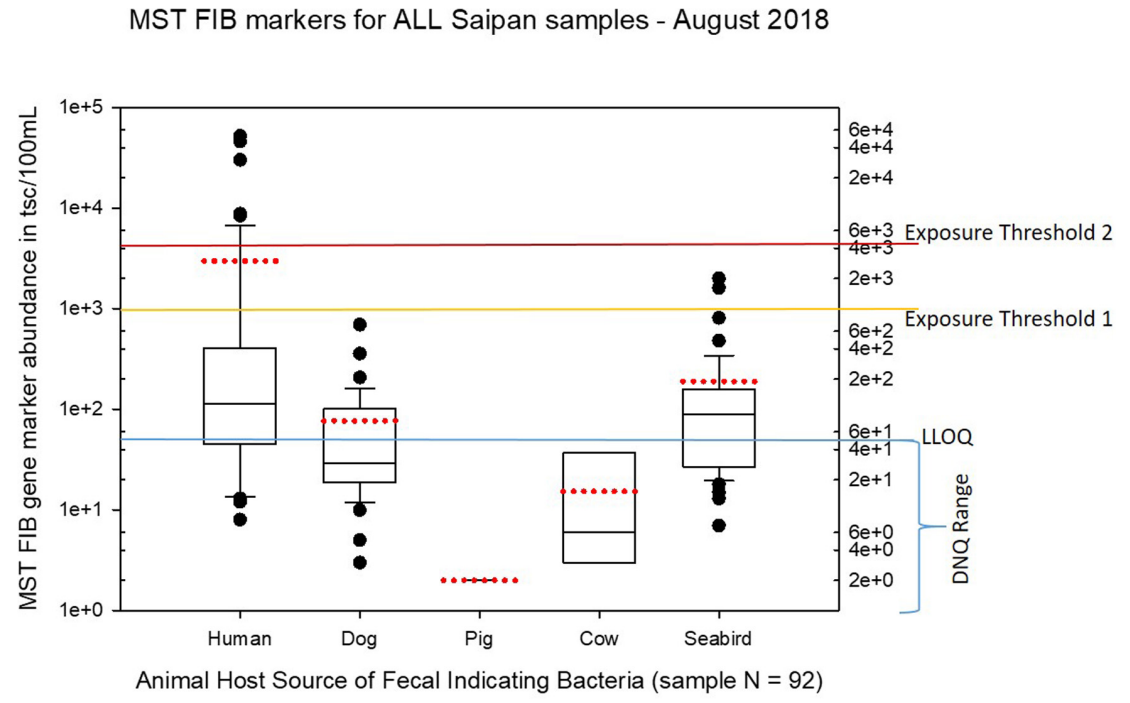


FIGURE 6 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for all Saipan coastal water samples collected in August of 2018. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age.

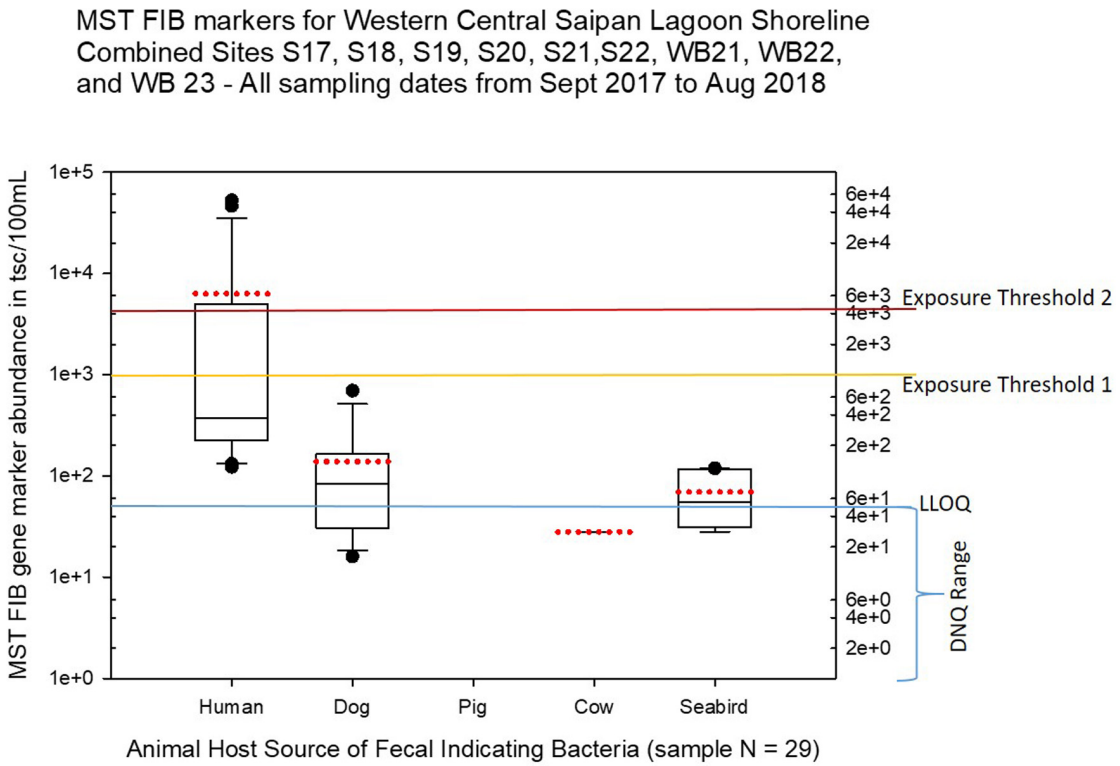


FIGURE 7 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for sample sites along the western central Saipan Lagoon shoreline, including sites S17, S18, S19, S20, S22, WB21, WB22, and WB23. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age. All exceedances for QMRA exposure threshold 2 come from sites S18, S20, and S21.

independent study in Saipan of nutrient measurements, nitrogen stable isotope source tracking for nutrient pollution, and tracking radon (Rn) and radium (Ra) isotopes as natural tracers of SGD (Knapp et al., 2020).

Our study found that the predominant inputs of FIB into the critical and protected Saipan Lagoon system and reef tract ecosystems were primarily from human and dog fecal sources. Birds were also a major contributor of fecal bacteria at certain sites, especially along the more sparsely settled northern and northeastern parts of Saipan. The distribution of dog fecal bacteria contamination was relatively ubiquitous around the island and was especially prevalent in the Tanapag area, the urbanized Garapan district, and along the central western coastline bordering Saipan Lagoon. In addition, this Garapan Lagoon region of the Saipan Lagoon system is subject to substantial submarine groundwater discharge (SGD), especially in the sections from sites S17 to S21, which also carry sewage-derived N into the coastal environment (**Figure 8**).

Two primary hotspots with high levels of human FIB were observed: a wide stretch of the urbanized and developed central western Saipan Lagoon shoreline in the Garapan Lagoon region, and also at the Grotto, an isolated coastal grotto swimming and scuba diving area on the northeast end of the island that attracts extremely large numbers of tourists. Both areas have also been noted as chronic hotspots for elevated enterococci during routine water quality monitoring by the BECQ, although there are other chronic enterococci hotspots previously observed by the BECQ around the island where we did not observe associated high levels of the human marker during this study.

For our study, it was assumed that human fecal contamination, particularly from municipal sewage with a large population of contributors, was likely to be the highest risk fecal input source for water contamination (Soller et al., 2010a,b, 2014). One could cautiously speculate that the other fecal bacterial MST markers used in this Saipan study might be assumed to have similar exposure thresholds of at least the same scope (or more likely even higher thresholds) to that of that human fecal marker, for a similar level of relative public health risk, since the animal fecal sources might potentially be less risky than the human sources. If the more protective human HF183 exposure threshold (i.e., sewage < 3 day old) level of 1000 copies/100 mL is accepted for a predicted illness rate of less than 30 illnesses per 1000 bathers, then one might also apply this same threshold to the interpretation of the other MST markers used in this study, and be reasonably confident of being conservatively protective of public health. However, it must be recognized that bacteria targeted by different fecal source markers will have different attenuation/die-off rates to the *Bacteroides* targeted by the HF183 marker, and some experiments have shown that the source of feces can affect die-off

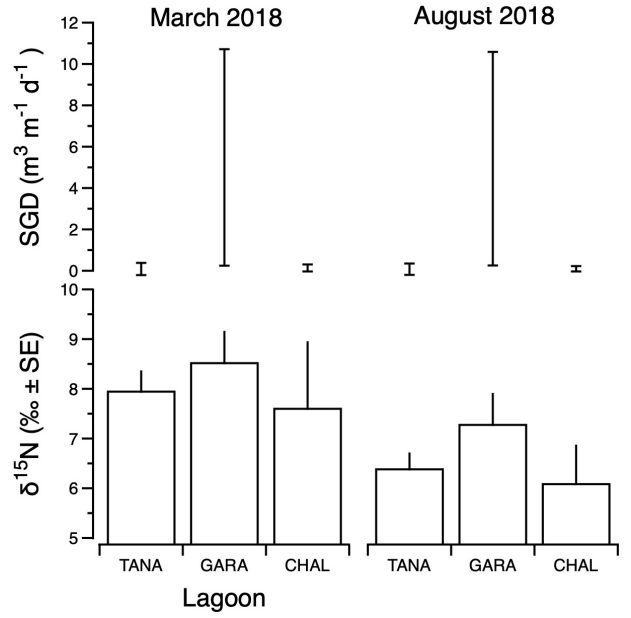


FIGURE 8 | Composite $\delta^{15}N$ values of benthic primary producers **(lower panels)** and submarine groundwater discharge (SGD) estimates **(upper panels)** for Tanapag (TANA), Garapan (GARA) and Chalan Kanoa (CHAL) Lagoons of the west coast lagoon system of Saipan. For SGD, upper and lower bounds of estimates are shown. Data for two sampling events are shown: March 2018 **(left panels)** and August 2018 **(right panels)**.

of fecal bacteria and the environmental persistence of fecal source tracking markers (Korajkic et al., 2013, 2019). We thus suggested that for purposes of interpreting MST results from this Saipan study, levels below this value of 1000 copies per 100 mL probably would not reflect a significant public health concern for any of the MST markers used in this study. Therefore, given a lack of epidemiological or QMRA data on the relationships of illness and other host-specific MST markers, we applied the same HF183 risk exposure guidelines as adapted from Boehm et al. (2018) to the other host-source MST markers used in this study. We believe this should be conservatively protective given the likely relative risk of the different fecal host sources.

Although there is a moderate degree of cattle ranching and pig farming on the island of Saipan, the relative levels of both cow and pig fecal Bacteroidales markers in the coastal waters tested were observed to be quite low, with the majority of samples being either non-detects or in the DNQ range. No samples exceeded the 1000 copies/100 mL exposure threshold for either cow or pig fecal Bacteroidales marker. The reason for the lower than expected detection of these agricultural markers is unclear. Previous studies have observed that dietary changes and the age of cattle can influence the abundance and distribution patterns of bacterial fecal source tracking markers in cattle, including both *Bacteroides* and *Enterococcus* species of fecal bacteria, as well as fecal pathogens (Busato et al., 1999; Shanks et al., 2010, 2014; Raith et al., 2013). It is possible that seawater dilution factors, as well as attenuation of marker signal with the distance traveled (via land runoff, stormwater discharge, etc.), might have been a factor in the lower detection of the agricultural fecal markers in coastal waters, especially if the original signal was relatively low in abundance. Also, the particular CowM2 marker used here specifically targets a bovine-host Bacteroidales single-copy gene sequence predicted to encoded for membrane-associated and secreted proteins (Shanks et al., 2006, 2008). Many host-specific MST qPCR assays target ribosomal genes, which are generally considered highly desirable genetic targets as they are highly conserved and each bacterial genome may contain many copies of the ribosomal gene target, making it easier to detect a small number of cells at low concentrations. However, the development of bovine-specific qPCR assays has been restricted in part by the limited amount of genetic variation in the rRNA genes of bovine-associated *Bacteroidales* (Shanks et al., 2008). The currently available host-specific qPCR methods targeting ribosomal genes designed to detect bovine fecal pollution can typically discriminate only between ruminant and non-ruminant sources (Layton et al., 2006; Reischer et al., 2006). The CowM2 target used here, putatively a single-copy gene target, exhibits high levels of host specificity for bovine-associated *Bacteroidales*, and high reaction sensitivity, but it may not be evenly distributed across cattle populations, and as a single-copy target it may have a substantially lower environmental detection sensitivity than the more general multi-copy ribosomal ruminant gene markers (Shanks et al., 2008). This might make the CowM2 marker more susceptible to target signal attenuation, dilution, and decay during transport to coastal waters. In addition, while the CowM2 and Pig2Bact markers have shown high reaction sensitivity and specificity in many studies in North America, Europe, and Australia, they have not always shown similar sensitivity in other geographic areas with cattle populations, such as in India (Odagiri et al., 2015). It might be that dietary or population characteristics of the particular livestock herds in Saipan may be affecting the sensitivity of these agricultural fecal marker assays. This is the first time these MST assays have been used for any livestock of the Mariana Islands archipelago. Therefore, it is likely that further optimization for detection sensitivity testing of the cow and pig markers for the particular local Saipan livestock populations may be warranted, as well as conducting qPCR testing for broader multi-copy ribosomal ruminant markers, such as the Rum2Bact assay (Mieszkin et al., 2010). We have recommended that BECQ follow up with future studies incorporating both a Saipan-optimized CowM2 assay and the more general Rum2Bact ruminant marker assay.

In examining Saipan waters, three samples showed the HF183 human FIB marker to be over 10,000 copies per 100 mL (i.e., substantially above the QMRA-based exposure threshold of 4100 copies/100 mL for human fecal contamination of uncertain age). The HF183 levels observed in these samples reflect a potentially serious exposure risk to public health and may be suggestive of significant leakage of sanitary infrastructure at those sites on those sample dates. It should be noted that these extreme exceedances of the HF183 marker occurred at sites relatively close to one another (sites S18 and S20) and repeatedly over the course of a few days in the region of the western central Saipan Lagoon shoreline. Two of these extreme exceedance samples were

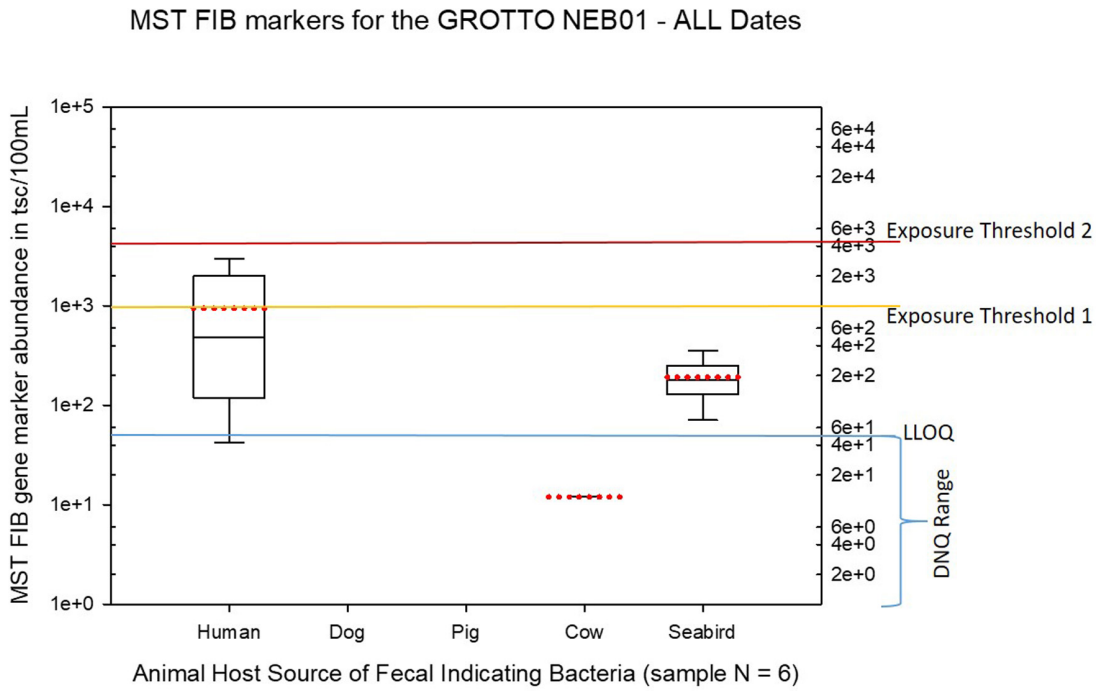


FIGURE 9 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for water sample sites collected from the Grotto (site NEB01) for all sample dates from March 2017 to August 2018. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age.

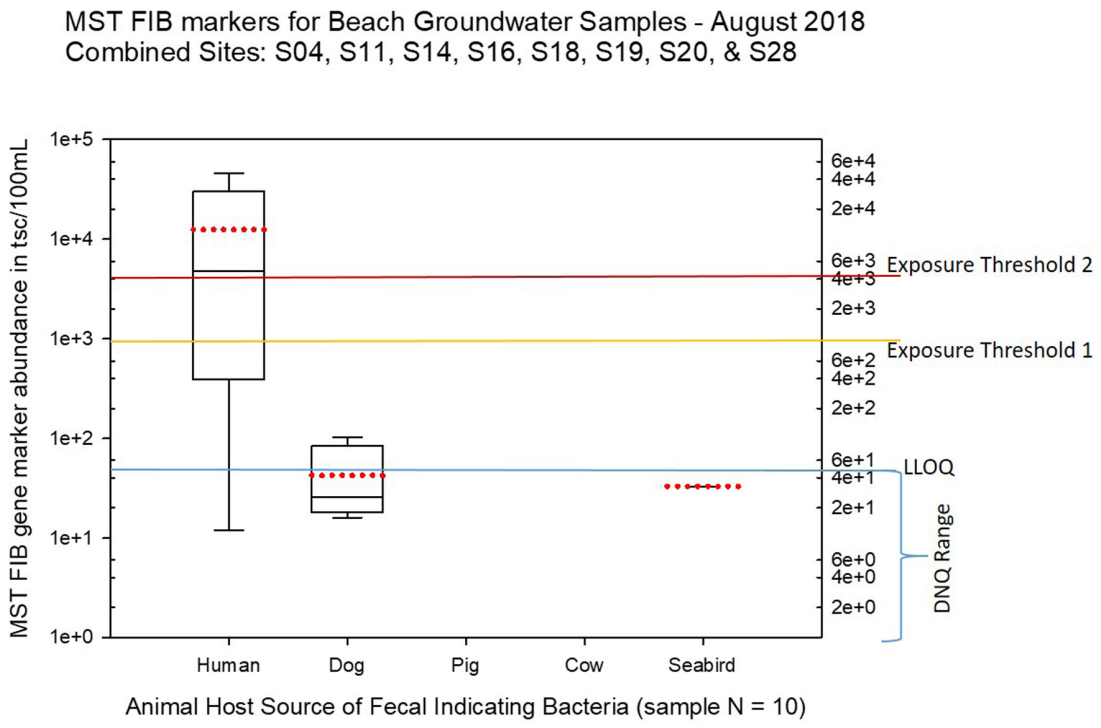


FIGURE 10 | Boxplots summarizing the statistical distribution of the relative abundance of host-specific MST FIB markers for groundwater sample sites along the western central Saipan Lagoon shoreline, including sites S17, S18, S19, S20, S22, WB21, WB22, and WB23. Exposure threshold 1 = QMRA suggested limit for HF183 marker from sewage contamination ≤ 3 days. Exposure threshold 2 = QMRA suggested limit for HF183 marker from sewage contamination of uncertain age. All of the exceedances of the QMRA exposure threshold 2 come from sites S18 and S19.

from groundwater samples at site S18, collected on August 3 and August 7 respectively, in 2018. The third extreme level of human FIB marker was observed in the shoreline sample of site S20 collected on August 2, 2018. S18 was consistently the site with the highest SGD, and S20 was also located within the high-SGD area of Garapan Lagoon. Additionally, SGD across all sites was higher in August (rainy season). This suggests that SGD may transport dangerous levels of human fecal pollution to the coast. All shore sites in this region of the Saipan Lagoon had elevated HF183 marker levels and accounted for the majority of the exceedances of the QMRA exposure thresholds for our data set (with the other primary exceedance site for HF183 being at the Grotto). The urban area of west central Saipan has substantial aging sanitary infrastructure, and it is probable that the observed elevations in human fecal marker were associated with compromised and leaking sanitary infrastructure sources in the area. Indeed, following the conclusion of this study, some compromised sanitary infrastructure in the region of highest human fecal markers (particularly in the groundwater) was later discovered and addressed for mitigation.

This region of the west central Saipan Lagoon shoreline also had elevated dog fecal bacterial marker in many of the samples. Dog fecal bacterial marker appeared to be relatively widespread about the island, with many sample sites having significant elevations, including the Tanapag Meeting Hall area, the GrandVrio and Fiesta hotels, the Central Repair Shop area, the Garapan Fishing Dock area, and throughout the western central region of the Saipan Lagoon shoreline. Dog fecal contamination has been shown to have a substantial impact on microbial water quality (Garfield and Walker, 2008; Wright et al., 2009; Cinquepalmi et al., 2013). Although previous studies have found that gene markers associated with dog feces may be present in some sewage influents, especially in some highly urbanized communities (Caldwell and Levine, 2009), plus the fact that the US EPA, the Humane Society, and other organizations have recommended flushing pet fecal waste as a safe disposal method, the collection and flushing of dog fecal waste in Saipan is not a typical practice. There is a very large outdoors free-roaming dog population (more than 21,000 at last survey in 2018) which is largest in urban areas (Nimer et al., 2018). Also, we did not detect significant levels of dog fecal marker in the ground water samples tested, whereas some groundwater samples had very high levels of human fecal marker (**Supplementary Table 5**), however we did detect elevated levels of dog marker in surface waters of this area. This is why we are suggesting that the observed dog fecal marker in Saipan coastal waters most likely represented stormwater/runoff transport mechanisms, while human fecal marker most likely represented sewage/septage inputs. In general, there were no effective dog fecal cleanup activities in place at beaches or inland areas, as well as no dog fecal hygiene policies, at the time of this study. As such, dog fecal contamination appears relatively commonplace and widespread in runoff and stormwater discharge, particularly in the more developed parts of the island.

The results of the nitrogen stable isotope nutrient tracking and Rn-derived SGD measurements from the Tanapag, Garapan, and Chalan Kanoa regions of the western Saipan Lagoon are generally consistent with the qPCR MST observations from this area. The $\delta^{15}N$ and SGD observations from this study have been separately published in more detail elsewhere (Knapp et al., 2020). In summary, three clusters of sites along the western coast were identified — around sites S8, S18, and S26 — that had $\delta^{15}N$ values greater than 10‰ in seagrasses and algae. These $\delta^{15}N$ values are consistent with either a sewage or manure N source to these parts of Saipan Lagoon, but sewage is more likely for several reasons. Agricultural practices that use or generate manure are not very prevalent in this area (United States Department of Agriculture [USDA], 2020), but human population centers with failing wastewater infrastructure are. Additionally, elevated $\delta^{15}N$ in algae was spatially correlated with human fecal HF183 MST markers, which would not originate from manure. In contrast, algae collected along the reef line had comparatively lower $\delta^{15}N$ values, indicating limited availability of sewage-derived nitrogen offshore. It was further observed that Rn measurements highlighted the spatial and temporal variability in the input of groundwater into Saipan Lagoon. As expected, groundwater inputs were higher during August (i.e., the rainy season) and were the most pronounced near site S18. When surface and groundwater were analyzed for nutrients, groundwater nitrate concentrations were nearly an order of magnitude higher than those in surface water, indicating groundwater flow was an important pathway of nitrogen pollution into Saipan Lagoon at the time of the study.

These observations of nutrients, stable nitrogen isotopes, and the natural SGD tracer Rn were consistent with observations of the patterns of human FIB input into this same area of Saipan Lagoon. Our MST study along with the stable nitrogen isotope and SGD study collectively observed the groundwater and shoreline samples from this area for sites S18–S22 to simultaneously have elevated human-source fecal bacteria, elevated nutrients, and elevated Rn, strongly implicating contaminated groundwater input to the Saipan Lagoon in this area. The greatest chronic export of high levels of FIB, and particularly high-risk human-source FIB appeared to come from groundwater discharge and shoreline runoff discharge along the western central Saipan Lagoon region stretching from roughly the Garapan Fishing Dock to Susupe, with the worst of the discharge in areas near sites S18 through S20. This is likely the LBSP microbial contaminant source causing the greatest risk exposure of LBSP pathogens to the coral reefs of Saipan Lagoon (at least at the time of our study). Reef waters just offshore from this area also showed the greatest abundance of LBSP microbial contaminants, particularly human FIB marker.

The other hotspot for the HF183 human fecal marker was at the Grotto, site NEB01, a popular tourist destination lacking sufficient sanitary infrastructure. There is only one restroom facility near a parking lot at the top of a cliff located a distance from the Grotto. The Grotto is protected from open ocean wave action, but is tidally flushed. It often experiences high densities of swimmers, bathers, scuba divers, and general tourists who frequently visit this natural attraction in large groups. The swimming areas of the Grotto can only be accessed by descending a very long, steep stairway cut into the cliffside. Large crowds

often move up and down the stairway, making transit into and out of the Grotto area difficult and time consuming. There are no restroom facilities in the Grotto area itself, and due to the challenging climb up the stairway, many visiting swimmers may either be unable or unwilling to access the restroom in the parking area at the top of the cliff.

Routine testing by the BECQ shows that enterococci levels at the Grotto frequently exceed the water quality criteria for the CNMI. The random but high densities of visiting swimmers at this site most probably have contributed to the pattern of FIBs observed there. While HF183 was not the only fecal marker detected in water samples from the Grotto, it was the dominant marker detected, although seabirds were also significant contributors to fecal contamination in the region. It is likely that bacterial shedding from the swimmers represents a primary source of the human-source fecal bacteria MST markers detected. Previous studies by others and ourselves have shown that bathers can shed large amounts of fecal and skin-associated bacteria, including *Enterococcus*, fecal *Bacteroides*, and *Staphylococcus aureus* among others (Gerba, 2000; Elmir et al., 2007, 2009; Plano et al., 2011). While there are undoubtedly some incidents of direct defecation into Grotto waters by certain individuals unwilling to hike back up the cliff to the only restroom facility, this human fecal marker signal is more likely attributable to unintentional routine bacterial shedding by general bathers. Previous studies have shown that bathers can potentially release large amounts of enterococci into the surrounding water just by normal bathing activities or even just soaking quietly, and that the degree of this bacterial shedding is highly variable between individuals and over time, and that some individuals may even be "super-shedders." Young children and toddlers especially can also potentially shed inordinately high levels of FIB when bathing or playing in water. Both pool studies and a beach epidemiological study where participants were directed to sample their own personal water space while bathing have demonstrated bacterial shedding phenomenon by bathers (Elmir et al., 2007, 2009; Shibata et al., 2010; Sinigalliano et al., 2010). The combination of large crowds of bathers in a relatively small water body with poor circulation, or limited flushing at a particular time may lead to elevated FIB levels with the crowd of people themselves as a point source. The high-density crowds at the Grotto have also been observed to potentially represent other physical health and safety risks, particularly during the descent and ascent of large numbers of people along the steep stairway leading from the parking area to the waters of the Grotto. More robust crowd control to limit bather density at any given time, along with allowing for periods of low or closed activity to permit the natural tidal flushing of local waters, are all measures that might serve to improve both general safety and water quality at this popular and unique Saipan attraction. Additional restroom facilities that are protected from discharge or leakage to coastal waters might also be considered.

Since the time of this qPCR MST study reported here, a new on-going study of the Grotto area has been initiated by a NOAA Coral Fellow working with the BECQ. This study is looking at enterococci exceedances compared to the daily bather numbers at the Grotto, with the idea of examining thresholds of user capacity that may trigger enterococci exceedances. This study has been conducted through the impacts of both Super Typhoon Tutu and the COVID-19 pandemic. Preliminary data through July 2020 shows that the percentage of Enterococci sample exceedances at the Grotto have been reduced by half from the 2019 levels and have decreased more than threefold from the 2018 levels. It is believed that this reduction may be attributed to extensive flushing of the Grotto area by Super Typhoon Yutu and to the substantially reduced tourism following Super Typhoon Yutu and the COVID-19 pandemic. The Grotto has been closed to the public since March 2020. This is still preliminary data, and the final results of the Grotto bather usage study will be published elsewhere (unpublished data, personal communication LM, BECQ). However, these later observations from the Grotto about the reductions in enterococci levels during the COVID-19 pandemic and beach closures are consistent with the human-marker MST data presented here, and further support the idea that bather shedding is a primary input of FIBs in the Grotto area.

Bird fecal input appeared to be a major contributor to other observed enterococci hotspots, with areas such as Bird Island and American Memorial Park Drainage routinely having regulatory enterococci exceedances and high levels of Gull2 seabird marker. No other significant sources of fecal marker were found in the water samples from Bird Island; it is therefore logical to assume that bird populations are the primary cause for the chronic enterococci water quality failures of this area. Based on our Gull2 qPCR results, birds also appeared to be major fecal contributors to other areas of chronic enterococci exceedance, such as Jeffrey's Beach and Old Man by the Sea. Thus, birds would appear to be significant contributors of FIB at areas such as the American Memorial Park Drainage, the Lao Lao Bay area, the Grotto, Jeffrey's Beach, Old Man by the Sea and, of course, Bird Island. Such bird fecal contamination is probably not manageable in a practical sense. However, the potential of bird fecal contributions and their relatively lower risk, should be considered when assessing observed exceedances of general fecal indicators such as enterococci or *E. coli*, especially as bird fecal contributions may serve to confound general water quality assessments. If significant bird contribution is determined to be a primary contributing factor in recreational water sites of chronic exceedance without any other high-risk fecal input sources, alternative site-specific water quality criteria might be considered. In this case, a combination of methods, including MST and Quantitative Microbial Risk Assessment, could be deployed to help confirm if any such alternative criteria promulgated by the BECQ would likely be equally protective of both public and environmental health.

In examining coral reef exposure, this study found that LBSP-derived microbial contaminants were being transported from the Saipan Lagoon coastline to the reef tract. Therefore, the reefs offshore from the Saipan Lagoon coast may also potentially be exposed to pathogens and other contaminants associated with this LBSP pollutant transport (likely including microbial, nutrient, and chemical contaminants as well). While reef waters were tested during this study, coral tissues were not. Consequently, it could not be determined whether microbial contaminants were being absorbed into or influencing the coral holobionts at these reefs. However, impacts of LBSP microbial

contaminants on coral microbiota have been documented at other coral reefs, including reefs in southeast Florida (Staley et al., 2017). A follow-up study to test coral tissue along the Saipan reef tract for the presence of fecal indicator markers and/or pathogens would be useful to further define reef exposure. From the pattern of results of this MST source tracking study, it would appear the most likely source of the exposure and impact to these reef sites is the western central Saipan Lagoon shoreline area, particularly the region from roughly sites S17 through S22 (i.e., roughly from the area of the Garapan Fishing Dock to Susupe).

Based on this MST study, the following recommendations have been made to the resource managers of the CNMI-BECQ:

(1) Conduct further investigations of sanitary infrastructure and potential leakages (including further investigation of potential groundwater contamination), especially along the west central Saipan Lagoon shoreline, and aggressively pursue the detection and repair of sanitary infrastructure in the region.
(2) Conduct further investigations of the potential contamination of stormwater runoff, especially in the regions of Tanapag, Garapan district (including American Memorial Park), Susupe, and the west central Saipan Lagoon shoreline.
(3) Implement more robust control measures for dealing with animal waste, particularly dog fecal contamination, especially along the western coast of Saipan Lagoon. A large number of dog fecal marker elevations suggest that dog fecal influence is widespread. Among other things, consider mandating dog fecal cleanup policies and provide the necessary resources to make dog fecal cleanup in public spaces easy and cost effective.
(4) Implement more robust bather crowd control measures at high density bathing beaches and recreational water areas, especially at the Grotto, along with more accurate bather density monitoring of crowds at the Grotto. Managing crowd size at the Grotto will allow more time for natural tidal flushing, which might be advantageous to maintaining better water quality at the site. It would also likely provide a benefit in increased physical safety and reduced accidents.
(5) Integrate periodic molecular MST assessments into the more routine water quality monitoring efforts of the BECQ to permit a better understanding of the seasonality, infrastructure, environmental, and social factors that influence source contributions of contaminants to local waters, and to help trouble-shoot any sites of chronic exceedances. BECQ is now also planning to expand MST assessments to other less populated islands of CNMI, including Tinian and Rota, where it is suspected that the predominant FIB sources may be dogs and birds, rather than humans as in Saipan (personal communication, LM, BECQ).
(6) Further test the CowM2 and Pig2Bact qPCR assays to determine whether local or regional factors (e.g., livestock diet, etc.) affect the detection sensitivity of these two livestock assays. If this turns out to be the case, modified or alternate MST protocols for livestock could be developed or adopted to boost the detection sensitivity for local Saipan livestock fecal contamination. Also, supplement testing with the use of a more general ruminant fecal marker such as the Rum2Bact assay.

The results from this study demonstrate the value of integrating molecular MST technology into the assessments of water quality in the CNMI, along with providing the first baseline assessment of host-specific fecal water quality indicators for the CNMI region.

CONCLUSION

In conclusion, we identified a need for molecular MST capability within CNMI regulatory agencies. We initiated an MST technology transition effort between NOAA-AOML and the CNMI-BECQ and used this new MST capacity at the BECQ to conduct the first baseline assessment of host-specific fecal water quality microbial indicators in the coastal waters and protected coral reefs of Saipan. We found that the west central shoreline region of Saipan Lagoon was the greatest contributor of human-specific fecal bacteria contamination to the lagoon and reef tract during the period of study. Our study also identified another hotspot of human fecal marker contamination at the popular recreational swimming and scuba diving site of the Grotto. We have suggested that the primary input of fecal bacteria for the Saipan Lagoon is most likely compromised sanitary infrastructure along the western Saipan Lagoon shoreline from Garapan Fishing Dock to Susupe, while at the Grotto the most likely cause is the high density of human bathers.

Birds also appeared to be significant contributors of FIB to areas such as the American Memorial Park Drainage, the Lao Lao Bay area, the Grotto, Jeffrey's Beach, Old Man by the Sea, and, of course, Bird Island. Such bird fecal input is probably not manageable in a practical sense, but it should be recognized that bird fecal contributions may serve to confound general water quality risk assessments such as the use of general culture-based enterococci tests. We also detected much less fecal marker associated with pigs and cows than expected, but we suspect the particular qPCR assays used (CowM2 and Pig2Bact) might have a lower sensitivity to the specific populations of Saipan livestock. We recommend that further investigation and optimization of MST methods for these specific populations of Saipan livestock may be warranted, as well as use of more general ruminant fecal qPCR assays. This study also found that moderate levels of dog fecal contamination were widespread on the island and have suggested a more robust control of animal fecal waste management. Overall, we believe the inclusion of molecular MST capability as demonstrated here can be a useful addition to the water quality toolbox of resource managers in the CNMI to enhance the ability of the BECQ to better protect public health and that of the environment.

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

CS conceived and designed the experiments, performed the experiments, contributed reagents, materials, and analysis tools, analyzed the data, prepared the figures and/or tables, authored or reviewed drafts of manuscript, approved the final draft, and serves as Principal Investigator and as corresponding author. KK conceived and designed the stable nitrogen isotope experiments, participated in sampling, analyzed the data, prepared the figures, authored or reviewed drafts of manuscript, and approved the final draft. MG conceived and designed the experiments, performed the experiments, contributed reagents, materials, and analysis tools, analyzed the data, prepared the figures and/or tables, authored or reviewed drafts of manuscript, and approved the final draft. KY conceived and designed the experiments, analyzed the data, prepared the figures and/or tables, authored or reviewed drafts of manuscript, and approved the final draft. KKn conceived and designed the SGD experiments, participated in sampling, analyzed the data, authored or reviewed drafts of manuscript, and approved the final draft. DP performed the experiments, contributed reagents, materials, and analysis tools, reviewed drafts of manuscript, and approved the final draft. CB performed the experiments, contributed reagents, materials, and analysis tools, authored or reviewed drafts of manuscript, and approved the final draft. AB performed the experiments, analyzed the data, reviewed drafts of manuscript, and approved the final draft. HL performed the experiments, analyzed the data, reviewed drafts of manuscript, and approved the final draft. LM provided oversight for BECQ activities for the project and the manuscript, contributed materials and analysis tools, reviewed drafts of manuscript, and approved the final draft.

FUNDING

This study was funded by NOAA's Coral Reef Conservation Program (CRCP Project # 31184 to CS and MG) and by the NOAA 'Omics Initiative Program (to CS and MG). Additional funding and support for the acquisition and operation of the Pall GeneDisc qPCR instrumentation at the CNMI-BECQ was provided by the US EPA. The research for the stable nitrogen isotope and SGD observations was supported by NOAA (award #NA17NOS4820082 to KK and KKn), NASA DC Space Grant (to Melissa A. Knapp and KKn), the Research Grants Council, Hong Kong and the HKU Faculty of Science SIRMS postdoctoral fellowship (T21-602/16-R for Naomi Geeraert). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We would like to give special recognition to the efforts of the staff of the Water Quality Surveillance/Non-Point Source, Planning, Permitting, and Marine Monitoring Team branches of the BECQ who aided in the collection and processing of water quality samples, without whom this study would not have been completed (listed in alphabetical order): David Benavente, Rodney Camacho, Erin Derrington, Katie Graziano, John Iguel, Joseph Ito, Lyza Johnston, Carlos Ketebengang, Shawn Masga, Vanessa Nogis, Emily Northrop, Olivia Tenorio, and KY. We would also like to especially recognize the staff of the BECQ laboratory who helped analyze samples and provided scientifically defensible data for water quality assessments (listed in alphabetical order): CB, Michael Kapileo, DP, Melvin Piteg, and Jaime Reyes. Finally, we acknowledge and thank the Pall Corporation, GeneDisc Division, for their collaboration and cooperation in adapting the GeneDisc instrumentation, software, and GeneDisc plate format to run the standard microbial source tracking qPCR assays that were transitioned in this project to operations at the BECQ laboratory. Pall Corporation donated personnel time to help in troubleshooting technology transition issues related to the GeneDisc thermocycler platform and provided this study with a supply of instrument-specific consumables such as the custom empty blank GeneDisc plates and associated GeneDisc mastermix reagent. We would especially like to acknowledge the help of Wayne Miller for his coordination efforts between Pall and our research project, and Dr. Stephane Bonilla of the Pall GeneDisc research and development team who helped with troubleshooting the adaptation and modification of the GeneDisc instrument and software to run these MST assays on the GeneDisc platform.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.596650/full#supplementary-material

REFERENCES

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

Boehm, A. E., Graham, K. E., and Jennings, W. C. (2018). Can we swim yet? Systematic review, meta-analysis, and risk assessment of aging sewage in surface waters. *Environ. Sci. Technol.* 52, 9634–9645. doi: 10.1021/acs.est.8b01948

Busato, A., Hofer, D., Lentze, T., Gaillard, C., and Burnens, A. (1999). Prevalence and infection risks of zoonotic enteropathogenic bacteria in Swiss

cow-calf farms. *Vet. Microbiol.* 69, 251–263. doi: 10.1016/s0378-1135(99)00119-4

Caldwell, J. M., and Levine, J. F. (2009). Domestic wastewater influent profiling using mitochondrial real-time PCR for source tracking animal contamination. *J. Microbiol. Methods* 77, 17–22. doi: 10.1016/j.mimet.2008.11.007

Campbell, A., Fleisher, J., Sinigalliano, C. D., White, J., and Lopez, J. V. (2015). Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of Southeast Florida. *Microbiologyopen* 4, 390–408. doi: 10.1002/mbo3.245

Carruth, R. L. (2003). *Ground-Water Resources of Saipan, Commonwealth of the Northern Mariana Islands*. Honolulu, HI: USGS.

Casciotti, K. L., Sigman, D. M., Hastings, M. G., Bohlke, J. K., and Hilkert, A. (2002). Measurement of the oxygen isotopic composition of nitrate in seawater and freshwater using the denitrifier method. *Anal. Chem.* 74, 4905–4912. doi: 10.1021/ac020113w

Cinquepalmi, V., Monno, R., Fumarola, L., Ventrella, G., Carlia, C., Greco, M. F., et al. (2013). Environmental contamination by dog's faeces: a public health problem? *Int. J. Environ. Res. Public Health* 10, 72–84. doi: 10.3390/ijerph10010072

Elmir, S. M., Shibata, T., Solo-Gabriele, H. M., Sinigalliano, C. D., Gidley, M. L., Miller, G., et al. (2009). Quantitative evaluation of Enterococci and *Bacteroidales* released by adults and toddlers in marine water. *Water Res.* 43, 4610–4616. doi: 10.1016/j.watres.2009.07.006

Elmir, S. M., Wright, M. E., Abdelzaher, A., Solo-Gabriele, H. M., Fleming, L. E., Miller, G., et al. (2007). Quantitative evaluation of bacteria released by bathers in a marine water. *Water Res.* 41, 3–10. doi: 10.1016/j.watres.2006.10.005

Garfield, L., and Walker, M. (2008). Microbial water quality and influences of fecal accumulation from a dog exercise area. *J. Environ. Health* 71, 24–29.

Gerba, C. P. (2000). Assessment of enteric pathogen shedding by bathers during recreational activity and its impact on water quality. *Quant. Microbiol.* 2, 55–68. doi: 10.1023/A:1010000230103

Griffith, J. F., Layton, B. A., Boehm, A. B., Holden, P. A., Jay, J. A., Hagedorn, C., et al. (2013). *The California Microbial Source Identification Manual: A Tiered Approach to Identifying Fecal Pollution Sources to Beaches*. Costa Mesa, CA: Southern California Coastal Water Research Project.

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

Knapp, M. A., Geeraert, N., Kim, K., and Knee, K. L. (2020). Submarine groundwater discharge (SGD) to coastal waters of Saipan (Commonwealth of the Northern Mariana Islands, USA): implications for nitrogen sources, transport and ecological effects. *Water* 12:3029. doi: 10.3390/w12113029

Korajkic, A., McMinn, B. R., Harwood, V. J., Shanks, O. C., Fout, G. S., and Ashbolt, N. J. (2013). Differential decay of enterococci and *Escherichia coli* originating from two fecal pollution sources. *Appl. Environ. Microbiol.* 79, 2488–2492. doi: 10.1128/aem.03781-12

Korajkic, A., Wanjugi, P., Brooks, L., Cao, Y., and Harwood, V. J. (2019). Persistence and decay of fecal microbiota in aquatic habitats. *Microbiol. Mol. Biol. Rev.* 83:e00005-19. doi: 10.1128/MMBR.00005-19

Kruger, J., Kumar, S., Damlamian, H., and Sharma, A. (2010). *Oceanographic Survey, Shoreline Mapping and Preliminary Hydrodynamic Modeling Report, Saipan, Commonwealth of the Northern Mariana Islands*. SOPAC Data Release Report No. 26. Suva: Pacific Islands Applied Geoscience Commission, 63.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., and Sayler, G. (2006). Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* 72, 4214–4224. doi: 10.1128/AEM.01036-05

Mieszkin, S., Yala, J. F., Joubrel, R., and Gourmelon, M. (2010). Phylogenetic analysis of *Bacteroidales* 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J. Appl. Microbiol.* 108, 974–984. doi: 10.1111/j.1365-2672.2009.04499.x

Nimer, A., Meneses, N., Watson, Z. D., Shuster, S. M., and Benford, R. (2018). Population survey and management strategies of free-roaming dogs (*Canis familiaris*) on Saipan, Commonwealth of the Northern Mariana Islands. *J. Appl. Anim. Welf. Sci.* 21, 170–184. doi: 10.1080/10888705.2017.1406801

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., et al. (2015). Validation of *Bacteroidales* quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* 502, 462–470. doi: 10.1016/j.scitotenv.2014.09.040

Perez, D. I., Phinn, S. R., Roelfsema, C. M., Shaw, E., Johnston, L., and Iguel, J. (2018). Primary production and calcification rates of algae-dominated reef flat and seagrass communities. *J. Geophys. Res. Biogeosci.* 123, 2362–2375. doi: 10.1029/2017JG004241

Perreault, J. A. (2007). *Reconnaissance Study of the Hydrology of American Memorial Park, Island of Saipan, Commonwealth of the Northern Mariana Islands*. Reston, VA: USGS.

Plano, L. R. W., Garza, A., Shibata, T., Elmir, S., Kish, J., Sinigalliano, C. D., et al. (2011). Shedding of *Staphylococcus aureus* and methicillin-resistant *Staphylococcus aureus* from adult and pediatric bathers in marine waters. *BMC Microbiol.* 11:5. doi: 10.1186/1471-2180-11-5

Raith, M. R., Kelty, C. A., Griffith, J. F., Schriewer, A., Wuertz, S., Mieszkin, S., et al. (2013). Comparison of PCR and quantitative real-time PCR methods for the characterization of ruminant and cattle fecal pollution sources. *Water Res.* 47, 6921–6928. doi: 10.1016/j.watres.2013.03.061

Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., and Farnleitner, A. H. (2006). Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72, 5610–5614. doi: 10.1128/AEM.00364-06

Shanks, O. C., Atikovic, E., Blackwood, A. D., Lu, J., Noble, R. T., Santo Domingo, J. W., et al. (2008). Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 74, 745–752. doi: 10.1128/aem.01843-07

Shanks, O. C., Kelty, C. A., Peed, L., Sivaganesan, M., Mooney, T., and Jenkins, M. (2014). Age-related shifts in the density and distribution of genetic marker water quality indicators in cow and calf feces. *Appl. Environ. Microbiol.* 80, 1588–1594. doi: 10.1128/aem.03581-13

Shanks, O. C., Santo Domingo, J. W., Lamendella, R., Kelty, C. A., and Graham, J. E. (2006). Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Appl. Environ. Microbioliol.* 72, 4054–4060. doi: 10.1128/aem.00023-06

Shanks, O. C., White, K., Kelty, C. A., Hayes, S., Sivaganesan, M., Jenkins, M., et al. (2010). Performance assessment PCR-based assays targeting *Bacteroidales* genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 76, 1359–1366. doi: 10.1128/aem.02033-09

Shibata, T., Solo-Gabriele, H. M., Sinigalliano, C. D., Gidley, M. B., Plano, L. R., Fleisher, J., et al. (2010). Evaluation of conventional and alternative monitoring methods for a recreational marine beach with non-point source of fecal contamination. *Environ. Sci. Technol.* 44, 8175–8181. doi: 10.1021/es100884w

Sigman, D. M., Casciotti, K. L., Andreani, M., Barford, C., Galanter, M., and Böhlke, J. K. (2001). A bacterial method for the nitrogen isotopic analysis of nitrate in seawater and freshwater. *Anal. Chem.* 73, 4145–4153. doi: 10.1021/ac010088e

Sinigalliano, C. D., Enochs, I. C., Stamates, S. J., Jones, J. R., Featherstone, C. M., Gidley, M. L., et al. (2019). *Water Quality and Coral Reef Monitoring Along the Southeast Florida Coast*. NOAA Technical Report OAR-AOML-47. Washington, DC: National Oceanic and Atmospheric Administration, 145. doi: 10.25923/aanj-0912

Sinigalliano, C. D., Ervin, J., Van de Werfhorst, J. L. C., Badgley, B., Ballesté, E., Bartkowiak, J., et al. (2013). Multi-laboratory evaluations of the performance of *Catellicoccus marimammalium* PCR assays developed to target gull fecal sources. *Water Res.* 47, 6883–6896. doi: 10.1016/j.watres.2013.02.059

Sinigalliano, C. D., Fleisher, J. M., Gidley, M. L., Solo-Gabriele, H. M., Shibata, T., Plano, L. R., et al. (2010). Traditional and molecular analyses for fecal indicator bacteria in non-point source subtropical recreational marine waters. *Water Res.* 44, 3763–3772. doi: 10.1016/j.watres.2010.04.026

Soller, J. A., Bartrand, T., Ashbolt, N. J., Ravenscroft, J., and Wade, T. J. (2010a). Estimating the primary etiologic agents in recreational freshwaters impacted by human sources of faecal contamination. *Water Res.* 44, 4736–4747. doi: 10.1016/j.watres.2010.07.064

Soller, J. A., Schoen, M. E., Bartrand, T., Ravenscroft, J. E., and Ashbolt, N. J. (2010b). 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

Soller, J. A., Schoen, M. E., Varghese, A., Ichida, A. M., Boehm, A. B., Eftim, S., et al. (2014). Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. *Water Res.* 66, 254–264. doi: 10.1016/j.watres.2014.08.026

Solo-Gabriele, H. M., Boehm, A. B., Scott, T. M., and Sinigalliano, C. D. (2011). "Beaches and coastal environments," in *Microbial Source Tracking: Methods, Applications, and Case Studies, Section IX: Case Studies*, 1st Edn, eds E. Hagedorn, A. Blanch, and V. J. Harwood (New York, NY: Springer-Verlag), 451–484. doi: 10.2112/08-1134.1

Staley, C., Kaiser, T., Gidley, M. L., Enochs, I. C., Jones, P. R., Goodwin, K. D., et al. (2017). Differential impacts of land-based sources of pollution on the microbiota of southeast Florida coral reefs. *Appl. Environ. Microbiol.* 83:e03378-16. doi: 10.1128/AEM.03378-16

Starmer, J., Asher, J., Castro, F., Gochfeld, D., Gove, J., Hall, A., et al. (2008). "The state of coral reef ecosystems of the U.S. commonwealth of the northern Mariana Islands," in *The State of Coral Reef Ecosystems of the United States and Pacific Freely Associated States: 2008. NOAA Technical Memorandum NOS NCCOS 73*, eds J. E. Waddell and A. M. Clarke (Silver Spring, MD: National Centers for Coastal Ocean Science), 437–463.

Symonds, E. M., Sinigalliano, C. D., Gidley, M. L., Ahmed, W., McQaig-Ulrich, S. M., and Breitbart, M. (2016). Fecal pollution along the southeastern coast of Florida and insight into the use of pepper mild mottle virus as an indicator. *J. Appl. Microbiol.* 121, 1469–1481. doi: 10.1111/jam.13252

United States Department of Agriculture [USDA] (2020). *Northern Mariana Islands (2018) Commonwealth and Island Data, In: USDA 2017 Census of Agriculture, Volume 1, Geographic Area Series Part 56. 2020. AC-17-A-56*. Available online at: https://www.nass.usda.gov/Publications/AgCensus/2017/Full_Report/Outlying_Areas/cnmi.pdf (accessed September 29, 2020).

US Environmental Protection Agency [US EPA] (2009). *Method 1600: Enterococci in Water by Membrane Filtration Using Membrane-Enterococcus Indoxyl-β-D-Glucoside Agar (mEI)*. Washington, DC: US Environmental Protection Agency, 42.

US Environmental Protection Agency [US EPA] (2012a). *Method 1611: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay*. Washington, DC: US Environmental Protection Agency.

US Environmental Protection Agency [US EPA] (2012b). *Recreational Water Quality Criteria (2012)*. Washington, DC: US EPA.

US Environmental Protection Agency [US EPA] (2019). *Method 1696: Characterization of Human Fecal Pollution in Water by HF183/BacR287 TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay*. Washington, DC: US Environmental Protection Agency, 31.

Wright, M. E., Solo-Gabriele, H. M., Elmir, S., and Fleming, L. E. (2009). Microbial load from animal feces at a recreational beach. *Mar. Pollut. Bull.* 58, 1649–1656. doi: 10.1016/j.marpolbul.2009.07.003

Yuknavage, K., Arriola, J., Benavente, D., Camacho, R., Chambers, D., Derrington, E., et al. (2018). *2018 Commonwealth of the Northern Mariana Islands 305(b) and 303(d), Water Quality Assessment Integrated Report*. Saipan: CNMI Bureau of Environmental and Coastal Quality.

Disclaimer: Neither the United States Government nor any of its employees, contractors, or their employees make any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use of apparatus, product, or process discussed in this document, or represents that its use by such party would not infringe on privately owned rights. Mention of any commercial entities, trade names, or commercial products in this document does not constitute any type of endorsement or recommendation for use.

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.

Copyright © 2021 Sinigalliano, Kim, Gidley, Yuknavage, Knee, Palacios, Bautista, Bonacolta, Lee and Maurin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 January 2021
doi: 10.3389/fmicb.2020.616692

Tracking Major Sources of Water Contamination Using Machine Learning

Jianyong Wu[1], Conghe Song[2], Eric A. Dubinsky[3] and Jill R. Stewart[1]*

[1] Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States, [2] Department of Geography, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States, [3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Current microbial source tracking techniques that rely on grab samples analyzed by individual endpoint assays are inadequate to explain microbial sources across space and time. Modeling and predicting host sources of microbial contamination could add a useful tool for watershed management. In this study, we tested and evaluated machine learning models to predict the major sources of microbial contamination in a watershed. We examined the relationship between microbial sources, land cover, weather, and hydrologic variables in a watershed in Northern California, United States. Six models, including K-nearest neighbors (KNN), Naïve Bayes, Support vector machine (SVM), simple neural network (NN), Random Forest, and XGBoost, were built to predict major microbial sources using land cover, weather and hydrologic variables. The results showed that these models successfully predicted microbial sources classified into two categories (human and non-human), with the average accuracy ranging from 69% (Naïve Bayes) to 88% (XGBoost). The area under curve (AUC) of the receiver operating characteristic (ROC) illustrated XGBoost had the best performance (average AUC = 0.88), followed by Random Forest (average AUC = 0.84), and KNN (average AUC = 0.74). The importance index obtained from Random Forest indicated that precipitation and temperature were the two most important factors to predict the dominant microbial source. These results suggest that machine learning models, particularly XGBoost, can predict the dominant sources of microbial contamination based on the relationship of microbial contaminants with daily weather and land cover, providing a powerful tool to understand microbial sources in water.

Keywords: machine learning, XGBoost, fecal contamination, microbial source tracking, land use, rainfall

OPEN ACCESS

Edited by:
Michèle Gourmelon, Institut Français de Recherche pour l'Exploitation de la Mer, France

Reviewed by:
Elisenda Balleste, University of Barcelona, Spain
Yi Xiong, Shanghai Jiao Tong University, China

****Correspondence:***
Jianyong Wu
Jianyong.wu@alumni.unc.edu

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *13 October 2020*
Accepted: *29 December 2020*
Published: *20 January 2021*

Citation:
Wu J, Song C, Dubinsky EA and Stewart JR (2021) Tracking Major Sources of Water Contamination Using Machine Learning. Front. Microbiol. 11:616692. doi: 10.3389/fmicb.2020.616692

INTRODUCTION

Understanding the sources of microbial contamination in drinking and recreational water is important for mitigating health risks of waterborne pathogens and protecting the public from waterborne diseases (Scott et al., 2002; Simpson et al., 2002; Harwood et al., 2014). Currently, multiple methods have been developed to track sources of microbial contamination, including matching phenotypic or genotypic characteristics of indicator bacteria, detecting host-specific

molecular markers, and identifying chemical indicators of wastewater (Scott et al., 2002; Simpson et al., 2002; Meays et al., 2004; Boehm et al., 2013; Harwood et al., 2014; Dubinsky et al., 2016). However, these laboratory-based methods have limitations including time and cost constraints and/or technique difficulty. More importantly, these methods only help to identify sources of contamination at the sites and time of sampling. The methods are limited in their ability to map microbial sources across space and time. Therefore, approaches to modeling and predicting sources of fecal contamination in unsampled locations and times are highly desirable.

To date, it is still difficult to determine optimal models and appropriate variables for predicting the sources of microbial contamination (Belanche-Muñoz and Blanch, 2008; Belanche and Blanch, 2011). This is due in part to the complexity of host-specific sources of microbial contaminants in water, which may originate from humans, birds, dogs, or other animals. It is also because factors affecting microbial sources are not fully understood. Several recent studies have revealed that land use/land cover (LULC) and weather significantly impact fecal contamination and its sources in water (Peed et al., 2011; Wu et al., 2011; Gentry-Shields et al., 2012; Haack et al., 2013; Jent et al., 2013; Liang et al., 2013; Staley et al., 2013; Verhougstraete et al., 2015; Wu, 2019). For example, in Jordan Lake, North Carolina, and land use components but not rainfall were found to associate with the concentrations of *M. smithii* (nifH) markers, an indicator of human-source contamination (Gentry-Shields et al., 2012). A study by Peed et al. (2011) reported that the abundance of human-associated genetic markers had a positive significant correlation with septic tank density following wet weather events. In urbanized coastal watersheds in Florida, microbial sources were strongly affected by the change in rainfall patterns (Shehane et al., 2005). Besides land cover and weather, microbial sources may also be affected by hydrological factors because hydrology strongly influences the transport and fate of fecal contaminants in water (Wilkes et al., 2009; Liao et al., 2015). For example, microorganisms can be transported from upstream to downstream, and between sediments and waterbodies (Wu et al., 2009). These examples and many other studies have substantiated that land use, weather, and hydrological factors play an important role in determining the sources of fecal contamination in water, and suggest that the relationship between these factors and microbial sources are complicated and non-linear.

Machine learning is a set of methods or algorithms to automatically find patterns and extract valuable information from data (Bishop, 1995; Hastie et al., 2009). Based on whether input data include a response variable (also called target variable, output variable, or label), machine learning algorithms are divided into two major categories: unsupervised learning and supervised learning (Bishop, 1995; Hastie et al., 2009). For unsupervised learning, a model makes inferences from datasets consisting of only features (predictive variables) but no labeled responses. K-means clustering, principal component analysis (PCA), and expectation–maximization algorithm (EM) are common unsupervised learning algorithms. In contrast, for supervised learning, a model is trained with input data that are composed of both features and a response variable. Common supervised learning algorithms include K-nearest neighbors (KNN), Naïve Bayes, support vector machine (SVM), neural network (NN), Random Forest, XGBoost, and others. These algorithms are often used for two purposes: classification (input data with discrete labels), and regression (input data with continuous labels). For all these algorithms, research is needed to inform selection and applicability of machine learning to help identify and remediate sources of surface water contamination.

The objectives of this study are (1) to examine the relationship of land cover, weather, and hydrologic variables with microbial sources, (2) to predict the major types of microbial sources based on these data inputs using machine learning, and (3) compare the performance of six machine learning algorithms in predicting the dominant source of microbial contamination. This study is the first to predict host-specific sources of fecal contamination based on land cover, weather, and hydrologic data using machine learning; the output of which provides useful information for making appropriate watershed management decisions.

MATERIALS AND METHODS

Microbial Source Tracking Data

Host-specific sources of fecal contamination in the Russian River watershed (**Figure 1**), Northern California, were investigated previously during 2011–2013 (Dubinsky et al., 2016). The Russian River is heavily used for recreational activities in the summer months including swimming, wading and paddling. The area has a diverse and mixed land use types, such as urbanized areas, open space, dairy farms and pastureland (Dubinsky et al., 2016). Agriculture is the major land use type near Sites 10, 21, and 40, where ruminants (cows and deer) and horses are often found. Forest land is the major land use type near Sites 24, 30, and 31, where wildlife, such as deer, raccoons, rodents, rabbits, coyotes, and birds are common.

During the original study (Dubinsky et al., 2016), 102 water samples were collected from 46 sites, in both dry and wet seasons (**Figure 1**). The sources of fecal bacteria in these water samples were tested using the PhyloChip microarray and classified into six major sources (human, bird, dog, horse, pig, and ruminant) using SourceTracker, a machine learning approach for classification. The existence of each host-specific source was indicated by the positive likelihood ratio from the SourceTracker test. Briefly, DNA was extracted from water samples and hybridized into the PhyloChip microarray overnight. After stained and scanned, the PhyloChip microarray provided raw data as fluorescent image files. The fluorescent image files which were considered as probe-quartet profiles, were analyzed using the probe quartet approach described previously (Cao et al., 2013). To assign each probe-quartet profile a specific source of microbial contamination, a machine learning method (SourceTracker) was used to classify the quartet data (Another machine learning approach, Random Forest, was also used. We chose the result from SourceTracker because of its better performance). The quartet profiles from 70 reference samples were used as the training data. The quartet profiles from the test data were classified into six different

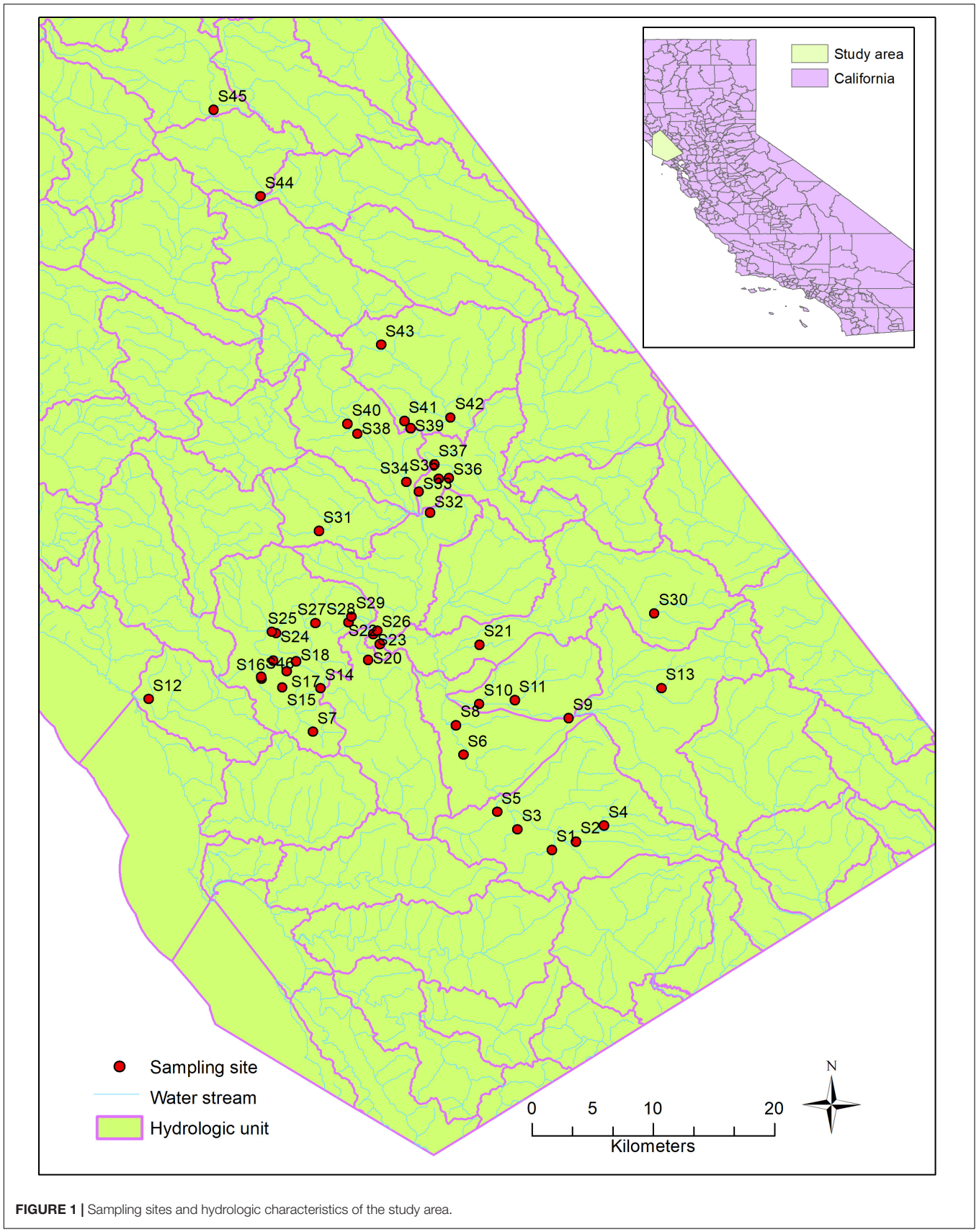


FIGURE 1 | Sampling sites and hydrologic characteristics of the study area.

source types: humans (stool, sewage, and septage), dogs and cats, pigs, ruminants (cows, elk, and deer), horses, and birds. In the classification, each source was given a probability that its microbial community DNA was found in the probe-quartet profile of each sample. The source with the highest probability was regarded as the major source. One limitation of the PhyloChip approach is that the method is not developed for source apportionment and quantification. Detailed information about microbial source tracking data and the relevant methods were described by Dubinsky et al. (2016).

For this study, the microbial sources previously described were reclassified into two categories: human vs. non-human sources. In this way, the modeling problem becomes a binary classification problem, which is simpler than a multi-class classification problem and an appropriate first step for applying machine learning to microbial source tracking.

Weather Data

Daily mean temperature and daily precipitation data during 2011–2013 were obtained from PRISM time-series datasets, which are available online[1]. Daily climate data provided by the PRIMS climate group have a spatial resolution of 4 km. The methods for generating the climate data were described previously (Daly et al., 2008). Based on the geographic coordinates of each water sampling site, a 4 km × 4 km grid from the climate data map was delineated, and the climate data of this grid was assigned to the sampling site. Four weather variables were created, including daily mean temperature on the sampling date ($Temp_0$), daily mean temperature on the day before sampling ($Temp_{-1}$), daily precipitation on the sampling date ($Prep_0$), and daily precipitation on the day before sampling ($Prep_{-1}$).

Hydrologic Data

The hydrologic information in the study area was obtained from National Hydrologic Dataset (NHD) plus version 2 (NHDplus V2) (Moore and Dewald, 2016). Developed by United States EPA and United States Geological Survey, NHDplus V2 consists of three major components: NHD, National Elevation Dataset (NED), and Watershed Boundary Dataset (WBD). The drainage network information including rivers, streams and ponds in the study area was provided by the NHD. From the WBD, a six-level hydrologic unit (the smallest classification unit of watershed) was obtained. Generally, a hydrologic unit was delineated to let the surface drainage within the unit converge at a single outlet. The NED was used to determine stream flow directions based on the assumption that water flows from high elevation to low elevation.

Land Cover Data and Processing

Land cover information was obtained from the National Land Cover Database (NLCD) 2011 (Homer et al., 2015). From the NLCD 2011, eight major types of land cover classes were identified in the study area, including water, developed land, barren land, forest, shrubland, grassland, agriculture, and wetland.

[1]http://prism.oregonstate.edu/

Since there are no established criteria to determine the scale of land cover that influences water quality in a watershed, two approaches, distance-based and hydrologic-based approaches, were used to calculate land cover components around a sampling point. For the distance-based approach, a circular 2 km buffer was created around each sampling site and the percentages of eight land cover classes in the buffer were calculated. We chose 2 km as the buffer distance because a pervious study showed that land cover in a 2 km buffer had strong associations with microbial water quality (Wu and Jackson, 2016). For the hydrologic-based approach, hydrologic characteristics of the watershed were taken into account. Specifically, the scale of the influential land cover of a sampling site was delineated based on the following criteria: (1) only the land cover in a single 12-digit hydrologic unit (sub-watershed level) was considered; (2) only land cover in the upstream contributing area of a sampling site was considered; and (3) when an upper stream and a downstream of a sampling site are very close and their contributing areas are difficult to distinguish, a midline was drawn to divide the land cover between the upper stream and the downstream into two parts: one is associated with the upper steams, and the other is associated with the downstream. Based on these criteria, a buffer of 2 km was drawn around the upstream in a single hydrologic unit where the sample site is located. Then, the percentages of land cover classes in the buffer were calculated. For a few sampling sites that are not on a stream network, land cover associated with these sites were calculated using the first approach, i.e., a circular 2 km buffer.

Exploratory Data Analysis

After these data were collected and processed, exploratory data analysis was conducted to examine the mean, standard deviation, and range of weather and land cover data as well as microbial source tracking data obtained for the all sampling sites. The relationships between microbial sources (the probabilities of specific microbial sources), weather, and land cover variables (the percentages of land use types) were examined by Spearman correlation analysis.

Machine Learning Models

Predictive Feature Selection

To predict the dominant source of microbial contamination, we selected two groups of predictive features to fit six supervised machine learning algorithms, respectively. Group 1 includes eight land cover variables measured by the distance-based approach (the percentages of water, developed, barren land, forest, shrubland, grassland, agriculture, and wetland), two weather variables (daily mean temperature and daily precipitation on the sampling date), and two hydrologic variables (elevation and flow accumulation). Because the flow accumulation variable has a large variability, we included it in the model after a square root transformation. The daily mean temperature and daily precipitation on the day before sampling were not included in the model because of high correlation with the measures taken on the day of sampling. Group 2 is similar as Group 1 except that land cover variables measured by the distance-based approach were replaced with those measured by the hydrologic-based approach.

Model Implementation

The dataset was randomly divided into two subsets in the ratios of 80 and 20% for training and testing the model, respectively. The training dataset was used to train the model and obtain the model parameters. The testing dataset was used to evaluate prediction performance. We used k-fold cross validation to tune the hyperparameters of the models for the training set. Specifically, the training set was randomly split into four subsets equally, three subsets were selected as the training data and the remaining one was used as the test data to calculate the model accuracy. This process was repeated four times, reserving a different subset for validation for each repetition. Then six models were used to predict the major source of microbial contamination (human source vs. non-human source), including KNN, Naïve Bayes, SVM, NN, Random Forest, and XGBoost. The hyper-parameters (e.g., the number of iterations) of each model were tuned using a random search approach. In the random search, the hyperparameters are randomly combined and used to find the optimal values for the model. The analyses were conducted with Python (v 3.7) programming language.

K-nearest neighbors: KNN is a simple non-parametric method for classification, which classifies a new data point in a category same as the majority of its k neighbor points (Altman, 1992). Here, we set k as five and used "KNeighborsClassifier"[2] from "scikit-learn" library for prediction.

Naïve Bayes: Naive Bayes is a probabilistic model based on Bayes' theorem. It calculates the conditional probability of each class that a data point belongs to given training features with a naïve assumption that the features are independent (Hastie et al., 2009). Here, we used "GaussianNB"[3] from the "scikit-learn" library for prediction and set the parameters as default.

Support vector machine: SVM is a method to classify data by find a hyperplane in a high dimensional feature space that has the maximum margin, namely, the largest distance to the nearest training data points. The train data points that are close or on the boundary of the hyperplane are called support vectors, which determine the classification (Hastie et al., 2009). In this study, we used "LinearSVC"[4] from the "scikit-learn" library for prediction.

Neural Network: NN is a model that attempts to mimic neurons to process information, often called artificial neural network (ANN). A typical NN comprises three layers, an input layer that receives information, a hidden layer that processes information, and an output layer that produces the model outcomes. Each layer is composed of some computing elements (called neurons or nodes), and neurons in different layers are connected by weights (Jain et al., 1996; Samarasinghe, 2016). During the learning process, the weights between neurons in different layers are adjusted to obtain optimal model outcomes. In this study, "MLPClassifier"[5] from the "scikit-learn" library were used for prediction.

Random forest: Random Forest is an ensemble method that combines the results from multiple decision trees to obtain a better performance (Breiman, 2001). Briefly, for the Random Forest algorithm, the dataset is sampled n data points multiple times with replacement. Each subset of samples is used to train each decision tree. For the classification, Random Forest will combine the results from multiple trees and determine the class with the highest number of votes from these trees. One advantage of this algorithm is that the method can rank the importance of features (predictors) (Genuer et al., 2010). Here, "RandomForestClassifier"[6] was used for prediction.

Extreme gradient boosting (XGBoost): XGBoost is another decision-tree based ensemble algorithm, which implements the gradient boosting method to find the best tree model (Chen and Guestrin, 2016). In traditional gradient boosting, each new tree specifically focuses on the error of the previous tree. XGBoost adds more regularization terms in the model to control model over-fitting, which makes the model have a better performance (Chen and Guestrin, 2016; Chen et al., 2020). In this study, "XGBClassifier" from "xgboost" library[7] was used for prediction.

Model Performance Evaluation

The performance of each model was evaluated by two metrics: accuracy and the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC). The accuracy was calculated using the total number of correctly classified samples divided by the number of total samples. The ROC curve illustrates the ability of a binary classifier system to separate two groups at various discrimination threshold settings. The curve is plotted with the true positive rate (sensitivity) against the false positive rate (1-specificity). The AUC value indicates degree or measure of separability. When the value is closer to 1, the performance is better.

RESULTS

Description of Microbial Source Tracking, Land Cover, and Weather Data

Microbial sources in the Russian Watershed were tracked with 102 water samples (Dubinsky et al., 2016) and indicated by categorical probability. When microbial sources were reclassified into two categories, the probability of human sources ranged from 0 to 0.81, and the probability of non-human (birds, ruminants, horses, pigs, and dogs) sources ranged from 0 to 0.21. When microbial sources were reclassified into three categories, the probability of human sources was the same as that in the two-way classification, while the probabilities of bird sources and animal sources ranged from 0 to 0.08 and from 0 to 0.20, respectively (**Table 1**).

Land cover components associated with the sampling sites were calculated by two approaches. The result based on circular 2 km buffer approach showed that forest, urban, and grassland

[2]https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

[3]https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

[4]https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

[5]https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[6]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[7]https://xgboost.readthedocs.io/en/latest/python/python_api.html

were the dominant land cover types, with the mean percentages of 45, 16, and 19%, respectively. Water, barren land, and wetland only accounted for small percentages, with the mean percentages less than 2%. Similar results were observed when land cover was tallied based on the upstream hydrologic unit. Forest, urban, and grassland were still the major land cover types, with the mean percentages of 39, 25, and 14%, respectively. Water, barren land, and wetland were no more than 2%. In terms of weather, the mean temperature on and before the sampling days ranged from 7.2 to 22.5°C and from 6.1 to 23.3°C, respectively. The daily precipitation on and before the sampling days ranged from 0 mm to 87.1 mm and from 0 to 62.7 mm, respectively. Among 102 samples, 61 samples were collected in wet weather and 41 samples were collected in dry weather (0 mm of precipitation) (**Table 1**).

TABLE 1 | Descriptive statistics of variables in this study.

Categories	Variables	*N*	Mean	Standard deviation	Minimum	Maximum
Microbial sources	Human	102	0.06	0.14	0.00	0.81
	Non-human	102	0.03	0.04	0.00	0.21
	Bird	102	0.01	0.01	0.00	0.08
	Animal	102	0.02	0.04	0.00	0.20
Land cover variables calculated based on hydrologic unit	Water	102	0.01	0.01	0.00	0.03
	Urban	102	0.16	0.17	0.00	0.85
	Barren land	102	0.00	0.01	0.00	0.06
	Forest	102	0.45	0.32	0.00	0.96
	Shrubland	102	0.11	0.09	0.00	0.40
	Grassland	102	0.19	0.20	0.01	0.86
	Agriculture	102	0.08	0.13	0.00	0.73
	Wetland	102	0.01	0.01	0.00	0.10
Land cover variables calculated based on 2 km buffer	Water	102	0.02	0.02	0.00	0.10
	Urban	102	0.25	0.20	0.01	0.98
	Barren land	102	0.00	0.01	0.00	0.04
	Forest	102	0.39	0.32	0.00	0.93
	Shrubland	102	0.07	0.07	0.00	0.38
	Grassland	102	0.14	0.13	0.01	0.54
	Agriculture	102	0.12	0.17	0.00	0.60
	Wetland	102	0.01	0.01	0.00	0.05
Weather variables	Temperature (t = 0), °C	102	13.33	4.29	7.17	22.50
	Temperature (t = −1), °C	102	13.48	4.70	6.11	23.39
	Precipitation (t = 0, mm)	102	17.02	25.24	0.00	87.12
	Precipitation (t = −1, mm)	102	13.23	19.53	0.00	62.74
Hydrologic variable	Elevation (m)	102	57	59	0	243
	Flow accumulation (m^2)	102	14245	48,592	0	2,48,045

TABLE 2 | Spearman correlation between microbial sources and weather and hydrologic variables.

		Human sources	Bird sources	Animal sources	Non-human sources
$Temp_0$	r	−0.119	0.135	−0.152	−0.051
	p	0.235	0.176	0.128	0.609
$Temp_{-1}$	r	−0.108	0.140	**−0.218**	−0.123
	p	0.280	0.161	**0.028**	0.216
$Precp_0$	r	0.128	−0.048	**0.372**	**0.304**
	p	0.198	0.632	**<0.001**	**0.002**
$Precp_{-1}$	r	−0.034	−0.047	**0.346**	**0.271**
	p	0.733	0.642	**<0.001**	**0.006**
Elevation	r	−0.137	0.056	−0.100	−0.061
	p	0.168	0.579	0.319	0.543
Flow accumulation	r	−0.214	−0.059	−0.064	−0.068
	p	0.0301	0.557	0.521	0.494

$Temp_0$, daily mean temperature in the sampling date; $Temp_{-1}$, daily mean temperature in the day before sampling, $Prep_0$, daily precipitation in the sampling date, $Prep_{-1}$, daily precipitation in the day before sampling. Bold fonts indicate the correlation is significant ($p < 0.05$).

TABLE 3 | Spearman correlation between land cover and microbial sources during wet weather and dry weather.

		Wet weather				Dry weather			
Land cover		**Human sources**	**Non-human sources**	**Animal source**	**Bird source**	**Human sources**	**Non-human sources**	**Animal source**	**Bird source**
Water[1]	r	0.163	−0.051	−0.015	−0.086	−0.029	0.186	0.366	−0.071
	p	0.210	0.694	0.910	0.508	0.856	0.245	0.019	0.661
Developed land[1]	r	0.217	0.087	0.103	**−0.254**	0.018	−0.090	−0.048	−0.132
	p	0.093	0.503	0.430	**0.048**	0.912	0.574	0.768	0.411
Barren land[1]	r	−0.117	−0.198	−0.208	−0.011	−0.070	0.092	0.136	0.053
	p	0.371	0.125	0.108	0.931	0.665	0.567	0.396	0.743
Forest[1]	r	−0.016	−0.121	−0.152	**0.280**	−0.123	0.096	0.196	−0.098
	p	0.902	0.351	0.243	**0.029**	0.445	0.549	0.220	0.542
Shrub land[1]	r	−0.108	**−0.370**	**−0.409**	−0.023	−0.098	0.033	−0.206	0.301
	p	0.408	**0.003**	**0.001**	0.863	0.543	0.838	0.195	0.056
Grassland[1]	r	0.109	0.047	0.028	−0.122	0.140	0.025	−0.129	0.248
	p	0.402	0.719	0.830	0.350	0.382	0.875	0.421	0.119
Agriculture[1]	r	−0.047	0.000	0.090	**−0.279**	0.186	−0.013	−0.243	0.246
	p	0.719	0.998	0.492	**0.030**	0.246	0.935	0.126	0.121
Wetland[1]	r	0.246	−0.093	0.002	**−0.273**	−0.076	0.300	0.312	0.168
	p	0.056	0.476	0.987	**0.033**	0.635	0.057	0.047	0.293
Water[2]	r	**0.290**	**−0.273**	**−0.314**	0.052	−0.075	0.067	0.088	−0.001
	p	**0.024**	**0.033**	**0.014**	0.688	0.640	0.675	0.584	0.995
Developed land[2]	r	**0.386**	0.047	0.079	−0.124	0.052	−0.228	−0.159	−0.161
	p	**0.002**	0.720	0.544	0.343	0.745	0.152	0.320	0.315
Barren land[2]	r	−0.080	**−0.385**	**−0.441**	0.051	−0.114	0.208	0.179	0.095
	p	0.539	**0.002**	**0.000**	0.694	0.477	0.191	0.263	0.555
Forest[2]	r	−0.058	−0.072	−0.106	**0.297**	−0.137	0.006	0.198	−0.245
	p	0.654	0.581	0.417	**0.020**	0.393	0.969	0.215	0.122
Shrub land[2]	r	−0.174	**−0.308**	**−0.383**	−0.139	−0.190	−0.031	−0.110	0.082
	p	0.181	**0.016**	**0.002**	0.287	0.235	0.847	0.495	0.612
Grassland[2]	r	−0.002	0.090	0.067	−0.223	0.078	0.074	−0.061	0.231
	p	0.989	0.490	0.607	0.084	0.628	0.647	0.705	0.147
Agriculture[2]	r	−0.085	**−0.303**	−0.220	**−0.333**	−0.014	−0.159	−0.382	0.197
	p	0.516	**0.018**	0.089	**0.009**	0.932	0.322	0.014	0.218
Wetland[2]	r	**0.279**	−0.200	−0.157	**−0.313**	−0.026	0.182	0.156	0.091
	p	**0.030**	0.123	0.226	**0.014**	0.870	0.255	0.331	0.570

[1] *Land cover components were calculated based on 2 km buffer.*
[2] *Land cover components were calculated based on hydrologic unit.*
Bold fonts indicate the correlation is significant ($p < 0.05$).

TABLE 4 | The performance of each machine learning algorithm for predicting major sources of fecal contamination.

Models	Predictor	Parameters	Accuracy	Precision	Recall	F1 score
KNN	Group 1	n_neighbors = 5	0.71	0.7	0.7	0.7
Naïve Bayes	Group 1	priors = None, var_smoothing = 1e-09	0.62	0.7	0.58	0.64
Support vector machine	Group 1	C = 1, max_iter = 2000, tol = 1e-4	0.67	0.5	0.71	0.59
Neural network	Group 1	solver = 'lbfgs', alpha = 1e-4, hidden_layer_sizes = (5, 3)	0.76	0.7	0.78	0.74
Random forest	Group 1	n_estimators = 20	0.81	0.7	0.88	0.78
XGBoost	Group 1	n_estimators = 30	0.86	0.9	0.82	0.86
KNN	Group 2	n_neighbors = 5	0.76	0.7	0.78	0.74
Naïve Bayes	Group 2	priors = None, var_smoothing = 1e-09	0.76	1.0	0.67	0.8
Support vector machine	Group 2	C = 1, max_iter = 2000, tol = 1e-4	0.71	0.5	0.83	0.63
Neural network	Group 2	solver = 'lbfgs', alpha = 1e-4, hidden_layer_sizes = (5, 2)	0.76	0.9	0.69	0.78
Random forest	Group 2	n_estimators = 20	0.81	0.7	0.88	0.78
XGBoost	Group 2	n_estimators = 10	0.90	0.9	0.9	0.9

Correlation Between Microbial Sources and Land Cover and Weather

The Spearman correlation analysis showed that precipitation on or 1 day before the sampling day had positive correlations with non-human sources ($p < 0.01$) but was not significantly correlated with human sources ($p > 0.05$). When microbial sources were classified into three categories, precipitation on or 1 day before the sampling day had positive correlations with animal sources, while temperature on the day before sampling had a negative correlation with animal sources. No significant correlations were found between these weather variables and bird and human sources (**Table 2**).

Weather also affected the correlations between microbial sources and land cover variables. No significant correlations were observed for the samples collected in dry weather. However, for samples collected during wet weather, correlations were found between microbial sources and land cover variables. For example, bird sources had a significant positive correlation with the percentage of forest, but negative correlations with the percentages of developed land, agriculture, and wetland when these land cover variables were calculated based on the circular 2 km-buffer approach. For land cover variables calculated based on hydrologic unit, more significant correlations between microbial sources and land cover variables were found. Bird sources still had a significant positive correlation with the percentage of forest but negative correlations with the percentages of agriculture and wetland. Human sources had significant positive correlations with the percentages of water area, developed land, and wetland, while animal sources had negative correlations with the percentages of water area, barren land, shrubland, and agriculture (**Table 3**).

Prediction of Microbial Sources Using Machine Learning

Six machine learning models were applied to predict whether the microbial source of a sample was from human or non-human sources. According to the average accuracy, XGBoost had the best performance and correctly predicted 88% of the samples. Random forest had the second best method with the accuracy of 81%. The other tested models included KNN (accuracy = 74%), Neural Network (accuracy = 76%), SVM (accuracy = 69%), and Naïve Bayes (accuracy = 69%). When the group of predictors were compared, the accuracies of the neural network and Random Forest models remained consistent for both groups. For the other models, the accuracies were much higher when the Group 2 predictors were used. The accuracy of XGBoost reached up to 90% with the Group 2 predictors (**Table 4**).

The performance of these models was further evaluated using the AUC of ROC, which showed similar results as measured for accuracy. XGBoost had the highest AUC value, followed by Random Forest, KNN, Neural Network, SVM, and Naïve Bayes. Similarly, the performance of the Group 2 predictors was much higher than that of the Group 1 predictors. The AUC of XGBoost using the Group 2 predictors was up to 92%, which was the highest among all models (**Figure 2**).

Importance Ranking for Predictors

According to the importance index calculated by the Random Forest model, precipitation and temperature are the two most important predictors, of which the importance index values were above 20%. Flow accumulation, elevation, the percentages of developed land, grassland, water, forest, and wetland were less important predictors, of which the importance index values were between 5% and 10%. The percentages of agriculture, shrubland and barren land were not important predictors, of which the importance index values were below 5% (**Figure 3**).

DISCUSSION

We examined the relationships of microbial sources with land cover and weather variables. Based on their relationships, we applied six machine learning models to predict microbial sources. Our results revealed that all models using this dataset performed well to predict microbial sources in two categories (human vs. animal) and XGBoost had the best performance. To our best knowledge, this is the first study to model and predict major microbial sources of fecal contamination in water based on land cover and weather data using machine learning. This

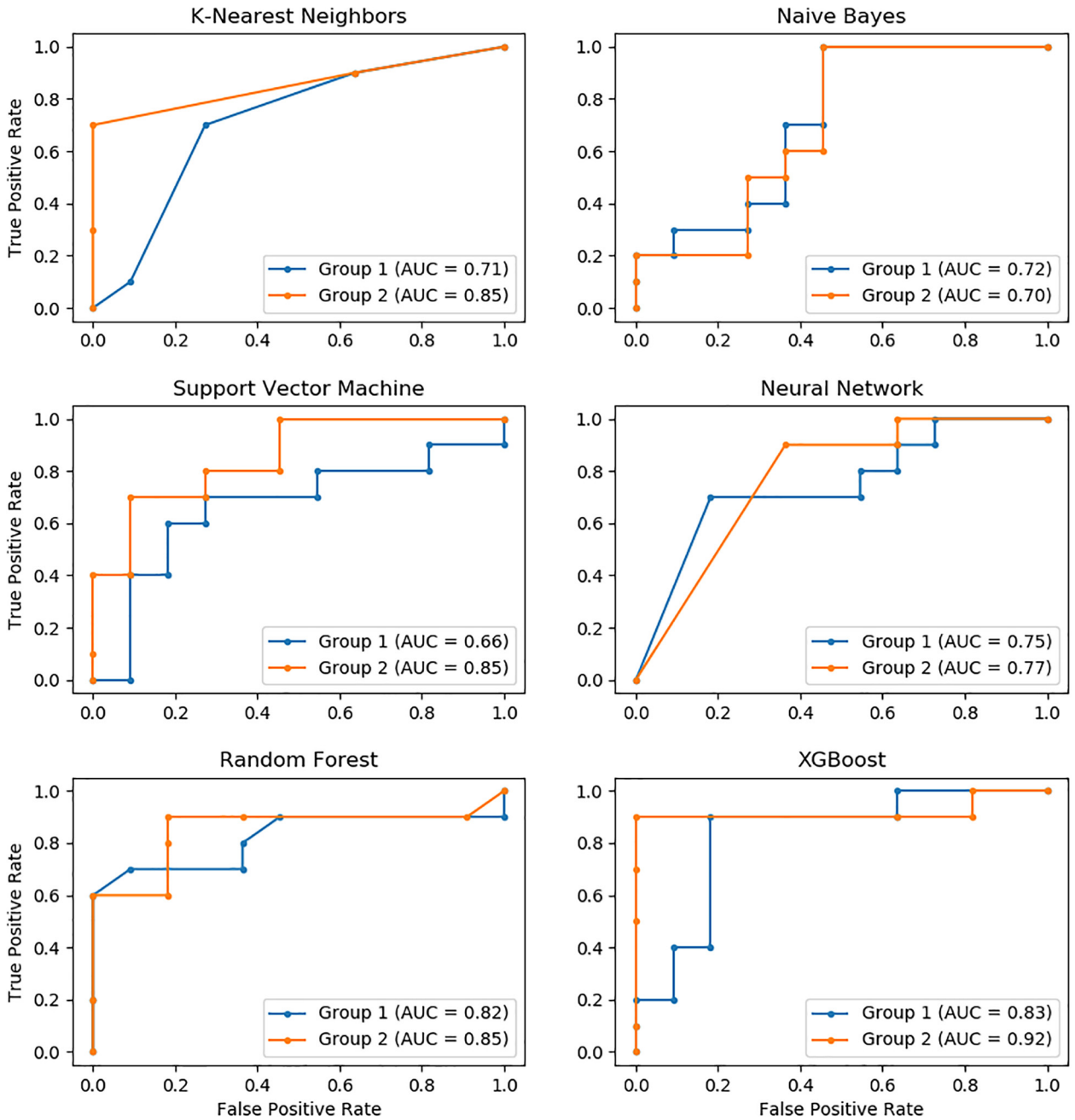


FIGURE 2 | The ROC curves of six models with Group 1 and Group 2 predictors.

modeling approach is a promising complement to current laboratory-based microbial source tracking methods because it can overcome some of their limitations. For example, the modeling approach enables estimation of microbial sources in space and time while fecal contamination in water is inevitably affected by weather, land cover, and other factors. As a result, modeling outputs can help mangers to better understand major microbial sources in water and make appropriate decisions to protect public health.

Land cover is regarded as a key factor to determine the sources of fecal contaminants because land cover is directly related to the activities and habitats of hosts of fecal contaminants (Kerr and Ostrovsky, 2003). For example, developed land, including residential, and commercial areas, are places mainly for human activities while other land covers tend to comprise habitats for birds, livestock, and wildlife. In this study, birds are often the major source in the area where forest is the primary land cover (e.g., Site 18), while in the area where grassland and agriculture

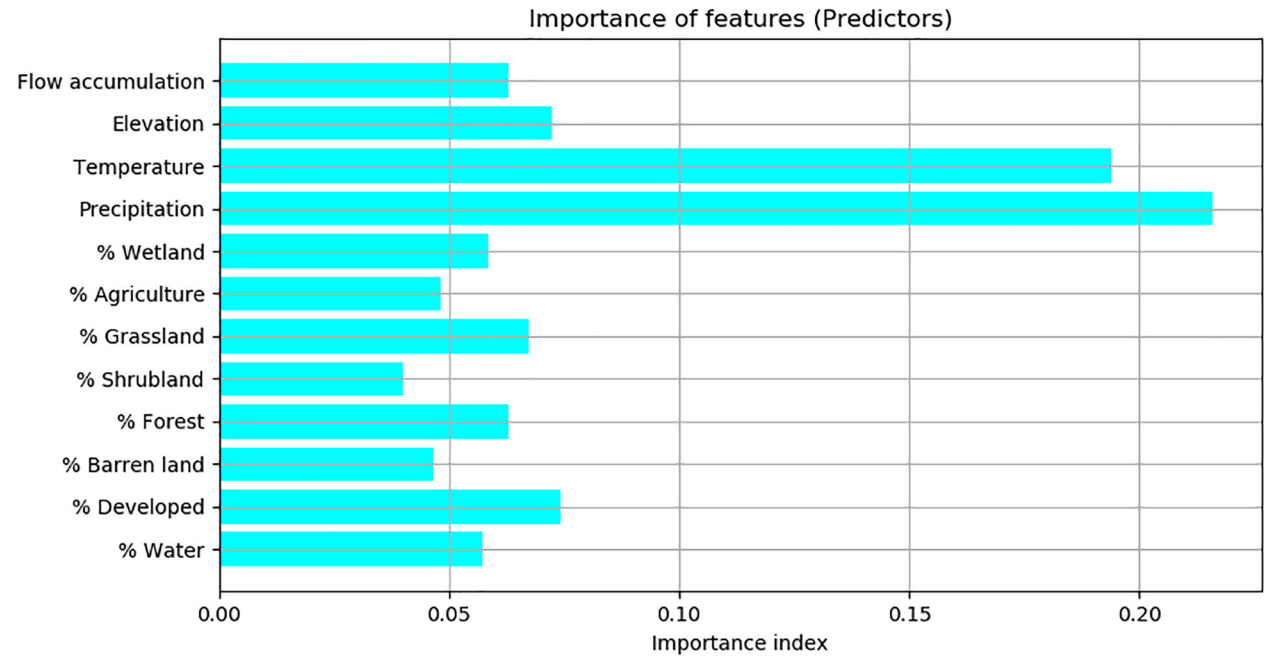


FIGURE 3 | The importance of predictors based on the importance index calculated by Random Forest.

dominate, the source of ruminant animals might have a higher probability (e.g., Site 3). Our Spearman correlation analysis supported these relationships as developed land had a significant positive correlation with human sources but a significant negative correlation with bird sources. The relationship between land cover and the sources of fecal contaminants is further influenced by wet weather because precipitation can facilitate the transport fecal contaminants to water (Wilkes et al., 2009; Wu et al., 2009, 2011). In wet weather, runoff from different types of land cover and overflow from sewer systems increase fecal contamination of water (McLellan et al., 2007; Wu et al., 2011), while in dry weather, microbes from land are transported to water to a lesser extent. In addition, solar radiation may inactivate fecal bacteria, thus reducing fecal contamination of water (Boehm et al., 2002). As a result, land cover is likely to have a closer relationship with microbial sources in wet weather than in dry weather. The result of Spearman correlation analysis corroborated this conclusion as no correlations between land cover and microbial sources were found in dry weather but some correlations (e.g., positive correlation between bird sources and forest land, negative correlation between animal sources and shrubland) were found in wet weather. We also found that animal sources had a significant and positive correlation with daily precipitation, which is consistent with results from other studies that have reported more frequent detection of fecal contaminants from animal sources during wet weather (Wu et al., 2009). Our model also showed that precipitation and temperature were two important predictors, suggesting that they have strong impact on the sources of microbial contaminants.

We also took hydrologic features of sampling sites into account in delineating influential land cover to predict microbial sources because hydrologic features might significantly influence the results of microbial source tracking (Reischer et al., 2008). We hypothesized that land cover in the catchment area upstream of a sampling site may influence microbial sources at that site because when fecal contaminants are flushed from land to water, they can be transported from upstream to downstream (Wu et al., 2009). Therefore, the components of influential land cover were calculated not only based on a circular 2 km buffer, but also based on upstream catchment area within a hydrologic unit. Our results clearly showed that the model performance was improved greatly when hydrologic features were taken into account. For example, the accuracies of KNN, Naïve Bayes, and SVM changed from 71, 62%, and 67 to 86%, 76 and 71%, respectively.

Our study showed XGBoost performed best when predicting microbial sources in two categories. The result is not surprising as other studies have shown that XGBoost has advantages over other models (Pan, 2018; Wang et al., 2020). Similar to Random Forest, it is an ensemble method that makes inferences based on multiple decision trees, thus reducing prediction errors. In addition, XGBoost can avoid model overfitting by adding additional regularization when the sample size is small, which is the case of this study. Among six models selected, Naïve Bayes had a relative low performance. This may be because that our predictors are not independent of each other, which does not meet the assumption of the model. Though Naïve Bayes did not perform well this study, it has worked well in spam filter and text classification (McCallum and Nigam, 1998; Metsis et al., 2006). As shown in **Table 5**, each model has its advantages and disadvantages. As a result, the selection of appropriate models depends on the specific problem to solve.

TABLE 5 | Comparison of the strength and weakness of different machine learning algorithms.

Algorithms	Strength	Weakness
KNN	It is intuitive and simple;	It is slow and not good for large dataset;
	It has only one hyper parameter;	It does not work well with high dimensional dataset;
	No training period is needed.	It is sensitive to noisy data, missing values and outliers.
Naive Bayes	It is simple and easy to implement;	It has the assumption of independent predictors, and the performance will be poor if the assumption is not met.
	It only requires only small amount of training data;	
	It works well with high dimensional dataset such as text classification.	
SVM	The training is relatively easy;	It is not suitable for a large dataset;
	It is effective in high dimensional data;	It requires feature scaling;
	It has no local optimal.	It difficult to choose appropriate kernel function.
Neural network	It has fault tolerance;	It is hard to interpret the model;
	It can model complicated relationship;	It is prone to overfitting;
		It is expensive to compute;
	It does not need any prior knowledge about the data.	It tends to end up in local minima.
Random forest	It does not require feature engineering;	It is hard to interpret the model.
	It does not need the assumptions on the distribution of the data;	
	It can handle collinearity;	
	It can rank variable importance.	
XGBoost	It has a high performance and accuracy as compared to other algorithms;	It needs to tune multiple parameters to get the optimal model.
	It does not require feature engineering;	
	It can rank variable importance.	

In this study, we did not predict microbial sources in more than three categories but we expect the performance will be reduced when multiple variables are predicted simultaneously. Several approaches might improve the prediction performance, for example, by increasing training data. Another limitation of this study is that the sample size is relatively small. The model trained on small data may not be generalized well. The performance of these models may vary when the models are applied to other environmental settings. Though the study has these limitations, it is expected that these models have promise to become a powerful tool to understand host-specific sources of fecal contamination in water.

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

JW conceived this study, conducted the data analysis, and drafted the manuscript. CS, ED, and JS advised and participated in discussion of the study and revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the Odum Institute at UNC-Chapel Hill for statistical consulting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.616692/full#supplementary-material

REFERENCES

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Statist.* 46, 175–185. doi: 10.2307/2685209

Belanche, L. A., and Blanch, A. R. (2011). "Statistical approaches for modeling in microbial source tracking," in *Microbial Source Tracking: Methods, Applications, and Case Studies*, eds V. J. Harwood, C. Hagedorn, and A. R. Blanch (Berlin: Springer), 207–227. doi: 10.1007/978-1-4419-9386-1_9

Belanche-Muñoz, L., and Blanch, A. R. (2008). Machine learning methods for microbial source tracking. *Environ. Model. Softw.* 23, 741–750. doi: 10.1016/j.envsoft.2007.09.013

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford university press.

Boehm, A. B., Grant, S., Kim, J., Mowbray, S., Mcgee, C., Clark, C., et al. (2002). Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environ. Sci. Technol.* 36, 3885–3892. doi: 10.1021/es020524u

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

Breiman, L. (2001). Random forests. *Machine Learn.* 45, 5–32.

Cao, Y., Van De Werfhorst, L. C., Dubinsky, E. A., Badgley, B. D., Sadowsky, M. J., Andersen, G. L., et al. (2013). Evaluation of molecular community analysis methods for discerning fecal sources and human waste. *Water Res.* 47, 6862–6872. doi: 10.1016/j.watres.2013.02.061

Chen, T., and Guestrin, C. (2016). "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, (California: SIGKDD), 785–794.

Chen, T., Wang, X., Chu, Y., Wang, Y., Jiang, M., Wei, D.-Q., et al. (2020). T4SE-XGB: interpretable sequence-based prediction of type IV secreted effectors using eXtreme gradient boosting algorithm. *Front. Microbiol.* 11:158253. doi: 10.1101/2020.06.18.158253

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* 28, 2031–2064. doi: 10.1002/joc.1688

Dubinsky, E. A., Butkus, S. R., and Andersen, G. L. (2016). Microbial source tracking in impaired watersheds using PhyloChip and machine-learning classification. *Water Res.* 105, 56–64. doi: 10.1016/j.watres.2016.08.035

Gentry-Shields, J., Rowny, J. G., and Stewart, J. R. (2012). HuBac and nifH source tracking markers display a relationship to land use but not rainfall. *Water Res.* 46, 6163–6174. doi: 10.1016/j.watres.2012.09.016

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognit. Lett.* 31, 2225–2236. doi: 10.1016/j.patrec.2010.03.014

Haack, S. K., Fogarty, L. R., Stelzer, E. A., Fuller, L. M., Brennan, A. K., Isaacs, N. M., et al. (2013). Geographic setting influences Great Lakes beach microbiological water quality. *Environ. Sci. Technol.* 47, 12054–12063. doi: 10.1021/es402299a

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

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Berlin: Springer Science & Business Media.

Homer, C., Dewitz, J., Yang, L. M., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information. *Photogramm. Engineer. Remote Sensing* 81, 345–354.

Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer* 29, 31–44. doi: 10.1109/2.485891

Jent, J. R., Ryu, H., Toledo-Hernandez, C., Domingo, J. W. S., and Yeghiazarian, L. (2013). Determining Hot Spots of Fecal Contamination in a Tropical Watershed by Combining Land-Use Information and Meteorological Data with Source-Specific Assays. *Environ. Sci. Technol.* 47, 5794–5802. doi: 10.1021/es304066z

Kerr, J. T., and Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299–305. doi: 10.1016/s0169-5347(03)00071-5

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

Liao, H., Krometis, L.-A. H., Hession, W. C., Benitez, R., Sawyer, R., Schaberg, E., et al. (2015). Storm loads of culturable and molecular fecal indicators in an inland urban stream. *Sci. Total Environ.* 530, 347–356. doi: 10.1016/j.scitotenv.2015.05.098

McCallum, A., and Nigam, K. (1998). "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, (New Jersey: Citeseer), 41–48.

McLellan, S. L., Hollis, E. J., Depas, M. M., Van Dyke, M., Harris, J., and Scopel, C. O. (2007). Distribution and fate of *Escherichia coli* in Lake Michigan following contamination with urban stormwater and combined sewer overflows. *J. Great Lakes Res.* 33, 566–580. doi: 10.3394/0380-1330(2007)33[566:dafoec]2.0.co;2

Meays, C. L., Broersma, K., Nordin, R., and Mazumder, A. (2004). Source tracking fecal bacteria in water: a critical review of current methods. *J. Environ. Manage.* 73, 71–79. doi: 10.1016/j.jenvman.2004.06.001

Metsis, V., Androutsopoulos, I., and Paliouras, G. (2006). "Spam filtering with naive bayes-which naive bayes?," in *Conference: CEAS 2006 - The Third Conference on Email and Anti-Spam*, (California: CEAS), 28–69.

Moore, R. B., and Dewald, T. G. (2016). The Road to Nhdplus - Advancements in Digital Stream Networks and Associated Catchments. *J. Am. Water Resour. Associat.* 52, 890–900. doi: 10.1111/1752-1688.12389

Pan, B. (2018). Application of XGBoost algorithm in hourly PM2. 5 concentration prediction. *IOP Confer. Ser. Earth Environ. Sci.* 113:012127. doi: 10.1088/1755-1315/113/1/012127

Peed, L. A., Nietch, C. T., Kelty, C. A., Meckes, M., Mooney, T., Sivaganesan, M., et al. (2011). Combining Land Use Information and Small Stream Sampling with PCR-Based Methods for Better Characterization of Diffuse Sources of Human Fecal Pollution. *Environ. Sci. Technol.* 45, 5652–5659. doi: 10.1021/es2003167

Reischer, G. H., Haider, J. M., Sommer, R., Stadler, H., Keiblinger, K. M., Hornek, R., et al. (2008). Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. *Environ. Microbiol.* 10, 2598–2608. doi: 10.1111/j.1462-2920.2008.01682.x

Samarasinghe, S. (2016). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*, Florida: CRC Press.

Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R., and Lukasik, J. (2002). Microbial source tracking: Current methodology and future directions. *Appl. Environ. Microbiol.* 68, 5796–5803. doi: 10.1128/aem.68.12.5796-5803.2002

Shehane, S. D., Harwood, V. J., Whitlock, J. E., and Rose, J. B. (2005). The influence of rainfall on the incidence of microbial faecal indicators and the dominant sources of faecal pollution in a Florida river. *J. Appl. Microbiol.* 98, 1127–1136. doi: 10.1111/j.1365-2672.2005.02554.x

Simpson, J. M., Santo Domingo, J. W., and Reasoner, D. J. (2002). Microbial source tracking: State of the science. *Environ. Sci. Technol.* 36, 5279–5288.

Staley, Z. R., Chase, E., Mitraki, C., Crisman, T. L., and Harwood, V. J. (2013). Microbial water quality in freshwater lakes with different land use. *J. Appl. Microbiol.* 115, 1240–1250. doi: 10.1111/jam.12312

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.* 112, 10419–10424. doi: 10.1073/pnas.1415836112

Wang, L., Wang, X., Chen, A., Jin, X., and Che, H. (2020). Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model. *Healthcare* 8:247.

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.

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.

Wu, J. Y., and Jackson, L. (2016). Association of land use and its change with beach closure in the United States, 2004-2013. *Sci. Total Environ.* 571, 67–76.

Wu, J. Y., Rees, P., Storrer, S., Alderisio, K., and Dorner, S. (2009). Fate and Transport Modeling of Potential Pathogens: The Contribution from Sediments. *J. Am. Water Resour. Associat.* 45, 35–44.

Wu, J., Rees, P., and Dorner, S. (2011). Variability of E. coli density and sources in an urban watershed. *J. Water Health* 9, 94–106.

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.

Copyright © 2021 Wu, Song, Dubinsky and Stewart. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fmicb.2021.647602

Human Fecal Pollution Monitoring and Microbial Risk Assessment for Water Reuse Potential in a Coastal Industrial–Residential Mixed-Use Watershed

Akechai Kongprajug[1†], Thammanitchpol Denpetkul[2†], Natcha Chyerochana[1], Skorn Mongkolsuk[1,3] and Kwanrawee Sirikanchana[1,3]*

[1] Research Laboratory of Biotechnology, Chulabhorn Research Institute, Bangkok, Thailand, [2] Department of Social and Environmental Medicine, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, [3] Center of Excellence on Environmental Health and Toxicology (EHT), Ministry of Education, Bangkok, Thailand

OPEN ACCESS

Edited by:
Georg H. Reischer,
Vienna University of Technology,
Austria

Reviewed by:
Anyi Hu,
Institute of Urban Environment (CAS),
China
Lucy Semerjian,
University of Sharjah, United Arab
Emirates

***Correspondence:**
Kwanrawee Sirikanchana
kwanrawee@cri.or.th

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 30 December 2020
Accepted: 15 March 2021
Published: 20 April 2021

Citation:
Kongprajug A, Denpetkul T, Chyerochana N, Mongkolsuk S and Sirikanchana K (2021) Human Fecal Pollution Monitoring and Microbial Risk Assessment for Water Reuse Potential in a Coastal Industrial–Residential Mixed-Use Watershed. Front. Microbiol. 12:647602. doi: 10.3389/fmicb.2021.647602

Rapid economic development has caused industrial expansion into residential communities, leading to higher fecal pollution loads that could be discharged into aquatic environments. However, little is known regarding the potential microbial impact on human health. This study investigated microbial contamination from coastal industrial–residential community areas in nine sampling sites in waterways during three dry events. A general microbial source tracking (MST) marker, GenBac3, was detected in all samples from all three events, indicating continuing fecal pollution in the area, mostly from human sewage contamination. This was shown by the human-specific genetic marker crAssphage (88.9%) and human polyomavirus (HPyVs; 92.6%) detection. Enteric human adenovirus (HAdV40/41) showed three positive results only from residential sites in the first event. No spatial difference was observed for MST markers and traditional fecal indicators (total coliforms and *Escherichia coli*) in each event. Still, a significantly lower abundance of GenBac3, HPyVs, and total coliforms in the first sampling event was detected. Spearman's rho analysis indicated a strong correlation among certain pairs of microbial parameters. Multivariate analysis revealed two clusters of samples separated by land use type (industrial vs. residential). According to factor analysis of mixed data, the land use parameter was more associated with physicochemical parameters (i.e., salinity, conductivity, water temperature, and dissolved oxygen). A Quantitative Microbial Risk Assessment (QMRA) was then conducted to estimate the annual infection risks of HAdV40/41 for non-potable water reuse purposes using predicted concentrations from crAssphage and HPyVs. The highest risks (95th percentiles) were ranked by food crop irrigation, aquaculture, and toilet flushing, at 10^{-1}, 10^{-2}, and 10^{-3} per person per year (pppy). Required treatment levels to achieve a 10^{-4} pppy annual infection risk were estimated. QMRA-based water treatment scenarios were suggested, including chlorination for toilet flushing reuse and

depth filtration prior to chlorination for aquaculture and food crop irrigation. Microbial monitoring combined with a QMRA could provide better insights into fecal pollution patterns and the associated risks, facilitating effective water quality management and appropriate prior treatments for water reuse.

Keywords: fecal indicator, microbial source tracking, industrial estate, land use, non-potable water reuse, freshwater, QMRA

INTRODUCTION

With rapid industrialization, the expansion of industries into peri-urban or rural communities has emerged in many geographical areas (Panyathanakun et al., 2013; Tian et al., 2017; Khan et al., 2020). The establishment of industries and industrial estates brings non-local labor, which promotes economic development in the communities. However, social, economic, and environmental components of the preceding residential communities could be transformed (Seemuang, 2018). The mismanagement of industrial pollution has also caused adverse health effects related to chemical exposure (García-Pérez et al., 2020; Khan et al., 2020; Lin et al., 2020). Efforts to reduce environmental and health impacts have been undertaken by promoting the mixed-use development concept, with careful consideration of safety to residents and environments (Altes and Tambach, 2008; Zagow, 2020). As summarized by the United Nations Industrial Development Organization (UNIDO), an eco-industrial park model is designed to consider the environmental dimension by encouraging the efficient use of resources, waste reduction and reuse, and chemical management (UNIDO, 2019). Many countries have adopted the industrial ecology concept (Piadeh et al., 2014; Industrial Estate Authority of Thailand, 2019; Susur et al., 2019; Hong and Gasparatos, 2020; Shah et al., 2020), and it was found to have successfully raised trust and improved relationship with adjacent communities (Yamsrual et al., 2019).

Although direct industrial pollutants comprise chemicals and hazardous materials, expanded populations in industrialized communities can also lead to higher fecal pollution loads that pose microbial risks to public health (Selvarajan et al., 2018). Higher total coliform counts have been reported in a river near an industrial area than in a river with no industry in both dry and wet weathers (Medeiros et al., 2017). Moreover, water quality standards for treated effluents from industries and industrial estates do not regulate bacterial parameters as microbial contamination indicators. However, some of the wastewater is from production, and some is from non-production activities, such as water use from workers (Ministry of Natural Resources and Environment, 2016; Ministry of Industry, 2017). Wastewater from toilets containing urine and fecal materials could carry waterborne pathogenic microorganisms, including protozoa, bacteria, and viruses that cause diseases in humans (Naidoo and Olaniran, 2014; Garcia-Aljaro et al., 2018). Inadequate treatment of wastewater could increase microbial contamination risks in receiving water bodies and limit their beneficial uses, such as aquaculture, recreation, and irrigation (Naidoo and Olaniran, 2014; Pandey et al., 2014). An explicit gap in the literature remains for investigations of microbial fecal pollution's contribution and impact in receiving waters in industrialized areas.

Fecal indicator bacteria (FIB) are gastrointestinal bacteria from warm-blooded animals that have long been used as traditional fecal indicators to represent the health risks associated with pathogenic microorganisms. FIB, such as total coliforms, fecal coliforms, *Escherichia coli*, and enterococci, have been regulated in surface and coastal water quality standards worldwide (Pollution Control Department (Pcd), 1994; European Union, 2006; US EPA, 2012). Recent advancements in fecal indication incorporate microbial source tracking (MST) indicators. MST markers are also gastrointestinal microorganisms, but their presence in only specific hosts (e.g., humans, horses, pigs, and cattle) is beneficial for pollution source discrimination (Sirikanchana et al., 2014; Wangkahad et al., 2015; Ahmed et al., 2016; Garcia-Aljaro et al., 2018; Holcomb et al., 2020).

Urban and industrial development causes higher water supply demand, which can initiate conflicts among beneficiaries. Therefore, water reuse is an effective solution to resolving water scarcity for industries and residential communities (Piadeh et al., 2014). Indirect wastewater reuse, a practice referring to the utilization of surface water contaminated by either treated or untreated wastewater, has been reported for agricultural irrigation (Jeong et al., 2016; Chacón et al., 2020; Jampani et al., 2020) and recreational activities (Crank et al., 2019). However, the protection of public health is an essential aspect when evaluating water reuse and reclamation. According to the Quantitative Microbial Risk Assessment (QMRA) frameworks, enteric waterborne pathogens' risks can be assessed. Public health risks associated with exposure to fecal microorganisms are calculated through hazard identification, exposure assessment, health effect evaluation, and risk characterization (World Health Organization, 2016). With a low abundance of waterborne pathogens, human sewage-specific MST markers have been used for QMRA analysis when the ratios of MST markers and the referenced pathogens are available (Zhang et al., 2019).

This study investigated water quality from receiving waters in industrial and residential zones to evaluate microbial impacts to public health and the water reuse potential for indirect wastewater reuse with prior treatment. The specific objectives were to (1) characterize the temporal and spatial abundance of microbial parameters (i.e., bacterial and viral fecal indicators and a pathogenic virus) in receiving waters in an industrial–residential coastal zone during base flow conditions, (2) investigate the correlations of microbial parameters with

physicochemical water quality parameters and land use type (i.e., industrial and residential), and (3) assess microbial risks, estimate required treatment levels, and evaluate QMRA-based water treatment scenarios for indirect wastewater reuse. The availability of this information will assist in the prioritization of water quality and risk management strategies that can be applied to industrial communities.

MATERIALS AND METHODS

Study Area and Sampling Locations

Sampling sites included nine locations (MP1–MP4 and MP6–MP10; MP5 was neglected after the pre-survey due to a dry canal) in four canals in the Map Ta Phut Sub-district, Rayong Province (**Figure 1** and **Supplementary Table 1**). The Chak Mak Canal (MP4 and MP6) connects to the coast through a private land area. The Lot Canal (MP7 and MP9) combines with the Nam Cha Canal (MP8, MP10, and MP3) before joining with the Nam Hu Canal (MP2) and exits to the coast (MP1). MP1–MP3, MP6, MP9, and MP10 are situated in a residential land use type, while MP4, MP7, and MP8 are in an industrial zone (The Eastern Economic Corridor Policy Committee, 2019). Established in 1989, the Map Ta Phut industrial estate currently owns 21.997 km^2 (Industrial Estate Authority of Thailand, 2020). This industrial park hosts seaports for transporting natural gas and goods and 57 industries comprising petrochemical, chemical and fertilizer, steel, oil, and power plants (Office of Map Ta Phut Industrial Estate, 2020). The Map Ta Phut Industrial Estate has been certified by the Industrial Estate Authority of Thailand as an eco-industrial town by matching the criteria of physical, economic, environmental, social, and managerial aspects (Industrial Estate Authority of Thailand, 2019). The industrial park coexists with residential areas and public facilities, including meeting halls, a library, and more (Industrial Estate Authority of Thailand, 2018, 2019). The industrial estate's wastewater treatment plant has a capacity of 4,000 m^3/day (Office of Map Ta Phut Industrial Estate, 2020). The adjacent coastal water in Pra Du Bay is a resource for fisheries and shellfish farming for residents (Seemuang, 2018).

Sample Collection

Three sampling campaigns from the nine sampling sites were conducted in December 2019 and February and March 2020, all of which were done in the dry season to mainly represent point source pollution. As retrieved from the Hydro and Agro Informatics Institute, the precipitation records from the nearest rainwater monitoring station reported no precipitation in Events 1 and 2 and very light rain (1.5 mm) within 5 days before sampling in Event 3. The samples were collected during the low tide to minimize a tidal effect. On-site measurements of pH and water temperature were made with the YSI 60 (YSI Inc., United States), and the YSI Pro2030 instrument (YSI Inc., United States) was employed to quantify the dissolved oxygen (DO) conductivity, and salinity. Two-liter samples were collected at 30 cm below the surface for microbiological analysis (i.e., total coliforms, *E. coli*, and DNA analyses of a pathogenic virus and MST markers), according to Thailand standard (National Environment Board, 1994). Another liter of samples was grabbed at mid-depth for physicochemical parameters: total suspended solids (TSS) and biochemical oxygen demand (BOD) (National Environment Board, 1994). The samples were kept in sterile plastic bottles and placed on ice during transportation to the laboratory within 6 h. As a quality control, three field blanks and three field duplicates, each collected at each sampling campaign, were utilized.



FIGURE 1 | Map of sampling sites.

Physicochemical and Microbiological Water Quality Parameters

TSS were analyzed with the drying method at 103–105°C (American Public Health Association et al., 2017a), and BOD was measured using the 5-day BOD test (American Public Health Association et al., 2017b). Total coliforms and *E. coli* were simultaneously detected using a membrane filtration method with 4-Methylumbelliferyl-%-D-Galactopyranoside-Indoxyl-%-D-Glucuronide (MI medium) (US EPA, 2002).

Water Filtration and DNA Extraction

One liter of samples (n = 27) was stored at 5°C for up to 3 days before a pre-acidification–filtration method, as previously described (Kongprajug et al., 2019a). Briefly, the sample pH was adjusted to 3.5 ± 0.2 using 2N hydrochloric acid before half a liter of the sample was separately filtered using a 0.45-μm-pore-size HAWP membrane (Merck Millipore, Germany). Filters were subsequently extracted with a ZymoBIOMICS DNA Microprep Kit (Zymo Research, United States), and extracts from the same samples were combined. The DNA concentration was measured with a NanoDrop 2000 Spectrophotometer (Thermo Scientific, United States). The DNA extracts were kept at −80°C until use. Three method blanks were analyzed by processing sterile laboratory water through all the steps as quality control measures.

Quantitative PCR (qPCR) Assays

A general fecal marker, GenBac3 (Siefring et al., 2008), two human-specific fecal markers, human polyomavirus BK and JC (HPyVs) (McQuaig et al., 2009) and crAssphage (Stachler et al., 2017), and adenovirus types 40/41 (HAdV40/41) (Ko et al., 2005) were used in this study (**Supplementary Table 2**). The abovementioned MST markers were selected because they have been investigated and validated in Thailand (Kongprajug et al., 2019a, 2020; Sangkaew et al., 2021), and HAdV40/41 was chosen because it has been identified as the second most prevalent following the subgroup C from child patients with acute gastroenteritis in Thailand (Kumthip et al., 2019). A 20-μl qPCR reaction contained 0.8 μl of each 10 μM forward and reverse primers, 0.4 μl of 10 μM probe, 2 μl of extracted DNA, 6 μl of 1 μg/μl BSA, and 10 μl of the 2X iTaq Universal Probes Supermix (Bio-Rad, United States). The qPCR cycling conditions comprised an initial denaturation at 95°C for 3 min followed by 40 cycles of a denaturation step at 95°C for 20 s and a combined annealing and elongation step for 1 min at 55°C for HPyVs and HAdV40/41 and at 60°C for GenBac3 and crAssphage. The qPCR reactions were performed with the QuantStudio™ 3 Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, United States), and the results were examined using the QuantStudio™ Design and Analysis Software with an automatic baseline and manual adjustment of the threshold values for GenBac3 (0.150), crAssphage (0.036), HPyVs (0.025), and HAdV40/41 (0.020). Environmental samples were run in duplicate, and the averaged C_q was used to calculate gene copy numbers when the standard deviation of C_q was less than 0.5; otherwise, an additional run was undertaken. The qPCR protocol was conducted according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2009). For each instrumental run containing environmental samples, the DNA standard, at a concentration of 5×10^4–5×10^5 copies/reaction, was run in triplicate as a calibration control according to a mixed model (Sivaganesan et al., 2010; Kongprajug et al., 2020). No-template controls (NTCs) in triplicate were also included in every qPCR instrumental run.

Standard Curves, Limits, and Inhibition Analysis

Standard curves were constructed using synthetic plasmid standards (Invitrogen, Thermo Fisher Scientific, United States) for crAssphage (Kongprajug et al., 2019b), and synthetic linear DNA fragments were used for GenBac3 (string 1) (Kongprajug et al., 2019a) and for HPyVs and HAdV40/41 (string 2) (**Supplementary Figure 1**). The standard curves were obtained from four replicates of individual instrumental runs according to the mixed model method, each with a triplicate of six 10-fold concentrations, ranging from 5×10^1 to 5×10^6 copies/reaction. The assay limit of detection (ALOD) was the lowest concentration in copies/reactions that showed positive detection in all 10 standard replicates. The assay limit of quantification (ALOQ) was considered to be the lowest concentration in copies/reactions of the target gene that could be correctly quantified—in this case, the lowest concentration in the standard curve with a standard deviation of C_q of less than 0.5 (Haugland et al., 2010; Kongprajug et al., 2020). The method limit of quantification (MLOQ) was calculated for each sample as copies/100 ml by incorporating the sample's filtration volume and DNA extracted volume. An inhibition analysis was performed with the dilution method using three dilutions (0.5, 1.0, and 2.0 μl) of the DNA templates in duplicates. The GenBac3 assay was administered, and C_q values for each dilution were plotted against the DNA concentration, and an R^2 of <0.90 suggested significant inhibition. Three field blanks and three method blanks were also investigated with HPyV and crAssphage assays (Sangkaew et al., 2021).

Statistical Analyses

A total of 13 water quality parameters for 27 samples were analyzed in R (R Core Team, 2019). Normality was assessed using the Shapiro–Wilk test. The two groups' significant differences were discerned using a t-test for normal data and the Mann–Whitney test for non-normal data. Significant differences for more than two groups were tested using one-way analysis of variance (ANOVA), with Tukey's multiple comparisons for normal data and Kruskal–Wallis test with Dunn's multiple comparisons for non-normal data. Significant difference tests for paired samples were carried out with a paired t-test for normal data sets and the Wilcoxon signed-rank test for non-normal data.

The data sets containing data below the MLOQ, so-called non-detects, were analyzed via non-parametric survival analyses (Helsel, 2012). The data sets' summary statistics, including non-detects, were calculated with the non-parametric Kaplan–Meier method, with Efron bias correction. The significance of the

differences for these data sets was examined via the generalized Wilcoxon test (Peto–Prentice test) for multiple comparisons, with Holm's bias correction. A paired sample comparison was also made with the paired Prentice–Wilcoxon test. Correlation analysis among multiple parameters was performed using Spearman's rho on *U*-Score rank.

To study the interaction between the pollution sources and anthropogenic activities, the types of land use (i.e., industrial and residential) were analyzed with fecal markers and water quality parameters using factor analysis of mixed data (FAMD), which is a type of principal component analysis (PCA) for examining a data set containing both quantitative and qualitative variables (Pages, 2004).

Risk Assessment and Sensitivity Analysis

A risk assessment following the QMRA framework comprises four basic steps: hazard identification, exposure assessment, dose–response model, and risk characterization (World Health Organization, 2016). A static QMRA was administered to evaluate infection risks from the HAdV40/41 pathogenic virus. First, in the hazard identification step, crAssphage and HPyVs were selected to estimate concentrations of HAdV40/41 pathogen using their detectable ratios in the receiving water. HAdV40/41 was the target pathogen for health risk assessment owing to the occurrence of waterborne illness, abundant in wastewater, causing gastrointestinal (GI) illness (Fong et al., 2010) and listed in the US EPA Contaminant Candidate List 4 (CCL4) (US EPA, 2016). The probability density functions (PDFs) of HAdV40/41, crAssphage, and HPyVs were modeled as a log-normal distribution from the measured concentrations. Then, the distribution ratios of HAdV40/41:crAssphage and HAdV40/41:HPyVs were used to simulate HAdV40/41 concentrations using the Monte Carlo approach at 10,000 iterations. The viable ratio of HAdV40/41 was assumed as 0.001 (Crank et al., 2019). In the second step, an exposure assessment was performed by considering three exposure scenarios: toilet flushing, aquaculture, and food crop irrigation, according to current practices in Map Ta Phut communities (Seemuang, 2018). Exposure factors, including exposure type, exposed water volume, and exposure frequency for three water reuse scenarios and a natural decay effect, are listed in **Supplementary Table 3**. A uniform distribution was considered for the water volume and the yearly frequency, and a triangular distribution for the natural decay was used, as previously reported (Chhipi-Shrestha et al., 2017). In the third step, the adenovirus dose–response relationship has been established and described by an exponential model for calculating the probability of daily infection (P_{inf}) as $1 - \exp(-0.4172d)$, where d represents the dose of viable pathogens (Chigor et al., 2014; Katukiza et al., 2014). Lastly, in the risk characterization step, the probability risk of infection was estimated by integrating hazard identification, exposure assessment, and dose–response model to characterize the risk. Risk characterization involves the determination of a health outcome with the risk of infection. The annual risk of infection (P_y) was calculated with the following equation: $1 - (1 - P_{inf})^{\wedge}f$, where f is the frequency of exposure per year. The Monte Carlo approach for simulation was run for 10,000 iterations for each scenario. The mean, median, 5th, 25th, 75th, and 95th percentiles of P_y were calculated using Oracle Crystal Ball v.11.1.2.4.850 software. Additional data visualization was completed in KaleidaGraph version 4.5.4. The estimated P_y was compared to the US EPA annual infection risk benchmark for finished drinking water of no more than 1 case per 10,000 persons per year (10^{-4} pppy) (US EPA, 1989). The required $\log_{10}$ reductions of HAdV40/41 were calculated for each water reuse scenario to achieve the US EPA risk benchmark. The recommended treatment technologies were evaluated using log-removal information, as reported previously (**Supplementary Table 4**).

Moreover, each input variable's effects to risk calculation were assessed using a sensitivity analysis with Oracle Crystal Ball software. Each parameter's significance was characterized by its correlation coefficient values with the probability risks, where a higher value indicated a more significant contribution of risks. Contribution to variance was calculated by squaring the rank correlation coefficient values and normalizing to 100%.

RESULTS

The qPCR Standard Curve Characteristics, Limits, and Controls

Standard curves for GenBac3, HPyVs, crAssphage, and HAdV40/41 were characterized, with PCR efficiencies ranging from 90.01 to 101.91% (**Supplementary Table 5**). The ALOD ranged from 20 copies/reaction (GenBac3 and crAssphage) to 40 copies/reaction (HPyV), while the ALOQ was 50 copies/reaction for all assays. The MLOQ in the negative samples were measured as 2.10–2.40, 2.40–3.00, and 2.40 $\log_{10}$ copies/100 ml for HPyVs (n = 2), crAssphage (n = 3), and HAdV40/41 (n = 24), respectively. Inhibition was detected in 14.8% of the samples (n = 4), each using a 0.5-μl DNA template to alleviate the inhibition effect. Laboratory reproducibility was examined when all four qPCR assays were analyzed in field duplicates, showing acceptable coefficients of variation in three representative samples (0.37–8.10%; **Supplementary Table 6**). NTCs were found to be negative with crAssphage (n = 12) and HPyV (n = 21). Two of 15 and 25 of 30 NTCs were positive for HAdV40/41 and GenBac3, respectively. However, their abundance of 1.83–2.02 copies/reaction and 1.30–1.41 copies/reaction was very low compared to those of the samples, demonstrating negligible effects. Furthermore, all the field and method blanks were negative, showing no contamination in the field and laboratory processing steps.

Physicochemical Water Quality Parameters

Seven physicochemical parameters were measured for nine sampling sites during three sampling events (n = 27) (**Table 1**). TSS was negative (<2.5 mg/l) in one sample, while the rest

TABLE 1 | Descriptive statistics of microbial and physicochemical water quality parameters.

Parameter	Normality[a]	Land use	Number of samples	Number of non-detects	Min[b]	25th percentile[b]	Median[b]	75th percentile[b]	Max[b]
GenBac3 (log_{10}	0.0956	Industrial	9	0	4.89	5.41	5.91	7.81	7.90
copies/100 ml)	0.7987	Residential	18	0	4.68	5.95	6.85	7.25	8.24
HPyVs (log_{10}	0.8151	Industrial	9	1	<2.40[c]	3.08	3.36	4.09	5.10
copies/100 ml)	0.9509	Residential	18	1	<2.10	3.04	3.64	4.41	5.07
crAssphage (log_{10}	0.7540	Industrial	9	2	<3.00	3.11	3.82	4.44	4.96
copies/100 ml)	0.3037	Residential	18	1	<2.40	3.89	4.32	4.56	5.21
HAdV40/41 (log_{10}	0.9669	Industrial	9	9	<2.40	<2.40	<2.40	<2.40	<2.40
copies/100 ml)	0.3037	Residential	18	15	<2.40	<2.40	<2.40	<2.40	3.87
Total coliforms	0.0013	Industrial	9	0	453	11,400	43,667	985,000	2,253,333
(CFU/100 ml)	<0.0001	Residential	18	0	11,200	50,167	72,167	829,167	7,333,333
E. coli (CFU/100 ml)	0.0002	Industrial	9	0	50	268	4650	701,667	1,713,333
	<0.0001	Residential	18	0	210	2450	23,250	279,167	3,513,333
Total suspended solid	0.0277	Industrial	9	1	<2.5	5.0	9.0	24.0	100.0
(mg/l)	<0.0001	Residential	18	0	3.0	8.8	16.5	29.3	307.0
BOD (mg/l)	0.0846	Industrial	9	0	2.5	7.3	9.6	25.6	39.6
	<0.0001	Residential	18	0	3.7	7.8	11.6	18.9	72.0
DO (mg/l)	0.9933	Industrial	9	0	4.6	6.7	8.8	10.1	11.7
	0.6842	Residential	18	0	3.5	5.4	6.6	7.7	9.6
Conductivity (mS/cm)	0.0025	Industrial	9	0	0.159	0.547	0.594	2.941	3.141
	0.0010	Residential	18	0	0.550	0.707	2.052	12.750	24.700
Salinity (ppt)	0.0007	Industrial	9	0	0.2	0.2	0.3	1.4	1.4
	0.0003	Residential	18	0	0.2	0.4	1.0	5.3	8.9
pH	0.1280	Industrial	9	0	5.93	6.99	8.58	8.87	9.14
	0.6778	Residential	18	0	6.87	7.18	7.39	7.63	8.24
Temperature (°C)	0.9268	Industrial	9	0	28.8	31.3	32.1	33.5	35.4
	0.0859	Residential	18	0	28.3	28.9	29.6	31.4	33.0

[a] *The Shapiro–Wilk normality test. P-values of less than 0.05 indicate non-normality.*
[b] *Summary statistics of data sets including non-detects were calculated with the non-parametric Kaplan–Meier method incorporating Efron bias correction.*
[c] *Data lower than the MLOQ.*

ranged from 3.0 to 307.0 mg/l. BOD ranged from 2.5 to 72.0 mg/l, DO from 3.5 to 11.7 mg/l, conductivity from 0.159 to 24.700 mS/cm, salinity from 0.2 to 8.9 ppt, pH from 5.93 to 9.14, and water temperature from 28.3 to 35.4°C. No spatial distributions, except conductivity and salinity, were significantly different among all the sampling sites ($p < 0.05$; one-way ANOVA with Tukey's multiple comparison test), with the highest values in MP1 and MP2 near the exits to the coast (**Supplementary Table 7**). For each sampling site, no temporal change was found for conductivity and salinity ($p > 0.05$; Wilcoxon signed-rank test) and temperature ($p > 0.05$; paired *t*-test) (**Supplementary Table 8**). The pH was higher in Event 1 than in the other two events, while BOD was significantly lower in Event 1 than in Event 3 ($p < 0.05$; Wilcoxon signed-rank test). DO was found to be higher in Event 1 than in Event 3 ($p < 0.05$; paired *t*-test), and TSS were higher in Event 3 than in Event 2 ($p < 0.05$; paired Prentice–Wilcoxon test).

MST Marker and Bacterial Monitoring

GenBac3 was detected in all samples with concentrations from 4.68 to 8.24 log_{10} copies/100 ml (**Table 1** and **Figure 2**). HPyVs were detected in all but two samples (MLOQs of 2.10 and 2.40 log_{10} copies/100 ml), with a maximum of 5.10 log_{10} copies/100 ml, while crAssphage was in all but three samples (MLOQs of 2.40 and 3.00 log_{10} copies/100 ml) at concentrations up to 5.21 log_{10} copies/100 ml, as previously reported (Sangkaew et al., 2021). HAdV40/41 was found only in three residential sites from Event 1 at MP1 (2.82 log_{10} copies/100 ml), MP3 (2.59 log_{10} copies/100 ml), and MP10 (3.87 log_{10} copies/100 ml), while the rest were non-detects with an MLOQ of 2.40 log_{10} copies/100 ml. GenBac3 was found to be most abundant in the same samples, followed by crAssphage, HPyVs, and HAdV40/41, respectively ($p < 0.05$; paired Prentice–Wilcoxon test). Total coliforms and *E. coli* were positive in all samples, with levels from 453 to 7,333,333 and from 50 to 3,513,333 CFU/100 ml, respectively. A total of 85.19% of the samples exceeded the Thailand

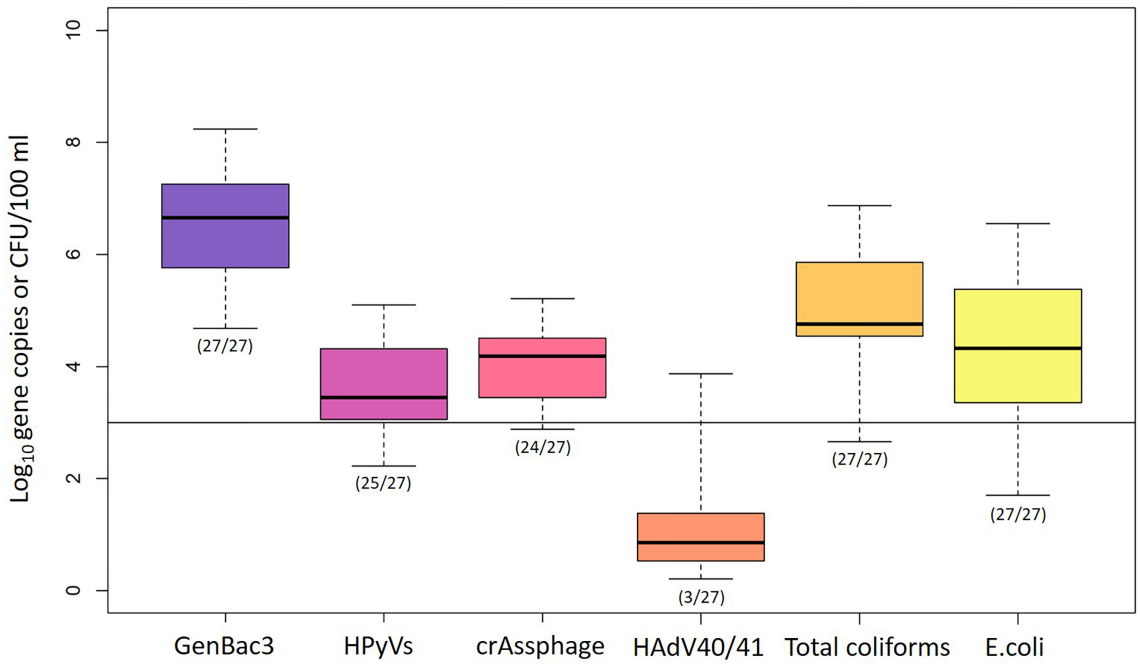


FIGURE 2 | Abundance of microbial water quality parameters. Box plots represent the estimated 25th to 75th percentiles, with the median between. The whiskers exhibit the maximum and minimum values. The numbers in parentheses describe the number of positive samples/the number of total samples. A solid line indicates the highest MLOQ of 3.00 $\log_{10}$ copies/100 ml.

surface water quality standard of 20,000 MPN/100 ml (National Environment Board, 1994) when the CFU unit was comparable to the MPN unit (Gronewold and Wolpert, 2008). Moreover, no difference for microbial parameters was observed between sampling sites, indicating continual microbial contamination in the study area (**Supplementary Table 7**). CrAssphage and *E. coli* presented no temporal differences among the three events at each sampling site, while GenBac3, HPyVs, and total coliforms all represented significantly lower concentrations in Event 1 ($p < 0.05$; **Supplementary Table 8**).

Correlation and Cluster Analyses

Quantitative correlation analysis demonstrated a significantly strong correlation for the microbial parameters in pairs of GenBac3 and HPyVs, crAssphage, total coliforms, and *E. coli* (Spearman's rhos = 0.64–0.80); a pair of HPyVs and crAssphage (rho = 0.74); and a pair of total coliforms and *E. coli* (rho = 0.84) (**Figure 3A**). Total coliforms and *E. coli* were also strongly correlated with TSS and BOD (rhos = 0.69–0.76). The physicochemical parameters were very strongly correlated between salinity and conductivity (rho = 0.97) and were strongly associated between pairs of BOD and TSS (rho = 0.76) and pH and DO (rho = 0.71).

A PCA was conducted, and the first two components displayed 42.4 and 20.7% of the variance (**Supplementary Figure 2A**). The variables contributing to the first two dimensions included GenBac3, salinity, total coliforms, BOD, HPyVs, conductivity, and *E. coli* (**Supplementary Figure 2B**). Two clusters of water quality parameters were demonstrated by the PCA (**Figure 3B**). Land use was more associated with physicochemical parameters in the first cluster (i.e., salinity, conductivity, water temperature, and DO). In contrast, the second cluster comprised physicochemical (i.e., pH, BOD, and TSS) and microbiological (i.e., GenBac3, HPyVs, total coliforms, *E. coli*, and crAssphage) variables (**Figure 3B**). When incorporating all the water quality parameters, the samples were clustered into two separate groups by land use type (**Figure 3C**). No significant differences in the abundance of MST and bacterial parameters between the two land use types were observed (**Supplementary Table 9**). DO and temperature were significantly higher and salinity was lower in industrial sites than in the residential sites ($p < 0.05$). At the same time, the rest of the physicochemical parameters provided no significant difference (**Supplementary Table 9**).

QMRA and Sensitivity Analysis

The infection risks of HAdV40/41 were assessed using measured concentrations in the three monitoring campaigns. Due to positively detected HAdV40/41 in only three samples from the residential sites, we decided to increase the data set's robustness by combining the data from all the sampling sites without segregating them into industrial and residential groups. The measured concentrations of HAdV40/41, crAssphage, and HPyVs were fitted to estimate the PDF in a log-normal distribution (**Supplementary Table 10**), and the predicted HAdV40/41 distribution was discerned from crAssphage and HPyVs. Due to higher exposure factors, the estimated Py increased from toilet flushing to aquaculture and food crop

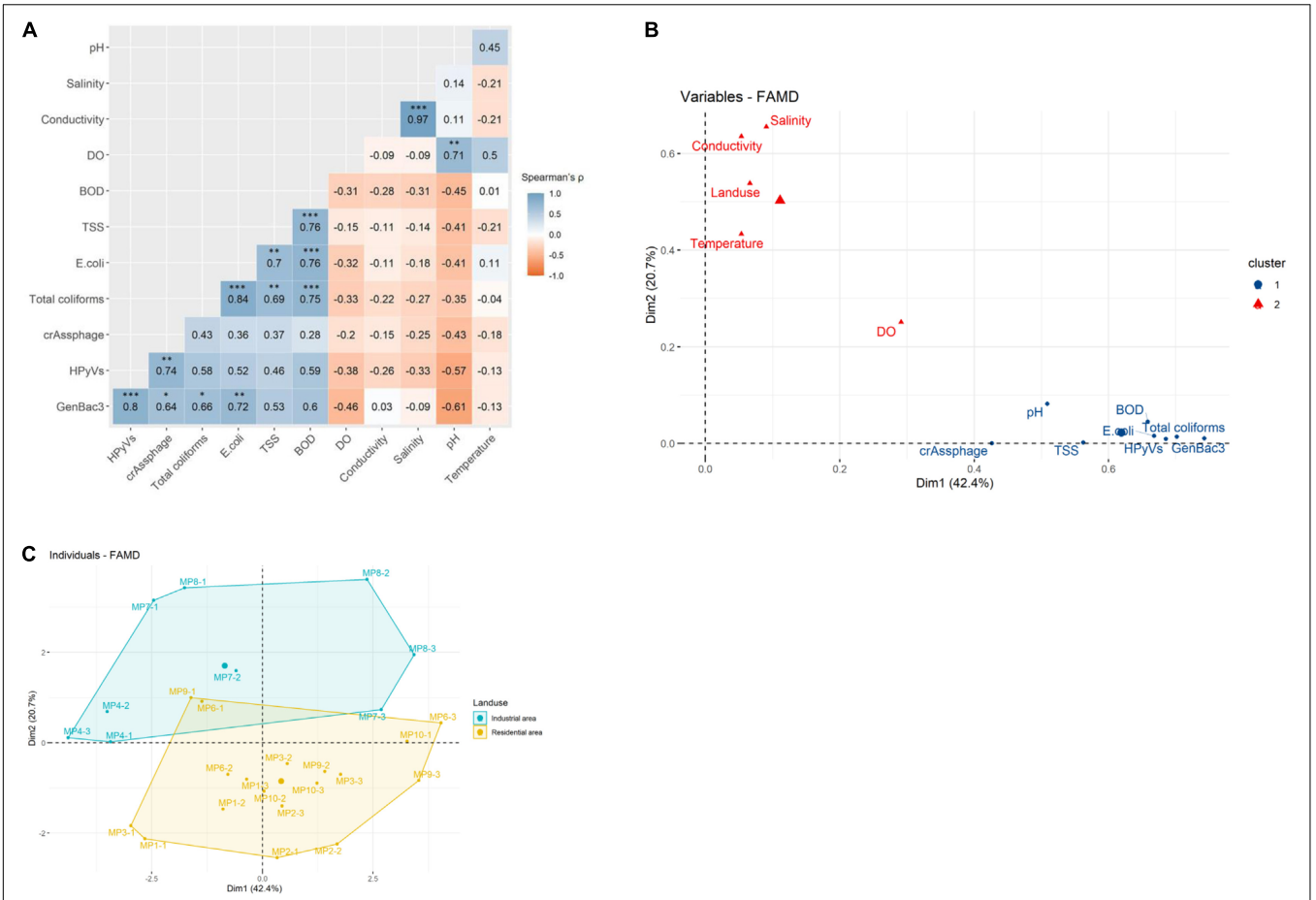


FIGURE 3 | Multivariate analysis for quantitative correlation using Spearman's rho on *U*-score rank with *, **, and *** indicating Holm's adjusted *P*-value < 0.05, 0.01, and 0.001, respectively **(A)**, clustered principal component analysis (PCA) using factor analysis of mixed data (FAMD) for water quality parameters **(B)**, and clustered PCA using FAMD for water samples **(C)**.

irrigation scenarios. The 95th percentiles of P_y predicted from crAssphage were 9.45×10^{-3}, 6.34×10^{-2}, and 2.63×10^{-1} pppy for toilet flushing, aquaculture, and agricultural irrigation, respectively (**Table 2**). In addition, the 95th percentiles of P_y predicted from HPyVs were 9.74×10^{-3}, 6.48×10^{-2}, and 2.70×10^{-1} pppy for toilet flushing, aquaculture, and irrigation, respectively (**Table 2**). The risk results from crAssphage and HPyVs were not significantly different due to a significant correlation between these two parameters. However, all the scenarios exceeded the US EPA risk benchmark at 10^{-4} pppy. Therefore, additional treatment was necessary. Based on the pooled detectable data in the dry season of the viruses, the requirements of additional treatment for HAdV40/41 to achieve an annual health risk of less than 10^{-4} pppy were estimated at 1.65–2.62, 2.48–3.46, and 3.15–4.13 $\log_{10}$ reduction (5th–95th percentiles) for toilet flushing, aquaculture, and food crop irrigation, respectively (**Table 3**). Various water treatment units were evaluated for each water reuse scenario (**Supplementary Table 4**). Chlorination was sufficient to achieve the 95th percentile ($\log_{10}$) reduction of HAdV40/41 for the toilet flushing reuse purpose. Moreover, the combination of depth filtration and chlorination was satisfactory to achieve the 95th percentile ($\log_{10}$) reduction of HAdV40/41 for aquaculture and food crop irrigation.

Furthermore, this study predicted the P_y from varying initial concentrations of crAssphage and HPyVs in the receiving water from 10^0 to 10^5 copies/100 ml (**Figure 4**). We found that the health risks expanded as the concentrations of markers increased. The 95th health risks of HAdV40/41 passed the US EPA annual infection risk benchmark of 10^{-4} pppy only in the following scenarios: (1) 10^0 copies/100 ml of crAssphage for all three water reuse schemes, and (2) 10^0 copies/100 ml of HPyVs for toilet flushing. Required log-removal can also be estimated from **Figure 4**. For instance, at the initial concentration of crAssphage at 10^4 copies/100 ml, the receiving water requires approximately 3.2 log reduction treatment to achieve the US EPA benchmark at 10^{-4} for toilet flushing purposes. This is a useful tool for evaluating water quality and selecting the appropriate treatment for water reuse.

A sensitivity analysis was conducted to investigate the effects of each input variable on the estimated probability

TABLE 2 | Yearly infection probability (P_y) of human adenovirus types 40 and 41 (HAdV40/41) predicted from crAssphage and HPyVs.

	Yearly infection probability (P_y) in pppy					
Reuse purposes	**5th percentile**	**25th percentile**	**Median**	**75th percentile**	**95th percentile**	**Mean**
Prediction from crAssphage						
Toilet flushing	8.85×10^{-4}	1.77×10^{-3}	2.89×10^{-3}	4.70×10^{-3}	9.45×10^{-3}	3.74×10^{-3}
Aquaculture	5.94×10^{-3}	1.20×10^{-2}	2.51×10^{-2}	3.14×10^{-2}	6.34×10^{-2}	2.51×10^{-2}
Food crop irrigation	2.72×10^{-2}	5.50×10^{-2}	8.83×10^{-2}	1.39×10^{-1}	2.63×10^{-1}	1.08×10^{-1}
Prediction from HPyVs						
Toilet flushing	8.53×10^{-4}	1.78×10^{-3}	2.93×10^{-3}	4.85×10^{-3}	9.74×10^{-3}	3.83×10^{-3}
Aquaculture	5.80×10^{-3}	1.21×10^{-2}	1.99×10^{-2}	3.23×10^{-2}	6.48×10^{-2}	2.56×10^{-2}
Food crop irrigation	2.69×10^{-2}	5.48×10^{-2}	8.94×10^{-2}	1.43×10^{-1}	2.70×10^{-1}	1.11×10^{-1}

risks. The main contributor to risk variability in all the scenarios was the concentrations of HAdV40/41 in the receiving water (85.10 to 86.60%), which were similar when predicting HAdV40/41 from crAssphage and HPyVs (**Supplementary Figure 3**). Natural decay rates also had a reverse relationship with calculated risks (−12.30 to −13.50%). Less contributing factors included water use frequency per year, the volume of water, and the crAssphage or HPyV marker for prediction.

DISCUSSION

The MST marker abundance was investigated in surface water near the mixed land use of industrial and residential areas. Fecal contamination was indicated by GenBac3 level in this study (four to eight orders of magnitude in $\log_{10}$ copies/100 ml), which were slightly lower than those found in untreated sewage in Thailand (five to nine orders in $\log_{10}$ copies/100 ml) (Kongprajug et al., 2020), but slightly higher than the Tha Chin River in Central Thailand (three to seven orders in $\log_{10}$ copies/100 ml) (Kongprajug et al., 2019a). These GenBac3 levels were aligned with levels found in other geographical regions in urban rivers (five to seven orders), urban recreational beaches (two to five orders), and karst springs (four to seven orders) (Molina et al., 2014; Ohad et al., 2015; Devane et al., 2019). Human fecal contamination in this study was identified by the HPyVs and crAssphage markers of up to five orders in $\log_{10}$ copies/100 ml in this study, which were relatively lower than those presented in untreated sewage in Thailand (three to six and five to seven orders, respectively) (Kongprajug et al., 2019b; Sangkaew et al., 2021) and in the Tha Chin river (three to seven orders for crAssphage) (Kongprajug et al., 2019b), but slightly higher than in beach water in Thailand (up to four and three orders, respectively) (Sangkaew et al., 2021). The levels of HPyVs and crAssphage in Thailand's freshwater were equivalent to those found in other geographical areas, as previously discussed (Sangkaew et al., 2021). Notably, marker abundance is also affected by the decay rate, which can be dependent on the water type (freshwater or marine water) and other environmental stressors (e.g., sunlight and indigenous microbiota) at different levels (Booncharoen et al., 2018; Korajkic et al., 2019).

TABLE 3 | Required $\log_{10}$ reduction of HAdV40/41 to achieve the US EPA annual infection risk benchmark of 10^{-4} per person per year as predicted by either crAssphage or HPyVs.

	Required $\log_{10}$ reduction			
	5th percentile	**Median**	**95th percentile**	**Mean**
Toilet flushing	1.65	2.13	2.62	2.14
Aquaculture	2.48	2.97	3.46	2.97
Food crop irrigation	3.15	3.63	4.13	3.64

The lack of spatial differences in the microbial indicators among the sampling sites indicated that the Nam Hu Canal passing through residential areas was affected by fecal contamination at levels similar to the Nam Cha and Lot Canals with upstream industrial sites (**Figure 1**). The Chak Mak Canal connects to the coast through a private land area, which disallowed sample collection at the downstream site. However, the Chak Mak Canal has been reported for its low flow and low current circulation in Pra Doo Bay (Singkran, 2012), leading to a minimal distribution of fecal pollution to the broader environment. This study emphasized the human fecal contamination impact on receiving water during dry weather, as has been previously reported (Sercu et al., 2009; Zimmer-Faust et al., 2018; Green et al., 2019; Gyawali et al., 2020). Sources of fecal contamination during dry weather could be from damaged sewer collection pipes, related infrastructure, or illicit discharges, which require further investigation. A total of 16,171 households are upstream of the Nam Cha and Lot Canals in the Huay Pong Sub-district, according to the 2019 National Population and Housing Census. The Chak Mak Canal runs through the Map Ta Phut Sub-district, which hosts 20,980 households, while the Nam Hu Canal passes by the Nean Phra Sub-district, which has 13,991 households. Stormwater runoff during rainfall events, even though not yet characterized in this study, potentially carries additional general or animal fecal microorganisms into the receiving water (Ahmed et al., 2019; Lee et al., 2020; Powers et al., 2020). However, the already-high level of fecal contamination in dry weather, as shown in this study, and especially the high percentage of samples exceeding the total coliform standards, has raised the attention for water

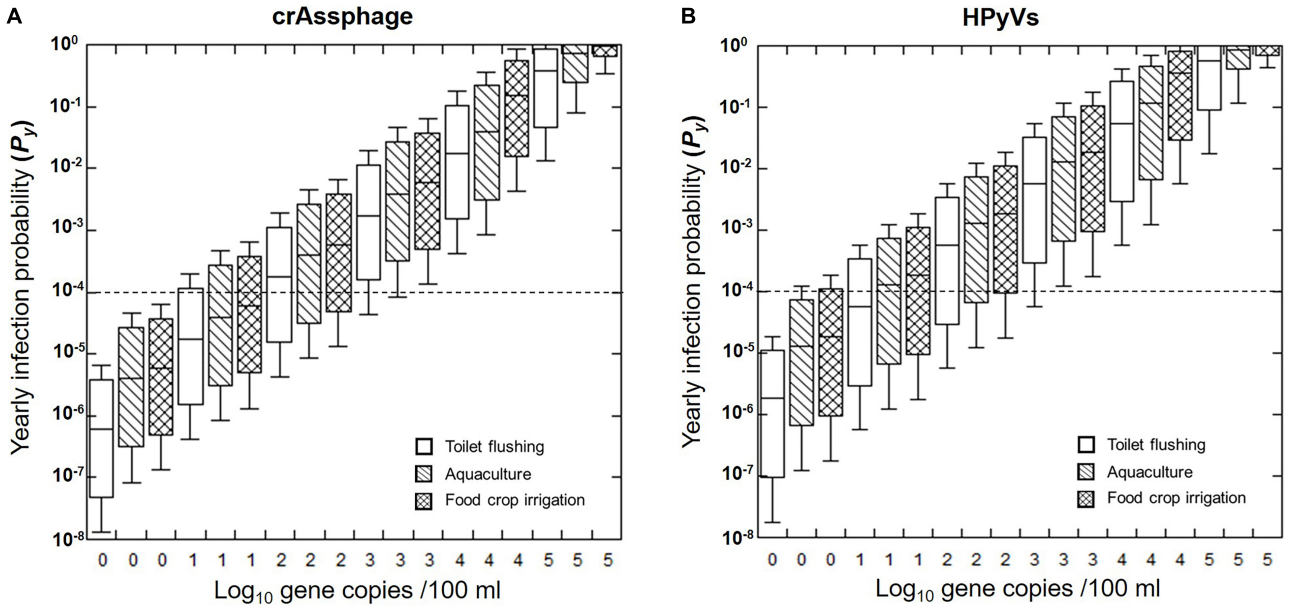


FIGURE 4 | Predicted annual probability of infection (*Py*) for HAdV40/41 based on crAssphage **(A)** and HPyVs **(B)**. Box plots represent the estimated 25th to 75th percentiles, with the median between. The whiskers exhibit the 5th and 95th percentile values. The dashed lines indicate the US EPA annual risk benchmark of 10^{-4} pppy.

quality mitigation strategies for expected higher pollution levels in the rainy season.

This study introduced a QMRA framework to promote a water reuse concept of receiving water in industrial communities. Adenoviruses have been widely used for QMRA in many applications, such as direct potable reuse (Amoueyan et al., 2019), indirect potable reuse to augment ground or surface water drinking sources (Amoueyan et al., 2019; Purnell et al., 2020), indirect wastewater reuse (Ahmed et al., 2018), biosolids for agricultural land applications (Hamilton et al., 2020), public drinking water supplies (Owens et al., 2020), natural recreational water (Federigi et al., 2019), stormwater runoff (Ahmed et al., 2019), and occupational exposure at wastewater treatment facilities (Carducci et al., 2018). In this study, HAdV40/41 distribution functions were estimated from crAssphage distributions, as supported by a significant correlation between these two DNA markers in wastewater (Farkas et al., 2019; Sirikanchana et al., 2020; Tandukar et al., 2020; Wu et al., 2020). Moreover, HPyVs and crAssphage have been reported for their significant correlation in wastewater and environmental waters (Sangkaew et al., 2021), thus supporting the fact that similar estimated risks of HAdV40/41 were observed in this study when predicted from either HPyVs or crAssphage. Human-specific MST markers, such as crAssphage (Crank et al., 2019) and HF183 (Boehm et al., 2015; Ahmed et al., 2018), have been used for QMRA; they provide the advantage of higher abundance than waterborne pathogens (Zhang et al., 2019).

In risk assessment, uncertainties are identified by a sensitivity analysis, and the uncertainties in this study were mainly due to HAdV40/41 abundance. With a relatively low positive percentage (11%) of HAdV40/41 DNA measurement, the simulated distribution functions from crAssphage and HPyVs prediction may have been confounded by different environmental conditions in each sampling event. Moreover, the annual risk overestimation could have occurred from substitution of MLOQ values for non-detects. Decreasing the MLOQ value would increase the possibility of detectable concentrations. Certain factors that could lower the detection limits include improved concentration recovery, larger processed water volume, and PCR inhibition alleviation (Kreader, 1996; Gibson et al., 2012; Petterson et al., 2015; Ahmed et al., 2020). Moreover, an assumption of infectious ratio values applied to the monitored DNA measurements (e.g., by molecular qPCR assays) contributes another uncertainty factor to the concentrations of infectious viruses to inform QMRA. Viral infectivity measurements, including cell culture and qPCR with azo-dye pretreatment methods (Ko et al., 2003; Leifels et al., 2019, 2021), could directly provide infectious concentrations of viruses for risk analysis. The differential decay characteristics of the HAdV40/41 and MST markers (i.e., crAssphage and HPyVs) in receiving water may also result in under- or over-estimation of the yearly infection risk (Greaves et al., 2020; McMinn et al., 2020). Therefore, future research is needed to investigate decay rates of the pathogens of interest in local environmental conditions and water matrices for accurate risk assessment. Studies to retrieve exposure factor values, including exposure volume and frequency, will also aid in precise assessment in local settings.

The US EPA drinking water annual risk benchmark of 10^{-4} pppy (US EPA, 1989) has been widely referenced in direct and indirect potable reuse (Amoueyan et al., 2019) and non-potable reuse (Hamilton et al., 2017; Simhon et al., 2020), although an

attempt has been made to loosen the risk standard for irrigation purposes (Lim and Jiang, 2013). Moreover, the disability-adjusted life year (DALY) scores could be further calculated and emphasize the need for regional-specific information, including the total incidence, odds of severity, and illness duration (Hamilton et al., 2019). This study proposed three water reuse schemes: toilet flushing, aquaculture, and food crop irrigation. Toilet flushing reuse has been encouraged to promote the concept of eco-industrial parks (UNIDO, 2019). This study demonstrated that chlorination treatment is sufficient to decrease the viral risks of the receiving water and can promote water reuse practice in factories in the Map Ta Phut industrial estate. Chlorination is also a most common disinfection practice in Thailand due to its low cost and wide availability. Moreover, food crop irrigation and aquaculture, such as shellfish farming, are currently practiced in the local community. Although suggested pretreatments of depth filtration and chlorination may pose certain complications to implement, the infection risks calculated in this study have raised awareness of the current public health risks that will require further corrective actions. Additionally, effects of storm events have shown to be area-specific as higher abundance of human-specific fecal markers were observed after rainfall events in some studies (McGinnis et al., 2018; Wu et al., 2018), but not in others (García-Aljaro et al., 2017; Staley et al., 2018; Powers et al., 2020). Future wet-weather MST monitoring in this study area is therefore needed in order to evaluate risks and propose appropriate treatment units to reduce public health risks after storm events.

CONCLUSION

Microbial fecal contamination from industrial and residential land uses in a coastal mixed-use area was investigated in this study using an integrated approach combining microbial water quality monitoring and the QMRA framework. Human fecal pollution as identified by sewage-specific MST markers was equally contributed by industrial and residential land uses. Annual infection risks for HAdV40/41 were assessed, and prior treatments were recommended before non-potable water reuse, including toilet flushing, aquaculture, and food crop irrigation. This first integrative microbial risk assessment approach in Thailand could enable effective preventive and mitigation measures for water quality management and water reuse in industrial and residential mixed land uses.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

AK performed the experiments and analyzed data. NC conducted the experiments. TD carried out data and risk analyses. SM and KS conceived and designed the study. All authors discuss the results and wrote the manuscript.

FUNDING

This work was financially supported by the Chulabhorn Research Institute (Grant No. 312/3057) and the Kurita Overseas Research Grant (Grant No. 19Pth007 to AK) provided by Kurita Water and Environment Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.647602/full#supplementary-material

REFERENCES

Ahmed, W., Bertsch, P. M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., et al. (2020). Comparison of virus concentration methods for the RT-qPCR- based recovery of murine hepatitis virus, a surrogate for SARS- CoV-2 from untreated wastewater warish. *Sci. Total Environ.* 739:139960. doi: 10.1016/j.scitotenv.2020.139960

Ahmed, W., Hamilton, K., Toze, S., Cook, S., and Page, D. (2019). A review on microbial contaminants in stormwater runoff and outfalls: potential health risks and mitigation strategies. *Sci. Total Environ.* 692, 1304–1321. doi: 10.1016/j.scitotenv.2019.07.055

Ahmed, W., Hamilton, K. A., Lobos, A., Hughes, B., Staley, C., Sadowsky, M. J., et al. (2018). Quantitative microbial risk assessment of microbial source tracking markers in recreational water contaminated with fresh untreated and secondary treated sewage. *Environ. Int.* 117, 243–249. doi: 10.1016/j.envint.2018.05.012

Ahmed, W., Hughes, B., and Harwood, V. J. (2016). Current status of marker genes of *bacteroides* and related taxa for identifying sewage pollution in environmental waters. *Water* 8:231. doi: 10.3390/w8060231

Altes, W. K. K., and Tambach, M. (2008). Municipal strategies for introducing housing on industrial estates as part of compact-city policies in the netherlands. *Cities* 25, 218–229. doi: 10.1016/j.cities.2008.04.005

American Public Health Association, and Water Environment Federation. (2017a). *2540 D. Total Suspended Solids Dried at 103 – 105°C," in Standard Methods for the Examination of Water & Wastewater*. Denver, CO: AWWA, 4–5.

American Public Health Association, and Water Environment Federation. (2017b). *5210 B. 5-Day BOD Test," in Standard Methods for the Examination of Water & Wastewater*. Denver, CO: AWWA, 1–6.

Amoueyan, E., Ahmad, S., Eisenberg, J. N. S., and Gerrity, D. (2019). Equivalency of indirect and direct potable reuse paradigms based on a quantitative microbial risk assessment framework. *Microb. Risk Anal.* 12, 60–75. doi: 10.1016/j.mran.2019.06.003

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

Booncharoen, N., Mongkolsuk, S., and Sirikanchana, K. (2018). Comparative persistence of human sewage-specific enterococcal bacteriophages in freshwater and seawater. *Appl. Microbiol. Biotechnol.* 102, 6235–6246.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J. F., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Carducci, A., Donzelli, G., Cioni, L., Federigi, I., Lombardi, R., and Verani, M. (2018). Quantitative microbial risk assessment for workers exposed to bioaerosol in wastewater treatment plants aimed at the choice and setup of safety measures. *Int. J. Environ. Res. Public Health* 15:1490. doi: 10.3390/ijerph15071490

Chacón, L., Barrantes, K., Santamaría-Ulloa, C., Solano, M., Reyes, L., Taylor, L., et al. (2020). A somatic coliphage threshold approach to improve the management of activated sludge wastewater treatment plant effluents in resource-limited regions. *Appl. Environ. Microbiol.* 86, 1–16. doi: 10.1128/AEM.00616-20

Chhipi-Shrestha, G., Hewage, K., and Sadiq, R. (2017). Microbial quality of reclaimed water for urban reuses: Probabilistic risk-based investigation and recommendations. *Sci. Total Environ.* 576, 738–751. doi: 10.1016/j.scitotenv.2016.10.105

Chigor, V. N., Sibanda, T., and Okoh, A. I. (2014). Assessment of the risks for human health of adenoviruses, hepatitis A virus, rotaviruses and enteroviruses in the buffalo river and three source water dams in the eastern cape. *Food Environ. Virol.* 6, 87–98. doi: 10.1007/s12560-014-9138-4

Crank, K., Petersen, S., and Bibby, K. (2019). Quantitative microbial risk assessment of swimming in sewage impacted waters using CrAssphage and pepper mild mottle virus in a customizable model. *Environ. Sci. Technol. Lett.* 6, 571–577. doi: 10.1021/acs.estlett.9b00468

Devane, M. L., Moriarty, E. M., Robson, B., Lin, S., Wood, D., Webster-Brown, J., et al. (2019). Relationships between chemical and microbial faecal source tracking markers in urban river water and sediments during and post-discharge of human sewage. *Sci. Total Environ.* 651, 1588–1604. doi: 10.1016/j.scitotenv.2018.09.258

European Union. (2006). *Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC*. Brussels: European Union.

Farkas, K., Adriaenssens, E. M., Walker, D. I., McDonald, J. E., Malham, S. K., and Jones, D. L. (2019). Critical evaluation of CrAssphage as a molecular marker for human-derived wastewater contamination in the aquatic environment. *Food Environ. Virol.* 11, 113–119. doi: 10.1007/s12560-019-09369-1

Federigi, I., Verani, M., Donzelli, G., Cioni, L., and Carducci, A. (2019). The application of quantitative microbial risk assessment to natural recreational waters: a review. *Mar. Pollut. Bull.* 144, 334–350. doi: 10.1016/j.marpolbul.2019.04.073

Fong, T. T., Phanikumar, M. S., Xagoraraki, I., and Rose, J. B. (2010). Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a michigan river. *Appl. Environ. Microbiol.* 76, 715–723. doi: 10.1128/AEM.01316-09

Garcia-Aljaro, C., Blanch, A. R., Campos, C., Jofre, J., and Lucena, F. (2018). Pathogens, faecal indicators and human-specific microbial source-tracking markers in sewage. *J. Appl. Microbiol.* 126, 701–717. doi: 10.1111/jam.14112

García-Aljaro, C., Martín-Díaz, J., Viñas-Balada, E., Calero-Cáceres, W., Lucena, F., and Blanch, A. R. (2017). Mobilisation of microbial indicators, microbial source tracking markers and pathogens after rainfall events. *Water Res.* 112, 248–253. doi: 10.1016/j.watres.2017.02.003

García-Pérez, J., Fernández de Larrea-Baz, N., Lope, V., Molina, A. J., O'Callaghan-Gordo, C., Alonso, M. H., et al. (2020). Residential proximity to industrial pollution sources and colorectal cancer risk: A multicase-control study (MCC-Spain). *Environ. Int.* 144:106055. doi: 10.1016/j.envint.2020.106055

Gibson, K. E., Schwab, K. J., Spencer, S. K., and Borchardt, M. A. (2012). Measuring and mitigating inhibition during quantitative real time PCR analysis of viral nucleic acid extracts from large-volume environmental water samples. *Water Res.* 46, 4281–4291. doi: 10.1016/j.watres.2012.04.030

Greaves, J., Stone, D., Wu, Z., and Bibby, K. (2020). Persistence of emerging viral fecal indicators in large-scale freshwater mesocosms. *Water Res. X* 9:135577. doi: 10.1016/j.scitotenv.2019.135577

Green, H., Weller, D., 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

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

Gyawali, P., Hamilton, K., Joshi, S., Aster, D., and Ahmed, W. (2020). Identification of reliable marker genes for the detection of canine fecal contamination in subtropical australia. *Sci. Total Environ.* 718:137246. doi: 10.1016/j.scitotenv.2020.137246

Hamilton, K., Reyneke, B., Waso, M., Clements, T., Ndlovu, T., Khan, W., et al. (2019). A global review of the microbiological quality and potential health risks associated with roof-harvested rainwater tanks. *npj Clean Water* 2, 1–18. doi: 10.1038/s41545-019-0030-5

Hamilton, K. A., Ahmed, W., Palmer, A., Smith, K., Toze, S., and Haas, C. N. (2017). Seasonal assessment of opportunistic premise plumbing pathogens in roof-harvested rainwater tanks. *Environ. Sci. Technol.* 51, 1742–1753.

Hamilton, K. A., Ahmed, W., Rauh, E., Rock, C., Mclain, J., and Muenich, R. L. (2020). Comparing microbial risks from multiple sustainable waste streams applied for agricultural use: Biosolids, manure, and diverted urine. *Curr. Opin. Environ. Sci. Heal.* 14, 37–50. doi: 10.1016/j.coesh.2020.01.003

Haugland, R. A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L., and Shanks, O. C. (2010). Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected bacteroidales species and human fecal waste by qPCR. *Syst. Appl. Microbiol.* 33, 348–357. doi: 10.1016/j.syapm.2010.06.001

Helsel, D. R. (2012). *Statistics for Censored Environmental Data Using Minitab and R*. second. Hoboken, NJ: John Wiley & Sons, Inc.

Holcomb, D. A., Stewart, J. R., and Stewart, J. R. (2020). Microbial indicators of fecal pollution: recent progress and challenges in assessing water quality. *Curr. Environ. Heal. Rep.* 7, 311–324.

Hong, H., and Gasparatos, A. (2020). Eco-industrial parks in china: key institutional aspects, sustainability impacts, and implementation challenges. *J. Clean. Prod.* 274:122853. doi: 10.1016/j.jclepro.2020.122853

Industrial Estate Authority of Thailand. (2018). *Feasibility study for the special economic zones in the Map Ta Phut Industrial Estate and Map Ta Phut Industrial Port*. Thailand: Industrial Estate Authority of Thailand.

Industrial Estate Authority of Thailand. (2019). *Eco Internal Audit Handbook for Eco-Champion*. Thailand: Industrial Estate Authority of Thailand.

Industrial Estate Authority of Thailand. (2020). Notification of the industrial estate authority of thailland: area modification of the map ta phut industrial estate (No. 7). *R. Gov. Gaz.* 137:62. ,

Jampani, M., Amerasinghe, P., Liedl, R., Locher-Krause, K., and Hülsmann, S. (2020). Multi-functionality and land use dynamics in a peri-urban environment influenced by wastewater irrigation. *Sustain. Cities Soc.* 62:102305. doi: 10.1016/j.scs.2020.102305

Jeong, H., Kim, H., Jang, T., and Park, S. (2016). Assessing the effects of indirect wastewater reuse on paddy irrigation in the osan river watershed in korea using the SWAT model. *Agric. Water Manag.* 163, 393–402. doi: 10.1016/j.agwat.2015.08.018

Katukiza, A. Y., Ronteltap, M., van der Steen, P., Foppen, J. W. A., and Lens, P. N. L. (2014). Quantification of microbial risks to human health caused by waterborne viruses and bacteria in an urban slum. *J. Appl. Microbiol.* 116, 447–463. doi: 10.1111/jam.12368

Khan, S. A., Muhammad, S., Nazir, S., and Shah, F. A. (2020). Heavy metals bounded to particulate matter in the residential and industrial sites of islamabad, pakistan: implications for non-cancer and cancer risks: Evaluation of heavy metals bounded to particulate matter for non-cancer and cancer risks. *Environ. Technol. Innov.* 19:100822. doi: 10.1016/j.eti.2020.100822

Ko, G., Cromeans, T. L., and Sobsey, M. D. (2003). Detection of infectious adenovirus in cell culture by mRNA reverse transcription-PCR. *Appl. Environ. Microbiol.* 69, 7377–7384. doi: 10.1128/AEM.69.12.7377-7384.2003

Ko, G., Jothikumar, N., Hill, V. R., and Sobsey, M. D. (2005). Rapid detection of infectious adenoviruses by mRNA real-time RT-PCR. *J. Virol. Methods* 127, 148–153. doi: 10.1016/j.jviromet.2005.02.017

Kongprajug, A., Chyerochana, N., Mongkolsuk, S., and Sirikanchana, K. (2020). Effect of quantitative polymerase chain reaction data analysis using sample amplification efficiency on microbial source tracking assay performance and source attribution. *Environ. Sci. Technol.* 54, 8232–8244.

Kongprajug, A., Chyerochana, N., Somnark, P., Kampaengthong, P. L., Mongkolsuk, S., and Sirikanchana, K. (2019a). Human and animal microbial source tracking in a tropical river with multiple land use activities. *Int. J. Hyg. Environ. Health* 222, 645–654. doi: 10.1016/j.ijheh.2019.01.005

Kongprajug, A., Mongkolsuk, S., and Sirikanchana, K. (2019b). CrAssphage as a potential human sewage marker for microbial source tracking in southeast asia. *Environ. Sci. Technol. Lett.* 6, 159–164. doi: 10.1021/acs.estlett.9b00041

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

Kreader, C. A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl. Environ. Microbiol.* 62, 1102–1106.

Kumthip, K., Khamrin, P., Ushijima, H., and Maneekarn, N. (2019). Enteric and non-enteric adenoviruses associated with acute gastroenteritis in pediatric patients in thailand, 2011 to 2017. *PLoS One* 14:220263. doi: 10.1371/journal.pone.0220263

Lee, S., Suits, M., Wituszynski, D., Winston, R., Martin, J., and Lee, J. (2020). Residential urban stormwater runoff: a comprehensive profile of microbiome and antibiotic resistance. *Sci. Total Environ.* 723:138033. doi: 10.1016/j.scitotenv.2020.138033

Leifels, M., Dan, C., Sozzi, E., Shoults, D. C., Wuertz, S., Mongkolsuk, S., et al. (2021). Capsid integrity quantitative PCR to determine virus infectivity in environmental and food applications – A systematic review. *Water Res. X* 11:100080. doi: 10.1016/j.wroa.2020.100080

Leifels, M., Shoults, D., Wiedemeyer, A., Ashbolt, N. J., Sozzi, E., Hagemeier, A., et al. (2019). Capsid integrity qPCR-an azo-dye based and culture-independent approach to estimate adenovirus infectivity after disinfection and in the aquatic environment. *Water* 11, 1–15. doi: 10.3390/w11061196

Lim, K.-Y., and Jiang, S. C. (2013). Reevaluation of health risk benchmark for sustainable water practice through risk analysis of rooftop-harvested rainwater. *Water Res.* 47, 7273–7286. doi: 10.1016/j.watres.2013.09.059

Lin, C. K., Hsu, Y. T., Brown, K. D., Pokharel, B., Wei, Y., and Chen, S. T. (2020). Residential exposure to petrochemical industrial complexes and the risk of leukemia: a systematic review and exposure-response meta-analysis. *Environ. Pollut.* 258:113476. doi: 10.1016/j.envpol.2019.113476

McGinnis, S., Spencer, S., Firnstahl, A., Stokdyk, J., Borchardt, M., McCarthy, D. T., et al. (2018). Human *Bacteroides* and total coliforms as indicators of recent combined sewer overflows and rain events in urban creeks. *Sci. Total Environ.* 630, 967–976. doi: 10.1016/j.scitotenv.2018.02.108

McMinn, B. R., Rhodes, E. R., Huff, E. M., and Korajkic, A. (2020). Decay of infectious adenovirus and coliphages in freshwater habitats is differentially affected by ambient sunlight and the presence of indigenous protozoa communities. *Virol. J.* 17, 1–11. doi: 10.1186/s12985-019-1274-x

McQuaig, S. M., Scott, T. M., Lukasik, J. O., Paul, J. H., and Harwood, V. J. (2009). Quantification of human polyomaviruses JC virus and BK virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples. *Appl. Environ. Microbiol.* 75, 3379–3388. doi: 10.1128/AEM.02302-08

Medeiros, A. C., Faial, K. R. F., do Carmo Freitas Faial, K., da Silva Lopes, I. D., de Oliveira Lima, M., Guimarães, R. M., et al. (2017). Quality index of the surface water of amazonian rivers in industrial areas in Pará brazil. *Mar. Pollut. Bull.* 123, 156–164. doi: 10.1016/j.marpolbul.2017.09.002

Ministry of Industry. (2017). Announcement of the ministry of industry entitled the treated effluent standards from factories. *R. Gov. Gaz.* 134, 11–15.

Ministry of Natural Resources and Environment. (2016). Notification of the ministry of natural resources and environment entitled the standards of treated effluent from industrial factories, industrial estates, and industrial areas. *R. Gov. Gaz.* 133, 17–21.

Molina, M., Hunter, S., Cyterski, M., Peed, L. A., Kelty, C. A., Sivaganesan, M., et al. (2014). Factors affecting the presence of human-associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. *Water Res.* 64, 196–208. doi: 10.1016/j.watres.2014.06.036

Naidoo, S., and Olaniran, A. O. (2014). Treated wastewater effluent as a source of microbial pollution of surface water resources. *Int. J. Environ. Res. Public Health* 11, 249–270. doi: 10.3390/ijerph110100249

National Environment Board. (1994). *Notification of the National Environmental Board: Surface Water Quality Standards, No. 8, B.E. 2537 (1994), issued under the Enhancement and Conservation of National Environmental Quality Act B.E.2535 (1992)*.

Office of Map Ta Phut Industrial Estate. (2020). *Map Ta Phut Industrial Estate*. Available Online at: http://www.mtpie.com/ [Accessed October 3, 2020]

Ohad, S., Vaizel-Ohayon, D., Rom, M., Guttman, J., Berger, D., Kravitz, V., et al. (2015). Microbial source tracking in adjacent karst springs. *Appl. Environ. Microbiol.* 81, 5037–5047. doi: 10.1128/AEM.00855-15

Owens, C. E. L., Angles, M. L., Cox, P. T., Byleveld, P. M., Osborne, N. J., and Rahman, M. B. (2020). Implementation of quantitative microbial risk assessment (QMRA) for public drinking water supplies: systematic review. *Water Res.* 174:115614. doi: 10.1016/j.watres.2020.115614

Pages, J. (2004). Analyse factorielle de donnees mixtes. *Rev. Stat. Appl.* 4, 93–111.

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.

Panyathanakun, V., Tantayanon, S., Tingsabhat, C., and Charmondusit, K. (2013). Development of eco-industrial estates in thailand: initiatives in the northern region community-based eco-industrial estate. *J. Clean. Prod.* 51, 71–79. doi: 10.1016/j.jclepro.2012.09.033

Petterson, S., Grøndahl-Rosado, R., Nilsen, V., Myrmel, M., and Robertson, L. J. (2015). Variability in the recovery of a virus concentration procedure in water: implications for QMRA. *Water Res.* 87, 79–86. doi: 10.1016/j.watres.2015.09.006

Piadeh, F., Moghaddam, M. R. A., and Mardan, S. (2014). Present situation of wastewater treatment in the Iranian industrial estates: recycle and reuse as a solution for achieving goals of eco-industrial parks. *Resour. Conserv. Recycl.* 92, 172–178. doi: 10.1016/j.resconrec.2014.06.004

Pollution Control Department (Pcd). (1994). Notification of the pollution control department entitled the surface water quality category in tha chin river B.E. 2537 (1994). *R. Gov. Gaz.* 111:14.

Powers, N., Wallgren, H. R., Marbach, S., and Turner, J. W. (2020). Relationship between rainfall, fecal pollution, antimicrobial resistance, and microbial diversity in an urbanized subtropical bay. *Appl. Environ. Microbiol.* 86, 1–15.

Purnell, S. E., Halliday, A., Newman, F., Sinclair, C., and Ebdon, J. E. (2020). Pathogen infection risk to recreational water users, associated with surface waters impacted by de facto and indirect potable reuse activities. *Sci. Total Environ.* 722:137799. doi: 10.1016/j.scitotenv.2020.137799

R Core Team. (2019). *R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria*. Available Online at: URL https//www.R-project.org/ (accessed March 30, 2019).

Sangkaew, W., Kongprajug, A., Chyerochana, N., Ahmed, W., Rattanakul, S., Denpetkul, T., et al. (2021). Performance of viral and bacterial genetic markers for sewage pollution tracking in tropical thailand. *Water Res.* 190:116706.

Seemuang, J. (2018). The changes of community's economy affected by the establishment of industrial estate: a case study of hat takuan-ao pradu fisheries community at map ta phut sub-district, muang district in rayong province (in thai). *J. Polit. Econ.* 5, 1–21.

Selvarajan, R., Sibanda, T., Venkatachalam, S., Kamika, I., and Nel, W. A. J. (2018). Industrial wastewaters harbor a unique diversity of bacterial communities revealed by high-throughput amplicon analysis. *Ann. Microbiol.* 68, 445–458. doi: 10.1007/s13213-018-1349-8

Sercu, B., Werfhorst, L. C., Van De Murray, J., and Holden, P. A. (2009). Storm drains are sources of human fecal pollution during dry weather in three urban southern california watersheds. *Environ. Sci. Technol.* 43, 293–298. doi: 10.1021/es801505p

Shah, I. H., Dong, L., and Park, H. S. (2020). Tracking urban sustainability transition: an eco-efficiency analysis on eco-industrial development in ulsan, korea. *J. Clean. Prod.* 262:121286. doi: 10.1016/j.jclepro.2020.121286

Siefring, S., Varma, M., Atikovic, E., Wymer, L., and Haugland, R. A. (2008). Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. *J. Water Health* 6, 225–237. doi: 10.2166/wh.2008.022

Simhon, A., Pileggi, V., Flemming, C. A., Lai, G., and Manoharan, M. (2020). Norovirus risk at a golf course irrigated with reclaimed water: should QMRA doses be adjusted for infectiousness? *Water Res.* 183:116121. doi: 10.1016/j.watres.2020.116121

Singkran, N. (2012). *Determining water patterns in Chakmak canal and Prodoo bay, Rayong province, Thailand. in Proceedings of the IASTED Asian Conference Modelling, Identification, and Control (AsiaMIC 2012)*. Phuket, Thailand: Mahidol University.

Sirikanchana, K., Patarapongsang, Y., Denpetchkul, T., Ratanakul, S., Kongprajug, A., Chyerochana, N., et al. (2020). *Final Report: Health Security Management in the Tourist Beach Environment (in Thai)*. Thailand: Thailand Science Research and Innovation.

Sirikanchana, K., Wangkahad, B., and Mongkolsuk, S. (2014). The capability of non-native strains of *Bacteroides* bacteria to detect bacteriophages as faecal indicators in a tropical area. *J. Appl. Microbiol.* 117, 1820–1829. doi: 10.1111/jam.12646

Sivaganesan, M., Haugland, R. A., Chern, E. C., and Shanks, O. C. (2010). Improved strategies and optimization of calibration models for real-time PCR absolute quantification. *Water Res.* 44, 4726–4735. doi: 10.1016/j.watres.2010.07.066

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative crassphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Staley, Z. R., Chuong, J. D., Hill, S. J., Grabuski, J., Shokralla, S., Hajibabaei, M., et al. (2018). Fecal source tracking and eDNA profiling in an urban creek following an extreme rain event. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-32680-z

Susur, E., Martin-Carrillo, D., Chiaroni, D., and Hidalgo, A. (2019). Unfolding eco-industrial parks through niche experimentation: insights from three italian cases. *J. Clean. Prod.* 239:118069. doi: 10.1016/j.jclepro.2019.118069

Tandukar, S., Sherchan, S. P., and Haramoto, E. (2020). Applicability of crAssphage, pepper mild mottle virus, and tobacco mosaic virus as indicators of reduction of enteric viruses during wastewater treatment. *Sci. Rep.* 10:3616. doi: 10.1038/s41598-020-60547-9

The Eastern Economic Corridor Policy Committee. (2019). Announcement of the eastern economic corridor policy committee on land use plan and plan for infrastructure development and utilities in the eastern economy corridor. *R. Gov. Gaz.* 136, 1–200.

Tian, L., Liang, Y., and Zhang, B. (2017). Measuring residential and industrial land use mix in the peri-urban areas of china. *Land Use Policy* 69, 427–438. doi: 10.1016/j.landusepol.2017.09.036

UNIDO. (2019). *Eco-industrial Parks - Achievements and Key Insights from the Global RECP programme 2012-2018*. Vienna, Austria: UNIDO.

US EPA. (1989). National primary drinking water regulations 40 CFR: filtration, disinfection, turbidity, giardia lamblia, viruses, legionella, and heterotrophic bacteria. final rule. *Fed. Regist.* 54:27486.

US EPA. (2002). *Method 1604: Total Coliforms and Escherichia coli in Water by Membrane Filtration Using a Simultaneous Detection Technique (MI Medium). EPA 821-R-02-024*. Available Online at: http://www.epa.gov/nerlcwww/1604sp02.pdf (accessed January 12, 2015).

US EPA. (2012). *Recreational Water Quality Criteria. Office of Water 820-F-12-058*. Washington, D.C: US EPA.

US EPA. (2016). Drinking water contaminant candidate list 4 - Final. *Fed. Regist.* 81, 81099–81114.

Wangkahad, B., Bosup, S., Mongkolsuk, S., and Sirikanchana, K. (2015). Occurrence of bacteriophages infecting *Aeromonas, Enterobacter*, and *Klebsiella* in water and association with contamination sources in Thailand. *J. Water Health* 13, 613–624. doi: 10.2166/wh.2014.204

World Health Organization. (2016). *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. Geneva, Switzerland: WHO.

Wu, H., Oun, A., Kline-robach, R., and Xagoraraki, I. (2018). Microbial pollution characterization at a TMDL site in Michigan: source identification. *J. Great Lakes Res.* 44, 412–420. doi: 10.1016/j.jglr.2018.02.006

Wu, Z., Greaves, J., Arp, L., Stone, D., and Bibby, K. (2020). Comparative fate of CrAssphage with culturable and molecular fecal pollution indicators during activated sludge wastewater treatment. *Environ. Int.* 136:105452. doi: 10.1016/j.envint.2019.105452

Yamsrual, S., Sasaki, N., Tsusaka, T. W., and Winijkul, E. (2019). Assessment of local perception on eco-industrial estate performances after 17 years of implementation in thailand. *Environ. Dev.* 32:100457. doi: 10.1016/j.envdev.2019.100457

Zagow, M. (2020). Does mixed-use development in the metropolis lead to less carbon emissions? *Urban Clim.* 34:100682. doi: 10.1016/j.uclim.2020.100682

Zhang, Q., Gallard, J., Wu, B., Harwood, V. J., Sadowsky, M. J., Hamilton, K. A., et al. (2019). Synergy between quantitative microbial source tracking (qMST) and quantitative microbial risk assessment (QMRA): a review and prospectus. *Environ. Int.* 130, 1–10. doi: 10.1016/j.envint.2019.03.051

Zimmer-Faust, A. G., Thulsiraj, V., Lee, C. M., Whitener, V., Rugh, M., Mendoza-Espinosa, L., et al. (2018). Multi-tiered approach utilizing microbial source tracking and human associated-IMS/ATP for surveillance of human fecal contamination in baja california. *Mexico. Sci. Total Environ.* 640–641, 475–484. doi: 10.1016/j.scitotenv.2018.05.172

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.

Copyright © 2021 Kongprajug, Denpetkul, Chyerochana, Mongkolsuk and Sirikanchana. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fmicb.2021.619495

Bacteriophages Are Good Estimators of Human Viruses Present in Water

*Elisenda Ballesté[1], Anicet R. Blanch[1], Javier Mendez[1], Laura Sala-Comorera[1], Leena Maunula[2], Silvia Monteiro[3], Andreas H. Farnleitner[4,5], Andreas Tiehm[6], Joan Jofre[1] and Cristina García-Aljaro[1]**

[1] Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Barcelona, Spain, [2] Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland, [3] Laboratório Analises, Instituto Superior Tecnico, Universidade Lisboa, Lisbon, Portugal, [4] Institute of Chemical, Environmental and Bioscience Engineering, Research Group Environmental Microbiology and Molecular Diagnostics 166/5/3, TU Wien, Vienna, Austria, [5] Research Division Water Quality and Health, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria, [6] Department of Microbiology and Molecular Biology, DVGW-Technologiezentrum Wasser, Karlsruhe, Germany

OPEN ACCESS

Edited by:
Eric Altermann, AgResearch Ltd., New Zealand

Reviewed by:
James Ebdon, University of Brighton, United Kingdom
James Lowther, Centre for Environment, Fisheries and Aquaculture Science (CEFAS), United Kingdom

***Correspondence:**
Cristina García-Aljaro
crgarcia@ub.edu

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *20 October 2020*
Accepted: *06 April 2021*
Published: *03 May 2021*

Citation:
Ballesté E, Blanch AR, Mendez J, Sala-Comorera L, Maunula L, Monteiro S, Farnleitner AH, Tiehm A, Jofre J and García-Aljaro C (2021) Bacteriophages Are Good Estimators of Human Viruses Present in Water. Front. Microbiol. 12:619495. doi: 10.3389/fmicb.2021.619495

The detection of fecal viral pathogens in water is hampered by their great variety and complex analysis. As traditional bacterial indicators are poor viral indicators, there is a need for alternative methods, such as the use of somatic coliphages, which have been included in water safety regulations in recent years. Some researchers have also recommended the use of reference viral pathogens such as noroviruses or other enteric viruses to improve the prediction of fecal viral pollution of human origin. In this work, phages previously tested in microbial source tracking studies were compared with norovirus and adenovirus for their suitability as indicators of human fecal viruses. The phages, namely those infecting human-associated *Bacteroides thetaiotaomicron* strain GA17 (GA17PH) and porcine-associated *Bacteroides* strain PG76 (PGPH), and the human-associated crAssphage marker (crAssPH), were evaluated in sewage samples and fecal mixtures obtained from different animals in five European countries, along with norovirus GI + GII (NoV) and human adenovirus (HAdV). GA17PH had an overall sensitivity of ≥83% and the highest specificity (>88%) for human pollution source detection. crAssPH showed the highest sensitivity (100%) and specificity (100%) in northern European countries but a much lower specificity in Spain and Portugal (10 and 30%, respectively), being detected in animal wastewater samples with a high concentration of fecal indicators. The correlations between GA17PH, crAssPH, or the sum of both (BACPH) and HAdV or NoV were higher than between the two human viruses, indicating that bacteriophages are feasible indicators of human viral pathogens of fecal origin and constitute a promising, easy to use and affordable alternative to human viruses for routine water safety monitoring.

Keywords: *Bacteroides*, bacteriophages, indicators, fecal pollution, crAssphage, enteric viruses, norovirus, adenovirus

INTRODUCTION

To date, over 150 human enteric viruses have been detected in water bodies (Rodríguez-Lázaro et al., 2012; Tran et al., 2015; Farkas et al., 2019), including adenoviruses, noroviruses, enteroviruses, sapoviruses, rotaviruses, and polyomaviruses, among others. As the assessment of all pathogenic viral strains in water is not feasible, indicators have been traditionally used as a proxy. The most used fecal indicators worldwide are Escherichia *coli* and enterococci, but they are less resistant to environmental stresses than viruses such as temperature, pH and sun irradiation and behave differently in the environment (Gerba et al., 1979; Wyer et al., 1995; Baggi et al., 2001; Borchardt et al., 2004; Harwood et al., 2005; Costán-Longares et al., 2008). As an alternative, the monitoring of reference viruses such as NoV and human adenovirus (HAdV) has been proposed (Hundesa et al., 2006; Ahmed et al., 2010; Rusiñol et al., 2014; Mayer et al., 2016). Enteroviruses are the most easily detectable enteric viruses by cell culture, and have been considered as potential indicators of the human enteric viruses (Grabow, 2007). Enteroviruses are recommended in the few water quality regulations that include virus-based criteria (EEC, 1976; USEPA, 1992). However, cell culture is expensive, time-consuming and needs specialized staff and equipment. Since the advent of genomic techniques, genome fragments of enteric viruses excreted by humans and animals have also been proposed as potential viral indicators (Hot et al., 2003; Albinana-Gimenez et al., 2009; Hunt et al., 2010; Silva et al., 2011; Wong et al., 2012). Additionally, viruses that exclusively infect humans, such as HAdV, have been postulated as microbial source tracking (MST) markers (Hundesa et al., 2006; Ahmed et al., 2010; Rusiñol et al., 2014).

However, human infectious viruses and human viral genomes both present serious drawbacks as indicators of waterborne enteric viruses. Firstly, the concentration of human viruses is variable in sewage and low in other types of water, requiring cumbersome, time-consuming and costly, concentration procedures (Hewitt et al., 2013; Haramoto et al., 2015). Secondly, there is no evidence that the presence and concentration of a given human virus in a water matrix unequivocally predicts the presence of other viruses (Irving and Smith, 1981; Chapron et al., 2000; Jiang et al., 2001). Finally, genomic methods have the inconvenience of not being able to distinguish between infectious and non-infectious viruses without additional unwieldy steps (Nuanualsuwan and Cliver, 2002; Jofre and Blanch, 2010). Such a distinction is essential when evaluating water treatments and natural inactivation and risk to human health.

Bacteriophages infecting bacteria have been proposed as efficient indicators of human viruses transmitted in water (IAWPRC, 1991; Grabow, 2004), and concurrence between enteroviruses and human-associated bacteriophages infecting *Bacteroides fragilis* has been demonstrated in a range of matrices including bivalve molluscs, treated wastewater, etc. (Tartera et al., 1988; Ebdon et al., 2007; Costán-Longares et al., 2008). Bacteriophages have been incorporated in recently developed guidelines for water quality monitoring (NHMRC, 2011; European Commission, 2018; Health Canada, 2019). In particular, somatic bacteriophages and F-RNA phages have proved to be reliable indicators of fecal viral pollution and suitable for use in water treatment processes (Jofre et al., 2016; Jebri et al., 2017). MST studies have used genogroups of F-specific RNA phages (Jofre et al., 2011) and bacteriophages infecting selected strains of *Bacteroides* (Tartera et al., 1989; Gómez-Dóñate et al., 2011; Jofre et al., 2014) to detect the source of fecal pollution in water environments. Recent studies have identified crAssphage (crAssPH) as the most abundant bacteriophage family in sewage, with *Bacteroides* as the putative host (Dutilh et al., 2014; Shkoporov et al., 2018), and its potential as a human-associated MST molecular marker is being evaluated (García-Aljaro et al., 2017; Ahmed et al., 2018; Ballesté et al., 2019). A study comparing the efficacy of human viruses with that of the molecular marker crAssPH and phages infecting the human-associated *Bacteroides* GA17 strain (GA17PH) to predict the origin of fecal pollution is therefore timely and of interest.

In this work, we determined the presence and abundance of genome fragments of HAdV, noroviruses GI and GII (NoV), crAssPH, and cultured phages infecting *Bacteroides thetaiotaomicron* strain GA17 (GA17PH), *B. fragilis* strain PG76 for porcine pollution (PGPH), and the sum of these two host-associated *Bacteroides* phages (BACPH) in human and animal wastewaters and fecal slurries in five European countries. The aim was to evaluate their sensitivity and specificity as MST markers for human pollution source detection and to assess the potential of the bacteriophages as indicators of the two human viruses (HAdV and NoV).

MATERIALS AND METHODS

Samples and Sampling Campaigns

A total of 120 sewage and wastewater samples and animal fecal slurries were collected from municipal wastewater treatment plants, abattoirs and farms in five different countries (Austria, Finland, Germany, Portugal, and Spain) from 2013 to 2014 within the framework of the European Project AQUAVALENS. These samples were of different origins: human (35), porcine (24), and other animals (61) [bovine (23), poultry (24), horse (7), dog (2), cat (2), goat (1), bird (1), and rabbit (1)]. The sewage samples came from communities with 2,100 to 4.0 million inhabitants. Wastewater was taken from abattoirs processing about 400 pigs, 8,000 ruminant animals and 100,000 poultry per day. Animal fecal slurries were obtained by mixing feces from at least 10 individual animals with sterile water. Details of each sample can be found in **Supplementary Material**. All the countries took samples of the different origins throughout the year to avoid any seasonality effect. They were collected in sterile containers and kept at 4°C while in transit to the laboratory. One hundred ml of each sample was sent to the other partner institutions at 4°C to perform the assigned parameter analysis (NoV, Finland; HAdV, Portugal; PGPH, crAssPH and GA17PH, Spain) and immediately processed.

Detection and Enumeration of MST Markers

Detection of Host-Associated *Bacteroides* Phages

Phages infecting human- and porcine-associated *Bacteroides* species were enumerated according to the ISO standard method 10705-4 (ISO, 2001). Plaque-forming units (PFU) of host-associated *Bacteroides* phages were enumerated by the double-agar-layer technique using the *B. thetaiotaomicron* strain GA17 to detect human pollution and *B. fragilis* strain PG76 for porcine pollution (Payan et al., 2005; Gómez-Dóñate et al., 2011). One ml of fresh samples was directly analyzed in triplicate with the corresponding host strain.

Detection of Human Viruses: Adenovirus and Norovirus

To detect HAdV, 200 μl of sewage samples were extracted directly using a QIAamp DNA Blood Mini Kit (Qiagen). HAdV sequences were amplified following a previously described protocol (Hernroth et al., 2002). Each run included the original sample, 10- and 100-fold dilutions of each sample, a standard curve and positive and negative controls.

Norovirus GI and GII (NoV) sequences were amplified following ISO/TS 15216-1 (ISO, 2013) with some modifications. Briefly, a sample volume of 250 μl was used for RNA extraction, which was performed using a NucliSens® Magnetic Extraction Kit and NucliSens® MiniMag® instrument (Biomerieux, Boxtel, Netherlands) according to the manufacturer's instructions. The sample was spiked with mengovirus to be used as a positive process control. Samples were amplified in a 20 μl-reaction using the QuantiTect Probe RT-PCR Kit (Qiagen, Hilden, Germany) and Rotorgene PCR cycler (Corbett) as described by Oristo et al. (2018), except using primer-probe sets as mentioned in ISO 15216: QNIF2 (FW) ATGTTCAGRTGGATGAGRTTCTCWGA, COG2R (REV) TCGACGCCATCTTCATTCACA, and QNIFs (PROBE) AGCACGTGGGAGGGCGATCG. Reverse transcrip tion- PCR runs included sample RNAs undiluted and diluted 1:10. For every set of samples, a negative extraction control, positive external RNA controls, and dilutions of purified plasmid dsDNA for construction of a standard curve were added.

Detection of Human-Associated crAssphage

DNA was extracted directly from 1 ml using the QIAamp DNA Blood Mini Kit (Qiagen). The abundance of crAssPH was analyzed with TaqMan Environmental Master Mix 2.0 (Applied Biosystems) using ABI StepOne Real-Time qPCR as previously described (García-Aljaro et al., 2017). DNA extraction controls were run together with the samples.

Statistical Analyses

Specificity and sensitivity of the different markers in detecting fecal pollution of known human or non-human origin were calculated as follows: specificity was defined as the proportion of negative samples in which the marker was not detected (true negatives/[true negatives + false positives]), whereas sensitivity was defined as the proportion of positive samples in which the marker was detected (true positives/[true positives + false negatives]). Spearman's correlation coefficients were used to study the relationship between the different markers. Traditional multidimensional scaling was carried out in R using the library stats (R Core Team, 2019). Receiver operating characteristic (ROC) curve analysis was performed using OptimalCutpoints R library (López-Ratón et al., 2014). ROC curve analysis was performed to define the cut-off levels for the predictors of NoV and HAdV. The criteria for obtaining the optimal cut-off point was based on the Youden index. The optimal cut-off point was defined as the point on the ROC curve nearest to the point where both the sensitivity and specificity were 1. Positive predictive values (PPV) for NoV and HAdV were calculated as the proportion of true positive results out of the number of samples with a positive result (true positives/[true positives + false positives]). Negative predictive values (NPV) were calculated as the proportion of true negative results out of the number of samples with a negative result (true negatives/[true negatives + false negatives]). The best cut-off points defined the thresholds, and thresholds were included in PPV and NPV calculations.

All the results were log transformed and the mean used for the calculations.

RESULTS

Bacteriophages as Human MST Trackers

Most samples of human origin tested positive for NoV, as did a similar proportion of porcine samples. Briefly, a total of 30 out of 35 human samples (85.7%) tested positive for norovirus by RT-PCR, compared to19 out of 24 porcine samples (79.2%) and only 2 out of 61 samples (3.3%) of other origins, including poultry and cow. The concentrations found in positive samples of different origin ranged from 4.05 to 6.01 $\log_{10}$ GC/10 ml (human), 3.20 to 6.04 $\log_{10}$ GC/10 ml (porcine), and 2.90 to 4.37 $\log_{10}$ GC/10 ml (other) (**Figure 1** and **Supplementary Table 1**). Important geographical differences were observed for the specificity and

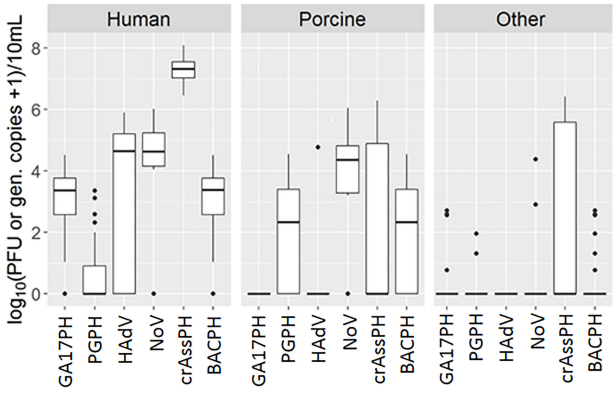


FIGURE 1 | Boxplot representation of GA17PH, phages infecting human-associated *Bacteroides thetaiotaomicron* strain GA17; PGPH, phages infecting porcine-associated *B. fragilis* strain PG76; HAdV, human Adenoviruses; NoV, GI + GII Noroviruses, crAssPH, crAssphage, BACPH, sum of GA17PH, and PGPH. Outliers are shown as black dots.

TABLE 1 | Specificity (sp) and sensitivity (se) of the different markers at the different geographical locations in detecting human pollution (GA17PH, crAssPH, HAdV, and NoV) and porcine pollution (PGPH).

	GA17PH		PGPH		crAssPH		HAdV		NoV	
	sp	se	sp	se	sp	se	sp	se	sp	se
Austria	1.00	0.88	0.79	0.00	1.00	1.00	1.00	0.88	1.00	1.00
Germany	0.93	1.00	0.93	0.00	1.00	1.00	0.93	0.60	0.71	0.80
Finland	1.00	0.86	0.94	1.00	1.00	1.00	1.00	0.71	0.82	1.00
Portugal	0.88	1.00	0.94	1.00	0.38	1.00	1.00	0.14	0.63	0.57
Spain	0.95	0.83	0.65	1.00	0.10	1.00	1.00	0.67	0.55	1.00

GA17PH, phages infecting human-associated Bacteroides thetaiotaomicron strain GA17; PGPH, phages infecting porcine-associated B. fragilis strain PG76; BACPH, sum of GA17PH and PGPH; crAssPH, crAssphage; HAdV, human Adenoviruses; and NoV, GI + GII Noroviruses.

sensitivity for human source detection. Whereas the NoV marker presented a sensitivity ≥80% for all the countries except Portugal, its specificity was low (71, 63, and 55% in Germany, Portugal, and Spain, respectively), mainly due to its detection in pig samples (**Table 1**); in contrast, specificity in Austria was 100%.

Human adenovirus was detected almost exclusively in samples of human origin, with the exception of 1 positive porcine sample out of 24. Nevertheless, only 20 out of 33 (60.6%) human samples were positive by qPCR (the remaining 2 samples up to 35 were not analyzed due to technical problems). The concentrations in the positive samples were high, ranging from 4.29 to 5.90 $\log_{10}$ GC/10 ml (**Figure 1** and **Supplementary Table 1**). Once again, geographical differences were observed. In this case, although the specificity of the HAdV marker for human source detection was very high in all the countries (≥90%), its sensitivity was low, being ≤80% in all the countries except Austria (**Table 1**).

The performance of two human-related bacteriophages (GA17PH and crAssPH) was also assessed in terms of specificity and sensitivity for human source detection. The GA17PH marker displayed a higher specificity and sensitivity for human sources than either NoV or HAdV. Accordingly, GA17PH was detected in most of the human samples (32 out of 35, 91.4%), not detected in any porcine sample, and rarely detected in the samples of mixed animal origin (4 out of 61). The concentration in the positive human source samples ranged from 1.00 to 4.52 $\log_{10}$ PFU/10 ml, whereas in the other animal wastewater samples it was 2 $\log_{10}$ units lower in those with a higher fecal load (**Figure 1** and **Supplementary Table 1**). No geographical differences were observed for GA17PH, with specificity ≥88% and sensitivity ≥83% for human source detection (**Table 1**). On the other hand, crAssPH was present at higher levels in human samples, which ranged from 6.45 to 8.10 $\log_{10}$ GC/10 ml, although it was also detected in 11 out of 24 samples of porcine origin and in 17 out of 61 samples of other animals (**Figure 1** and **Supplementary Table 1**). This marker exhibited a notable geographical variability in specificity, which was 100% for the Northern European countries, yet very low in Portugal and Spain (38 and 10%, respectively; **Table 1**).

As stated above, most false positives in the NoV assay were due to interference from NoV of porcine origin. These viruses may also infect humans and therefore, it could be advantageous to have a suite of markers able to predict the presence of NoV of both origins. In this study, the sum of the human-associated (GA17PH) and porcine-associated (PGPH) *Bacteroides* phages was assessed as an indicator for the presence of NoV of either human or porcine origin. It should be noted that the results of the PGPH marker showed considerable geographical variability (**Table 1**), with a sensitivity of 0% in Austria and Germany, where it was not detected, and a specificity of 65% in Spain. The concentration of PGPH in positive samples of wastewater and slurries ranged from 0.70 to 4.55 $\log_{10}$ PFU/10 ml.

Bacteriophages as Indicators of HAdV and NoV

A significant correlation was observed between GA17PH and HAdV ($p < 0.01$, Spearman correlation coefficient rho = 0.66), as well as between crAssPH and both human viruses ($p < 0.01$, rho = 0.55, and 0.54 for HAdV and NoV, respectively; **Figure 2**). The sum of the two *Bacteroides* markers (GA17PH + PGPH, indicated as BACPH) gave a higher correlation with NoV (rho = 0.7) but lower with HAdV (rho = 0.44). NoV correlated more strongly with BACPH than with HAdV (rho = 0.44), indicating a weak correlation between the viruses.

In view of these results, the concentration limits of the different bacteriophage markers that would be required to predict the presence of the enteric viruses were assessed statistically with the analysis of ROC curves (**Figure 3**). Based on the ROC curve analyses, the minimum concentration of crAssPH and GA17PH needed to predict HAdV was 1.00 and 1.54 $\log_{10}$ units, respectively, which was slightly higher than the respective 1.00 and 1.04 $\log_{10}$ units required to predict NoV. GA17PH showed a PPV of 63% and NPV of 99% for HAdV; for NoV a higher PPV of 79% and a lower NPV of 66% were obtained (**Table 2**). The sum of the markers PGPH and GA17PH (BACPH) increased the PPV and NPV for NoV to 81 and 86%, respectively, but gave a lower PPV of 42% for HAdV. In the case of crAssPH, a higher NPV of 79% was obtained for NoV, but the PPV was lower for both HAdV and NoV, being 33 and 67%, respectively. In summary, the highest PPV and NPV for HAdV were obtained with GA17PH (63 and 99%, respectively) followed by crAssPH (33 and98%, respectively), whereas for NoV the most acceptable PPV and NPV were obtained with BACPH (81 and 86%, respectively).

DISCUSSION

In the current work, we assessed the suitability of human viruses (NoV and HadV) as markers of fecal viral pollution

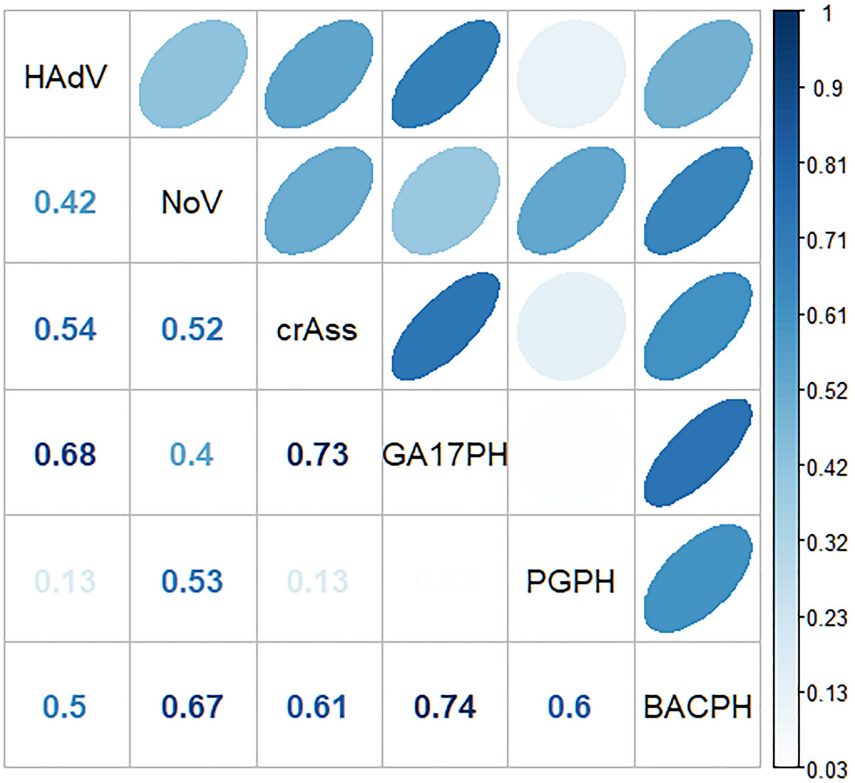


FIGURE 2 | Spearman's correlation coefficients between the different markers (GA17PH, phages infecting human-associated *Bacteroides thetaiotaomicron* strain GA17; PGPH, phages infecting porcine-associated *B. fragilis* strain PG76; BACPH, sum of GA17PH and PGPH; crAssPH, crAssphage; HAdv, human Adenoviruses; and NoV, GI + GII Noroviruses).

(Albinana-Gimenez et al., 2009; Rusiñol et al., 2014), and compared their performance with that of different source-associated *Bacteroides* phages (Gómez-Dóñate et al., 2011), including the recently reported human bacteriophage crAssPH (García-Aljaro et al., 2017; Stachler et al., 2017; Ballesté et al., 2019).

In this study, HAdV was detected in 60% of the human sewage samples, NoV in 80%, GA17PH in 91%, and crAssPH in 100%, indicating that the bacteriophage markers were more sensitive than the enteric viruses. Comparable with our results, studies in Japan and New Zealand did not detect NoV in 10–40% of raw sewage samples, and also report a variable detection (Wolf et al., 2010; Hewitt et al., 2013; Haramoto et al., 2015). On the other hand, high sensitivity and specificity for human source detection has been generally observed for HAdV (Albinana-Gimenez et al., 2009; Ahmed et al., 2010; Rodríguez-Lázaro et al., 2012; Hewitt et al., 2013; Rusiñol et al., 2014; Haramoto et al., 2015), leading to its proposal as a human MST marker. Here, although HAdV specificity was high, sensitivity was only 60%, with variability among geographical areas. Another study has also reported lower specificity in individual human feces or septic tanks and sewage, which could be attributed to the variable prevalence of HAdV in the population pool, and the variety of HAdV species associated with outbreaks (Wolf et al., 2010). Another explanation for the differences among studies could be the different volumes and concentrating methods used. In our study, HAdV was measured directly from a low volume (0.2 ml) of sewage, but as no concentration method was required, DNA loss due to concentration was avoided. It has to be pointed out that in our study GA17PH and PGPH were detected by culture-based methods, whereas NoV and AdV were detected by qPCR. In theory, qPCR analysis of the two *Bacteroides* phages would provide useful data, however, it was not possible to perform because there are not available qPCR methods for these phages since they are a very heterogeneous group and difficult to analyze them by this technique.

The arithmetic mean of NoV GI and GII combined in samples was 4.79 $\log_{10}$ GC/10 ml. Values reported in the literature are highly variable, ranging from 1 to 7 $\log_{10}$ GC/10 ml (Kitajima et al., 2012; Hewitt et al., 2013; Haramoto et al., 2015; Katayama and Vinjé, 2017). Such variability hampers NoV detection in the environment and its application as an MST marker, another disadvantage being its presence in pig feces in some areas in this study. The mean of HAdV in raw sewage (5.13 $\log_{10}$ GC/10 ml) was found to be higher than in other studies (about 3–4 $\log_{10}$ GC/1,000 ml; Hewitt et al., 2013; Rusiñol et al., 2014; Haramoto et al., 2015), although a considerable variability (2–7 $\log_{10}$ GC/10 ml) between geographical areas has also been observed (Allard and Vantarakis, 2017). As a marker, HAdV

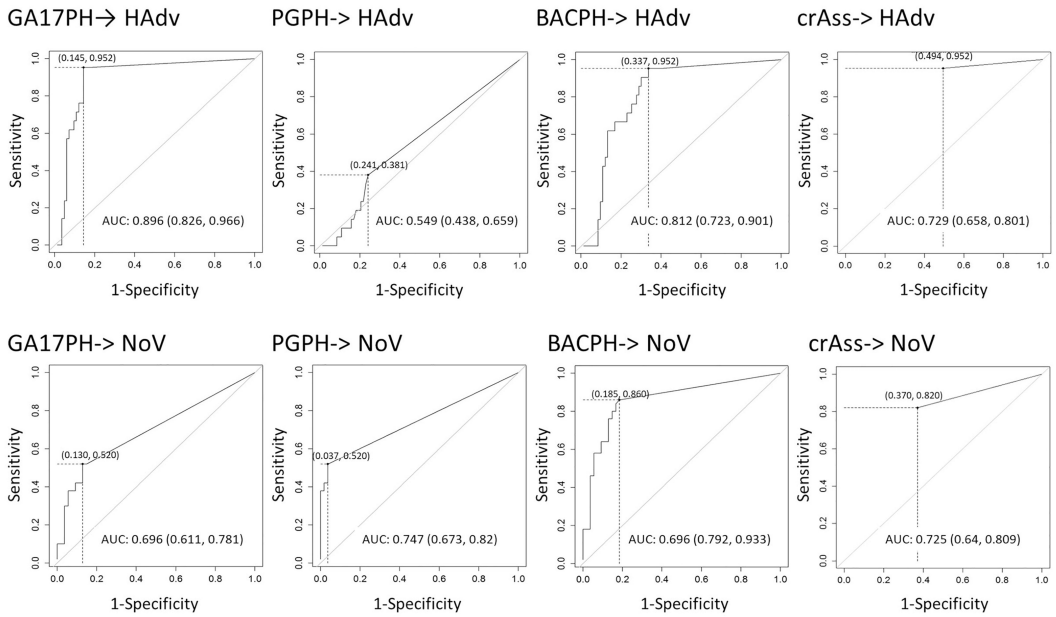


FIGURE 3 | ROC curve analysis with the prediction of HAdV and NoV by the different markers (GA17PH, phages infecting human-associated *Bacteroides thetaiotaomicron* strain GA17; PGPH, phages infecting porcine-associated *B. fragilis* strain PG76; BACPH, sum of GA17PH and PGPH; crAssPH, crAssphage; HAdv, human Adenoviruses; and NoV, GI + GII Noroviruses). AUC with the log of the minimal concentration (in brackets) needed to predict HAdV and NoV are indicated in the table below.

TABLE 2 | Predictive capacity of the different markers for HAdV and NoV.

Markers	GA17PH		PGPH		BACPH		crAssPH	
	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
HAdV	0.63	0.99	0.29	0.83	0.42	0.98	0.33	0.98
NoV	0.79	0.66	0.93	0.68	0.81	0.86	0.67	0.79

Positive predictive value (PPV), negative predictive value (NPV). GA17PH, phages infecting human-associated Bacteroides thetaiotaomicron strain GA17; PGPH, phages infecting porcine-associated B. fragilis strain PG76; BACPH, sum of GA17PH and PGPH; crAssPH, crAssphage; HAdV, human Adenoviruses; and NoV, GI + GII Noroviruses.

showed good specificity for human source detection, being found mainly in human fecal samples, but its low sensitivity limits its use for MST unless combined with more sensitive markers (Blanch et al., 2006). To overcome the limitations of viruses as MST markers, host-associated bacteriophages have been proposed as MST tools, such as human-associated *Bacteroides* strains GA17 and GB124, the animal-associated PG76 (Ebdon et al., 2007; Gómez-Dóñate et al., 2011), or crAssPH (García-Aljaro et al., 2017), recently related to human pollution. In this study, we assessed the capacity of human phages infecting the human *B. thetaiotaomicron* GA17 strain as well as the porcine strain PG76 to predict the presence of human viruses. The GA17PH marker displayed the highest sensitivity and specificity, giving values from 1 to 4.52 $\log_{10}$ PFU/10 ml, in accordance with previous reports of 1.69 to 5.84 PFU/100 ml in Spain, Sweden, France, United Kingdom, Tunisia, and Colombia (Payan et al., 2005; Blanch et al., 2006; Venegas et al., 2015; Yahya et al., 2015). CrAssPH was sensitive, although its specificity varied according to location. The concentration of crAssPH found in our study was around 10^7 GC/10 ml, which was 2.5 $\log_{10}$ units higher compared to NoV. These results are in accordance with those of Farkas and co-workers, who compared crAssPH with several enteric viruses (NoVGI, NoVGII, SaV, AdV, and JCV; Farkas et al., 2019). However, their crAssPH numbers only correlated with JCV and not with any of the human viruses analyzed this study. The low concentration of GA17PH in raw sewage may be a problem for the analysis of environmental samples where fecal pollution is diluted. However, as the concentration of crAssPH in river water is 2–3 log10 units higher than GA17PH, the two markers could be used in combination in certain geographical locations or in catchments not containing large numbers of pigs or abattoirs (Ballesté et al., 2019).

It should be noted that most of the false positives of NoV as an MST marker of human fecal contamination were due to the cross-reaction of the assay with NoV of porcine origin. A possible explanation is that NoV GII real-time RT-PCR primers and probes could detect porcine GII NoV strains, since primers and probes used in the RT-qPCR assay are highly similar to the sequences available in GeneBank for porcine NoV (LC509111, AB126329, HQ392821, and AY823305). In fact, strong homologies between swine NoVs genomic sequences and human primer sequences have been reported previously (L'Homme et al., 2009). Although human infections with porcine GII NoV have not been documented (Villabruna et al., 2019), they cannot be excluded. On the other hand, pigs may be susceptible to infection with human NoV strains, which have been detected in porcine feces (Mattison et al., 2007; Sisay et al., 2016), although how commonly this occurs is not known. Given the risk of these

viruses infecting humans, it would be of interest to have a suite of markers able to predict the presence of NoV of both origins (Mattison et al., 2007). In this case, the combination of GA17PH and PGPH (BACPH) could be used as an indicator for the presence of NoV of either human or porcine origin. Nevertheless, it should be noted that the PGPH marker showed geographical variability, as reported for other *Bacteroides* phages, and thus a *Bacteroides* pig-associated strain may be isolated in each targeted area (Payan et al., 2005) or other MST marker capable of identifying porcine sources such as Bacteroidales PG markers.

In summary, the results presented here show that bacteriophages infecting human-associated *Bacteroides* strains are a promising alternative for the prediction of human viruses of fecal origin as they outperform human viruses such as AdV and NoV, and provide an easier and more affordable technique for routine monitoring, avoiding the need to look for pathogenic viruses. Moreover, they also provide useful information on infectiousness/viability of the phages which is useful when establishing risk to human health.

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, LS-C, CG-A, SM, and LM participated in the sampling and carried out the experiments. JM performed the statistical analyses. SM, LM, AB, AF, and AT conceived and supervised the sampling experiments in the different laboratories. EB, JJ, and CG-A contributed to the data analysis and writing of the manuscript. All authors revised the manuscript.

FUNDING

This project was supported by the European Union AQUAVALENS project, European Union (FP7-KBBE.2012.2.5–01, grant agreement No.: 311846), and the Catalan government, Spain (2017 SGR 170). AF was additionally supported by the Austrian Science Fund (FWF) as part of the Vienna Doctoral Program on Water Resource Systems (W1219-N22) and the Project AQUASAFE (SC15-016) funded by the Niederösterreichische Forschungs- und Bildungsgesellschaft (NFB).

ACKNOWLEDGMENTS

The authors are very grateful to Satu Oristo and Kirsi Söderberg for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.619495/full#supplementary-material

REFERENCES

Ahmed, W., Goonetilleke, A., and Gardner, T. (2010). Human and bovine adenoviruses for the detection of source-specific fecal pollution in coastal waters in Australia. *Water Res.* 44, 4662–4673. doi: 10.1016/j.watres.2010.05.017

Ahmed, W., Lobos, A., Senkbeil, J., Peraud, J., Gallard, J., and Harwood, V. J. (2018). Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida. *Water Res.* 131, 142–150. doi: 10.1016/j.watres.2017.12.011

Albinana-Gimenez, N., Miagostovich, M. P., Calgua, B., Huguet, J. M., Matia, L., and Girones, R. (2009). Analysis of adenoviruses and polyomaviruses quantified by qPCR as indicators of water quality in source and drinking-water treatment plants. *Water Res.* 43, 2011–2019. doi: 10.1016/j.watres.2009.01.025

Allard, A., and Vantarakis, A. (2017). "Adenoviruses," in *Water and Sanitation for the 21st Century: Health and Microbiological Aspects of Excreta and Wastewater Management (Global Water Pathogen Project). (J.S Meschke, and R. Girones (eds), Part 3: Specific Excreted Pathogens: Environmental and Epidemiology Aspects - Section 1: Viruses) Michigan State University*, eds J. B. Rose and B. Jiménez-Cisneros (E. Lansing, MI: UNESCO). doi: 10.14321/waterpathogens.11

Baggi, F., Demarta, A., and Peduzzi, R. (2001). Persistence of viral pathogens and bacteriophages during sewage treatment: lack of correlation with indicator bacteria. *Res. Microbiol.* 152, 743–751. doi: 10.1016/S0923-2508(01)01255-4

Ballesté, E., Pascual-Benito, M., Martín-Díaz, J., Blanch, A. R., Lucena, F., Muniesa, M., et al. (2019). Dynamics of crAssphage as a human source tracking marker in potentially faecally polluted environments. *Water Res.* 155, 233–244. doi: 10.1016/j.watres.2019.02.042

Blanch, A. R., Belanche-Munoz, L., Bonjoch, X., Ebdon, J., Gantzer, C., Lucena, F., et al. (2006). Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Appl. Environ. Microbiol.* 72, 5915–5926. doi: 10.1128/aem.02453-05

Borchardt, M. A., Haas, N. L., and Hunt, R. J. (2004). Vulnerability of drinking-water wells in La Crosse, Wisconsin, to enteric-virus contamination from surface water contributions. *Appl. Environ. Microbiol.* 70, 5937–5946. doi: 10.1128/AEM.70.10.5937-5946.2004

Chapron, C. D., Ballester, N. A., Fontaine, J. H., Frades, C. N., and Margolin, A. B. (2000). Detection of astroviruses, enteroviruses, and adenovirus types 40 and 41 in surface waters collected and evaluated by the information collection rule and an integrated cell culture-nested PCR procedure. *Appl. Environ. Microbiol.* 66, 2520–2525. doi: 10.1128/AEM.66.6.2520-2525.2000

Costán-Longares, A., Montemayor, M., Payán, A., Méndez, J., Jofre, J., Mujeriego, R., et al. (2008). Microbial indicators and pathogens: removal, relationships and predictive capabilities in water reclamation facilities. *Water Res.* 42, 4439–4448. doi: 10.1016/j.watres.2008.07.037

EEC (1976). *The Council of European Economic Communities Directive of 8 December 1975 Concerning the Quality of Bathing Waters. Official J. the European Communities, Directive no. 76/160/EEC.* Brussels: Council of European Economic Communities.

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G. Z., Boling, L., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 37–43. doi: 10.1038/ncomms5498

Ebdon, J., Muniesa, M., and Taylor, H. (2007). The application of a recently isolated strain of *Bacteroides* (GB-124) to identify human sources of faecal pollution in a temperate river catchment. *Water Res.* 41, 3683–3690. doi: 10.1016/j.watres.2006.12.020

European Commission (2018). *Directive of the European Parliament and of the Council on the Quality of Water Intended for Human Consumption. COM/2017/0753 final – 2017/0332 (COD).* Brussels: European Commission.

Farkas, K., Adriaenssens, E. M., Walker, D. I., McDonald, J. E., Malham, S. K., and Jones, D. L. (2019). Critical evaluation of crAssphage as a molecular marker for human-derived wastewater contamination in the aquatic environment. *Food Environ. Virol.* 11, 113–119. doi: 10.1007/s12560-019-09369-1

García-Aljaro, C., Ballesté, E., Muniesa, M., and Jofre, J. (2017). Determination of crAssphage in water samples and applicability for tracking human faecal pollution. *Microb. Biotechnol.* 10, 1775–1780. doi: 10.1111/1751-7915.12841

Gerba, C. P., Goyal, S. M., LaBelle, R. L., Cech, I., and Bodgan, G. F. (1979). Failure of indicator bacteria to reflect the occurrence of enteroviruses in marine waters. *Am. J. Public Health* 69, 1116–1119. doi: 10.2105/AJPH.69.11.1116

Gómez-Dóñate, M., Payán, A., Cortés, I., Blanch, A. R., Lucena, F., Jofre, J., et al. (2011). Isolation of bacteriophage host strains of *Bacteroides* species suitable for tracking sources of animal faecal pollution in water. *Environ. Microbiol.* 13, 1622–1631. doi: 10.1111/j.1462-2920.2011.02474.x

Grabow, W. (2004). Bacteriophages: update on application as models for viruses in water. *Water SA* 27, 251–268. doi: 10.4314/wsa.v27i2.4999

Grabow, W. O. K. (2007). "Chapter 1 Overview of health-related water virology," in *Human Viruses in Water: Perspectives in Medical Virology*, ed. A. Bosch (Amsterdam: Elsevier B.V.), 1–25. doi: 10.1016/S0168-7069(07)17001-4

Haramoto, E., Fujino, S., and Otagiri, M. (2015). Distinct behaviors of infectious F-specific RNA coliphage genogroups at a wastewater treatment plant. *Sci. Total Environ.* 520, 32–38. doi: 10.1016/j.scitotenv.2015.03.034

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. doi: 10.1128/AEM.71.6.3163

Health Canada (2019). *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document – Enteric Viruses*. Ottawa, ON: Health Canada.

Hernroth, B. E., Conden-Hansson, A.-C., Rehnstam-Holm, A.-S., Girones, R., and Allard, A. K. (2002). Environmental factors influencing human viral pathogens and their potential indicator organisms in the Blue mussel, *Mytilus edulis*: the first Scandinavian report. *Appl. Environ. Microbiol.* 68, 4523–4533. doi: 10.1128/AEM.68.9.4523-4533.2002

Hewitt, J., Greening, G. E., Leonard, M., and Lewis, G. D. (2013). Evaluation of human adenovirus and human polyomavirus as indicators of human sewage contamination in the aquatic environment. *Water Res.* 47, 6750–6761. doi: 10.1016/j.watres.2013.09.001

Hot, D., Legeay, O., Jacques, J., Gantzer, C., Caudrelier, Y., Guyard, K., et al. (2003). Detection of somatic phages, infectious enteroviruses and enterovirus genomes as indicators of human enteric viral pollution in surface water. *Water Res.* 37, 4703–4710. doi: 10.1016/S0043-1354(03)00439-1

Hundesa, A., Maluquer de, M. C., Bofill-Mas, S., Binana-Gimenez, N., and Girones, R. (2006). Identification of human and animal adenoviruses and polyomaviruses for determination of sources of fecal contamination in the environment. *Appl. Environ. Microbiol.* 72, 7886–7893. doi: 10.1128/aem.01090-06

Hunt, R. J., Borchardt, M. A., Richards, K. D., and Spencer, S. K. (2010). Assessment of sewer source contamination of drinking water wells using tracers and human enteric viruses. *Environ. Sci. Technol.* 44, 7956–7963. doi: 10.1021/es100698m

IAWPRC (1991). Bacteriophages as model viruses in water quality control. *Water Res.* 25, 529–545. doi: 10.1016/0043-1354(91)90126-B

Irving, L. G., and Smith, F. A. (1981). One-year survey of enteroviruses, adenoviruses, and reoviruses isolated from effluent at an activated-sludge purification plant. *Appl. Environ. Microbiol.* 41, 51–59. doi: 10.1128/AEM.41.1.51-59.1981

ISO (2013). *Microbiology of Food and Animal Feed – Horizontal method for Determination of Hepatitis A Virus and Norovirus in Food Using Real-Time RT-PCR – Part 1: Method for Quantification*. Geneva: ISO.

ISO (2001). *ISO 10705-4: Water Quality – Detection and Enumeration of Bacteriophages. Part 4: Enumeration of Bacteriophages Infecting Bacteroides* fragilis. Geneva: ISO.

Jebri, S., Muniesa, M., and Jofre, J. (2017). "General and host-associated bacteriophage indicators of faecal pollution," in *Global Water Pathogen Project. http://www.waterpathogens.org (Farnleitner, A, Blanch, A. (eds) Part 2 Indicators and Microbial Source Tracking Markers) http://www.waterpathogens.org/book/coliphage Michigan State University*, eds J. B. Rose and B. Jiménez-Cisneros (E. Lansing, MI: UNESCO), doi: 10.14321/waterpathogens

Jiang, S., Noble, R., and Chu, W. (2001). Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. *Appl. Environ. Microbiol.* 67, 179–184. doi: 10.1128/AEM.67.1.179-184.2001

Jofre, J., Lucena, F., Blanch, A. R., and Muniesa, M. (2016). Coliphages as model organisms in the characterization and management of water resources. *Water (Switzerland)* 8:199. doi: 10.3390/w8050199

Jofre, J., and Blanch, A. R. (2010). Feasibility of methods based on nucleic acid amplification techniques to fulfil the requirements for microbiological analysis of water quality. *J. Appl. Microbiol.* 109, 1853–1867. doi: 10.1111/j.1365-2672.2010.04830.x

Jofre, J., Blanch, A. R., Lucena, F., and Muniesa, M. (2014). Bacteriophages infecting *Bacteroides* as a marker for microbial source tracking. *Water Res.* 55, 1–11. doi: 10.1016/j.watres.2014.02.006

Jofre, J., Stewart, J. R., and Grabow, W. (2011). "Phage methods," in *Microbial Source Tracking: Methods, Applications, and Case Studies*, eds C. Hagedorn, A. R. Blanch, and V. J. Harwood (New York, NY: Springer New York), 137–156. doi: 10.1007/978-1-4419-9386-1_6

Katayama, H., and Vinjé, J. (2017). "Norovirus and other caliciviurses," in *Global Water Pathogens Project. Michigan State University*, eds J. B. Rose and B. Jimenez-Cisneros (East Lansing, MI: UNESCO).

Kitajima, M., Haramoto, E., Phanuwan, C., Katayama, H., and Furumai, H. (2012). Molecular detection and genotyping of human noroviruses in influent and effluent water at a wastewater treatment plant in Japan. *J. Appl. Microbiol.* 112, 605–613. doi: 10.1111/j.1365-2672.2012.05231.x

L'Homme, Y., Sansregret, R., and Simard, C. (2009). Broad range RT-PCR assays targeting human noroviruses also detect swine noroviruses. *Food Microbiol.* 26, 552–555. doi: 10.1016/j.fm.2009.03.011

López-Ratón, M., Rodríguez-Álvarez, M. X., Suárez, C. C., and Sampedro, F. G. (2014). OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i08

Mattison, K., Shukla, A., Cook, A., Pollari, F., Friendship, R., Kelton, D., et al. (2007). Human noroviruses in swine and cattle. *Emerg. Infect. Dis.* 13, 1184–1188. doi: 10.3201/eid1308.070005

Mayer, R. E., Bofill-Mas, S., Egle, L., Reischer, G. H., Schade, M., Fernandez-Cassi, X., et al. (2016). Occurrence of human-associated *Bacteroidetes* genetic source tracking markers in raw and treated wastewater of municipal and domestic origin and comparison to standard and alternative indicators of faecal pollution. *Water Res.* 90, 265–276. doi: 10.1016/j.watres.2015.12.031

NHMRC (2011). *Australian Drinking water Guidelines Paper 6 National Water Quality Management Strategy*. Canberra, NSW: NHMRC.

Nuanualsuwan, S., and Cliver, D. O. (2002). Pretreatment to avoid positive RT-PCR results with inactivated viruses. *J. Virol. Methods* 104, 217–225. doi: 10.1016/S0166-0934(02)00089-7

Oristo, S., Lee, H.-J., and Maunula, L. (2018). Performance of pre-RT-qPCR treatments to discriminate infectious human rotaviruses and noroviruses from heat-inactivated viruses: applications of PMA/PMAxx, benzonase and RNase. *J. Appl. Microbiol.* 124, 1008–1016. doi: 10.1111/jam.13737

Payan, A., Ebdon, J., Taylor, H., Gantzer, C., Ottoson, J., Papageorgiou, G. T., et al. (2005). Method for isolation of *Bacteroides* bacteriophage host strains suitable for tracking sources of fecal pollution in water. *Appl. Environ. Microbiol.* 71, 5659–5662. doi: 10.1128/AEM.71.9.5659-5662.2005

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Rodríguez-Lázaro, D., Cook, N., Ruggeri, F. M., Sellwood, J., Nasser, A., Nascimento, M. S. J., et al. (2012). Virus hazards from food, water and other contaminated environments. *FEMS Microbiol. Rev.* 36, 786–814. doi: 10.1111/j.1574-6976.2011.00306.x

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

Shkoporov, A. N., Khokhlova, E. V., Fitzgerald, C. B., Stockdale, S. R., Draper, L. A., Ross, R. P., et al. (2018). ΦCrAss001 represents the most abundant

bacteriophage family in the human gut and infects *Bacteroides* intestinalis. *Nat. Commun.* 9:4781. doi: 10.1038/s41467-018-07225-7

Silva, H. D., García-Zapata, M. T. A., and Anunciação, C. E. (2011). Why the use of adenoviruses as water quality virologic marker? *Food Environ. Virol.* 3, 138–140. doi: 10.1007/s12560-011-9069-2

Sisay, Z., Djikeng, A., Berhe, N., Belay, G., Abegaz, W. E., Wang, Q. H., et al. (2016). First detection and molecular characterization of sapoviruses and noroviruses with zoonotic potential in swine in Ethiopia. *Arch. Virol.* 161, 2739–2747. doi: 10.1007/s00705-016-2974-9

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Tartera, C., Jofre, J., and Lucena, F. (1988). Relationship between numbers of enteroviruses and bacteriophages infecting bacterowes fragilis in different environmental samples. *Environ. Technol. Lett.* 9, 407–410. doi: 10.1080/09593338809384584

Tartera, C., Lucena, F., and Jofre, J. (1989). Human origin of *Bacteroides* fragilis bacteriophages present in the environment. *Appl. Environ. Microbiol.* 55, 2696–2701. doi: 10.1128/AEM.55.10.2696-2701.1989

Tran, N. H., Gin, K. Y.-H., and Ngo, H. H. (2015). Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. *Sci. Total Environ.* 538, 38–57. doi: 10.1016/j.scitotenv.2015.07.155

USEPA (1992). *Environmental Regulations and Technology: Control of Pathogens and Vector Attraction in Sewage Sludge Under 40 CFR Part 503.* Washington, DC: U.S. Environmental Protection Agency.

Venegas, C., Diez, H., Blanch, A. R., Jofre, J., and Campos, C. (2015). Microbial source markers assessment in the Bogotá River basin (Colombia). *J. Water Health* 13, 801–810. doi: 10.2166/wh.2015.240

Villabruna, N., Koopmans, M. P. G., and de Graaf, M. (2019). Animals as Reservoir for human norovirus. *Viruses* 11:478. doi: 10.3390/v11050478

Wolf, S., Hewitt, J., and Greening, G. E. (2010). Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. *Appl. Environ. Microbiol.* 76, 1388–1394. doi: 10.1128/AEM.02249-09

Wong, K., Fong, T.-T., Bibby, K., and Molina, M. (2012). Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environ. Int.* 45, 151–164. doi: 10.1016/j.envint.2012.02.009

Wyer, M. D., Fleisher, J. M., Gough, J., Kay, D., and Merrett, H. (1995). An investigation into parametric relationships between enterovirus and faecal indicator organisms in the coastal waters of England and Wales. *Water Res.* 29, 1863–1868. doi: 10.1016/0043-1354(95)00003-4

Yahya, M., Hmaied, F., Jebri, S., Jofre, J., and Hamdi, M. (2015). Bacteriophages as indicators of human and animal faecal contamination in raw and treated wastewaters from Tunisia. *J. Appl. Microbiol.* 118, 1217–1225. doi: 10.1111/jam.12774

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.

The reviewer, JL, declared a past collaboration with one of the authors, LM.

Copyright © 2021 Ballesté, Blanch, Mendez, Sala-Comorera, Maunula, Monteiro, Farnleitner, Tiehm, Jofre and García-Aljaro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fmicb.2021.659784

Human Pathogenic Bacteria Detected in Rainwater: Risk Assessment and Correlation to Microbial Source Tracking Markers and Traditional Indicators

*Julia K. Denissen[1], Brandon Reyneke[1], Monique Waso[2], Sehaam Khan[2] and Wesaal Khan[1]**

[1] Department of Microbiology, Faculty of Science, Stellenbosch University, Stellenbosch, South Africa, [2] Faculty of Health Sciences, University of Johannesburg, Doornfontein, South Africa

OPEN ACCESS

Edited by:
Georg H. Reischer,
Vienna University of Technology, Austria

Reviewed by:
Svetoslav Todorov,
University of São Paulo, Brazil
Jørgen J. Leisner,
University of Copenhagen, Denmark
Julia Derx,
Vienna University of Technology, Austria

***Correspondence:**
Wesaal Khan
wesaal@sun.ac.za

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 28 January 2021
Accepted: 09 April 2021
Published: 07 May 2021

Citation:
Denissen JK, Reyneke B, Waso M, Khan S and Khan W (2021) Human Pathogenic Bacteria Detected in Rainwater: Risk Assessment and Correlation to Microbial Source Tracking Markers and Traditional Indicators. Front. Microbiol. 12:659784. doi: 10.3389/fmicb.2021.659784

Roof-harvested rainwater (RHRW) was investigated for the presence of the human pathogenic bacteria *Mycobacterium tuberculosis* (*M. tuberculosis*), *Yersinia* spp. and *Listeria monocytogenes* (*L. monocytogenes*). While *Yersinia* spp. were detected in 92% (n = 25) of the RHRW samples, and *L. monocytogenes* and *M. tuberculosis* were detected in 100% (n = 25) of the samples, a significantly higher mean concentration (1.4×10^3 cells/100 mL) was recorded for *L. monocytogenes* over the sampling period. As the identification of appropriate water quality indicators is crucial to ensure access to safe water sources, correlation of the pathogens to traditional indicator organisms [*Escherichia coli* (*E. coli*) and *Enterococcus* spp.] and microbial source tracking (MST) markers (*Bacteroides* HF183, adenovirus and Lachnospiraceae) was conducted. A significant positive correlation was then recorded for *E. coli* versus *L. monocytogenes* (r = 0.6738; p = 0.000), and *Enterococcus* spp. versus the *Bacteroides* HF183 marker (r = 0.4071; p = 0.043), while a significant negative correlation was observed for *M. tuberculosis* versus the *Bacteroides* HF183 marker ($r = -0.4558$; $p = 0.022$). Quantitative microbial risk assessment indicated that the mean annual risk of infection posed by *L. monocytogenes* in the RHRW samples exceeded the annual infection risk benchmark limit (1×10^{-4} infections per person per year) for intentional drinking ($\sim 10^{-4}$). In comparison, the mean annual risk of infection posed by *E. coli* was exceeded for intentional drinking ($\sim 10^{-1}$), accidental consumption ($\sim 10^{-3}$) and cleaning of the home ($\sim 10^{-3}$). However, while the risk posed by *M. tuberculosis* for the two relevant exposure scenarios [garden hosing ($\sim 10^{-5}$) and washing laundry by hand ($\sim 10^{-5}$)] was below the benchmark limit, the risk posed by adenovirus for garden hosing ($\sim 10^{-3}$) and washing laundry by hand ($\sim 10^{-3}$) exceeded the benchmark limit. Thus, while the correlation analysis confirms that traditional indicators and MST

markers should be used in combination to accurately monitor the pathogen-associated risk linked to the utilisation of RHRW, the integration of QMRA offers a more site-specific approach to monitor and estimate the human health risks associated with the use of RHRW.

Keywords: microbial source tracking markers, human pathogenic bacteria, traditional indicator organisms, QMRA, rainwater

INTRODUCTION

A conservative estimate predicts that 4% of deaths worldwide and 5.7% of the global burden of disease in disability-adjusted life years (DALYs) could be attributed to water, sanitation and hygiene (WASH) related infectious diseases (Yang et al., 2012). The primary aim of Sustainable Development Goal 6 (SDG6) is thus to ensure that the world population has access to safe and affordable water and adequate sanitation services by 2030 (United Nations, 2018). In an effort to achieve this aim, roof-harvested rainwater (RHRW) is being investigated and applied as an alternative, supplementary water source in many countries around the world, including South Africa (Ahmed et al., 2008b; De Kwaadsteniet et al., 2013). However, while RHRW may be used to augment current water supplies (Gould and Nissen-Petersen, 1999), rainwater is exposed to various contamination sources as it traverses the air (e.g., bioaerosols) and during the harvesting process (e.g., debris and animal faecal matter on the catchment surface) (Hamilton et al., 2019).

It is thus well documented that the microbiological quality of RHRW is sub-standard and numerous research groups have reported on the detection of traditional indicator organisms (using culture-based analysis) such as total coliforms, faecal coliforms, *Escherichia coli* (*E. coli*) and enterococci species in rainwater (Kaushik et al., 2014; Waso et al., 2018; Hamilton et al., 2019). While the analysis of the indicator groups is routine in water quality monitoring, amongst other pitfalls, a poor correlation has been recorded between traditional indicator organisms and potentially pathogenic microorganisms (Harwood et al., 2005, 2014; Field and Samadpour, 2007). Researchers are thus investigating the use of microbial source tracking (MST) markers to monitor and detect faecal contamination within environmental water samples (Ahmed et al., 2016). Microbial source tracking methods have greatly improved the capacity to detect microorganisms that are host-specific to animals, that occur in water and sediments and unlike faecal indicator bacteria, MST markers are able to differentiate between several sources of faecal contamination (Bradshaw et al., 2016). A few of the common MST markers include the *Enterococcus esp* gene, enterovirus, *Bifidobacterium* spp., human-specific *Bacteroides* HF183, human adenovirus and polyomavirus. Ideal characteristics of these MST markers include: specificity to the target host-group; the marker should be geographically and temporally stable in the target host-group, and; the decay rates of the markers and pathogens present in the relevant water sources should correlate (Ahmed et al., 2015). Moreover, advances in water quality monitoring techniques such as molecular viability and whole community analysis, renders culture-based analyses superfluous and ensures the accurate detection and quantification of MST markers and waterborne pathogens. Using quantitative PCR (qPCR), Savichtcheva et al. (2007) observed a significant positive correlation between the human *Bacteroides* MST marker versus *Salmonella*, whilst Viau and Boehm (2011) observed a significant correlation between the human *Bacteroides* MST marker and *Leptospira*.

Using molecular detection methods, numerous studies have also identified a variety of opportunistic and pathogenic microorganisms in RHRW. These frequently detected microorganisms include *Legionella pneumophila*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* and *Salmonella* spp. (Hamilton et al., 2019). Ahmed et al. (2017); Strauss et al. (2019) and Reyneke et al. (2020) then employed Illumina amplicon-based sequencing and ethidium monoazide bromide (EMA)-Illumina analysis, to investigate the whole bacterial community in RHRW. Correspondingly, many of the frequently detected bacterial genera such as *Legionella*, *Pseudomonas* etc. formed part of this indigenous or core microbial rainwater group. However, a high frequency of detection percentage was also obtained for genera such as *Mycobacterium*. This genus includes pathogens known to cause serious disease in humans, which is concerning as the utilisation of RHRW as an alternative water source may pose a health risk to the end-user communities. *Mycobacterium tuberculosis* (*M. tuberculosis*) was thus predominantly focused on in the current study and as *Yersinia* spp. and *Listeria monocytogenes* (*L. monocytogenes*) have previously been detected in rainwater, these bacterial pathogens were also included (Dobrowsky et al., 2014; Jongman and Korsten, 2016).

The Gram-negative coccobacillus *Yersinia* genus forms part of the *Yersiniaceae* family and is comprised of 11 species, three of which are pathogenic to humans: *Y. pseudotuberculosis*, *Y. enterocolitica* and *Y. pestis*. *Y. pestis* is responsible for causing three different forms of the highly infectious disease known as the plague (pneumonic, bubonic and septicaemic) (Pechous et al., 2016). The plague is transmitted by fleas to various hosts by blood feeding, or by regurgitative transmission of the bacteria (once it has grown in the form of a cohesive biofilm) from the flea host into the bite of the receiving organism and can persist for extended periods of time at low levels in enzootic cycles (Eisen et al., 2006; Hinnebusch et al., 2017; Bosio et al., 2020). Although limited research is available on the presence of *Yersinia* spp. in RHRW, Dobrowsky et al. (2014) identified the *Yersinia* genus in rainwater using conventional PCR analysis. There is thus a high probability that these species are present in RHRW systems as bird and rodent faecal matter are often detected on roofing systems and might wash into the harvesting tanks during

a rain event. Similarly, *M. tuberculosis*, which causes tuberculosis, can survive and adapt to hostile or extreme environmental conditions (Cook et al., 2009). It is also capable of adapting to a variety of intracellular human systems such as dendritic cells and macrophages (Cook et al., 2009). However, despite its poor geographical characterisation in terms of abundance in various environments (King et al., 2017), as previously indicated, using Illumina and EMA-Illumina whole-community analysis, the *Mycobacterium* genus was identified as one of the primary frequently detected genera in rainwater (Ahmed et al., 2017; Strauss et al., 2019; Reyneke et al., 2020). This is a matter of serious concern as 360 000 people were diagnosed with tuberculosis in South Africa in 2019 [World Health Organisation (WHO), 2019] and 781 tuberculosis cases are reported amongst every 100 000 individuals each year (Tadokera et al., 2020). *L. monocytogenes* is a Gram-positive, psychotropic bacteria that has been detected in the environment and typically occurs in most raw foods. It is responsible for causing listeriosis and from 2017 to 2018, *L. monocytogenes* sequence type 6 was associated with a listeriosis outbreak in South Africa, that was described by the WHO as the biggest outbreak ever recorded worldwide (Smith et al., 2019). Jongman and Korsten. (2016) then detected *L. monocytogenes* (using selective culture-based analysis combined with matrix-assisted laser desorption/ionisation time of flight mass spectrometry) in RHRW samples collected from three rural South African villages which rely on RHRW as an alternative water source.

Roof-harvested rainwater thus has the potential to expose vulnerable end-user communities to a myriad of microbial pathogens and opportunistic pathogens (Low, 2002). Quantitative microbial risk assessment (QMRA) can therefore be implemented as a health risk assessment tool and is a technique that has been used for more than two decades to estimate the pathogen-associated risk in drinking water (Pecson et al., 2017). This technique is comprised of four stages, namely (i) hazard identification, (ii) exposure assessment, (iii) dose-response modelling and (iv) risk characterisation. The combination of these stages allows for the construction of a prediction-based analysis model that elucidates the potential health risk associated with specific pathogens based on exposure scenarios associated with a particular environment or activity (Owens et al., 2020). Moreover, using a QMRA framework, it is envisaged that the gap between the level of pathogens in a water source and the treatment required to effectively reduce pathogen-associated risk, can be bridged (Sano et al., 2019).

The primary aim of this study was thus to explore a consortium of RHRW tanks located in a sustainable housing project in Kleinmond, South Africa for the human pathogenic bacterial species *M. tuberculosis*, *Yersinia* spp. and *L. monocytogenes*. Additionally, the presence of the human bacterial pathogens (*M. tuberculosis*, *Yersinia* spp. and *L. monocytogenes*) was correlated to the presence of traditional indicator organisms (i.e., *E. coli* and *Enterococcus* spp.) and MST markers (i.e., *Bacteroides* HF183, adenovirus and Lachnospiraceae). The *Bacteroides* HF183 marker was selected as the HF183 primer set occurs in all *Bacteroides* strains of human origin. This marker thus has high specificity for the detection of human faecal matter and sewage in environmental waters (Harwood et al., 2014). Similarly, both Lachnospiraceae and adenovirus have been investigated as indicators of sewage and faecal pollution in environmental waters and could therefore provide valuable information on the health risks associated with the use of various water sources (Newton et al., 2011; McLellan et al., 2013; Sidhu et al., 2013; Rusiñol et al., 2016; Waso et al., 2018). Quantitative PCR analyses were used to identify and quantify the target bacterial pathogens, MST markers and indicator groups, whereafter the health risk associated with the utilisation of the RHRW for potable and various domestic activities (e.g., cleaning the house, laundry, etc.), in the target community, was determined.

MATERIALS AND METHODS

Sampling Site

The Kleinmond Housing Scheme, Western Cape, South Africa (GPS co-ordinates: 34°20.11′ 81″ S; 19° 00.59′ 74″ E), was used as the sampling site. The sustainable housing project was conceptualised in 2007 in a collaboration between the Overstrand Local Municipality, the Council for Scientific and Industrial Research, the Department of Science and Technology and the Western Cape Provincial Department of Human Settlements. Within this community are 411 houses (40 m^2 each), each fitted with a 2000 L aboveground polyethylene rainwater harvesting tank. A random sampling technique was implemented to select five houses, designated A to E (with functioning rainwater harvesting tanks installed), from a collection of houses used in the Dobrowksy et al. (2014) study. Sampling was conducted once a week for five consecutive weeks (August to September 2020), with 5 L of rainwater collected from each RHRW tank (n = 25) using sterile polypropylene bottles as previously described by Waso et al. (2018).

The temperature of each sample was measured on-site using a hand-held mercury thermometer; the pH, total dissolved solids, and electrical conductivity were determined using a hand-held Milwaukee Instruments MI806 meter, and the dissolved oxygen was measured using a hand-held Milwaukee Instruments M600 meter (Spraytech, South Africa). All physico-chemical parameters are outlined in **Supplementary Table 1** (**Supplementary Material**). The daily ambient temperature and rainfall data for the duration of the sampling period was obtained from the South African Weather Services (Pretoria, South Africa). A visual representation of the daily ambient temperature and rainfall data is shown in **Supplementary Figure 1** (**Supplementary Material**).

Rainwater Concentration, EMA Treatment and DNA Extraction for the Detection of Target Pathogens and Indicator Organisms

One litre of each RHRW sample (n = 25) was subjected to flocculation as previously described by Dobrowsky et al. (2015). The flocculated samples were filtered through non-charged

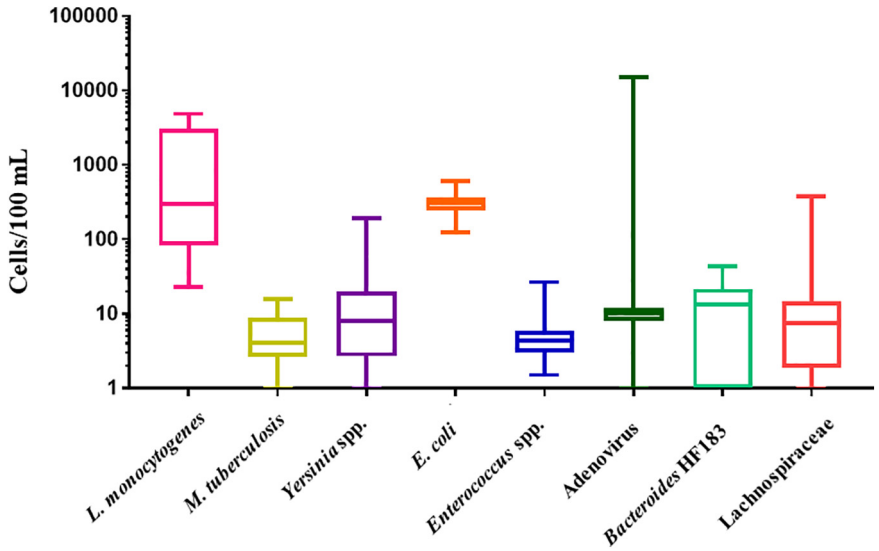


FIGURE 1 | Box and whiskers plot of the concentration (cells/100 mL) for *L. monocytogenes*, *M. tuberculosis*, *Yersinia* spp., *E. coli*, *Enterococcus* spp., adenovirus, *Bacteroides* HF183, and Lachnospiraceae. The whiskers illustrate the minimum and maximum, the outer box illustrates the 1st and 3rd quartiles, and the inner line illustrates the median.

mixed ester membrane filters with a pore size of 0.45 μM (Merck, Millipore, Billerica, MA, United States) and each filter was placed in a 9 cm petri dish containing 1.5 mL citrate buffer (0.3 M, pH 3.5) (Saarchem, Durban, South Africa) and gently agitated using an orbital platform shaker to remove the cells from the filter. Thereafter, each 1.5 mL concentrated sample was centrifuged at 16,000 × *g* for 5 min and the resulting pellet was subjected to 6 μM EMA (Biotium, Hayward, CA, United States) treatment as previously described by Reyneke et al. (2017). Ethidium monoazide bromide intercalates with the DNA of cells with compromised membranes or with extracellular DNA. Upon photoactivation, EMA covalently binds to the DNA, which inhibits their amplification in quantification assays (e.g., qPCR) (Emerson et al., 2017). Treatment with EMA was thus done so that the detected gene copies for the pathogens and indicator organisms could be converted into viable cells (may be detected using culture-based analysis) for use in the QMRA analysis. Following EMA treatment, the supernatant containing residual EMA was removed and total genomic DNA extractions (i.e., DNA from intact and presumed viable cells) were performed on the remaining pellet for each of the RHRW samples (n = 25) using the *Quick*-DNA™ Fecal/Soil Microbe Miniprep Kit (Zymo Research, Irvine, CA, United States) according to the manufacturer's instructions.

Rainwater Concentration and DNA Extraction for the Detection of MST Markers

Preliminary comparative analysis of EMA-qPCR (intact and presumed viable cells) versus qPCR (whole or total DNA, without EMA), indicated that the EMA treatment may have influenced the detection of the MST markers in the RHRW. Therefore, conventional qPCR was used for the detection of the MST markers (i.e., *Bacteroides* HF183, Lachnospiraceae, adenovirus) within the RHRW samples. One litre of each RHRW sample (n = 25) was concentrated as described in Section "Rainwater Concentration, EMA Treatment and DNA Extraction for the Detection of Target Pathogens and Indicator Organisms." Thereafter, each 1.5 mL concentrated sample was centrifuged at 16,000 × *g* for 5 min and total genomic DNA extractions were performed on the remaining pellet for each of the RHRW samples (n = 25) using the *Quick*-DNA Fecal/Soil Microbe Miniprep Kit (Zymo Research) according to the manufacturer's instructions.

Quantitative PCR Analyses

All qPCR analyses were conducted using a LightCycler®96 Instrument (Roche Diagnostics, Mannheim, Germany) and the FastStart Essential DNA Green Master/FastStart Essential DNA Probes Master (Roche Diagnostics) in order to quantify gene copies of the target pathogens *M. tuberculosis*, *Yersinia* spp. and *L. monocytogenes*, as well as the MST markers *Bacteroides* HF183, adenovirus and Lachnospiraceae, and the indicator organisms *E. coli* and *Enterococcus* spp. in the RHRW samples. The primers and cycling parameters for each target organism are outlined in **Table 1**. Each qPCR assay was performed in duplicate. The reaction mixture (final volume of 20 μL) for all the qPCR assays, except the Lachnospiraceae assay, consisted of 10 μL FastStart Essential DNA Green Master (1X), 0.4 μL of the forward and reverse primers (0.2 μM) and 5 μL template DNA. For the Lachnospiraceae assay, the reaction mixture consisted of 10 μL (1X) FastStart Essential DNA Probes Master, 2 μL primer-probe mixture [1 μM of each primer and 0.08 μM of the probe], 3 μL PCR-grade water and 5 μL template DNA. All DNA samples were diluted 10-fold before analysis with the respective qPCR assays in order to minimise PCR inhibitors (Reyneke et al., 2017).

For each qPCR reaction, a negative control of sterile milliQ was included in the analysis, while melt curve analysis was included for all SYBR® Green qPCR assays to verify the specificity of the primer sets (temperature increase from 65 to 97 °C at 0.2°C/s and continuous fluorescent signal acquisition at 5 readings/°C) (Reyneke et al., 2017).

Standard curves for the qPCR assays were generated using the methodology outlined in Reyneke et al. (2017). Briefly, conventional PCR assays (**Table 1**) were performed to amplify the respective target genes using positive control DNA [*E. coli* ATCC 13706, *Enterococcus faecalis* (clinical isolate), lyophilised adenovirus (Coris Bioconcept, Gembloux, Belgium), *L. monocytogenes* ATCC 13932 and *Yersinia enterocolitica* subsp. *enterocolitica* ATCC 27729] or DNA extracted from an influent sewage sample collected from a local wastewater treatment plant (*M. tuberculosis,* Lachnospiraceae and *Bacteroides* HF183) (**Supplementary Table 2**, **Supplementary Material**). A negative control of sterile milliQ was utilised for each PCR reaction.

A standard curve was generated by preparing serial 10-fold dilutions (10^9 to 10^0 gene copies/μL) of the PCR products. The lower limit of detection (LLOD) for each qPCR assay was reported as the lowest number of gene copies that was consistently detected in the standard curve (Dobrowsky et al., 2016). The Roche LightCycler®96 Software version 1.1 was utilised for the analysis of the qPCR performance characteristics of the assays.

Quantitative Microbial Risk Assessment

Hazard Identification and Quantification of Target Pathogens

Based on the detection frequency of the target pathogens, indicators and MST markers obtained using qPCR analysis (Section "Quantitative PCR Analyses") and the availability of applicable dose-response models, *E. coli* (representative traditional indicator organism), adenovirus (representative MST marker), *L. monocytogenes* and *M. tuberculosis* were selected as the target organisms for the health risk assessment of the RHRW. For use in the QMRA analyses, the detected gene copies (Section "Quantitative PCR Analyses") were converted to gene copies/100 mL of the original RHRW sample as outlined by Waso et al. (2018). The gene copies/100 mL were then converted to cell equivalents (cells/100 mL) by utilising the number of copies of the target gene present within the host (**Table 1**). All final concentrations for the EMA-qPCR (intact and presumed viable cells) and conventional qPCR (whole or total DNA) analyses are thus presented as equivalent cells/100 mL original RHRW sample.

Exposure Assessment

The major exposure routes associated with the use of RHRW for several domestic activities in the Kleinmond Housing Scheme site were identified by consulting social survey data reported by Dobrowksy et al. (2014). These activities include washing laundry by hand, cleaning of the home, garden hosing, garden work, washing/bathing, intentional drinking and accidental consumption. The various exposure scenarios (which includes the exposure volume and frequency of occurrence) that were evaluated in the present study are outlined in **Supplementary Table 3** (**Supplementary Material**).

The formulae used for the calculation of ingestion/inhalation dose, as well as descriptions of the various exposure routes are outlined in Reyneke et al. (2020). The concentration of each respective target organism was obtained from the results of the qPCR analyses. Based on the detection frequency of pathogenic *E. coli* in RHRW from previous studies, the fraction of *E. coli* presumed to be human infectious was set at 0.005–0.1 (Reyneke et al., 2020). Additionally, the fraction of detected adenovirus, *L. monocytogenes* and *M. tuberculosis* assumed to be infectious to humans were 5.88×10^{-4}, 1.00 and 0.66–1.00, respectively (Buchanan et al., 1997; World Health Organisation, 2004; Lyautey et al., 2007; Schijven et al., 2019).

Dose Response

Dose-response models are a set of mathematical expressions which illustrate the probability that an individual will experience an adverse health effect (e.g., infection, death) following exposure to an infectious organism. These models are specifically fitted to the adverse health effects observed in animals or humans that have been exposed to varying doses of infectious microorganisms (Haas et al., 1999; Jones and Su, 2015). Two of the most commonly used dose-response models are the exponential and beta-poisson models. The exponential model assumes that the probability of an organism causing infection is independent of organism dosage, whereas the beta-poisson model assumes that infectivity of an organism is dependent on dose (Buchanan et al., 2000).The exponential dose-response model (Eq. 1) was used to calculate the risk of infection linked to the presence of pathogenic adenovirus (inhalation exposure), *M. tuberculosis* (inhalation exposure) and *L. monocytogenes* (ingestion exposure) within the RHRW samples for various exposure scenarios (Couch et al., 1966; Buchanan et al., 1997; Jones et al., 2009):

$$P_{\text{inf}} = 1 - \exp^{(-kd)} \tag{1}$$

where P_{inf} is the probability of infection following a single exposure, k is the parameter which describes the probability of a pathogen surviving the host defence to initiate infection and d is the dose of microorganisms (number of microorganisms that is inhaled/ingested).

The beta-Poisson dose-response model (Eq. 2) was then employed to calculate the risk of infection linked to the presence of enteroinvasive *E. coli* (ingestion exposure) within the RHRW samples for various exposure scenarios (Haas et al., 1999; Ryan et al., 2014):

$$P_{inf} = 1 - \left(1 + \frac{d}{N50}\left(2^{\frac{1}{\alpha}} - 1\right)\right)^{-\alpha} \tag{2}$$

where P_{inf} is the probability of getting infected following a single exposure event, d is the dosage of microorganisms (number of microorganisms that is ingested), α is a shape factor and N_{50} is the median infective dosage. All parameters associated with the dose-response models are outlined in **Table 2**.

TABLE 1 | Conventional PCR and qPCR primers, cycling parameters and PCR product size of the organisms screened for in the RHRW samples.

Organism	Primers	Primer sequence (5′-3′)	[Primer] for conventional PCR	Conventional PCR cycling parameters	qPCR cycling parameters	Target gene (bp)	Mean copies per cell	References
Mycobacterium tuberculosis	MTP t8-F MTP t9-R	GTGCGGATGGTCGCAGAGAT CTCGATGCCCTCACGGTTCA	0.2 μM	3 min at 95°C; 40 cycles of 94°C for 1.5 min, 65°C for 2 min, 72°C for 3 min	3 min at 95°C; 40 cycles of 94°C for 1.5 min, 65°C for 2 min, 72°C for 3 min	*orfA* (541)	15[a]	Kox et al., 1994
Yersinia spp.	227Fmod 669R	GTCTGGGCTTTGCTGGTC GCGTCGTATTTAGCACCAACG	0.8 μM	5 min at 95°C; 40 cycles of 94°C for 20 s, 60°C for 20 s, 72°C for 15 s	5 min at 95°C; 40 cycles of 94°C for 20 s, 60°C for 20 s, 72°C for 15 s	*ompF* (465)	1[b]	Stenkova et al., 2008
Listeria monocytogenes	LIS-F LIS-R	TCATCGACGGCAACCTCGG TGAGCAACGTATCCTCCAGAGT	0.3 μM	7 min at 95°C; 40 cycles of 50 s at 95°C, 40 s at 54°C, 50 s at 72°C, 5 min at 72°C	7 min at 95°C; 40 cycles of 95°C for 50 s, 54°C for 40 s, 72°C for 50 s	*prfA* (217)	2[b]	Germini et al., 2009
Bacteroides HF183	HF183F HF183R	ATCATGAGTTCACATGTCCG TACCCCGCCTACTATCTAATG	0.25 μM	95°C for 4 min; 40 cycles of 95°C for 30 s, 53°C for 1 min, 72°C for 2 min; final elongation at 72°C for 10 min	95°C for 10 min; 40 cycles of 95°C for 30 s, 53°C for 1 min, 60°C for 1 min	16 S rRNA (86)	7[c]	Seurinck et al., 2005
E. coli	784F 866R	GTGTGATATCTACCCGCTTCGC AGAACGGTTTGTGGTTAATCAGGA	0.5 μM	95°C for 10 min; 50 cycles of 95°C for 15 s, 60°C for 1 min, final elongation at 72°C for 10 min	95°C for 10 min; 50 cycles of 95°C for 15 s, 60°C for 1 min	*uidA* (80)	1[d]	Frahm and Obst, 2003
Enterococcus spp.	ECST784F ENC854R	AGAAATTCCAAACGAACTTG CAGTGCTCTACCTCCATCATT	0.5 μM	95°C for 10 min; 50 cycles of 95°C for 15 s, 60°C for 1 min, final elongation at 72°C for 10 min	95°C for 10 min; 50 cycles of 95°C for 15 s, 60°C for 1 min	23 S rRNA (80)	6[c]	Frahm and Obst, 2003
Adenovirus	AQ1 AQ2	GCCACGGTGGGGTTTCTAAACTT GCCCCAGTGGTCTTACATGCACATC	0.3 μM	94 °C for 2 min; 35 cycles of 94 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min; final elongation at 72 °C for 7 min	95 °C for 10 min; 55 cycles of 95 °C for 3 s, 55 °C for 10 s, 65 °C for 1 min	Hexon (110)	1[b]	Heim et al., 2003
Lachnospiraceae	Lachno2 FWD Lachno2 REV Lachno2 probe	TTCGCAAGAATGAAACTCAAAG AAGGAAAGATCCGGTTAAGGATC 6-carboxyfluoroscein (6-FAM)-ACCAAGTCTTGACATCCG – minor groove binder (MGB)	200 μM	95 °C for 10 min; 40 cycles of 95 °C for 15 s, 60 °C for 1 min, 72 °C for 1 min; final elongation at 72 °C for 10 min	50 °C for 2 min, 95 °C for 10 min; 55 cycles of 95 °C for 15 s, 60 °C for 1 min	16S rRNA (144)	5[c]	Newton et al., 2011

[a]*Wall et al. (1999).* [b]*National Center for Biotechnology Information (NCBI) (2020).* [c]*Stoddard et al. (2015).* [d]*Reyneke et al. (2020).*

Risk Characterisation

Lastly, risk characterisation was conducted, whereby the likelihood of infection was calculated for each target pathogen and the corresponding exposure routes as described in **Table 2**. This was expressed as likely numbers of infections per 10 000 persons per year as previously described by Haas et al. (1999), using Eq. 3:

$$P = 1 - (1 - P_{\text{inf}})^n \quad (3)$$

where P is the probability of infection following n exposure events per year, based on the previously calculated exposure probability of infection (P_{inf}).

Each exposure scenario was simulated using Monte Carlo analysis in RStudio (version 1.0.153) using 500 000 iterations. Throughout the analyses, the different dose parameters [e.g., pathogen concentrations (**Table 2**) and ingestion volumes (**Supplementary Table 3**), amongst others] and exposure events per year (**Supplementary Table 3**) were sampled randomly based on the corresponding distribution of each parameter. However, the annual risk of infection for adenovirus and *M. tuberculosis* were only determined for two exposure scenarios: garden hosing (aerosol inhalation) and washing laundry by hand (aerosol ingestion), as inhalation of these organisms is the primary route of infection (World Health Organisation, 2005; Fennelly and Jones-López, 2015).

Statistical Analyses

The relationships between the detected bacterial pathogens, indicator organisms and MST markers enumerated using qPCR analysis, were investigated using Pearson's correlation analysis and further investigated using Cluster analysis with Ward's method in Statistica™ version 12.5 (2014). The Cluster analysis with Ward's method was specifically applied to visualise the relatedness of the detected bacterial pathogens, indicator organisms and MST markers in the RHRW samples (Waso et al., 2018). Cluster analysis is used to illustrate correlations between organisms. A stronger positive correlation (i.e., a higher correlation coefficient between two organisms), will be represented by a lower linkage distance on a dendrogram (Tilevik, 2017).

RESULTS

Molecular Viability Quantification (EMA-qPCR) of the Target Pathogens and Indicator Organisms in the RHRW Samples

The quantification of intact *L. monocytogenes*, *M. tuberculosis* and *Yersinia* spp. cells as well as the indicator organisms *E. coli* and *Enterococcus* spp. in the RHRW samples was investigated using EMA-qPCR (intact and presumed viable cells) analysis. The respective performance characteristics of the EMA-qPCR analyses are outlined in **Supplementary Table 4** (**Supplementary Material**).

The LLOD for *L. monocytogenes* (*prfA* gene) was 6 gene copies/μL. *L. monocytogenes* was detected in all RHRW samples (100%, $n = 25$) at a mean concentration of 1.4×10^3 cells/100 mL (**Figure 1**). The lowest concentration of *L. monocytogenes* was detected in sampling 2 with the cell counts in the RHRW tanks ranging from 23 cells/100 mL to 1.8×10^3 cells/100 mL, while the highest concentration of *L. monocytogenes* was detected in sampling 3, with the cell counts ranging from 2.5×10^3 cells/100 mL to 4.9×10^3 cells/100 mL. The LLOD for *M. tuberculosis* (*orfA* gene) was 3 gene copies/μL. *M. tuberculosis* was detected in all RHRW samples (100%, $n = 25$) at a mean concentration of 6 cells/100 mL (**Figure 1**). The lowest concentration of *M. tuberculosis* was detected in sampling 5, with the cell counts in the RHRW tanks ranging from 1 cell/100 mL to 4 cells/100 mL, while the highest concentration of *M. tuberculosis* was detected in sampling 2 with the cell counts ranging from 4 cells/100 mL to 16 cells/100 mL. The LLOD for *Yersinia* spp. (*ompF* gene) was 11 gene copies/μL.

TABLE 2 | Monte Carlo simulation dose response input parameters for the target pathogens.

Organism	Variables and distribution*	Dose response model (DRM)	DRM Background	References
Adenovirus	B_C: $\beta = 0.351$; $\eta = 41.458$ (Weibull) $IF_{\%}$: 5.88×10^{-4} (Uniform)	Exponential $k = 6.07 \times 10^{-1}$	M: Human Ex: Inhalation R: Infection	Couch et al., 1966
E. coli (Enteroinvasive *E. coli*)	B_C: $\mu = 3.102$; $\sigma = 1.070$ (Normal) $IF_{\%}$: 0.005 to 0.10 (Uniform)	Beta-Poisson $N_{50} = 2.11 \times 10^6$ $\alpha = 1.55 \times 10^{-1}$	M: Human Ex: Ingestion R: Infection with positive stool isolation	Haas et al., 1999; Reyneke et al., 2020; Ryan et al., 2014
L. monocytogenes	B_C: $\mu = 2.253$; $\sigma = 0.902$ (Lognormal) $IF_{\%}$: 1.00 (Point)	Exponential $k = 1.18 \times 10^{-10}$	M: Human Ex: Ingestion R: Infection	Buchanan et al., 1997; Lyautey et al., 2007
M. tuberculosis	B_C: $\mu = -3.110$; $\sigma = 0.652$ (Lognormal) $IF_{\%}$: 0.66 to 1.00 (Uniform)	Exponential $k = 2.18 \times 10^{-2}$	M: Human Ex: Inhalation R: Infection	Jones et al., 2009

B_C, bacterial concentration; $IF_{\%}$, human infectious fraction of target organism; M, model; Ex, exposure; R, response.
For the QMRA analysis, the Minitab® 19 statistical software package was used to calculate the distribution of each target pathogens' concentration based on the EMA-qPCR and conventional qPCR data (Figure 1**).*

Yersinia spp. were detected in 23 of the RHRW samples (92%, $n = 25$) at a mean concentration of 24 cells/100 mL (**Figure 1**). The lowest concentration of *Yersinia* spp. cells was detected in sampling 5 with the cell counts in the RHRW tanks ranging from 2 cells/100 mL to 11 cells/100 mL, while the highest concentration of *Yersinia* spp. was detected in sampling 3 with the cell counts ranging from 10 cells/100 mL to 1.5×10^2 cells/100 mL. The LLOD for *E. coli* (*uid*A gene) was 2 gene copies/μL. *E. coli* was detected in all RHRW samples (100%, $n = 25$) at a mean concentration of 3.1×10^2 cells/100 mL (**Figure 1**). The lowest concentration of *E. coli* was detected in sampling 2 with the cell counts in the RHRW tanks ranging from 1.4×10^2 cells/100 mL to 2.9×10^2 cells/100 mL, while the highest concentration of *E. coli* was detected in sampling 3 and ranged from 3.3×10^2 cells/100 mL to 6.1×10^2 cells/100 mL. The LLOD for *Enterococcus* spp. (23S rRNA gene) was 1 gene copy/μL. *Enterococcus* spp. were detected in all RHRW samples (100%, $n = 25$) at a mean concentration of 6 cells/100 mL (**Figure 1**). The lowest concentration of *Enterococcus* spp. was detected in sampling 2 with the cell counts in the RHRW tanks ranging from 2 cells/100 mL to 5 cells/100 mL, while the highest concentration of *Enterococcus* spp. was detected in sampling 5 and ranged from 2 cells/100 mL to 27 cells/100 mL.

Molecular Quantification (qPCR) of the MST Markers in the RHRW Samples

The performance characteristics of the qPCR (whole or total DNA) analyses of the MST markers adenovirus, Lachnospiraceae and *Bacteroides* HF183 are outlined in **Supplementary Table 4** (**Supplementary Material**).

The LLOD for *Bacteroides* HF183 (16S rRNA gene) was 3 gene copies/μL. *Bacteroides* HF183 was detected in 19 of the RHRW samples (76%, $n = 25$) at a mean concentration of 13 cells/100 mL (**Figure 1**). The lowest concentration of *Bacteroides* HF183 was detected in sampling 2 with the cell counts in the RHRW tanks ranging from 0 cells/100 mL to 16 cells/100 mL, while the highest concentration of *Bacteroides* HF183 cells was detected in sampling 5 and ranged from 0 cells/100 mL to 44 cells/100 mL. The LLOD for adenovirus (Hexon gene) was 5 gene copies/μL. Adenovirus was detected in 24 of the RHRW samples (96%, $n = 25$) at a mean concentration of 6.6×10^2 cells/100 mL (**Figure 1**). The lowest concentration of adenovirus was detected in sampling 1 with the cell counts in the RHRW tanks ranging from 0 cells/100 mL to 9 cells/100 mL, while the highest concentration of adenovirus cells was detected in sampling 3 and ranged from 9 cells/100 mL to 1.5×10^4 cells/100 mL. The LLOD for Lachnospiraceae (16S rRNA gene) was 1 gene copy/μL. Lachnospiraceae was detected in 20 of the RHRW samples (80%, $n = 25$) at a mean concentration of 24 cells/100 mL (**Figure 1**). The lowest concentration of Lachnospiraceae was detected in sampling 4 with the cell counts in the RHRW tanks ranging from 0 cells/100 mL to 10 cells/100 mL, while the highest concentration of Lachnospiraceae cells was detected in sampling 2 and ranged from 2 cells/100 mL to 3.7×10^2 cells/100 mL.

Correlation Between the Target Pathogens, MST Markers and Indicator Organisms

Pearson's correlation and Cluster analysis was used to correlate and visualise the relatedness of the pathogenic bacterial species, indicator organisms and MST markers detected in the RHRW samples (**Table 3** and **Figure 2**). Results indicated that a significant positive correlation was recorded for *E. coli* versus *L. monocytogenes* ($r = 0.6738$; $p = 0.000$); and *Enterococcus* spp. versus the *Bacteroides* HF183 marker ($r = 0.4071$; $p = 0.043$), while a significant negative correlation was recorded for *M. tuberculosis* versus the *Bacteroides* HF183 marker ($r = -0.4558$; $p = 0.022$) (**Table 3**). However, despite no significant correlation being observed, based on the cluster analysis, adenovirus was then related to *E. coli* ($r = 0.1938$; $p = 0.353$) and *L. monocytogenes* ($r = 0.2517$; $p = 0.225$) (**Figure 2**). Additionally, the MST marker Lachnospiraceae clustered with the target pathogens *M. tuberculosis* ($r = -0.0445$; $p = 0.833$) and *Yersinia* spp. ($r = -0.0810$; $p = 0.700$) (**Figure 2**).

Health Risk Associated With Utilising the RHRW

The annual infection risks linked to the utilisation of untreated RHRW for each exposure scenario, based on the presence of pathogenic adenovirus, *L. monocytogenes*, *M. tuberculosis* and *E. coli* (**Figure 3**) were determined. This was done by comparing all annual risks to a hypothetical benchmark value of 1×10^{-4} which represents the benchmark annual risk of infection for drinking water (Regli et al., 1991).

For all the sampling sessions, the mean annual risk of infection posed by *E. coli* for garden hosing ($\sim 10^{-4}$), garden work ($\sim 10^{-5}$), washing laundry by hand (10^{-8}) and washing/bathing ($\sim 10^{-4}$) was below the annual infection risk benchmark limit (**Figures 3A,B**). Similarly, the mean annual risk of infection posed by *L. monocytogenes* for garden hosing ($\sim 10^{-7}$), garden work ($\sim 10^{-8}$), washing laundry by hand (10^{-11}) and washing/bathing ($\sim 10^{-7}$) was below the annual infection risk benchmark limit (**Figures 3A,B**). In comparison, while the annual risk of infection posed by *L. monocytogenes* for accidental consumption and cleaning of the home were below ($\sim 10^{-7}$) the benchmark limit, the annual risk of infection posed by *E. coli* for these two exposure scenarios exceeded the recommended benchmark limit ($\sim 10^{-3}$). Additionally, the annual risk of infection posed by the intentional drinking of RHRW, was exceeded for both *E. coli* and *L. monocytogenes* (**Figure 3A**). The annual risk of infection for *M. tuberculosis* and adenovirus was only determined for two exposure scenarios: garden hosing (aerosol inhalation) and washing laundry by hand (aerosol ingestion volume was used as a substitute for aerosol inhalation volume to represent a worst-case scenario). Analysis of the mean annual risk of infection posed by *M. tuberculosis* in the RHRW samples indicated that both garden hosing ($\sim 10^{-5}$) and washing laundry by hand ($\sim 10^{-5}$) were below the annual infection risk benchmark limit (**Figure 3B**). In contrast, analysis of the mean annual risk of infection posed by adenovirus in the RHRW samples indicated that both garden hosing ($\sim 10^{-3}$) and washing

TABLE 3 | Summary of the correlations observed between the target pathogens, MST markers and indicator organisms detected in the RHRW samples.

Variable	*Bacteroides* HF183	*E. coli*	*Enterococcus* spp.	*L. monocytogenes*	*M. tuberculosis*	*Yersinia* spp.	Adenovirus	Lachnospiraceae
Bacteroides HF183	NA							
	NA							
E. coli	−0.2321	NA						
	$p = 0.264$	NA						
Enterococcus spp.	**0.4071**	0.0105	NA					
	$p = 0.043$	$p = 0.960$	NA					
L. monocytogenes	−0.1605	**0.6738**	0.1412	NA				
	$p = 0.444$	**$p = 0.000$**	$p = 0.501$	NA				
M. tuberculosis	**−0.4558**	0.1377	−0.0045	0.0055	NA			
	$p = 0.022$	$p = 0.512$	$p = 0.983$	$p = 0.979$	NA			
Yersinia spp.	−0.2169	0.0813	0.0137	0.3001	0.0352	NA		
	$p = 0.298$	$p = 0.699$	$p = 0.948$	$p = 0.145$	$p = 0.867$	NA		
Adenovirus	−0.0032	0.1938	−0.0705	0.2517	−0.0315	−0.0237	NA	
	$p = 0.988$	$p = 0.353$	$p = 0.738$	$p = 0.225$	$p = 0.881$	$p = 0.910$	NA	
Lachnospiraceae	−0.0909	−0.3368	−0.1186	−0.1856	−0.0445	−0.0810	−0.0269	NA
	$p = 0.665$	$p = 0.100$	$p = 0.572$	$p = 0.374$	$p = 0.833$	$p = 0.700$	$p = 0.898$	NA

Significant correlations ($p < 0.05$) are indicated in red and bold.

laundry by hand ($\sim 10^{-3}$) exceeded the annual infection risk benchmark limit (**Figure 3B**).

DISCUSSION

Ethidium monoazide bromide-qPCR analyses indicated that intact (viable) cells of the human pathogens *L. monocytogenes* and *M. tuberculosis* were detected in all (100%) the RHRW samples, while *Yersinia* spp. were detected in 92% of the samples. *L. monocytogenes* has frequently been detected in soil, plant and surface water samples, as well as in sewage, slaughterhouse waste, human and animal faeces (Weis and Seeliger, 1975; Farber and Peterkin, 1991). Correspondingly, in a study conducted by Lyautey et al. (2007), *L. monocytogenes* was isolated repeatedly from surface water over a 5-month sampling period using a culture-based selective enrichment and isolation procedure. Similarly, Colburn et al. (1990) isolated *L. monocytogenes* from 62% ($n = 37$) of samples extracted from fresh or low-salinity river water, which drained into the Humboldt-Arcata Bay (California). These results emphasise the ability of *L. monocytogenes* to survive in environments that are characterised by physical (e.g., solar irradiation, temperature) and chemical (e.g., pH, oxygen concentration, nutrients) variation. Additionally, Linke et al. (2014) observed a significant link between the presence of *Listeria* spp. in soil samples (collected from 12 different areas in Austria from 2007 to 2009) and the abiotic conditions of the soil (e.g., pH, moisture, type of soil) and found that *Listeria* spp. were more regularly isolated from soil samples characterised by neutral pH, low moisture, or having a consistency made up of sand and humus. Moreover, the same authors noticed that seasonal changes had an effect on the prevalence of *Listeria* spp. in soil, with the lowest cell counts recorded in July. Several strains of *L. monocytogenes* have also been found to survive for months to several years in food processing plants, including those used to produce dairy, meat, fish and ready-to-eat products (Ferreira et al., 2014). It is thus hypothesised that the high frequency of detection of the ubiquitous organism *L. monocytogenes* in the rainwater samples, may be due to its ability to survive across a wide range of temperatures and resist several environmental stresses (Lyautey et al., 2007).

Similar to *L. monocytogenes*, *Yersinia* spp. are extensively distributed in the environment with common reservoirs identified as wild rodents, livestock, wild animals, water and soil (Kapperud, 1975; Mollaret et al., 1979). In 2014 and 2019 two outbreaks of yersiniosis were caused by *Y. enterocolitica* O9 (Norway) and *Y. enterocolitica* O3 (Sweden and Denmark), respectively, with both outbreaks linked to the consumption of fresh salad/vegetables (MacDonald et al., 2016; Espenhain et al., 2019). Traceback investigations into the outbreak linked to *Y. enterocolitica* O9 indicated that the factory did not regularly change the water in the rinsing tanks, used for the processing of the salad mixes, which was subsequently identified as the likely cause of the *Y. enterocolitica* O9 contamination of the food products (MacDonald et al., 2016). Interestingly, *Yersinia* spp. are carried by most mammals, but generally do not cause serological or histopathological responses in these hosts (Pocock et al., 2001).

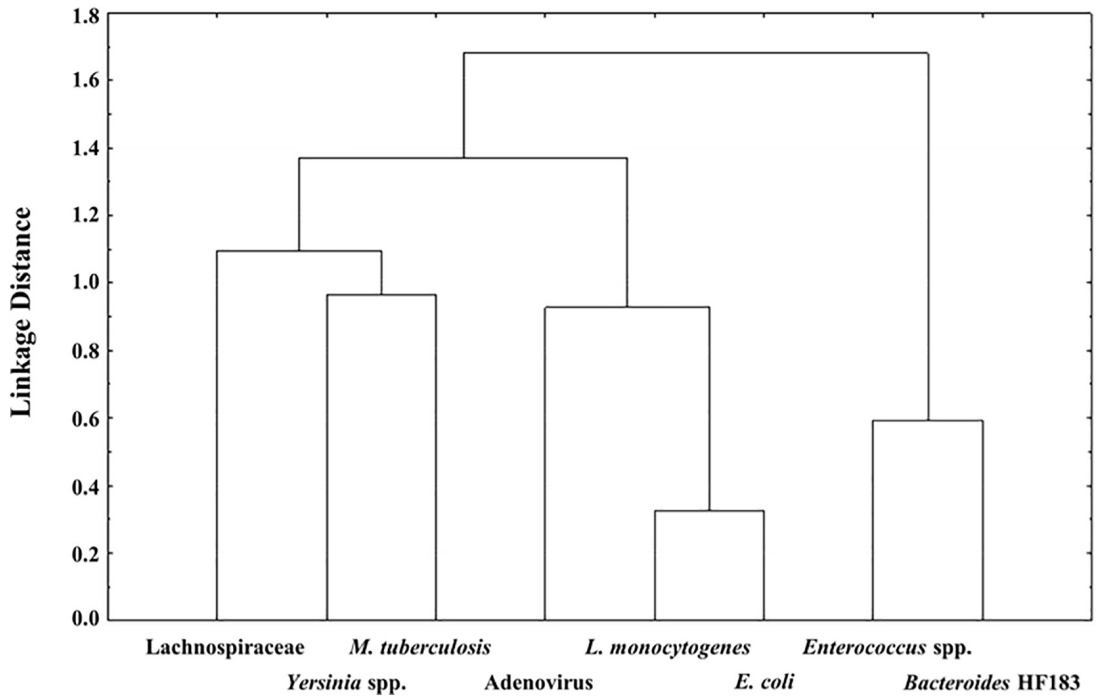


FIGURE 2 | Dendrogram of the Cluster Analysis with Ward's Methods of target pathogens versus the MST markers and indicator organisms detected in the RHRW samples.

However, close contact between rodents (e.g., house mice) and humans or livestock, has resulted in rodents being identified as significant vectors of *Yersinia* spp. infection (Pocock et al., 2001). Therefore, the detection of *Yersinia* spp. in the RHRW samples is hypothesised to have originated from mice and other rodents' faecal matter or bodily fluids being deposited on the catchment surface or in the rainwater harvesting tanks.

In comparison, while *M. tuberculosis* is capable of adapting to hostile or extreme environmental conditions (Cook et al., 2009), the presence and persistence of this bacterium in the environment, and its possible role in the cause and distribution of community-acquired tuberculosis, has been a continuous debate since the start of the 20th century (Velayati et al., 2015). However, it is known that transmission of this bacterium from the environment is possibly due to its ability to persist under various environmental conditions. For example, tuberculosis bacilli have been isolated from wooden tongue depressors over an 88-day time-period, woollen household carpet over a 19-day sampling period, and both dry and moist soil for up to 4 weeks post initial contamination (Velayati et al., 2015). Additionally, it has been suggested that water and soil can become contaminated with *M. tuberculosis* through sputum from infected individuals (coughing sputum) (Velayati et al., 2015). The presence of *M. tuberculosis* within the rainwater tank samples may thus have resulted from sputum contaminated soil, or other debris, being deposited into the rainwater harvesting tanks. This is a cause for serious concern as South Africa forms part of the top six countries around the world that are burdened by a high incidence of tuberculosis (Tadokera et al., 2020). Within the Overberg District Municipality (region where the sampling site is located), from 2011 to 2015, tuberculosis-related deaths were reported as the number one cause of mortality for individuals between 25 and 64 years of age (Health Systems Trust, 2019). Additionally, individuals living in poverty-stricken areas have been identified as being at higher risk of contracting tuberculosis, as these individuals generally live in crowded conditions and lack access to basic healthcare (Foster et al., 2015).

Escherichia coli and *Enterococcus* spp. are generally employed as indicators of faecal pollution by warm-blooded animals (Field and Samadpour, 2007; Harwood et al., 2014) and based on the high frequency of detection (100 and 99%, respectively) of these indicator organisms in the RHRW samples, the hypothesis that the rainwater may be contaminated with the faecal matter of rodents and animals, amongst others, that was deposited on the rooftops or in the gutter systems, is thus confirmed. However, as previously indicated, MST markers are frequently used in combination with traditional indicator organisms, as these markers are able to differentiate between several sources of faecal contamination. Amongst the most promising MST markers are members of the *Bacteroides* spp. as these organisms are limited to the digestive tract of both humans and warm-blooded animals, where they dominate in the natural gut microflora (Ravaliya et al., 2014) and are subsequently detected in high concentrations in host faecal matter (Fogarty et al., 2003; Ge et al., 2010). Of particular interest is the HF183 marker, which is conserved among *Bacteroides* strains of human origin and has exhibited high specificity for the detection of human faecal matter and sewage

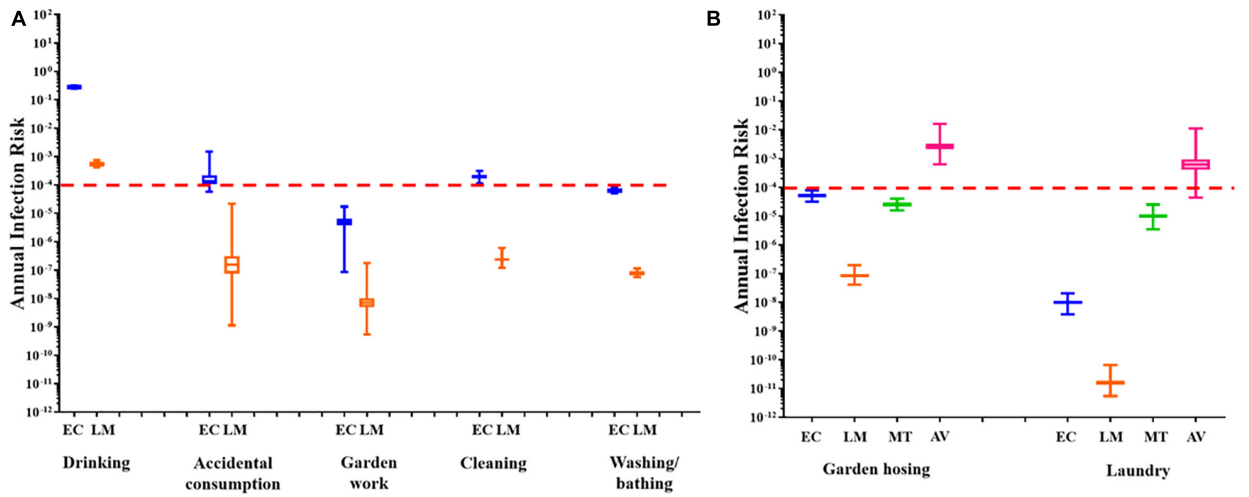


FIGURE 3 | (A) Annual health risk associated with the use of RHRW in the community for ingestion scenarios based on the presence of *E. coli* (EC; **blue**) and *L. monocytogenes* (LM; **orange**). **(B)** Annual health risk associated with the use of RHRW in the community for inhalation scenarios based on the presence of *E. coli* (EC; **blue**), *L. monocytogenes* (LM; **orange**), *M. tuberculosis* (MT; **green**) and adenovirus (AV; **pink**). The whiskers illustrate the minimum and maximum, the outer box illustrates the 1st and 3rd quartiles, and the inner line illustrates the median. The benchmark limit (1×10^{-4}) is indicated by the dashed red line.

contamination in environmental waters (Harwood et al., 2014). A high frequency of the *Bacteroides* HF183 marker (76%, $n = 25$) was subsequently detected within the RHRW samples in the current study. Personal communication with a few residents of the Kleinmond Housing Scheme site indicated that in order to prevent pets from scavenging household waste, garbage bags are regularly placed on top of the rainwater harvesting tanks (Waso et al., 2016). It is thus hypothesised that household waste stored on top of the tanks could potentially have introduced human faecal matter (e.g., from babies nappies/diapers) into the rainwater tanks. Similarly, adenovirus and Lachnospiraceae are ubiquitously distributed in the environment, particularly in areas contaminated with sewage or human faeces (World Health Organisation, 2005; Newton et al., 2011). Therefore, the detection of adenovirus and Lachnospiraceae in the RHRW samples (96 and 80%, respectively) may have also occurred through the introduction of household waste into the rainwater tanks.

The MST markers (*Bacteroides* HF183, adenovirus and Lachnospiraceae) and traditional indicator organisms (*E. coli* and *Enterococcus* spp.) were then statistically correlated to the human pathogenic species (*M. tuberculosis*, *Yersinia* spp. and *L. monocytogenes*) detected in the rainwater. Results showed significant positive correlations for *E. coli* versus *L. monocytogenes* ($r = 0.6738$; $p = 0.000$); and *Enterococcus* spp. versus the *Bacteroides* HF183 marker ($r = 0.4071$; $p = 0.043$), while a significant negative correlation was observed for *M. tuberculosis* versus the *Bacteroides* HF183 marker ($r = -0.4558$; $p = 0.022$) (**Table 3**). The significant positive correlation recorded between *E. coli* and *L. monocytogenes*, could be explained by the fact that these organisms share several common reservoirs (e.g., water, soil, human and animal faeces) and could thus have entered the rainwater tank via a common source. Based on the cluster analysis (**Figure 2**), adenovirus was then also related to *L. monocytogenes* and *E. coli* (albeit not significantly), which is hypothesised to be due to the common occurrence of all three groups in faecal matter (Weis and Seeliger, 1975; Farber and Peterkin, 1991; Pocock et al., 2001; Waso et al., 2018). Similarly, the significant correlation and clustering observed between *Enterococcus* spp. and the *Bacteroides* HF183 marker confirms results of previous studies where indicator organisms positively correlated with MST markers (Korajkic et al., 2018; Waso et al., 2018). This is hypothesised to be due to the common occurrence of these indicators and MST markers in the gut of humans and warm-blooded animals, and consequently, in host faecal matter (Field and Samadpour, 2007; Ahmed et al., 2008a; Harwood et al., 2014). Interestingly, a significant negative correlation was observed between *M. tuberculosis* and the *Bacteroides* HF183 marker ($r = -0.4558$; $p = 0.022$). Research has indicated that during tuberculosis infection and the implementation of subsequent treatment strategies, the gut microbiota is altered significantly (Namasivayam et al., 2018). Consequently, a decrease in the diversity of *Bacteroides* spp. present in the gut has been observed during *M. tuberculosis* infection (Namasivayam et al., 2018), which could possibly elucidate the significant negative correlation observed between *M. tuberculosis* and the *Bacteroides* HF183 marker in the current study.

A QMRA framework was then applied to assess the health risk associated with the consumption of RHRW containing pathogenic *E. coli*, adenovirus, *L. monocytogenes* and *M. tuberculosis* for potable and several domestic activities (exception of *M. tuberculosis* and adenovirus where only two potential inhalation exposure scenarios were assessed). Results of the QMRA for *L. monocytogenes* indicated that the annual benchmark for infection risk was only exceeded for intentional drinking, while the risk associated with the use of the rainwater contaminated with *L. monocytogenes*

for each of the remaining domestic activities was below the annual infection risk benchmark limit ($<1 \times 10^{-4}$). Similarly, for *E. coli*, the risk associated with the use of the RHRW for the domestic activities garden hosing, garden work, washing laundry by hand and washing/bathing, was below the annual infection risk benchmark limit. In contrast, the risk associated with intentional drinking, accidental consumption and cleaning of the home exceeded the annual infection risk benchmark limit (1×10^{-4}) for untreated rainwater and thus posed a possible risk of infection by *E. coli*. This is concerning as results of a social survey conducted by Dobrowksy et al. (2014) indicated that 70% of individuals residing in the Kleinmond Housing Scheme site use the RHRW for cleaning, whilst 24% use it for drinking (without treatment). Consumption of rainwater contaminated with enteroinvasive *E. coli* pathotypes and *L. monocytogenes* could therefore significantly increase the occurrence of gastrointestinal disease in developing countries, particularly in sub-Saharan Africa (Ryan et al., 2014; Robins-Browne et al., 2016). Moreover, it should be noted that while the *L. monocytogenes* sequence type 6 subtype (predominantly associated with the major listeriosis outbreak in South Africa in 2017 and 2018) (Smith et al., 2019) was not analysed for in the current study, consumption of untreated RHRW could potentially lead to an increase in the number of listeriosis cases in the end-user community.

The annual risk of infection for adenovirus and *M. tuberculosis* was only calculated for two exposure scenarios (i.e., garden hosing and washing laundry by hand) which are linked to the inhalation of water particles, as infection with human adenovirus and *M. tuberculosis* bacilli primarily results in respiratory infections rather than gastrointestinal illness (World Health Organisation, 2005; Fennelly and Jones-López, 2015). Although it is possible to contract *M. tuberculosis* by consuming water contaminated with this bacterium, tuberculosis infection is initiated when droplet nuclei containing *M. tuberculosis* are inhaled and reach the alveoli of the lungs (Centers for Disease Control and Prevention, 2016). While the QMRA for *M. tuberculosis*, for garden hosing and washing laundry by hand, was below the annual infection risk benchmark limit, the QMRA for adenovirus exceeded the annual infection risk benchmark limit for both garden hosing ($\sim 10^{-3}$) and washing laundry by hand ($\sim 10^{-3}$). Adenovirus was selected for the QMRA analysis as this group of viruses are prevalent in high numbers in a wide range of water environments and have shown to be highly resistant to processes of disinfection and purification (Van Heerden et al., 2003). However, while the results obtained for *M. tuberculosis* are similar to data obtained by Hamilton et al. (2017), who observed that the annual risk of infection posed by the *Mycobacterium avium* complex (a group of bacteria related to *M. tuberculosis*), was below the benchmark value of 1×10^{-4}, the health risks posed by *M. tuberculosis* in rainwater need to be further investigated. A significantly high incidence of tuberculosis is reported for the Western Cape region of South Africa and immune-compromised individuals have been identified as highly vulnerable to infection with *M. tuberculosis* (Sester et al., 2014). The *M. tuberculosis* QMRA results obtained in the current study may thus be an underestimation of the risk associated for immune-compromised individuals residing in the end-user communities, who rely on RHRW as a primary water source.

CONCLUSION

The frequent detection of *L. monocytogenes*, *M. tuberculosis* and *Yersinia* spp. in the RHRW samples verifies that human pathogenic species are able to survive in rainwater which can pose a serious health risk to low- and middle-income communities, who routinely utilise RHRW as a sustainable water source. In addition, results of the correlation analysis confirm that traditional indicator organisms and MST markers should be used in combination to monitor RHRW quality, as both indicator groups correlated with the human pathogens (i.e., *E. coli* versus *L. monocytogenes*; *M. tuberculosis* versus *Bacteroides* HF183*)*, as well as with each other (i.e., *Enterococcus* spp. versus *Bacteroides* HF183). Nonetheless, additional research should be conducted to assess the correlation of a broader range of human pathogenic species to the presence of several indicator organism groups (e.g., total coliforms, faecal coliforms) and MST markers (e.g., polyomavirus, *Bifidobacterium* spp., human mitochondrial DNA), in order to fully elucidate the environmental distribution and relationships between the various indicator groups and human pathogens.

The QMRA analysis then indicated that the use of RHRW containing *L. monocytogenes*, adenovirus, and *E. coli* poses a health risk to end-user communities, particularly when used for intentional drinking (*E. coli* and *L. monocytogenes*), cleaning of the home (*E. coli*), garden hosing (adenovirus), washing laundry by hand (adenovirus), or when accidentally consumed (*E. coli*). However, while the QMRA results indicated that the concentration of *M. tuberculosis* obtained in the current study did not pose a health-risk to the end-user community, further research should be conducted, taking into consideration the approximate percentage of immune-compromised individuals living in South Africa and who utilise RHRW, in order to accurately estimate the risk associated with the use of RHRW for potable and domestic activities. This can ultimately determine or predict the potential of various available point-of-use treatment technologies (e.g., filtration, solar disinfection, chlorination, solar pasteurisation) to effectively reduce the estimated health risk to within the benchmark limit.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JD, BR, and WK conceived and designed the experiments. JD performed the experiments. JD, BR, and MW analysed the data. WK and SK contributed reagents, materials, and analysis tools. JD, BR, MW, and WK compiled the manuscript. SK edited the

manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project has received funding from the European Union's Horizon 2020 – Research and Innovation Programme under grant agreement no. 688928 (WATERSPOUTT H2020-Water-5c). The authors acknowledge the financial assistance provided by the National Research Foundation (NRF) of South Africa toward the student scholarship. Opinions communicated and all conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.659784/full#supplementary-material

REFERENCES

Ahmed, W., Goonetilleke, A., and Gardner, T. (2008a). Alternative indicators for detection and quantification of faecal pollution. *Water (Australia)* 39, 46–49. doi: 10.HH/j.l472-765X.2007.02287.x

Ahmed, W., Hamilton, K., Gyawali, P., Toze, S., and Haas, C. (2016). Evidence of avian and possum fecal contamination in rainwater tanks as determined by microbial source tracking approaches. *Appl. Environ. Microbiol.* 82, 4379–4386. doi: 10.1128/AEM.00892-16

Ahmed, W., Harwood, V., Nguyen, K., Young, S., Hamilton, K., and Toze, S. (2015). Utility of *Helicobacter* spp. associated GFD markers for detecting avian faecal pollution in natural waters of two continents. *Water Res.* 88, 613–622. doi: 10.1016/j.watres.2015.10.050

Ahmed, W., Huygens, F., Goonetilleke, A., and Gardner, T. (2008b). Real-time PCR detection of pathogenic microorganisms in roof-harvested rainwater in Southeast Queensland. Australia. *Appl. Environ. Microbiol.* 74, 5490–5496. doi: 10.1128/AEM.00331-08

Ahmed, W., Staley, C., Hamilton, K. A., Beale, D. J., Sadowsky, M. J., Toze, S., et al. (2017). Amplicon-based taxonomic characterization of bacteria in urban and peri-urban roof-harvested rainwater stored in tanks. *Sci. Total Environ.* 576, 326–334. doi: 10.1016/j.scitotenv.2016.10.090

Bosio, C. F., Jarrett, C. O., Scott, D. P., Fintzi, J., and Hinnebusch, J. (2020). Comparison of the transmission efficiency and plague progression dynamics associated with two mechanisms by which fleas transmit *Yersinia pestis*. *PLoS Pathogens* 16:e1009092. doi: 10.1371/journal

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

Buchanan, R. L., Damert, W. G., Whiting, R. C., and van Schothorst, M. (1997). Use of epidemiologic and food survey data to estimate a purposefully conservative dose-response relationship for *Listeria monocytogenes* levels and incidence of listeriosis. *J. Food Prot.* 60, 918–922. doi: 10.4315/0362-028x-60.8.918

Buchanan, R. L., Smith, J. L., and Long, W. (2000). Microbial risk assessment: dose-response relations and risk characterisation. *Int. J. Food Microbiol.* 58, 159–172. doi: 10.1016/s0168-1605(00)00270-1

Centers for Disease Control and Prevention (2016). *Latent Tuberculosis Infection Resources*. Available online at: https://www.cdc.gov/tb/. (Accessed March 13, 2021)

Colburn, K. G., Kaysner, C. A., Abeyta, C., and Wekell, M. M. (1990). *Listeria* species in a California coast estuarine environment. *Appl. Environ. Microbiol.* 56, 2007–2011. doi: 10.1128/AEM.56.7.2007-2011.1990

Cook, G. M., Berney, M., Gebhard, S., Heinemann, M., Cox, R. A., Danilchanka, O., et al. (2009). Physiology of *Mycobacteria*. *Adv. Microb. Physiol.* 55, 81–182,318–9. doi: 10.1016/S0065-2911(09)05502-7

Couch, R. B., Cate, T. R., Douglas, R. G., Gerone, P. J., and Knight, V. (1966). Effect of route of inoculation on experimental respiratory viral disease in volunteers and evidence for airborne transmission. *Microbiol. Mol. Biol. Rev.* 30, 517–529.

De Kwaadsteniet, M., Dobrowsky, P. H., van Deventer, A., Khan, W., and Cloete, T. E. (2013). Domestic rainwater harvesting: Microbial and chemical water quality and point-of-use treatment systems. *Water Air Soil Pollut.* 224, 961–971. doi: 10.1007/s11270-013-1629-7

Dobrowksy, P. H., Mannel, D., De Kwaadsteniet, M., Prozesky, H., Khan, W., and Cloete, T. E. (2014). Quality assessment and primary uses of harvested rainwater in Kleinmond. South Africa. *Water SA* 40, 401–406. doi: 10.4314/wsa.v40i3.2

Dobrowsky, P. H., De Kwaadsteniet, M., Cloete, T. E., and Khan, W. (2014). Distribution of indigenous bacterial pathogens and potential pathogens associated with roof-harvested rainwater. *Appl. Environ. Microbiol.* 80, 2307–2316. doi: 10.1128/AEM.04130-13

Dobrowsky, P. H., Khan, S., Cloete, T. E., and Khan, W. (2016). Molecular detection of *Acanthamoeba* spp., *Naegleria fowleri* and *Vermamoeba* (*Hartmannella*) *vermiformis* as vectors for *Legionella* spp. in untreated and solar pasteurized harvested rainwater. *Parasit. Vectors* 9:539. doi: 10.1186/s13071-016-1829-2

Dobrowsky, P. H., Lombard, M., Cloete, W. J., Saayman, M., Cloete, T. E., Carstens, M., et al. (2015). Efficiency of microfiltration systems for the removal of bacterial and viral contaminants from surface and rainwater. *Water Air Soil Pollut.* 226:33. doi: 10.1007/s11270-015-2317-6

Eisen, R. J., Bearden, S. W., Wilder, A. P., Montenieri, J. A., Antolin, M. F., and Gage, K. L. (2006). Early-phase transmission of *Yersinia pestis* by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15380–15385. doi: 10.1073/pnas.0606831103

Emerson, J. B., Adams, R. I., Betancourt Román, C. M., Coil, D. A., Dahlhausen, K., Ganz, H. H., et al. (2017). Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome.* 5, 86. doi: 10.1186/s40168-017-0285-3

Espenhain, L., Riess, M., Müller, L., Colombe, S., Ethelberg, S., Litrup, E., et al. (2019). Cross-border outbreak of *Yersinia enterocolitica* O3 associated with imported fresh spinach, Sweden and Denmark, March 2019. *Euro Surveill.* 24:1900368. doi: 10.2807/1560-7917.ES.2019.24.24.1900368

Farber, J. M., and Peterkin, P. I. (1991). *Listeria monocytogenes*, a food-borne pathogen. *Microbiol. Rev.* 55, 476–511. doi: 10.1016/B978-0-12-811444-5.00006-3

Fennelly, K. P., and Jones-López, E. (2015). Quantity and quality of inhaled dose predicts immunopathology in tuberculosis. *Front. Microbiol.* 6:313. doi: 10.3389/fimmu.2015.00313

Ferreira, V., Wiedmann, M., Teixeira, P., and Stasiewicz, W. J. (2014). *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J. Food Prot.* 77, 150–170. doi: 10.4315/0362-028X.JFP-13-150

Field, K. G., and Samadpour, M. (2007). Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res.* 41, 3517–3538. doi: 10.1016/j.watres.2007.06.056

Fogarty, L., Haack, S., Wolcott, M., and Whitman, R. (2003). Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull feces. *J. Appl. Microbiol.* 94, 865–878. doi: 10.1046/j.1365-2672.2003.01910.x

Foster, N., Vassal, A., Cleary, S., Cunnama, L., Churchyard, G., and Sinanovic, E. (2015). The economic burden of TB diagnosis and treatment in South Africa. *Soc. Sci. Med.* 130, 42–50. doi: 10.1016/j.socscimed.2015.01.046

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

Ge, Z., Nevers, M. B., Schwab, D. J., and Whitman, R. L. (2010). Coastal loading and transport of *Escherichia coli* at an embayed beach in Lake Michigan. *Environ. Sci. Technol.* 44, 6731–6737. doi: 10.1021/es100797r

Germini, A., Masola, A., Carnevali, P., and Marchelli, R. (2009). Simultaneous detection of *Escherichia coli* O175:H7. *Salmonella* spp., and Listeria monocytogenes by multiplex PCR. *Food Control* 20, 733–738. doi: 10.1016/j.foodcont.2008.09.010

Gould, J., and Nissen-Petersen, E. (1999). *Rainwater Catchment Systems for Domestic Supply*. London: Intermediate Technology Publications.

Haas, C. N., Rose, J. B., and Gerba, C. P. (1999). *Quantitative Microbial Risk Assessment*, 2nd Edn. New York, NY: John Wiley & Sons, Inc.

Hamilton, K., Reyneke, B., Waso, M., Clements, T., Ndlovu, T., Khan, W., et al. (2019). A global review of the microbiological quality and potential health risks associated with roof-harvested rainwater tanks. *npj* Clean Water 2:7. doi: 10.1038/s41545-019-0030-5

Hamilton, K. A., Ahmed, W., Toze, S., and Haas, C. N. (2017). Human health risks for *Legionella* and *Mycobacterium avium* complex (MAC) from potable and non-potable uses of roof-harvested rainwater. *Water Res.* 119, 288–303. doi: 10.1016/j.watres.2017.04.004

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

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

Health Systems Trust. (2019). *Western Cape Province: Population Distribution, Sub-district Boundaries and Health Facility Locations*. Available online at: https://www.hst.org.za/publications/District%20Health%20Barometers/16%20(Section%20B)%20Western%20Cape%20Province%20DHP20172018.pdf (Accessed March 18, 2021)

Heim, A., Ebnet, C., Harste, G., and Pring-Åkerblom, P. (2003). Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J. Med. Virol.* 70, 228–239. doi: 10.1002/jmv.10382

Hinnebusch, B. J., Jarrett, C. O., and Bland, D. M. (2017). "Fleaing" the plague: adaptations of *Yersinia pestis* to its insect vector that lead to transmission. *Annu. Rev. Microbiol.* 71, 215–232. doi: 10.1146/annurev-micro-090816-093521

Jones, M. R., Masago, Y., Bartrand, T., Haas, C. N., Nicas, M., and Rose, J. B. (2009). Characterising the risk of infection from *Mycobacterium tuberculosis* in commercial passenger aircraft using quantitative microbial risk assessment. *Risk Anal.* 29, 355–365. doi: 10.1111/j.1539-6924.2008.01161.x

Jones, R. M., and Su, Y. (2015). Dose-response models for selected respiratory infectious agents: *Bordetella pertussis*, group a *Streptococcus*, rhinovirus and respiratory syncytial virus. *BMC Infect. Dis.* 15:90. doi: 10.1186/s12879-015-0832-0

Jongman, M., and Korsten, L. (2016). Microbial quality and suitability of roof-harvested rainwater in rural villages for crop irrigation and domestic use. *J. Water Health* 14, 961–971. doi: 10.2166/wh.2016.058

Kapperud, G. (1975). *Yersinia enterocolitica* in small rodents from Norway, Sweden and Finland. *APMIS* 83, 335–342. doi: 10.1111/j.1699-0463.1975.tb00110.x

Kaushik, R., Balasubramanian, R., and Dunstan, H. (2014). Microbial quality and phylogenetic diversity of fresh rainwater and tropical freshwater reservoir. *PLoS One* 9:e100737. doi: 10.1371/journal.pone.0100737

King, H. C., Kera-Butler, T., James, P., Oakley, B. B., Erenso, G., Aseffa, A., et al. (2017). Environmental reservoirs of pathogenic *Mycobacteria* across the Ethiopian biogeographical landscape. *PLoS One* 12:e0173811. doi: 10.1371/journal.pone.0173811

Korajkic, A., McMinn, B. R., and Harwood, V. J. (2018). Relationships between microbial indicators and pathogens in recreational water settings. *Int. J. Environ. Res. Public Health* 15:2842. doi: 10.3390/ijerph15122842

Kox, L. F. F., Rhienthong, D., Miranda, A. M., Udomsantisuk, N., Ellis, K., van Leeuwen, J., et al. (1994). A more reliable PCR for detection of *Mycobacterium tuberculosis* in clinical samples. *J. Clin. Microbiol.* 32, 672–678. doi: 10.1128/jcm.32.3.672-678.1994

Linke, K., Ruckerl, I., Brugger, K., Karpiskova, R., Walland, J., Muri-Klinger, S., et al. (2014). Reservoirs of *Listeria* species in three environmental ecosystems. *Appl. Environ. Microbiol.* 80, 5583–5592. doi: 10.1128/AEM.01018-14

Low, C. S. (2002). *Appropriate Microbial Indicator Tests for Drinking Water in Developing Countries and Assessment of Ceramic Water Filters*. Ph.D. dissertation. Massachusetts, MA: Massachusetts Institute of Technology.

Lyautey, E., Lapen, D. R., Wilkes, G., McCleary, K., Pagotto, F., Tyler, K., et al. (2007). Distribution and characteristics of *Listeria monocytogenes* isolates from surface waters of the South Nation River Watershed, Ontario, Canada. *Appl. Environ. Microbiol.* 73, 5401–5410. doi: 10.1128/AEM.00354-07

MacDonald, E., Einöder-Moreno, M., Borgen, K., Brandal, L. T., Diab, L., and Fossli, O. (2016). National outbreak of *Yersinia enterocolitica* infections in military and Civilian populations associated with consumption of mixed salad, Norway, 2014. *Euro Surveill.* 21:34. doi: 10.2807/1560-7917.ES.2016.21.34.30321

McLellan, S. L., Newton, R. J., Vandewalle, J. L., Shanks, O. C., Huse, S. M., Eren, A. M., et al. (2013). Sewage reflects the distribution of human faecal Lachnospiraceae. *Environ. Microbiol.* 15, 2213–2227. doi: 10.1111/1462-2920.12092

Mollaret, H. H., Bercovier, H., and Alonso, J. M. (1979). Summary of data received at the WHO reference centre for *Yersinia enterocolitica*. *Contrib. Microbiol. Immunol.* 5, 174–184.

Namasivayam, S., Sher, A., Glickman, M. S., and Wipperman, M. F. (2018). The microbiome and tuberculosis: early evidence for cross talk. *mBio* 9, e01420-18. doi: 10.1128/mBio .01420-18

National Center for Biotechnology Information (NCBI) (2020). *COVID-19 is an Emerging, Rapidly Evolving Situation*. Available online at: https://www.ncbi.nlm.nih.gov/ (Accessed October 28, 2020)

Newton, R. J., Vandewalle, J. L., Borchardt, M. A., Gorelick, M. H., and McLellan, S. L. (2011). Lachnospiraceae and Bacteroidales alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. *Appl. Environ. Microbiol.* 77, 6972–6981. doi: 10.1128/aem.05480-11

Owens, C. E. L., Angles, M. L., Cox, P. T., Byleveld, P. M., Osborne, N. J., and Rahman, B. (2020). Implementation of quantitative microbial risk assessment (QMRA) for public drinking water supplies: systematic review. *Water Res.* 174:115614. doi: 10.1016/j.watres.2020.115614

Pechous, R. D., Sivaraman, V., Stasulli, N. M., and Goldman, W. E. (2016). Pneumonic plague: the darker side of *Yersinia pestis*. *Trends Microbiol.* 24, 190–197. doi: 10.1016/j.tim.2015.11.008

Pecson, B. M., Triolo, S. C., Olivieri, S., Chen, E. C., Pisarenko, A. N., Yang, C., et al. (2017). Reliability of pathogen control in direct potable reuse: performance evaluation and QMRA of a full-scale 1 MGD advanced treatment train. *Water Res.* 122, 258–268. doi: 10.1016/j.watres.2017.06.014

Pocock, M. J. O., Searle, J. B., Betts, W. B., and White, P. C. L. (2001). Patterns of infection by *Salmonella* and *Yersinia* spp. in commensal house mouse (*Mus musculus domesticus*) populations. *J. Appl. Microbiol.* 90, 755–760. doi: 10.1046/j.1365-2672.2001.01303.x

Ravaliya, K., Gentry-Shields, J., Garcia, S., Heredia, N., Fabiszewski de Aceituno, A., Bartz, F. E., et al. (2014). Use of Bacteroidales microbial source tracking to monitor fecal contamination in fresh produce production. *Appl. Env. Microbiol.* 80, 612–617. doi: 10.1128/AEM.02891-13

Regli, S., Rose, J. B., Haas, C. N., and Gerba, C. P. (1991). Modelling the risk from *Giardia* and viruses in drinking water. *J. Am. Water Works Assoc.* 83, 76–84.

Reyneke, B., Hamilton, K. A., Fernández-Ibáñez, P., Polo-López, M. I., McGuigan, K. G., Khan, S., et al. (2020). EMA-Illumina 16S metagenomic sequencing informs risk assessment analysis of water treatment systems. *Sci. Total Environ.* 743:140717. doi: 10.1016/j.scitotenv.2020.140717

Reyneke, B., Ndlovu, T., Khan, S., and Khan, W. (2017). Comparison of EMA-, PMA- and DNase qPCR for the determination of microbial cell viability. *Appl. Microbiol. Biotechnol.* 101, 7371–7383. doi: 10.1007/s00253-017-8471-6

Robins-Browne, R. M., Holt, K. E., Ingle, D. J., Hocking, D. M., Yang, J., and Tauschek, M. (2016). Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front. Cell. Infect. Microbiol.* 6:141. doi: 10.3389/fcimb.2016.00141

Rusiñol, M., Moriarty, E., Lin, S., Bofill-Mas, S., and Gilpin, B. (2016). Human-, ovine-, and bovine-specific viral source tracking tools to discriminate between the major fecal sources in agricultural waters. *Food Environ Virol.* 8, 34–45. doi: 10.1007/s12560-015-9223-3

Ryan, M. O., Haas, C. N., Gurian, P. L., Gerba, C. P., Panzl, B. M., and Rose, J. B. (2014). Application of quantitative microbial risk assessment for selection of

microbial reduction targets for hard surface disinfectants. *Am. J. Infect. Control.* 42, 1165–1172. doi: 10.1016/j.ajic.2014.07.024

Sano, D., Haas, C. N., and Rose, J. B. (2019). *A QMRA Framework for Sanitation Treatment Decisions*. Available online at: https://www.waterpathogens.org/book/a-QMRA-framework-for-sanitation-treatment -decisions (Accessed March 12, 2020)

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

Schijven, J., Teunis, P., Suylen, T., Ketelaars, H., Hornstra, L., and Rutjes, S. (2019). QMRA of adenovirus in drinking water at a drinking water treatment plant using UV and chlorine dioxide disinfection. *Water Res.* 158, 34–45. doi: 10.1016/j.watres.2019.03.090

Sester, M., van Leth, F., Bruchfeld, J., Bumbacea, D., Cirillo, D. M., Dilektasli, A. G., et al. (2014). Risk assessment of tuberculosis in immunocompromised patients–A TBNET study. *Am. J. Respir. Crit.Care Med.* 190:10. doi: 10.1164/rccm.201405-0967OC

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human fecal pollution in freshwater. *Environ. Microbiol.* 7, 249–259. doi: 10.1111/j.1462-2920.2004.00702.x

Sidhu, J. P., Ahmed, W., Gernjak, W., Aryal, R., McCarthy, D., Palmer, A., et al. (2013). Sewage pollution in urban stormwater runoff as evident from the widespread presence of multiple microbial and chemical source tracking markers. *Sci. Total Environ.* 463, 488–496. doi: 10.1016/j.scitotenv.2013.06.020

Smith, A. M., Tau, N. P., Smouse, S. L., Allam, M., Ismail, A., Ramalwa, N. R., et al. (2019). Outbreak of *Listeria monocytogenes* in South Africa, 2017-2018: laboratory activities and experiences associated with whole-genome sequencing analysis of isolates. *Foodborne Pathog. Dis.* 16, 524–530. doi: 10.1089/fpd.2018.2586

Stenkova, A. M., Issaeva, M., and Rasskazov, V. A. (2008). Development of a multiplex PCR procedure for detection of *Yersinia* genus with identification of pathogenic species (*Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*). *Mol. Gen. Microbiol. Virol.* 23, 119–125. doi: 10.3103/S0891416808030038

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., and Schmidt, T. M. (2015). *rrn*DB: Improved Tools for Interpreting rRNA Gene Abundance in Bacteria and Archaea and a new Foundation for Future Development. Available online at: https://rrndb.umms.med.umich.edu (Accessed October 15, 2020).

Strauss, A., Reyneke, B., Waso, M., Ndlovu, T., Brink, C. J., Khan, S., et al. (2019). EMA-amplicon-based taxonomic characterisation of the viable bacterial community present in untreated and SODIS treated roof-harvested rainwater. *Environ. Sci. Water Res. and Technol.* 5, 91–101. doi: 10.1039/C8EW00613J

Tadokera, R., Bekker, L., Kreiswirth, B. N., Mathema, B., and Middelkopp, K. (2020). TB transmission is associated with prolonged stay in a low-socioeconomic high burdened TB and HIV community in Cape Town, South Africa. *BMC Infect. Dis.* 20:120. doi: 10.1186/s12879-020-4828-z

Tilevik, A. (2017). *Evaluation of Clustering Methods for Analyzing Drug Cytokine profiles.* Ph.D. thesis. Sweden, SE: Karlstad University.

United Nations (2018). *Sustainable Development Goal 6: Synthesis Report on Water and Sanitation*. New York, NY: United Nations Publications.

Van Heerden, J., Ehlers, M. M., Van Zyl, W. B., and Grabow, W. O. K. (2003). Incidence of adenoviruses in raw and treated water. *Water Res.* 37, 3704–3708. doi: 10.1016/S0043-1354(03)00245-8

Velayati, A. A., Farnia, P., Mozafari, M., Malekshahian, D., Farahbod, A. M., Seif, S., et al. (2015). Identification and genotyping of *Mycobacterium tuberculosis* isolated from water and soil samples of a metropolitan city. *Chest* 147, 1094–1102. doi: 10.1378/chest.14-0960

Viau, E. J., and Boehm, A. B. (2011). Quantitative PCR-based detection of pathogenic *Leptospira* in Hawai'ian coastal streams. *J. Water Health* 9, 637–646. doi: 10.2166/wh.2011.064

Wall, S., Ghanekar, K., McFadden, J., and Dale, J. W. (1999). Context-sensitive transposition of *IS6110* in *Mycobacteria*. *Microbiology* 145, 3169–3176. doi: 10.1099/00221287-145-11-3169

Waso, M., Khan, S., and Khan, W. (2018). Microbial source tracking markers associated with domestic rainwater harvesting systems: correlation to indicator organisms. *Environ. Res.* 161, 446–455. doi: 10.1016/j.envres.2017.11.043

Waso, M., Ndlovu, T., Dobrowsky, P. H., Khan, S., and Khan, W. (2016). Presence of microbial and chemical source tracking markers in roof-harvested rainwater and catchment systems for the detection of fecal contamination. *Environ. Sci. Pollut. Res.* 23, 16987–17001. doi: 10.1007/s11356-016-6895-7

Weis, J., and Seeliger, H. P. R. (1975). Incidence of *Listeria monocytogenes* in nature. *Appl. Microbiol.* 30, 29–32.

World Health Organisation (WHO) (2004). "Pathogenic *Mycobacteria* in water–a guide to public health consequences, monitoring and management," in *Emerging Issues in Water and Infectious Disease Series*, eds S. Pedley, J. Bartram, G. Rees, A. Dufour, and J. Cotruvo (London: IWA Publishing).

World Health Organisation (WHO) (2005). *Recreation and Disease. Plausibility of Associated Infections: Acute Effects, Sequelae and Mortality*. Geneva: World Health Organisation.

World Health Organisation (WHO) (2019). *Global Tuberculosis Report 2019*. Geneva: World Health Organization.

Yang, K., LeJeune, J., Alsdorf, D., Lu, B., Shum, C. K., and Liang, S. (2012). Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl. Trop. Dis.* 6:2. doi: 10.1371/journal.pntd.000148

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.

Copyright © 2021 Denissen, Reyneke, Waso, Khan and Khan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

METHODS
published: 20 May 2021
doi: 10.3389/fmicb.2021.673604

Evaluation of Low-Cost Phage-Based Microbial Source Tracking Tools for Elucidating Human Fecal Contamination Pathways in Kolkata, India

*Renuka Kapoor[1†], James Ebdon[2†], Ashutosh Wadhwa[1], Goutam Chowdhury[3], Yuke Wang[1], Suraja J. Raj[1], Casey Siesel[1], Sarah E. Durry[1], Wolfgang Mairinger[1], Asish K. Mukhopadhyay[3], Suman Kanungo[3], Shanta Dutta[3] and Christine L. Moe[1]**

[1] Center for Global Safe Water, Sanitation and Hygiene, Rollins School of Public Health, Emory University, Atlanta, GA, United States, [2] Environment and Public Health Research and Enterprise Group (EPHREG), University of Brighton, Brighton, United Kingdom, [3] ICMR – National Institute of Cholera and Enteric Diseases (NICED), Kolkata, India

OPEN ACCESS

Edited by:
Anicet Ramon Blanch, University of Barcelona, Spain

Reviewed by:
Maite Muniesa, University of Barcelona, Spain
Adelaide Almeida, University of Aveiro, Portugal

***Correspondence:**
Christine L. Moe
clmoe@emory.edu

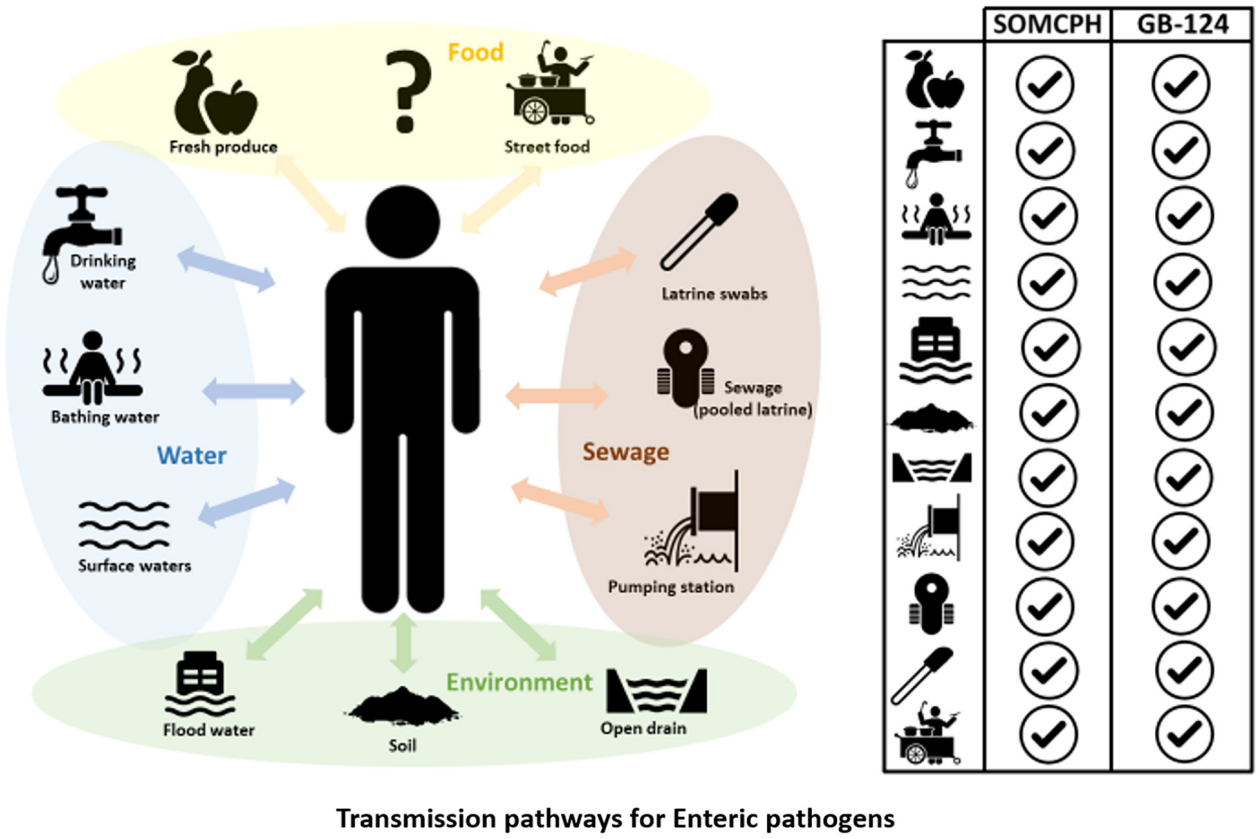
[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 27 February 2021
Accepted: 15 April 2021
Published: 20 May 2021

Citation:
Kapoor R, Ebdon J, Wadhwa A, Chowdhury G, Wang Y, Raj SJ, Siesel C, Durry SE, Mairinger W, Mukhopadhyay AK, Kanungo S, Dutta S and Moe CL (2021) Evaluation of Low-Cost Phage-Based Microbial Source Tracking Tools for Elucidating Human Fecal Contamination Pathways in Kolkata, India. Front. Microbiol. 12:673604. doi: 10.3389/fmicb.2021.673604

Phages, such as those infecting *Bacteroides* spp., have been proven to be reliable indicators of human fecal contamination in microbial source tracking (MST) studies, and the efficacy of these MST markers found to vary geographically. This study reports the application and evaluation of candidate MST methods (phages infecting previously isolated *B. fragilis* strain GB-124, newly isolated *Bacteroides* strains (K10, K29, and K33) and recently isolated *Kluyvera intermedia* strain ASH-08), along with non-source specific somatic coliphages (SOMCPH infecting strain WG-5) and indicator bacteria (*Escherichia coli*) for identifying fecal contamination pathways in Kolkata, India. Source specificity of the phage-based methods was first tested using 60 known non-human fecal samples from common animals, before being evaluated with 56 known human samples (municipal sewage) collected during both the rainy and dry season. SOMCPH were present in 40-90% of samples from different animal species and in 100% of sewage samples. Phages infecting *Bacteroides* strain GB-124 were not detected from the majority (95%) of animal samples (except in three porcine samples) and were present in 93 and 71% of the sewage samples in the rainy and dry season (Mean = 1.42 and 1.83 $\log_{10}$PFU/100mL, respectively), though at lower levels than SOMCPH (Mean = 3.27 and 3.02 $\log_{10}$PFU/100mL, respectively). Phages infecting strain ASH-08 were detected in 89 and 96% of the sewage samples in the rainy and dry season, respectively, but were also present in all animal samples tested (except goats). Strains K10, K29, and K30 were not found to be useful MST markers due to low levels of phages and/or co-presence in non-human sources. GB-124 and SOMCPH were subsequently deployed within two low-income neighborhoods to determine the levels and origin of fecal contamination in 110 environmental samples. *E. coli,* SOMCPH, and phages of GB-124 were detected in 68, 42, and 28% of the samples, respectively. Analyses of 166 wastewater samples from shared community toilets and 21 samples



GRAPHICAL ABSTRACT | Phages infecting *Bacteroides fragilis* strain GB-124 and non-source specific somatic coliphages (SOMCPH) were deployed for identification of fecal contamination pathways in Kolkata, India. Analyses of environmental samples representative of nine different exposure pathways, pooled sewage from shared community toilets and sewage from pumping stations showed the presence of both SOMCPH and GB-124 phages in all the sample types.

from sewage pumping stations from the same districts showed that SOMCPH were present in 100% and GB-124 phages in 31% of shared toilet samples (Median = 5.59 and <1 $\log_{10}$ PFU/100 mL, respectively), and both SOMCPH and GB-124 phages were detected in 95% of pumping station samples (Median = 5.82 and 4.04 $\log_{10}$ PFU/100 mL, respectively). Our findings suggest that GB-124 and SOMCPH have utility as low-cost fecal indicator tools which can facilitate environmental surveillance of enteric organisms, elucidate human and non-human fecal exposure pathways, and inform interventions to mitigate exposure to fecal contamination in the residential environment of Kolkata, India.

Keywords: fecal contamination, environmental surveillance, bacteriophages, GB-124, exposure pathways, transmission, low-income

INTRODUCTION

Exposure to enteric pathogens is associated with a significant health burden, largely borne by young children living in low-income settings (Kosek et al., 2003). Globally, enteric infections represent the third leading cause of death among children under five, accounting for approximately 589,000 deaths in 2017 (Global Burden of Disease Collaborative Network, 2018). Advances in methods to characterize human exposure to enteric pathogens have lagged behind, partly due to non-availability of resources in the settings where enteric diseases carry a disproportionally high burden of disease (Goddard et al., 2020).

In order to design prevention and control strategies necessary to reduce exposure to fecal contamination and enteric pathogens, there is a need to improve enteric disease surveillance strategies and understanding of the role of environment in the transmission of enteric pathogens. While clinical surveillance

can be used to monitor the transmission of infection within a community, effective implementation of prevention and control strategies requires understanding the environmental transmission pathways for these pathogens. In addition to food and drinking water, soil, surface water, open drains, and sewage-irrigated raw produce can serve as vehicles of exposure to enteric pathogens (Wang et al., 2017; Bauza et al., 2020; Holcomb et al., 2020). However, all of these vehicles can become contaminated with fecal matter from a variety of sources (Bauza et al., 2020). This is especially true in high density, low-resource settings in low- and middle-income countries, where poor sanitation infrastructure and fecal sludge management, coupled with unsuitable urban animal husbandry practices and stray animal populations, contribute to fecal contamination of the environment by both human and non-human sources.

The majority of enteric disease surveillance programs focus on clinical surveillance of diarrheal cases. However, in the recent years, there has been an increasing paradigm shift towards improving environmental surveillance and running it in parallel with clinical surveillance (Asghar et al., 2014; Gumede et al., 2018; Andrews et al., 2020; Bivins et al., 2020; Matrajt et al., 2020). Samples from wastewater collection systems and treatment plants offer a means to effectively assess pathogen burden and transmission at a population level (Gardy and Loman, 2018).

Historically, fecal indicator bacteria (FIB), such as *E. coli* and enterococci have been used to evaluate both the sanitary quality of water and the performance of wastewater and drinking water treatment processes (World Health Organization (WHO), 2017). However, whilst FIB have been shown to exhibit broadly similar levels of inactivation by treatment processes and environment stresses as some bacterial pathogens, they are less indicative of more resistant microorganisms (e.g., viruses or protozoa). This has called into question their value as universal surrogate indicators for enteric pathogens (AWPRC Study Group on Health-Related Water Microbiology, 1991; Grabow et al., 2001; Jiang et al., 2001; McQuaig et al., 2006; Field and Samadpour, 2007; Stoeckel and Harwood, 2007; World Health Organization (WHO), 2017). Moreover, as FIB are widely distributed in both human and animal feces, they do not distinguish the source of contamination (Malakoff, 2002) and have even been shown to multiply in certain environmental niches (Solo-Gabriele et al., 2000; Vanden Heuvel et al., 2010; Sadowsky and Whitman, 2011). Therefore, FIB may not correlate well with the magnitude of human enteric pathogens in environmental samples, particularly when non-human fecal sources are present (Odagiri et al., 2016).

Despite recent advances in molecular-based diagnostic techniques for rapid genomic surveillance of specific enteric pathogens (Diemert and Yan, 2019; Saha et al., 2019) and more recently SARS-CoV-2 (Ahmed et al., 2020; Medema et al., 2020; Wu et al., 2020), technological difficulties associated with low or highly variable concentrations in samples (compared to gut commensal microbes), continues to make routine pathogen detection challenging and potentially costly, even in high-income countries. This is largely due to the need for concentration (composite sampling), enrichment, and selection steps prior to identification, combined with the fact that many infective agents display seasonal and/or spatial variations in their prevalence and individual variation in shedding rates (Sinclair et al., 2008; Gerba et al., 2017). Therefore, although pathogens can be detected using molecular methods, challenges relating to the complexity of sample matrices (e.g., water, sludge, sediments), representative sampling, and technological complexity (e.g., DNA/RNA extraction, processing, sequencing and analytical protocols), makes their deployment in many low-resource settings highly infeasible for practical and routine purposes (Raes and Bork, 2008; Kyle Bibby and Peccia, 2013), as such settings often lack basic laboratory capacity even for indicator monitoring or pathogen surveillance (e.g., lack of cold chain, electrical power, reliable reagent supplies etc.).

Limitations associated with existing FIB, combined with technical challenges and costs associated with the detection of specific pathogens has driven the development of microbial source tracking (MST) methods to distinguish sources of fecal contamination and understand possible exposure to pathogens in the environment and through food and water. This includes a range of non-molecular approaches, such as those involving the detection of specific bacteriophages (Puig et al., 1999; Payan et al., 2005; Ebdon et al., 2007, 2012; Wicki et al., 2011; Jofre et al., 2014; McMinn et al., 2014; Diston and Wicki, 2015; Dias et al., 2018) and molecular-based markers (e.g., Bacteroidales and crAssphage qPCR markers) which have shown considerable promise as MST markers (Reischer et al., 2013; Stachler et al., 2017; Ahmed et al., 2018a,b; Mayer et al., 2018). However, despite the fact that studies of pollution sources have been conducted in a range of rural and urban low-income settings (e.g., Kenya, India, Mozambique, Bangladesh, Chile) most MST studies have involved shipping samples to laboratories in high-income countries for analysis (Jenkins et al., 2009; Odagiri et al., 2015, 2016; Harris et al., 2016; Bauza et al., 2019; Fuhrmeister et al., 2019, 2020; Holcomb et al., 2020; Jennings et al., 2020).

This study sought to develop MST indicators for environmental samples in Kolkata, India so that the presence, magnitude, and origin of fecal contamination could be rapidly identified (within 18-24 h) by staff at the National Institute of Cholera and Enteric Diseases (NICED). This involved determining the potential suitability of non-molecular MST tools based on the detection of phages capable of infecting certain bacteria such as *Bacteroides* spp. Although the host-range of *Bacteroides* is not restricted to humans, phages infecting certain strains of *B. fragilis and B. thetaiotaomicron*, such as HB-13, HSP-40, GA-17, have been reported to be found almost exclusively in fecal material of human origin (Payan et al., 2005; Jofre et al., 2014). However, research has also shown that phages that are capable of lysing a specific host strain of *Bacteroides* in one part of the world, are not necessarily detected in samples of similar origin (or at similar levels) in other parts of the world (Kator and Rhodes, 1992; Chung et al., 1998; Puig et al., 1999; Payan et al., 2005; Blanch et al., 2006), suggesting that different host strains may be needed for meaningful MST studies in different catchments or geographic regions. However, phages infecting certain *B. fragilis* strains (e.g., GB-124) have been shown to be restricted to human hosts (Ogilvie et al., 2012, 2018) and have been previously reported in Europe and North and South America as a potentially low-cost human fecal indicator

(Payan et al., 2005; Ebdon et al., 2007, 2012; McMinn et al., 2014, 2017; Prado et al., 2018). Geographic variations in the efficacy of MST markers have been reported in the extant literature (Ebdon and Taylor, 2006; Harwood et al., 2013) due to potential differences in host genetics, antibiotic usage, immunological factors and dietary differences. Consequently, it is essential to first validate the suitability of any MST approach in a given region, prior to deployment.

Here we report the development, evaluation, and suitability of low-cost candidate phage-based MST methods and non-source specific indicators for identifying fecal exposure pathways and sources in low-income urban districts of Kolkata. This study formed part of a larger SaniPath Typhoid project (grant number OPP1150697). The objectives of this study were to: (i) determine the ability of a well-established *Bacteroides* host strain (*B. fragilis* GB-124), less well-established *Kluyvera* host strain (*K. intermedia* ASH-08 - which had recently displayed promising human specificity in South East England) and newly-isolated *Bacteroides* hosts (K-10, K-29, and K-33) to sensitively and specifically quantify human fecal contamination through culture-based detection of phages infecting these strains in a range of environmental samples in Kolkata, India; (ii) deploy the most promising human-specific MST tool(s) and non-source specific indicators to identify fecal contamination sources in drinking water, bathing water, flood water, drain water, surface water, soil, raw produce, street food, and on swabs from shared community toilets in two low-income urban districts; and (iii) explore the influence of precipitation/season on the presence and levels of these phages in samples of municipal and domestic sewage and impacted surface waters from across the city.

MATERIALS AND METHODS

Study Setting

Kolkata, capital of the State of West Bengal in India, is divided into 15 boroughs and 141 wards (administrative units) under the Kolkata Municipal Corporation (KMC) area. It is estimated that 32.4% of the total population of Kolkata lives in 5,500 overcrowded low-income areas (Ray, 2017) where inhabitants do not have access to basic amenities including improved water and sanitation. The study area included low-income neighborhoods within wards 58 and 59 in the eastern part of KMC; these areas are representative of other low-income parts of the city and are characterized by poor housing conditions, intermittent supply of piped municipal water, and multiple households sharing pour-flush toilets and water taps. Surface water bodies, such as canals and ponds, are commonly used for bathing and domestic hygiene. Livestock ownership (74% households, unpublished findings) is chiefly comprised of poultry, cattle, goats and pigs. Most household wastewater is collected by an underground piped sewerage system, and some is collected in open drains that overflow during the rainy season. Underground potable water distribution lines and sewerage pipes often lie close to each other and are prone to leakage and cross-contamination (Singh et al., 2015). Other factors known to facilitate the transmission of diarrheal pathogens, such as poor hygiene, shared sanitation facilities, and the lack of adequate hand washing/facilities are prevalent. Vending of prepared street food and fresh produce (cut fruit), is widely practiced, and young children and adults reported frequent consumption of these foods.

Sample Collection

A total of 413 samples were collected and tested from a variety of known and unknown sources (animal feces, sewage, municipal wastewater, exposure pathways) over a 36-month period from July 2017 to March 2020.

Sewage Samples

In an attempt to isolate new *Bacteroides* host strains potentially suitable for phage detection in India and to establish whether the existing *B. fragilis* strain GB-124 and more recently isolated *K. intermedia* strain ASH-08 (which was isolated from raw municipal United Kingdom sewage and which appears to be capable of recovering high levels of phages in samples impacted by human fecal contamination in in South East England)could be used in this study setting, 100 mL sewage effluent samples were collected in 1 L sterile polyethylene bottles from the open canals, underground sewer manholes and open drains in the two study areas. Sample sites that were known to receive human fecal inputs were selected based on expert opinions of the field staff from NICED. All samples were kept on ice and transported to the lab for further processing within 4 h. Unfiltered sewage samples (n = 4, collected in Aug 2017) (serially diluted) were used for the isolation of potential *Bacteroides* host strains, whereas sewage samples [n = 56, collected between July-Aug 2017 (rainy) and March-May 2018 (dry season)] filtered through 0.22 μm polyvinylidene difluoride (PVDF) (or alternatively PES) membrane syringe filters (Millipore, USA) were used for evaluating the presence of phages against *B. fragilis* (GB-124) and *K. intermedia* (ASH-08) and SOMCPH(WG-5).

As part of our environmental surveillance study, two types of sewage samples were also collected, 1) wastewater accessed from manholes adjacent to shared toilets (n = 166 collected between Sept 2019 – March 2020) used by ∼ 100-250 residents in the two study neighborhoods, and 2) wastewater from selected pumping stations (n = 21 collected between Nov 2019 – March 2020) located across the city and representing excreta from >100,000 inhabitants. All samples were kept on ice and transported to the lab for further processing within 4 h of collection for enumeration of fecal indicator bacteria (FIB) (*E. coli*) and phages as described in sections "Enumeration of FIB (*E. coli*)" and "Enumeration of Bacteriophages".

Animal Fecal Samples

A total of 60 fecal samples, comprised of 10 samples each from five common types of animals in the study neighborhoods (cattle, chicken, goats, pigs, and dogs) and an additional 10 'mixed animal' fecal samples (containing samples of each of the five species) were analyzed to check the host specificity of *B. fragilis* (GB-124) and *K. intermedia* (ASH-08). On each occasion, two grams (wet weight) of animal feces was diluted 1:10 (w:v) in sterile DI water, vortexed until visible clumps were dispersed, and centrifuged at 4,500 x g for 10 min. The

supernatant was filtered using a 0.22 μm PVDF (or alternatively PES) membrane syringe filter and analyzed for the presence of phages in accordance with standardized methods (section "Isolation of new Bacteroides hosts" below).

Environmental Samples

A key informant interview (KII) was conducted in each of the wards with a community leader (i.e., local political leaders or NGO representatives) who had a good understanding about the WASH facilities, environmental characteristics, and how community members interacted with their environment. The information from the KIIs, as well as self-reported exposure behavior information collected from people in the study neighborhoods, was used to select environmental samples representative of identified exposure pathways (along which fecal contaminants may be transmitted). Potential fecal exposure pathways included: (1) municipal drinking water supplied by KMC and accessed through public taps (both direct supply and stored household water), (2) water from both municipal and non-municipal supplies that was stored and used for bathing, (3) surface water from community ponds, (4) open drain water carrying liquid and solid waste, including rainwater, flood water, sewage and water from household activities, (5) flood water that remained stagnant for at least one hour after rain, (6) soil from communal areas where people gathered and/or children played, (7) swabs from walls and door handles of shared community toilets, (8) fresh produce that was commonly consumed raw as salad or garnish including tomato, cucumber, coriander leaves, etc. (9) street food sold on the streets and commonly consumed by children and adults including *phuchka* (a round, hollow pastry filled with a mixture of flavored water, tamarind sauce, chickpea, potato and onion), *chow-mein* (stir-fried noodles with vegetables) and cut fruit (e.g., water melon and pineapple). Sample collection was performed according to protocols described in the SaniPath tool that has previously been used to assess risk of exposure to fecal contamination across multiple pathways in several cities (Raj et al., 2020). All samples were kept on ice and transported to the lab for further processing within 4 h of collection for enumeration of FIB and phages as described in sections "Enumeration of FIB (*E. coli*)" and "Enumeration of Bacteriophages."

Isolation of New Bacteroides Hosts

The procedure to isolate new *Bacteroides* spp. was performed as described in Payan et al., 2005 and Livingston et al., 1978. 100 μL of unfiltered sewage was serially diluted and plated onto *Bacteroides* bile esculin (BBE) agar (Livingston et al., 1978) and incubated at 37°C overnight in anaerobic jars where anaerobiosis was achieved using commercial anaerobic generators (Anaerogen, Oxoid, United Kingdom). Esculin-positive (black) colonies were picked and plated for pure culture on BBE agar, both under aerobic and anaerobic conditions. A Gram stain test of isolates that grew under strict anaerobic conditions only was conducted to ensure that only Gram-negative obligate anaerobic rods were selected for subsequent phage testing. All potential host strains were grown in the *Bacteroides* phage recovery medium (BPRM) broth and tested for their ability to recover phage from known human and non-human fecal samples.

Enumeration of FIB (*E. coli*)

All the environmental samples were processed within 6 h of collection and analyzed for *E. coli* using membrane filtration and m-ColiBlue24 broth (Hach, CO., United States) according to the manufacturer's instructions. Each sample was analyzed in duplicate without dilution (varying volumes) and/or after dilution (1 ml volume) in phosphate buffered saline (PBS). Drinking and bathing water were analyzed without dilution in 100 mL and 10 mL volumes. Surface water and flood water samples were tested without dilution (10 mL, 1 mL) and at 1:10 and $1:10^2$ dilutions and drain water at $1:10^3$, $1:10^4$, and $1:10^5$ dilutions. Soil samples were mixed (10g/20 mL) in PBST (PBS with 0.04% of Tween 80) and homogenized with a shaker (Innova 2,300 platform shaker, Eppendorf, Inc., United States) for 30 min. The homogenized samples were allowed to settle for another 30 min, and the soil suspension was tested without dilution (1 mL) and at 1:10 and $1:10^2$ dilutions in PBS. Swabs (2 per sampling site) from community toilets shared by multiple households were washed in PBST (2 x 10 mL for each swab) and homogenized with a vortexer (Corning vortex mixer, Fisher Scientific, United States) for 30 s, followed by incubation at room temperature for 5 min and mixed again for another 30 s. After mixing, 1- and 10-mL volumes of the swab eluates were analyzed by membrane filtration. Fresh produce samples were rinsed with 500 mL PBST in a 2 L Whirl-Pak™ bag and incubated at 37°C for 10 min, after which the bag was shaken manually for 30 s and was then massaged gently to dislodge the surface contamination. After shaking for an additional 30 s, the produce was removed from the Whirl-Pak™ bag, weighed, and the eluate was analyzed without dilution (1 mL) and at 1:10 dilution in PBS. Street food samples were homogenized in PBST (25 g/225 mL) for 1 min using a sterile BagMixer bag and BagMixer 400 CC (Interscience Laboratory Inc., Woburn, MA) at speed 4 with gap at-3 mm. The homogenate was analyzed without dilution (1 mL) and at 1:10 dilution in PBS.

After processing, all the samples were filtered through 47 mm, 0.45 μm membrane filter (Millipore, United States) using a vacuum manifold (Tarsons, Kolkata, India). The filter was then placed in a 50 mm Petri dish containing a filter pad soaked with 2 ml of m-ColiBlue broth (Hach, CO., United States). The plates were incubated facing up at 37°C and examined after 24 h to record number of blue colonies (*E. coli*), reported as CFU (colony forming units). All water samples were reported as CFU/100 mL, soil and street food were reported as CFU/gram, fresh produce were reported as CFU/single serving, and shared household toilet swabs were reported as CFU/swab. Colony counts in the range of 1-200 were considered valid, and counts >200 were recorded as TNTC (too numerous to count). Plates showing smudged growth obscured by sample sediment were recorded as TDTC (too difficult to count), and plates with no growth on the membrane were recorded as ND (non-detectable).

Enumeration of Bacteriophages

Samples were tested for phages infecting the newly isolated presumptive *Bacteroides* hosts, *B. fragilis* (GB-124), and

SOMCPH (WG-5) using the double-agar-layer method as described in ISO-10705-4 (Anon, 2001) and ISO-10705-2 (Anon, 2000), respectively. *K. intermedia* (ASH-08) was cultured in accordance with ISO-10705-4 (Anon, 2001) and phages enumerated using the double-agar-layer method.

Preparation of Host Strain Cultures

Tubes with approximately 15 mL of BPRM broth (BPRMB) were inoculated with 1 mL of overnight culture of strains GB-124 and ASH-08. The tubes were filled completely with broth to ensure anaerobic conditions and incubated at 37°C until the culture reached an optical density of 0.33 at 620 nm. Similarly, tubes containing 10 mL modified Scholtens' (MS) broth were inoculated with 1 mL of strain WG-5 (SOMCPH) and incubated at 37°C for growth to optical density of 0.33 at 600 nm. The cultures were placed on ice and used within 4 h.

Processing of Sewage and Environmental Samples

The sewage and aqueous environmental samples (drinking water, bathing water, surface water, open drain water, and flood water) were tested without any prior processing or dilution. Samples of fresh produce, street food and swabs from shared community toilets, were processed as described in section "Enumeration of FIB (*E. coli*)" and tested without dilution. For soil samples, 50 grams of soil was mixed with 100 mL of 10% beef extract (pH 7.2) and homogenized with a shaker (Innova 2,300 platform shaker, Eppendorf, Inc., United States) for 30 min. The homogenized samples were centrifuged at 1,500 *g* for 15 min, and the supernatant was filter sterilized (using 0.22 μm PVDF (or PES) membrane syringe filter).

Each sample was analyzed in duplicate (at either 1 mL and 100 μL volumes) and added to 1 mL of the host bacterium in a sterile 10 mL disposable test tube containing 2.5 mL of semi-solid BPRM agar (BPRMA) (for phages against GB-124 and ASH-08) or MS agar (MSA) (for SOMCPH). The contents of the tubes were gently vortexed and poured onto the surface of BPRMA plates supplemented with kanamycin monosulfate (100μg/mL) and MSA plates supplemented with nalidixic acid (100μg/mL). In order to increase the sensitivity of the assay in samples exhibiting low concentrations of phages, an increased sample volume (5 mL) was added to 5 mL of the *Bacteroides* and *Kluyvera* host strains and 7 mL of full strength BPRMA which was then poured onto a monolayer of BPRMA in a 90 mm Petri plate. All plates were left to set at room temperature, inverted and incubated at 37°C for 18-24 h either aerobically (MSA) or anaerobically (BPRMA) using anaerobic jars and anaerobic generators (Anaerogen, Oxoid, United Kingdom). Visible plaques (zones of lysis) in a confluent lawn of the host bacterium were counted as plaque forming units (PFU). PFU counts <300 were considered countable, counts >300 were recorded as TNTC and complete clearing of the host bacterial lawn was recorded as "complete lysis."

Phage Pre-enrichment

The presence/absence of phages infecting *B. fragilis* (GB-124) and SOMCPH in certain environmental samples where levels can be low (e.g., drinking water or street food), was also determined following a pre-enrichment step. However, whilst pre-enrichment is useful for detecting the presence/absence of a human signal in such samples where levels can be low, the presence/absence of SOMCPH is less useful compared to quantitative results for this general fecal indicator. Therefore, wherever possible it is better to use the quantitative, non-enriched counts for SOMCPH. Pre-enrichment was not used on environmental samples such as sewage from shared toilets and from pumping stations.

For pre-enrichment a 50 mL volume of the processed environmental sample was added to a 125 mL screw-capped Schott bottle (Millipore Sigma, United States) containing 60 mL double strength (ds) BPRMB and 15 mL of GB-124 culture (grown to optical density of 0.33 at 620 nm as described in section "Enumeration of Bacteriophages"). The bottle was filled completely with dsBPRMB to ensure anaerobic conditions, capped tightly, and incubated at 37°C for 18-21 h. Following overnight incubation, 5 mL of the enrichment culture was filtered using the 0.22 μm PVDF (or PES) membrane syringe filter into a centrifuge tube and centrifuged at 3,000 *g* for 5 min. 1 mL and 100 μL of the supernatant from the centrifuge tube was carefully added to 1 mL the host bacterium in a sterile 10 mL disposable test tubes containing 2.5 mL of semi-solid BPRMA (for GB-124) or MSA (for SOMCPH).

The contents of the test tubes were gently vortexed and poured onto the surface of BPRMA plates supplemented with kanamycin monosulfate (100 μg/mL) and MSA plates supplemented with nalidixic acid (100 μg/mL). The plates were left to set, inverted and incubated at 37°C for 18-24 h anaerobically (BPRMA) using commercial anaerobic jars and generators (Anaerogen, Oxoid, United Kingdom) or aerobically (MSA). The presence of phages resulted in the production of visible plaques (zones of lysis) in a confluent lawn of the host bacterium and the results were expressed as positive or negative for presence of plaques.

Quality Control

Quality control measures were employed during both sample collection and lab analyses. One field blank of DI water was collected and processed on each sample collection trip. The sample collectors filled one 100 mL Whirl-Pak™ bag with sterile distilled water at the sample collection site to assess their aseptic technique. This blank was tested in the laboratory for *E. coli*, and presence of any growth was considered as an indication of cross-contamination during sample collection process. One laboratory blank (Phosphate Buffered Saline) per laboratory technician per day, one negative control (sterile DI water) per batch of m-coliBlue, and one positive control for phage testing (B124-21, a purified *Siphoviridae* phage specifically infecting *Bacteroides* GB-124 strain) per day were processed for quality control. About 6.6% (28/426) of the field blanks were positive for *E. coli*, and these samples were removed from subsequent analysis. However, successful positive and negative phage controls (100% compliance) indicated that the NICED laboratory procedures met the study QA/QC criteria.

Data Analysis

A value of 300 PFU was assigned to samples recorded as TNTC or "complete lysis" in phage enumeration assays. A value of 0.5 CFU was assigned for samples recorded as non-detectable and 200 CFU for samples recorded as TNTC or TDTC in the *E. coli* enumeration assay. The *E. coli* concentration was calculated using the recorded CFU and corresponding dilution factors. The following criteria was used for calculating the *E. coli* test results: (i) if the *E. coli* counts of all three dilutions were <1 CFU, the lowest diluted sample was used to estimate the concentration, (ii) if the *E. coli* counts of all three dilutions were >200 CFU, the highest dilution was used to estimate the concentration, (iii) if at least one *E. coli* count was within the detection limit (from 1 to 200 CFU) the average concentration of *E. coli* was calculated, ignoring the censored (out of detection limit) *E. coli* counts. Phage and *E. coli* concentrations for water and sewage samples were normalized to 100 mL and all concentrations were $\log_{10}$-tranformed before analyses. The Wilcoxon rank sum test was used to compare the differences in concentration of the three phage types by season, and the Spearman-rank test was used to examine correlation between the concentrations of phages (SOMCPH, phages of GB-124 and phages of ASH-08) in sewage samples. Logistic regression was used to test the association between *E. coli* concentration and presence/absence of SOMCPH and phages of *Bacteroides* strain GB-124 in different sample types. Results were considered significant at p-value <0.05. All statistical analyses were conducted using R software version 4.0.2 (R Core Team, 2020).

Ethics

The study was reviewed and approved by the institutional review board at Emory University and the ethics committee at NICED. Written informed consent was obtained from the respondents at the time of surveys.

RESULTS

Isolation of New Bacteroides Hosts

Processing and culture of four sewage samples from the city of Kolkata yielded three potentially useful anaerobic Gram-negative hosts (K-10, K-29 and K-33) from 33 presumptive *Bacteroides* isolates (**Supplementary Information Table 1**). However, further analysis revealed that isolates K-10, K-29, and K-33 were not capable of consistently recovering phages in 28 sewage samples (of known human origin) at concentrations similar to, or greater than strain GB-124 and so were not used beyond this point.

Analyses of Fecal Samples From Animal Sources

To determine the host specificity of GB-124 and ASH-08, a total of 60 animal fecal samples were analyzed for phages against GB-124 and ASH-08 along with SOMCPH. SOMCPH were detected in a majority (40 to 90%) of animal fecal samples at concentrations ranging from 3.47-7.02 $\log_{10}$ PFU/gm of feces, while phages against ASH-08 were detected in 20 to 80% of animal samples (except goat) at concentrations ranging from 2.30-7.21 $\log_{10}$ PFU/gm of feces. Phages infecting GB-124 strain were not detected in any of the animal fecal samples, except for three pig feces samples (Mean = 3.64 $\log_{10}$ PFU/gm of feces) (**Table 1**).

Analyses of Sewage Samples From Known Human Sources

A total of 56 sewage samples, 28 each during the rainy season (July-Aug 2017) and dry season (March to May 2018) were collected from the two wards in Kolkata. SOMCPH were detected in all samples, regardless of season (**Table 1** and **Figure 1A**), whereas phages infecting ASH-08 were detected in 89% (25/28) and 96% (27/28) of samples in rainy and dry season, respectively. Similarly, phages against GB-124 were detected in 93% (26/28) and 71% (20/28) of the samples in rainy and dry season, respectively (**Figure 1A**).

In the sewage samples, SOMCPH concentrations [(mean (range)] were highest [3.15 $\log_{10}$ PFU/100 mL (1.08 to 5.58)], followed by phages infecting ASH-08 [3.00 $\log_{10}$ PFU/100 mL (1.60 to 5.15)] (**Table 1**). The overall concentrations of phages of GB-124 [1.59 $\log_{10}$ PFU/100 mL (0.60 to 3.6)] were significantly (p value < 0.001) lower than those of the non-source-specific phages (SOMCPH and phages against ASH-08). Further analyses showed a seasonal variation in the concentrations of all the three phage types (**Figure 1B**). Concentrations of SOMCPH were significantly higher (p value = 0.01) during rainy season [3.27 $\log_{10}$ PFU/100 mL (2.30 to 4.03)] compared to dry season [3.02 $\log_{10}$ PFU/100 mL (1.08 to 5.58)]. Concentration of phages infecting ASH-08 were also higher in the rainy season [3.15 $\log_{10}$ PFU/100 mL (2.30 to 4.70)] compared to the dry season [2.86 $\log_{10}$ PFU/100 mL (1.60 to 5.15)]. The concentrations of phages of GB-124 were consistently the lowest when compared to SOMCPH and ASH-08 phages, both during rainy [1.42 $\log_{10}$ PFU/100 mL (0.60 to 2.35) and dry season [1.83 $\log_{10}$ PFU/100 mL (0.60 to 3.60)]. There was no significant correlation between concentrations of the three phage groups.

Analyses of Environmental and Sewage Samples

FIB (*E. coli*)

A total of 110 environmental samples (ranging from 2 to 22 samples per pathway) from the nine exposure pathways were collected over a period of 11 months (from May 2019 to March 2020). *E. coli* were detected in all samples of drain water and flood water, the majority (92%) of the swabs from shared community toilets, 79% of surface water, nearly 70% of soil, produce and street food samples, 27% of bathing water, and 17% of drinking water samples (**Table 2**). As expected, drain water had the highest level of *E. coli* [mean (SD)] [7.07 $\log_{10}$ CFU/100 mL (1.25)], followed by high concentrations in flood water [4.63 $\log_{10}$ CFU/100 mL (0.46)] and surface water [4.2 $\log_{10}$ CFU/100 mL (1.25)]. The samples of drinking and bathing water had the lowest frequency of *E. coli* detection and the lowest mean concentrations [0.92 $\log_{10}$ CFU/100 mL (1.5)] and [0.93 $\log_{10}$ CFU/100 mL (0.58), respectively]. Fresh produce displayed substantial fecal

TABLE 1 | Analyses of samples from animal and human origin for phages against *B. fragilis*, *K. intermedia*, and somatic coliphages (SOMCPH) in Kolkata (n = 116).

Source	N	*Bacteroides* phages (strain GB-124)			*K. intermedia* phages (strain ASH-08)			SOMCPH (strain WG-5)		
		Log_{10} PFU per gram of feces		% of positive samples	Log_{10} PFU per gram of feces		% of positive samples	Log_{10} PFU per gram of feces		% of positive samples
		Mean	Range		Mean	Range		Mean	Range	
Cattle	10	ND	ND	0	3.95	2.30-4.34	40	4.14	2.70-4.83	60
Goat	10	ND	ND	0	ND	ND	0	3.47	2.60-3.78	40
Pig	10	3.64	3.15-4.32	30	6.38	3.70-7.21	80	6.41	5.48-7.00	70
Dog	10	ND	ND	0	4.54	3.00-4.83	20	7.02	2.00-7.67	90
Chicken	10	ND	ND	0	5.63	5.21-5.95	50	6.20	3.23-7.04	90
Pooled animal*	10	3.95	–	10	6.08	5.60-6.43	100	6.58	5.91-6.86	100
Human (Sewage)	56	1.59	0.60-3.60	79	3.00	1.60-5.15	93	3.15	1.08-5.58	100

*PFU = Plaque Forming units; N, number of samples tested; ND, not detected (i.e., below detection threshold (<1 PFU/gram of feces). *fecal samples from the five animal species were mixed to generate pooled samples.*

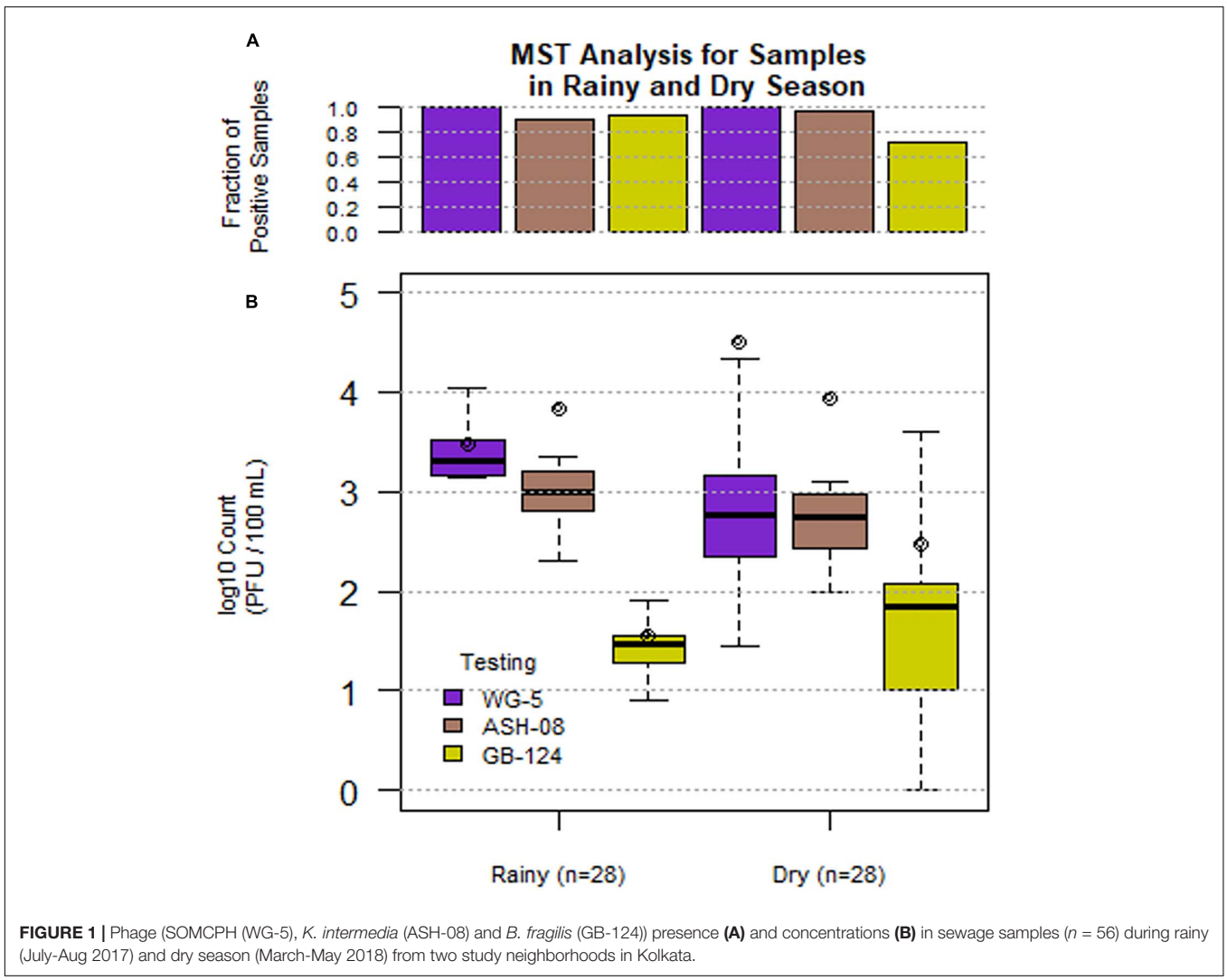


FIGURE 1 | Phage (SOMCPH (WG-5), *K. intermedia* (ASH-08) and *B. fragilis* (GB-124)) presence **(A)** and concentrations **(B)** in sewage samples (n = 56) during rainy (July-Aug 2017) and dry season (March-May 2018) from two study neighborhoods in Kolkata.

contamination as indicated by mean *E. coli* concentration of 3.7 $\log_{10}$ CFU/serving but street food was less contaminated [1.61 $\log_{10}$ CFU/gram (1.26)]. Soil samples from public areas where people gathered and/or children played were also highly contaminated with feces [mean *E. coli* 3.37 $\log_{10}$ CFU/gram]. However, swabs from surfaces likely to be touched in shared community toilets had only moderate *E. coli* levels [mean 1.78 $\log_{10}$ CFU/swab]. Of the 187 sewage samples collected as part of the SaniPath Typhoid environmental surveillance program over a period of 8 months (from August 2019 to March 2020), 95% of pumping station and 89% of the shared community toilet samples had similar high concentrations of *E. coli* [6.37 $\log_{10}$ CFU/100 mL (0.67) and 6.79 $\log_{10}$ CFU/100 mL (1.11), respectively] (**Table 2**).

Phages (SOMCPH and GB-124)

A total of 104 environmental samples from potential exposure pathways were tested for the presence of SOMCPH and phages infecting strain GB-124. Overall, SOMCPH were detected in 36% of these samples, while 15% tested positive for phages infecting GB-124 without pre-enrichment. The highest concentrations [(mean (range)] of SOMCPH and phages of GB-124 were detected in samples from open drain [5.29 $\log_{10}$ PFU/100 mL (4.74 to 5.48)] and [5.06 $\log_{10}$ PFU/100 mL (4.22 to 5.48), respectively] whilst the lowest concentrations of SOMCPH and phages of GB-124 were detected in samples of soil [1.96 $\log_{10}$ PFU/gram (0.30 to 3.29) and [1.67 $\log_{10}$ PFU/gram (0.40 to 3.78), respectively]. In order to increase the phage detection rates, a pre-enrichment step, followed by a concentration process (as described in section "Phage Pre-enrichment") was applied to a set of 69 samples. These overall detection rates using this enhanced detection approach increased from 36 to 42% and 15 to 28% for SOMCPH and GB-124, respectively, and both phage groups were detected in samples from all potential fecal exposure pathways except for GB-124 phage in produce (**Table 2**).

A total of 187 sewage samples were analyzed without the enrichment step because we expected both SOMCPH and GB-124 phages to be present in high concentrations. All of the samples from shared community toilets were positive for SOMCPH, but only 31% tested positive for phages infecting strain GB-124 (Median = 5.59 $\log_{10}$ PFU/100 mL and < 1 $\log_{10}$ PFU/100 mL, respectively). Almost all (95%) of the pumping station samples tested positive for both SOMCPH and phages of GB-124 (Median = 5.82 $\log_{10}$ PFU/100 mL and 4.04 $\log_{10}$ PFU/100 mL, respectively) (**Table 2**).

Association Between FIB and Phages

Logistic regression analyses indicated that there was no significant association between the concentration of *E. coli* and the odds of detecting SOMCPH in any of the environmental samples, except soil [Odds ratio (OR) (95% CI) = 3.32 (1.13, 9.75); p-value = 0.029]. Similarly, there was no significant association between *E. coli* concentrations and detection of phages of GB-124, with the exception of sewage from shared community toilets [OR (95% CI) = 1.17 (1.22, 2.36); p-value = 0.002] (**Supplementary Information Table 2**).

DISCUSSION

The overall goal of this study was to establish low-cost, geographically suitable tools to identify potential pathways of exposure to human fecal contamination that could serve as transmission routes of typhoid fever in Kolkata. Whilst phages against *Bacteroides* strain GB-124 had previously been successfully detected in municipal sewage (untreated and treated) from England, Cuba, Ireland, France, Portugal, Denmark, Brazil, Spain, Italy, United States, and Uganda (Ebdon et al., 2007, 2012; Vijayavel et al., 2010; McMinn et al., 2014; Dias et al., 2015, 2018; Purnell et al., 2016; Prado et al., 2018), the suitability of phages infecting GB-124 as indicators of human fecal contamination in India had not been characterized. The consistent detection of phages of GB-124 in sewage and a wide range of fecally contaminated environmental matrices, and its absence from pooled cattle, chicken, goat and dog feces suggest its potential as a culture-based indicator of human fecal contamination in Kolkata. Whilst this study did not yield new additional *Bacteroides* hosts suitable for recovering phages in Kolkata, the results of GB-124 deployment in this setting were promising for many types of environmental samples.

Specificity: Phages infecting GB-124 host strain were detected in the majority of the environmental samples with known human fecal contamination sources (sewage samples) and they were absent from the majority of fecal samples from common animal species in this setting (with the exception of three porcine fecal samples), despite close co-habitation of humans and animals in the study neighborhoods. Our findings are consistent with other studies where GB-124 phages were either not detected or detected at very low levels (close to the limit of detection) in non-human samples (Payan et al., 2005; Ebdon et al., 2007, 2012; Ogilvie et al., 2012; Harwood et al., 2013; McMinn et al., 2014; Diston and Wicki, 2015). These findings are also in line with the observations of Jebri et al. (2017), who suggest that source specificity is not absolute and, though seldom, *Bacteroides* host strains detect very low numbers of phages in the non-corresponding sources. Another possible reason for the presence of the GB-124 phages in a few porcine fecal samples could be the result of cross-contamination of porcine and human feces at the sample collection sites which were located in densely-populated household compounds. Interestingly, phages against *K. intermedia* (ASH-08) were also detected in non-human sources in Kolkata (with the exception of goats), which was not the case when this host was originally developed by the authors in the United Kingdom (data not shown), indicating a lack of discriminatory power (source specificity) and unsuitability for MST in Kolkata. On the other hand, SOMCPH were consistently detected in both human and non-human fecally contaminated samples (e.g., sewage and animal feces) indicating its utility as indicator of general (non-source specific) fecal contamination.

Sensitivity: Our analyses showed variable detection of phages in certain samples such as individual fecal samples (**Table 1**) as observed in other studies (Tartera and Jofre, 1987). This is not entirely unexpected since an individual fecal sample taken at a single point in time (during which phages may or may not be present), is unlikely to be representative of the total

intestinal microbiota, or an accurate assessment of community carriage and shedding. This is not in detriment of the prevalence of the phages but suggests that pooled samples and sewage (mixtures of numerous individuals) are more consistently likely to contain phages and as such may be more reliable MST targets. Our analyses of sewage samples showed the presence of all the three phage types, with levels of SOMCPH and phages of GB-124 similar to those reported earlier from untreated sewage (Ebdon et al., 2012; Dias et al., 2015, 2018; Purnell et al., 2015, 2016; McMinn et al., 2017). As expected, phages of GB-124 were present at lower concentrations than SOMCPH in sewage (Payan et al., 2005; Blanch et al., 2006; Ebdon et al., 2007, 2012; Nnane et al., 2012; Harwood et al., 2013; McMinn et al., 2014; Diston and Wicki, 2015) and were consistent with concentrations reported from Europe and United States of 3.90 $\log_{10}$ PFU/100 mL and 2.58 $\log_{10}$ PFU/100mL, respectively. The SOMCPH concentrations we detected in sewage in Kolkata were also similar to the concentrations observed in Europe [approx. 5.95 $\log_{10}$PFU/100mL] and the US [approx. 3.71 $\log_{10}$PFU/100mL] (Ebdon et al., 2007, 2012; McMinn et al., 2014, 2017; Dias et al., 2015; Purnell et al., 2015, 2016). The higher concentrations of SOMCPH are partly due to the fact that they are indicators of general fecal contamination and are not restricted to only human hosts, but all warm-blooded animals. Lower *Bacteroides* concentrations in some environmental sample types, such as sewage impacted surface waters, may be associated with increased rates of bacterial die-off due to greater sensitivity to temperature, the presence of oxygen and grazing by predators in certain environmental matrices (Ballesté and Blanch, 2010). No statistically significant correlation between concentrations of the three phage groups (SOMCPH, ASH-08 and GB-124) was found during this study. These findings are consistent with those of McMinn et al. (2014), who did not detect any significant correlation between concentrations of phages of GB-124 and SOMCPH during year-long study of wastewater effluents across the US. The seasonal variation in phage concentrations observed in Kolkata was similar for SOMCPH, but for GB-124 phages, contrasted with the findings of Ebdon et al. (2007), who reported higher concentrations of both SOMCPH and GB-124 in river water during high rainfall events as compared to low rainfall events. In Kolkata, the concentration of human fecal contamination in sewage may become more dilute because of the continued input of large volumes of rainwater into the sewers and canals during monsoon period. Whereas in some other settings, and especially during the 'first flush' rainfall may wash additional fecal contamination from land run-off into surface waters.

Lower concentrations of phages of GB-124 and SOMCPH compared to *E. coli* observed in certain matrices drove the need for the pre-enrichment and concentration process. This method modification appeared to enhance overall detection rates of SOMCPH and phages infecting GB-124, particularly in some specific types of samples (e.g., drinking water, surface water, open drain, flood water, street food, soil, and swabs from shared community toilets for phages infecting GB-124; and drinking water, surface water, street food, soil and swabs from shared community toilets for SOMCPH) as compared to phage detection without enrichment. However, the small sample sizes for many of the different environmental matrices make it difficult to draw conclusions about the impact of this method modification for specific types of samples. The pre-enrichment step is therefore

TABLE 2 | Fecal indicator bacteria (*E. coli*) and phages (SOMCPH and GB-124) detection (with and without enrichment) in environmental samples from two study neighborhoods and selected municipal pumping station samples in Kolkata (*n* = 297).

Sample type	Total (N)	n/N (%) positive for *E. coli*		n/N (%) positive for SOMCPH (WG-5)		n/N (%) positive for *Bacteroides* phage (GB-124)	
		n (%) positive	$\log_{10}$ CFU/100 mL Mean (SD)	Without enrichment	With enrichment	Without enrichment	With enrichment
Drinking water	11	3 (27)	0.92 (1.50)	1/11 (9)	1/8 (13)	0/11 (0)	1/8 (13)
Bathing water	6	1 (17)	0.93 (0.58)	1/6 (17)	NT	1/6 (17)	NT
Surface water	14	11 (79)	4.20 (1.25)	10/14 (71)	7/8 (87)	6/14 (43)	5/8 (62)
Open drain	6	6 (100)	7.07 (1.25)	4/6 (67)	4/6 (67)	2/6 (33)	3/6 (50)
Flood water	2	2 (100)	4.63 (0.46)	2/2 (100)	2/2 (100)	0/2 (0)	2/2 (100)
Produce	22	15 (68)	3.70 (1.80)	6/22 (27)	3/14 (21)	1/22 (5)	0/14 (0)
Street food	15	10 (67)	1.61 (1.26)	3/12 (25)	4/13 (31)	1/12 (8)	3/13 (23)
Soil	21	15 (71)	3.37 (1.12)	10/21 (48)	7/14 (50)	5/21 (24)	4/14 (29)
Swabs from shared toilets	13	12 (92)	1.78 (1.36)	0/10 (0)	1/4 (25)	0/10 (0)	1/4 (25)
Total	110	75 (68)		37/104 (36)	29/69 (42)	16/104 (15)	19/69 (28)
Sewage from shared toilets	166	148 (89)	6.79 (1.11)	166/166 (100)	NT	51/166 (31)	NT
Sewage from pumping station	21	20 (95)	6.37 (0.67)	20/21 (95)	NT	20/21 (95)	NT
Total	187	168 (90)		186/187 (99)		71/187 (38)	

N, Total number; n, number positive; CFU, Colony forming units; SD, standard deviation.
NT, not tested.

useful to determine presence or absence of MST markers (human signal) but might not be useful for SOMCPH when quantification of these indicators is required to ascertain the actual levels of fecal contamination in a given sample.

Detection of phages infecting GB-124 from municipal sewage samples collected at pumping stations was far greater (95%) than detection from wastewater samples collected from manholes adjacent to shared community toilets (31%), although the samples from the shared toilets were clearly from human fecal sources and spent less time in transit through sewage lines. This finding could possibly be due to the lower number of individuals contributing to the wastewater near the shared community toilets compared to municipal sewage from large catchment populations. Also, difficulty obtaining sufficient volumes of wastewater from the manholes meant that sample collectors sometimes added additional flush water to the nearest toilet, and this may have impacted the quality of the sample.

Phage-based MST approaches: Bacteriophages are attractive fecal indicators and MST tools for low-resource settings because they are relatively simple, low-cost, and do not require advanced laboratory conditions like PCR-based MST, or pathogen detection. Whilst SOMCPH have successfully been used as surrogates of pathogenic viruses with which to assess the efficacy of riverbank filtration of highly polluted surface waters in Delhi (Sprenger et al., 2014), this is the first study to describe the successful use of *Bacteroides* GB-124 phages for MST in India. Previous MST studies from India have used either antibiotic resistance analysis and DNA fingerprinting-based methods (Murugan et al., 2012), or *Bacteroidales* qPCR using human BacHum and animal markers (BacCow for ruminants, BacCan for canines) (Odagiri et al., 2015, 2016; Schriewer et al., 2015). For example, Odagiri et al. (2015, 2016) used molecular-based *Bacteroidales* fecal markers to measure human and domestic animal fecal contamination in community tube wells and ponds (n = 301) and to assess multiple exposure pathways in homes (n = 354), in rural settings in Odisha. However, whilst fecal indicator bacteria were detected and enumerated in country in this cross-sectional study of 60 villages in India, the quantitative PCR assays for the MST indicators (BacUni, BacHum and BacCow, BacCan) were performed on samples that were stored on ice for up to 2 months, prior to being sent to the United States for analysis. Similarly, molecular-based MST studies that have been conducted in a variety of rural and urban low-income settings, including Kenya, Mozambique, Bangladesh, and Chile, relied on shipping the samples to a lab in a high-income country to perform the analyses (Jenkins et al., 2009; Harris et al., 2016; Bauza et al., 2019; Fuhrmeister et al., 2019, 2020; Holcomb et al., 2020; Jennings et al., 2020). The findings of this Kolkata study are significant as they demonstrate that a phage-based MST assay offers a low-cost, rapid (18 h) MST approach that laboratories in the region can employ to elucidate human and non-human fecal contamination sources and transmission pathways in complex urban environments such as those present in Kolkata, India. Furthermore, phage-based MST methods such as those reported here are based on the detection of infectious viral particles within a sample (rather than fragments of DNA/RNA) and hence may provide additional information for making inferences about the potential viability and infectivity of other microorganisms (e.g., enteric pathogens) also present.

CONCLUSION

This is the first published study on the use of *Bacteroides* host strain GB-124 as an indicator of human fecal contamination in India. GB-124-based MST was found to be suitable for identifying potential exposure to human fecal contamination in densely populated, low-resource urban settings. Culture-based detection of GB-124 phage, coupled with the pre-enrichment and concentration steps described in this study, can be used by trained staff to routinely detect human fecal contamination in environmental samples, and can serve as a low-cost screening tool to enable more efficient environmental surveillance for human enteric pathogens by identifying subsets of samples that should be targeted for more advanced analyses for the detection of human-specific pathogens such as *S.* Typhi. The low-cost phage-based tools (GB-124 and SOMCPH) demonstrated here have potential to support fecal exposure assessments and evaluations of community-based sanitation and fecal sludge management interventions.

DATA AVAILABILITY STATEMENT

Raw data was generated at the National Institute of Cholera and Enteric Diseases (NICED), Kolkata, India. Derived data supporting the findings of this study are available from the corresponding author (CM) upon reasonable request.

AUTHOR CONTRIBUTIONS

RK and JE conceived and designed the experiments and prepared the manuscript. JE, RK, AW, and GC conducted the validation studies and trained the lab team. SK facilitated the study site selection and supervised field activities and sample collection. AM supervised the lab analyses of the field samples. SR, CS, and SED trained the field team in sample collection. WM designed data collection forms and conducted data analyses. WM and CS performed data monitoring. YW conducted data analyses. CM and SD conceived of and supervised the overall study. All authors contributed to the article and approved the submitted version.

FUNDING

This study was part of the SaniPath Typhoid Project funded by the Bill & Melinda Gates Foundation (Grant No. OPP1150697).

ACKNOWLEDGMENTS

We are grateful for the efforts of NICED field and lab personnel in sample collection and analyses. The environmental sample

collection was conducted by Suvankar Mandal, Biplap Jash, Buddhadeb Samaddar, Somnath Sarkar, Ashish Saha, Debabrata Maity, and Abhijit Guha. The lab analyses were conducted by Biswajit Sharma, Rinki Karmakar, Rajkumar, Sristi Biswas, Rajarshi Roy, Soma Banerjee, and Abhishek Das.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.673604/full#supplementary-material

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., et al. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci. Total Environ.* 728:138764. doi: 10.1016/j.scitotenv.2020.138764

Ahmed, W., Lobos, A., Senkbeil, J., Peraud, J., Gallard, J., and Harwood, V. J. (2018a). Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida. *Water Res.* 131, 142–150. doi: 10.1016/j.watres.2017.12.011

Ahmed, W., Payyappat, S., Cassidy, M., Besley, C., and Power, K. (2018b). Novel crAssphage marker genes ascertain sewage pollution in a recreational lake receiving urban stormwater runoff. *Water Res.* 145, 769–778. doi: 10.1016/j.watres.2018.08.049

Andrews, J. R., Yu, A. T., Saha, S., Shakya, J., Aiemjoy, K., Horng, L., et al. (2020). Environmental surveillance as a tool for identifying high-risk settings for typhoid transmission. *Clin. Infect. Dis.* 71(Suppl. 2), S71–S78. doi: 10.1093/cid/ciaa513

Anon (2000). *ISO 10705-2. Detection and Enumeration of Bacteriophages. Part 2: Enumeration of Somatic Coliphages*. Geneva: International Standardisation Organisation.

Anon (2001). *ISO 10705-4. Detection and Enumeration of Bacteriophages Enumeration of Bacteriophages Infecting Bacteroides fragilis*. Geneva: International Standardisation Organisation.

Asghar, H., Diop, O. M., Weldegebriel, G., Malik, F., Shetty, S., El Bassioni, L., et al. (2014). Environmental surveillance for polioviruses in the global polio eradication initiative. *J. Infect. Dis.* 210(Suppl. 1), S294–S303. doi: 10.1093/infdis/jiu384

AWPRC Study Group on Health-Related Water Microbiology (1991). Bacteriophages as model viruses in water quality controlag. *Water Res.* 25, 529–545. doi: 10.1016/0043-1354(91)90126-B

Ballesté, E., and Blanch, A. R. (2010). Persistence of *Bacteroides* species populations in a river as measured by molecular and culture techniques. *Appl. Environ. Microbiol.* 76, 7608–7616. doi: 10.1128/AEM.00883-10

Bauza, V., Madadi, V., Ocharo, R. M., Nguyen, T. H., and Guest, J. S. (2019). Microbial source tracking using 16S rRNA amplicon sequencing identifies evidence of widespread contamination from young children's feces in an urban slum of Nairobi, Kenya. *Environ. Sci. Technol.* 53, 8271–8281. doi: 10.1021/acs.est.8b06583

Bauza, V., Madadi, V., Ocharo, R., Nguyen, T. H., and Guest, J. S. (2020). Enteric pathogens from water, hands, surface, soil, drainage ditch, and stream exposure points in a low-income neighborhood of Nairobi, Kenya. *Sci. Total Environ.* 709:135344. doi: 10.1016/j.scitotenv.2019.135344

Bibby, K., and Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* 47, 1945–1951. doi: 10.1021/es305181x

Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., et al. (2020). Wastewater-based epidemiology: global collaborative to maximize contributions in the fight against COVID-19. *Environ. Sci. Technol.* 54, 7754–7757. doi: 10.1021/acs.est.0c02388

Blanch, A. R., Belanche-Muñoz, L., Bonjoch, X., Ebdon, J., Gantzer, C., Lucena, F., et al. (2006). Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Appl. Environ. Microbiol.* 72, 5915–5926. doi: 10.1128/AEM.02453-05

Chung, H., Jaykus, L. A., Lovelace, G., and Sobsey, M. D. (1998). Bacteriophages and bacteria as indicators of enteric viruses in oysters and their harvest waters. *Water Sci. Technol.* 38, 37–44. doi: 10.1016/S0273-1223(98)00797-5

Dias, E., Ebdon, J. E., and Taylor, H. (2015). The application of removal coefficients for viruses in different wastewater treatment processes calculated using stochastic modelling. *Water Sci. Technol.* 71, 1382–1388. doi: 10.2166/wst.2015.086

Dias, E., Ebdon, J., and Taylor, H. (2018). The application of bacteriophages as novel indicators of viral pathogens in wastewater treatment systems. *Water Res.* 129, 172–179. doi: 10.1016/j.watres.2017.11.022

Diemert, S., and Yan, T. (2019). Clinically unreported salmonellosis outbreak detected via comparative genomic analysis of municipal wastewater *Salmonella* isolates. *Appl. Environ. Microbiol.* 85:e00139-19. doi: 10.1128/AEM.00139-19

Diston, D., and Wicki, M. (2015). Occurrence of bacteriophages infecting *Bacteroides* host strains (ARABA 84 and GB-124) in fecal samples of human and animal origin. *J. Water Health* 13, 654–661. doi: 10.2166/wh.2014.199

Ebdon, J., Muniesa, M., and Taylor, H. (2007). The application of a recently isolated strain of *Bacteroides* (GB-124) to identify human sources of faecal pollution in a temperate river catchment. *Water Res.* 41, 3683–3690. doi: 10.1016/j.watres.2006.12.020

Ebdon, J. E., Sellwood, J., Shore, J., and Taylor, H. D. (2012). Phages of *Bacteroides* (GB-124): a novel tool for viral waterborne disease control? *Environ. Sci. Technol.* 46, 1163–1169. doi: 10.1021/es202874p

Ebdon, J. E., and Taylor, H. D. (2006). Geographical stability of enterococcal antibiotic resistance profiles in Europe and its implications for the identification of fecal sources. *Environ. Sci. Technol.* 40, 5327–5332. doi: 10.1021/es0607180

Field, K. G., and Samadpour, M. (2007). Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res.* 41, 3517–3538. doi: 10.1016/j.watres.2007.06.056

Fuhrmeister, E. R., Ercumen, A., Pickering, A. J., Jeanis, K. M., Ahmed, M., Brown, S., et al. (2019). Predictors of enteric pathogens in the domestic environment from human and animal sources in rural Bangladesh. *Environ. Sci. Technol.* 53, 10023–10033. doi: 10.1021/acs.est.8b07192

Fuhrmeister, E. R., Ercumen, A., Pickering, A. J., Jeanis, K. M., Crider, Y., Ahmed, M., et al. (2020). Effect of sanitation improvements on pathogens and microbial source tracking markers in the rural Bangladeshi household environment. *Environ. Sci. Technol.* 54, 4316–4326. doi: 10.1021/acs.est.9b04835

Gardy, J. L., and Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20. doi: 10.1038/nrg.2017.88

Gerba, C. P., Betancourt, W. Q., and Kitajima, M. (2017). How much reduction of virus is needed for recycled water: a continuous changing need for assessment? *Water Res.* 108, 25–31. doi: 10.1016/j.watres.2016.11.020

Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017 (GBD 2017)*. Seattle, WA: Institute for Health Metrics and Evaluation (IHME).

Goddard, F., Ban, R., Barr, D. B., Brown, J., Cannon, J., Colford, J. M. Jr., et al. (2020). Measuring environmental exposure to enteric pathogens in low-income settings: review and recommendations of an interdisciplinary working group. *Environ. Sci. Technol.* 54, 11673–11691. doi: 10.1021/acs.est.0c02421

Grabow, W. O., Taylor, M. B., and de Villiers, J. C. (2001). New methods for the detection of viruses: call for review of drinking water quality guidelines. *Water Sci. Technol.* 43, 1–8.

Gumede, N., Okeibunor, J., Diop, O., Baba, M., Barnor, J., Mbaye, S., et al. (2018). Progress on the implementation of environmental surveillance in the African region, 2011-2016. *J. Immunol. Sci.* (Suppl. 4), 24–30.

Harris, A. R., Pickering, A. J., Harris, M., Doza, S., Islam, M. S., Unicomb, L., et al. (2016). Ruminants contribute fecal contamination to the urban household environment in Dhaka, Bangladesh. *Environ. Sci. Technol.* 50, 4642–4649. doi: 10.1021/acs.est.5b06282

Harwood, V. J., Boehm, A. B., Sassoubre, L. M., Vijayavel, K., Stewart, J. R., Fong, T. T., et al. (2013). Performance of viruses and bacteriophages for fecal source determination in a multi-laboratory, comparative study. *Water Res.* 47, 6929–6943. doi: 10.1016/j.watres.2013.04.064

Holcomb, D. A., Knee, J., Sumner, T., Adriano, Z., de Bruijn, E., Nalá, R., et al. (2020). Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in Maputo, Mozambique. *Int. J. Hyg. Environ. Health* 226:113496. doi: 10.1016/j.ijheh.2020.113496

Jebri, S., Muniesa, M., and Jofre, J. (2017). "General and host-associated bacteriophage indicators of faecal pollution," in *Global Water Pathogen Project. (A Farnleitner, and A. Blanch (eds) Part 2 Indicators and Microbial Source Tracking Markers)*, eds J. B. Rose, and B. Jiménez-Cisneros (E. Lansing, MI: Michigan State University). doi: 10.14321/waterpathogens.7

Jenkins, M. W., Tiwari, S., Lorente, M., Gichaba, C. M., and Wuertz, S. (2009). Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res.* 43, 4956–4966. doi: 10.1016/j.watres.2009.07.028

Jennings, W. C., Gálvez-Arango, E., Prieto, A. L., and Boehm, A. B. (2020). CrAssphage for fecal source tracking in Chile: covariation with norovirus, HF183, and bacterial indicators. *Water Res. X* 9:100071. doi: 10.1016/j.wroa.2020.100071

Jiang, S., Noble, R., and Chu, W. (2001). Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. *Appl. Environ. Microbiol.* 67, 179–184. doi: 10.1128/AEM.67.1.179-184.2001

Jofre, J., Blanch, A. R., Lucena, F., and Muniesa, M. (2014). Bacteriophages infecting *Bacteroides* as a marker for microbial source tracking. *Water Res.* 55, 1–11. doi: 10.1016/j.watres.2014.02.006

Kator, H., and Rhodes, M. (1992). *Evaluation of Bacteroides fragilis Bacteriophages, A Candidate Human-Specific Indicator of Fecal Contamination from Shellfish Growing Waters. Final Report*. Gloucester Point, VA: Virginia Institute of Marine Sciences, College of Williams and Mary.

Kosek, M., Bern, C., and Guerrant, R. L. (2003). The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull. World Health Organ.* 81, 197–204.

Livingston, S. J., Kominos, S. D., and Yee, R. B. (1978). New medium for selection and presumptive identification of the *Bacteroides fragilis* group. *J. Clin. Microbiol.* 7, 448–453.

Malakoff, D. (2002). Water quality. Is *E. coli* distinct enough to join the hunt? *Science* 295:2353. doi: 10.1126/science.295.5564.2353

Matrajt, G., Lillis, L., and Meschke, J. S. (2020). Review of methods suitable for environmental surveillance of *Salmonella* Typhi and Paratyphi. *Clin. Infect. Dis.* 71(Suppl. 2), S79–S83. doi: 10.1093/cid/ciaa487

Mayer, R. E., Reischer, G. H., Ixenmaier, S. K., Derx, J., Blaschke, A. P., Ebdon, J. E., et al. (2018). Global distribution of human-associated fecal genetic markers in reference samples from six continents. *Environ. Sci. Technol.* 52, 5076–5084. doi: 10.1021/acs.est.7b04438

McMinn, B. R., Ashbolt, N. J., and Korajkic, A. (2017). Bacteriophages as indicators of faecal pollution and enteric virus removal. *Lett. Appl. Microbiol.* 65, 11–26. doi: 10.1111/lam.12736

McMinn, B. R., Korajkic, A., and Ashbolt, N. J. (2014). Evaluation of *Bacteroides fragilis* GB-124 bacteriophages as novel human-associated faecal indicators in the United States. *Lett. Appl. Microbiol.* 59, 115–121. doi: 10.1111/lam.12252

McQuaig, S. M., Scott, T. M., Harwood, V. J., Farrah, S. R., and Lukasik, J. O. (2006). Detection of human-derived fecal pollution in environmental waters by use of a PCR-based human polyomavirus assay. *Appl. Environ. Microbiol.* 72, 7567–7574. doi: 10.1128/AEM.01317-06

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., and Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. *Environ. Sci. Technol. Lett.* 7, 511–516. doi: 10.1021/acs.estlett.0c00357

Murugan, K., Prabhakaran, P., Al-Sohaibani, S., and Sekar, K. (2012). Identification of source of faecal pollution of Tirumanimuttar River, Tamilnadu, India using microbial source tracking. *Environ. Monit. Assess.* 184, 6001–6012. doi: 10.1007/s10661-011-2398-7

Nnane, D. K., Ebdon, J., and Taylor, H. (2012). The dynamics of feacal indicator organisms in a temperate river during storm conditions. *J. Water Clim. Change* 3, 139–150. doi: 10.2166/wcc.2012.145

Odagiri, M., Schriewer, A., Daniels, M. E., Wuertz, S., Smith, W. A., Clasen, T., et al. (2016). Human fecal and pathogen exposure pathways in rural Indian villages and the effect of increased latrine coverage. *Water Res.* 100, 232–244. doi: 10.1016/j.watres.2016.05.015

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., et al. (2015). Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* 502, 462–470. doi: 10.1016/j.scitotenv.2014.09.040

Ogilvie, L. A., Caplin, J., Dedi, C., Diston, D., Cheek, E., Bowler, L., et al. (2012). Comparative (meta)genomic analysis and ecological profiling of human gut-specific bacteriophage φB124-14. *PLoS One* 7:e35053. doi: 10.1371/journal.pone.0035053

Ogilvie, L. A., Nzakizwanayo, J., Guppy, F. M., Dedi, C., Diston, D., Taylor, H., et al. (2018). Resolution of habitat-associated ecogenomic signatures in bacteriophage genomes and application to microbial source tracking. *ISME J.* 12, 942–958. doi: 10.1038/s41396-017-0015-7

Payan, A., Ebdon, J., Taylor, H., Gantzer, C., Ottoson, J., Papageorgiou, G. T., et al. (2005). Method for isolation of *Bacteroides* bacteriophage host strains suitable for tracking sources of fecal pollution in water. *Appl. Environ. Microbiol.* 71, 5659–5662. doi: 10.1128/AEM.71.9.5659-5662.2005

Prado, T., Bruni, A. C., Barbosa, M., Bonanno, V., Garcia, S. C., and Sato, M. (2018). Distribution of human fecal marker GB-124 bacteriophages in urban sewage and reclaimed water of São Paulo city, Brazil. *J. Water Health* 16, 289–299. doi: 10.2166/wh.2017.011

Puig, A., Queralt, N., Jofre, J., and Araujo, R. (1999). Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater. *Appl. Environ. Microbiol.* 65, 1772–1776. doi: 10.1128/AEM.65.4.1772-1776.1999

Purnell, S., Ebdon, J., Buck, A., Tupper, M., and Taylor, H. (2015). Bacteriophage removal in a full-scale membrane bioreactor (MBR) – implications for wastewater reuse. *Water Res.* 73, 109–117. doi: 10.1016/j.watres.2015.01.019

Purnell, S., Ebdon, J., Buck, A., Tupper, M., and Taylor, H. (2016). Removal of phages and viral pathogens in a full-scale MBR: implications for wastewater reuse and potable water. *Water Res.* 100, 20–27. doi: 10.1016/j.watres.2016.05.013

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raes, J., and Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6, 693–699. doi: 10.1038/nrmicro1935

Raj, S. J., Wang, Y., Yakubu, H., Robb, K., Siesel, C., Green, J., et al. (2020). The SaniPath exposure assessment tool: a quantitative approach for assessing exposure to fecal contamination through multiple pathways in low resource urban settlements. *PLoS One* 15:e0234364. doi: 10.1371/journal.pone.0234364

Ray, B. (2017). Quality of life in selected slums in Kolkata: a step forward in the era of pseudo-urbanization. *Local Environ.* 22, 365–387. doi: 10:1080/13549839.2016.1205571

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Aström, J., et al. (2013). Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* 47, 8548–8556. doi: 10.1021/es304367t

Saha, S., Tanmoy, A. M., Andrews, J. R., Sajib, M., Yu, A. T., Baker, S., et al. (2019). Evaluating PCR-based detection of *Salmonella* Typhi and Paratyphi A in the environment as an enteric fever surveillance tool. *Am. J. Trop. Med. Hyg.* 100, 43–46. doi: 10.4269/ajtmh.18-0428

Sadowsky, M. J., and Whitman, R. L. (2011). *The Fecal Bacteria*. Washington DC: ASM Press.

Schriewer, A., Odagiri, M., Wuertz, S., Misra, P. R., Panigrahi, P., Clasen, T., et al. (2015). Human and animal fecal contamination of community water sources, stored drinking water and hands in rural India measured with validated microbial source tracking assays. *Am. J. Trop. Med. Hyg.* 93, 509–516. doi: 10.4269/ajtmh.14-0824

Sinclair, R. G., Choi, C. Y., Riley, M. R., and Gerba, C. P. (2008). Pathogen surveillance through monitoring of sewer systems. *Adv. Appl. Microbiol.* 65, 249–269. doi: 10.1016/S0065-2164(08)00609-6

Singh, R., Haque, S., and Grover, A. (2015). Drinking water, sanitation and health in Kolkata metropolitan city: contribution towards urban sustainability. *Geogr. Environ. Sustain.* 8, 64–81. doi: 10.15356/2071-9388_04v08_2015_07

Solo-Gabriele, H. M., Wolfert, M. A., Desmarais, T. R., and Palmer, C. J. (2000). Sources of *Escherichia coli* in a coastal subtropical environment. *Appl. Environ. Microbiol.* 66, 230–237. doi: 10.1128/aem.66.1.230-237.2000

Sprenger, C., Lorenzen, G., Grunert, A., Ronghang, M., Dizer, H., Selinka, H. C., et al. (2014). Removal of indigenous coliphages and enteric viruses during riverbank filtration from highly polluted river water in Delhi (India). *J. Water Health* 12, 332–342. doi: 10.2166/wh.2014.134

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Stoeckel, D. M., and Harwood, V. J. (2007). Performance, design, and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* 73, 2405–2415. doi: 10.1128/AEM.02473-06

Tartera, C., and Jofre, J. (1987). Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters. *Appl. Environ. Microbiol.* 53, 1632–1637. doi: 10.1128/AEM.53.7.1632-1637.1

Vanden Heuvel, A., McDermott, C., Pillsbury, R., Sandrin, T., Kinzelman, J., Ferguson, J., et al. (2010). The green alga, *Cladophora*, promotes *Escherichia coli* growth and contamination of recreational waters in Lake Michigan. *J. Environ. Qual.* 39, 333–344. doi: 10.2134/jeq2009.0152

Vijayavel, K., Fujioka, R., Ebdon, J., and Taylor, H. (2010). Isolation and characterization of *Bacteroides* host strain HB-73 used to detect sewage specific phages in Hawaii. *Water Res.* 44, 3714–3724. doi: 10.1016/j.watres.2010.04.012

Wang, Y., Moe, C. L., Null, C., Raj, S. J., Baker, K. K., Robb, K. A., et al. (2017). Multipathway quantitative assessment of exposure to fecal contamination for young children in low-income urban environments in Accra, Ghana: the SaniPath analytical approach. *Am. J. Trop. Med. Hyg.* 97, 1009–1019. doi: 10.4269/ajtmh.16-0408

World Health Organization (2017). *Sanitation Fact Sheet*. Geneva: World Health Organization.

Wicki, M., Auckenthaler, A., Felleisen, R., Tanner, M., and Baumgartner, A. (2011). Novel *Bacteroides* host strains for detection of human- and animal-specific bacteriophages in water. *J. Water Health* 9, 159–168. doi: 10.2166/wh.2010.165

Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., et al. (2020). SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *mSystems* 5:e00614-20. doi: 10.1128/mSystems.00614-20

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.

Copyright © 2021 Kapoor, Ebdon, Wadhwa, Chowdhury, Wang, Raj, Siesel, Durry, Mairinger, Mukhopadhyay, Kanungo, Dutta and Moe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 June 2021
doi: 10.3389/fmicb.2021.673306

The Use of Ribosomal RNA as a Microbial Source Tracking Target Highlights the Assay Host-Specificity Requirement in Water Quality Assessments

Annastiina Rytkönen[1], Ananda Tiwari[1], Anna-Maria Hokajärvi[1], Sari Uusheimo[2†], Asko Vepsäläinen[3], Tiina Tulonen[2] and Tarja Pitkänen[1,4]*

[1] Expert Microbiology Unit, Finnish Institute for Health and Welfare, Kuopio, Finland, [2] Lammi Biological Station, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland, [3] Environmental Health Unit, Finnish Institute for Health and Welfare, Kuopio, Finland, [4] Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer, France

Reviewed by:
Patricia Holden, University of California, Santa Barbara, United States
Rulong Liu, Shanghai Ocean University, China

***Correspondence:**
Annastiina Rytkönen
annastiina.rytkonen@thl.fi

†Present address:
Sari Uusheimo, Finnish Environment Institute, Helsinki, Finland

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *27 February 2021*
Accepted: *19 April 2021*
Published: *02 June 2021*

Citation:
Rytkönen A, Tiwari A, Hokajärvi A-M, Uusheimo S, Vepsäläinen A, Tulonen T and Pitkänen T (2021) The Use of Ribosomal RNA as a Microbial Source Tracking Target Highlights the Assay Host-Specificity Requirement in Water Quality Assessments. Front. Microbiol. 12:673306. doi: 10.3389/fmicb.2021.673306

For microbial source tracking (MST), the 16S ribosomal RNA genes (rDNA) of host-specific bacteria and mitochondrial DNA (mtDNA) of animal species, known to cause fecal contamination of water, have been commonly used as molecular targets. However, low levels of contamination might remain undetected by using these DNA-based qPCR assays. The high copy numbers of ribosomal RNA (rRNA) could offer a solution for such applications of MST. This study compared the performance of eight MST assays: GenBac3 (general *Bacteroidales*), HF183 (human), BacCan (dog), Rum-2-Bac (ruminant), Pig-2-Bac (swine), Gull4 (gull), GFD, and Av4143 (birds) between rRNA-based and rDNA-based approaches. Three mtDNA-based approaches were tested: DogND5, SheepCytB, and HorseCytB. A total of 151 animal fecal samples and eight municipal sewage samples from four regions of Finland were collected for the marker evaluation. The usability of these markers was tested by using a total of 95 surface water samples with an unknown pollution load. Overall, the performance (specificity, sensitivity, and accuracy) of mtDNA-based assays was excellent (95–100%), but these markers were very seldom detected from the tested surface water samples. The rRNA template increased the sensitivity of assays in comparison to the rDNA template. All rRNA-based assays (except Av4143) had more than 80% sensitivity. In contrast, only half (HF183, Rum-2-Bac, Pig-2-Bac, and Gull4) of rDNA-based assays reached this value. For markers targeted to bird feces, the use of the rRNA-based assay increased or at least did not change the performance. Regarding specificity, all the assays had >95% specificity with a DNA template, except the BacCan assay (71%). While using the RNA template for the assays, HF183 and BacCan exhibited only a low level of specificity (54 and 55%, respectively). Further, the HF183 assay amplified from multiple non-targeted animal fecal samples with the RNA template and the marker showed cross-amplification with the DNA template as well. This study recommends using the rRNA-based approach for MST assays targeting bird fecal contamination. In the case of

mammal-specific MST assays, the use of the rRNA template increases the sensitivity but may reduce the specificity and accuracy of the assay. The finding of increased sensitivity calls for a further need to develop better rRNA-based approaches to reach the required assay performance.

Keywords: microbial source tracking, performance analysis, field validation, ribosomal RNA, RT-qPCR, fecal contamination, surface water

INTRODUCTION

Fecal contamination of surface water from human and animal sources causes a public health risk when the water is used for drinking or food production, but also recreational, such as swimming and diving (Soller et al., 2010; Kauppinen et al., 2019). In many cases, the discharges of non-disinfected municipal wastewater effluents are considered as the main sources of fecal pathogens in watersheds (Hokajärvi et al., 2013; Anza et al., 2014; Kauppinen et al., 2014), and also urban and agricultural runoffs are known to contain fecal pathogens (Uusi-Kämppä and Heinonen-Tanski, 2008; Rankinen et al., 2016). The most prevalent causes of waterborne infections in Finland, zoonotic fecal bacterial pathogen *Campylobacter* spp. and human-specific pathogenic noroviruses, are known to retain their pathogenicity in the cold conditions in water environments very well (Hörman et al., 2004; Hokajärvi et al., 2013; Kauppinen et al., 2014; Guzman-Herrador et al., 2015).

Fecal indicator bacteria (FIB), *Escherichia coli*, and intestinal enterococci are used for monitoring fecal contamination levels in surface waters. However, the current approach of monitoring FIB cannot differentiate the source of contamination. It assigns equal waterborne health risk levels for fecal contamination despite that the occurrence of pathogens is often source-dependent. For example, contamination from human and cattle sources in recreational water may cause a higher gastrointestinal illness risk for swimmers than the contamination from gull, chicken, or swine feces (Soller et al., 2010). In addition to needs from precise human health risk assessment, source differentiation between human, animal, or persisted environmental contamination is a prerequisite for mitigation of contamination sources, i.e., the causes of increasing FIB counts observed during regulatory monitoring (Sinigalliano et al., 2013; Tiwari et al., 2018). Animal-specific markers have been developed, for example, for swine, cattle, and birds, which are animal hosts known to carry zoonotic pathogens (Green et al., 2012; Ryu et al., 2012; Boehm et al., 2013). Over the recent decades, the DNA-based quantitative PCR (qPCR) of molecular MST markers of a variable region of the 16S rRNA gene of host-specific microbes have been developed and applied worldwide (Mieszkin et al., 2009; Haugland et al., 2010; Ryu et al., 2012; Boehm et al., 2013).

The environmental RNA can be changed into complementary DNA (cDNA) with the reverse transcriptase process and can be amplified with the same primers and probes, as done in the DNA-based method. Earlier studies reported that the rRNA-based method is more sensitive than the DNA-based method (Matsuda et al., 2012; Pitkänen et al., 2013; Kapoor et al., 2014). Our study hypothesizes that the high sensitivity of the rRNA assays may improve the MST efficiency in water samples even during a low level of fecal pollution. To our knowledge, the performance characteristics of such rRNA approaches have not been described before. Further, mitochondrial DNA (mtDNA)-targeted marker assays are an interesting option for MST due to their high host specificity (Caldwell and Levine, 2009; Malla and Haramoto, 2020). The mtDNA assays detecting epithelial cells defoliated from the intestinal tract of the hosts have been applied elsewhere for MST, but not tested before for environmental water samples in Finland.

Sensitivity, specificity, and accuracy of assays have been used for the characterization of the performance of different microbial methods (Boehm et al., 2013; Tiwari et al., 2018; Ballesté et al., 2020). Among such criteria, specificity is the primarily important character for any given host-specific MST assay. A false-positive MST assay result may lead to incorrect measures when the source tracking is utilized to reduce contamination of water areas or as a risk assessment tool (Tiwari et al., 2018). Ideally, MST markers should be highly specific to targeted hosts, and the markers should exist with high copy numbers in fecal materials to enable detection even after a dilution of fecal material in environmental waters. This study evaluates, for the first time, the performance characteristics of the rRNA-based template for MST. A collection of animal feces, sewage effluents, and surface water samples in different geographical locations of Finland were analyzed to determine the specificity, sensitivity, and accuracy of GenBac3 (general fecal contamination), HF183 (human), Rum-2-Bac (ruminant), Pig-2-Bac (swine), Gull4 (gull), GFD (birds), Av4143 (birds), DogND5 (dog), SheepCytB (sheep), and HorseCytB (horse) to be applied for use in MST investigations in watersheds with different levels of contamination.

MATERIALS AND METHODS

Sampling Locations and Sampling

Sample materials were collected between June and October in 2018 from six cities of four different geographical regions in Finland: Northern Ostrobothnia (sites 1–6), Northern Savonia (sites 7–8), Pirkanmaa (sites 9–13), and Kanta-Häme (sites 14–20) (**Figure 1**). A total of 95 surface water samples were collected from 33 water sampling sites, of which nine were from rural and 12 from urban areas, and five from public bathing areas (**Supplementary Material 1** and **Supplementary Table 1**). Besides three sewage treatment plants, runoff water from two horse farms and a garden irrigation water site was sampled. In addition to the secondary (activated sludge) treated



FIGURE 1 | Water sample collection sites. Sites 1–6: Northern Ostrobothnia; sites 7–8: Northern Savonia; sites 9–13: Pirkanmaa; sites 14–20: Kanta-Häme (Map: National Land Survey of Finland; Sea area: Statistics Finland, Esri Finland). Each compartments on the map denotes 18 political regions of Finland, among them our study sampling covers four political regions.

TABLE 1 | Summary of fecal samples collected and analyzed.

Animal species	Sampling site	Sample count	Total
Cattle (*Bos taurus*)	2	16	16
Sheep (*Ovis aries*)	2	19	19
Bird, unknown species	2	27	34
	6	1	
	14	6	
Horse (*Equus caballus*)	6	1	19
	8	2	
	Other, Northern Savonia region	16	
Dog (*Canis lupus familiaris*)	6	1	21
	14	20	
Goose (*Anserinae* spp.)	9	13	13
Duck (*Anatidae* spp.)	13	1	2
	14	1	
Waterfowl*	14	2	2
Gull (*Laurus* spp.)	Other, Northern Ostrobothnia, previously stored	17	17
Swine (*Sus scrofa domesticus*)	Other, previously stored	6	6
Hare (*Lepus europaeus*)	14	2	2
Total			151

**Unspecified species.*

sewage effluent samples, waste water samples treated with LED-ultraviolet light (LED-UV, Led Future Inc., Kuopio, Finland) or exposed to wetland treatment were included (Uusheimo et al., 2018; Pitkänen et al., 2019). Half of the irrigation water samples were treated with LED-UV as well. Out of the 95 water samples, 85 were surface water and 10 were sewage effluent. Water samples of about 1 L were collected as grab samples into sterile plastic bottles.

Fecal samples for method development were collected nearby the water sampling sites within the same time frame. In addition, previously collected and stored gull and swine fecal material was used. Altogether, 151 fecal samples were used for method development (**Table 1**). The fecal samples were collected with a non-sterile disposable plastic spoon into a non-sterile re-sealable plastic bag.

Sample Transportation and Preservation

Samples were transported with sample coolers to the laboratory and processed within 24 h of sample collection. The water samples were filtrated onto 0.4-μm polycarbonate filters (as large volume as possible, 40–250 ml of effluents, and 50–600 ml of surface water) (Whatman Nuclepore Track-Etched Membranes, Sigma-Aldrich, United States). The membranes were frozen immediately after filtration and stored at −75°C or lower. A volume of 100 ml sterile-filtered water treated with diethyl pyrocarbonate (Invitrogen, Thermo Fisher Scientific, United States) was filtrated as negative filtration control. The fecal samples were distributed into 250-mg aliquots, frozen immediately, and stored at −75°C or lower.

Nucleic Acid Extraction and Processing

The nucleic acids from the water samples were extracted using a Chemagic DNA Plant Kit (Perkin Elmer, United States) as previously described in Inkinen et al. (2019). An All Prep DNA/RNA Mini Kit (Qiagen, United States) was used for RNA and DNA extraction from fecal samples according to Pitkänen et al. (2013). Negative extraction controls with extraction reagents only and the negative filtration controls were processed alongside the samples. DNA concentrations were measured using Qubitds DNA HS assay kits and the Qubit 3.0 Fluorometer (Life Technologies, Thermo Fisher Scientific, United States). Immediately after the extraction, RNA aliquots were further purified using a TURBO DNA-free DNase kit, following the manufacturer's instructions (Invitrogen, Thermo Fisher Scientific, United States). After purification, the RNA concentrations were measured using Qubit RNA HS assay kit and the Qubit 3.0 Fluorometer (Life Technologies, Thermo Fisher Scientific, United States). Following the extractions on the

same day, the purified RNA was converted into complementary DNA (cDNA) by using the SuperScript IV VILO Master Mix system for RT-PCR, following the manufacturer's instructions (Invitrogen, Thermo Fisher Scientific, United States) producing a total of 20 μl of each cDNA aliquot. To overcome the possible effect of reverse transcription inhibitors, the cDNA synthesis was performed using 8 μl as undiluted and 0.8 μl as 10-fold dilution of the total RNA. The total RNA was stored at −75°C or lower, while cDNA and DNA solutions were stored at −20°C until qPCR analysis.

Quantitative Polymerase Chain Reaction

The performance of a total of 11 bacterial or mitochondrial marker assays (**Table 2**) was tested against the collection of fecal or wastewater samples, using qPCR assays with cDNA and DNA extracts as templates. The qPCR assays were performed using the QuantStudio 6 Flex real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, United States). The TaqMan Environmental Master Mix 2.0 (Applied Biosystems, Thermo Fisher Scientific, United States) was used for TaqMan assays, and the Power SYBR Green PCR Master Mix (Applied Biosystems, Thermo Fisher Scientific, United States) was used for SYBR Green assays. Both DNA and cDNA were used as templates for most of the assays, but for assays targeting mtDNA, i.e., DogND5, HorseCytB, and SheepCytB assays, only a DNA template was used. The qPCR conditions are presented in **Supplementary Table 2**. The annealing temperatures described in the original assay publications (**Table 2**) were in accordance with the recommendations of the TaqMan and SYBR Green assay reaction mix manufacturers; thus, other cycling temperatures were not tested. The exception was optimization carried out for the assays GFD, HorseCytB, and SheepCytB, where a lower annealing temperature, 57°C, was considered. The performance characteristics (range of blanks, the limit of detection (LOD), amplification efficiency, R^2-value, range of quantification, sensitivity, and specificity) remained indifferent between the tested annealing temperatures or were better with the higher annealing temperature of 60°C (**Supplementary Table 2**).

The gBlocks Gene Fragments (Integrated DNA Technologies, United States), generated using reference sequences of the target sequences, selected by using the NCBI Nucleotide BLAST program (National Center for Biotechnology Information, United States National Library of Medicine), and including the exact primer and probe binding areas of the assays, were used for generating the standard curves. Ten-fold serial dilutions of these fragments were run with every assay with a total of 10 standard reactions per plate: 10^0, 2×10^1, 2×10^2, 2×10^3, 2×10^4, and 10^5 copies/μl. No template control (NTC) was run in duplicate with every standard set.

Undiluted and 1:10 and 1:100 diluted DNA and cDNA preparations in HyClone Water (GE Healthcare, Life Sciences, United Kingdom) were used to detect PCR inhibition. If inhibition was detected, the diluted samples were used for qPCR data generation. The limit of detection (LOD) was set as three copies per reaction, as suggested by Bustin et al. (2009). Background signals detected from negative extraction and filtration controls and LOD values were subtracted from all the results (clean NA) to generate the final data for the assay (**Supplementary Table 3**). Sample amount and dilution events from extraction, cDNA synthesis, and qPCR reaction (NA factor) were acknowledged. If the NA values were below the limit of quantification (LOQ), the result was treated as a present, but not quantitative, and therefore the value was set to half of the (0.5×) LOQ. The final data was calculated following the equations presented in **Supplementary Material 2**.

Consecution of the Assay Performance Analysis

The genetic materials extracted and purified from the fecal samples of the selected host animals were amplified against the tested MST assays, with both RNA-based and DNA-based approaches. The amplification of the assay on targeted hosts was reported as true positive, and no amplification from the non-targeted hosts was considered as true negative. The amplification from the samples of non-targeted hosts was reported as a false-positive detection, and no amplification from the samples of the targeted host was reported as a false-negative result. The performance of the assays was evaluated by calculating sensitivity and specificity and accuracy with the following formulas 1–3 (Boehm et al., 2013; Tiwari et al., 2018; Ballesté et al., 2020):

$$\text{Sensitivity} = \frac{TP\ (True\ Positive)}{TP + FN\ (False\ Negative)} \quad (1)$$

$$\text{Specificity} = \frac{TN\ (True\ Negative)}{TN + FP\ (False\ Positive)} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

The sensitivity, specificity, and accuracy values of a marker of more than 80% are considered as reliable and acceptable (US Environmental Protection Agency, 2005). Further, the accuracy of the performance was predicted with the Bayes theorem as done earlier (Kildare et al., 2007) by using formula (4). As the prior probabilities were unknown in this case, the posterior probabilities were calculated by varying the prior probability from the worst-case scenario (negative signals in all samples or probability = 0) to the best-case scenario (positive signals in all samples or probability = 1) as described by Lamendella et al. (2009).

$$P\ (A/B) = \frac{P\,(B/A) * P(A)}{P(B)} \quad (4)$$

In our case $P(B) = P(B/A)^* P(A) + P(B/A')^* P(A')$, where

A = originated from a targeted host, B = test positive with the source. The probability of recording contamination from a certain source, when there is truly contamination, is defined by the following equation of the Bayesian theorem. *P(B/A)* is the test positive when there is contamination (True positive); *P(B)* is the total number of positive-tested cases that can be truly positive [P(B/A)] and false-positive [P(B/A')].

Statistical Analysis

All data above the LOD was logarithmic transformed (Log_{10}) before further statistical analysis, as the original data did

TABLE 2 | The qPCR-assays used in the study.

Target (assay)	Primer or probe sequence (5′ to 3′)	Chemistry	Annealing temp. (°C)	Size (bp)	References
General *Bacteroidales* GenBac3	GenBactF3: GGGGTTCTGAGAGGAAGGT GenBactR4: CCGTCATCCTTCACGCTACT GenBactP2: 6-FAM-CAATATTCCTCACTGCTGCCTCCCGTA-ZEN/IBFQ	TaqMan	60	129	Siefring et al., 2008
Human-specific *Bacteroidalesdorei* HF183	HF183-1: ATCATGAGTTCACATGTCCG HfBthetR1: CGTAGGAGTTTGGACCGTGT HfBthetP1: 6-FAM-CTGAGAGGAAGGTCCCCCACATTGGA-ZEN/IBFQ	TaqMan	60	167	Haugland et al., 2010
Dog-specific *Bacteroidales* BacCan	BacCan-545f1: GGAGCGCAGACGGGTTTT BacUni-690r2: AATCGGAGTTCCTCGTGATATCTA BacUni-656p: 6-FAM-TGGTGTAGCGGTGAAA-BHQ1	TaqMan	60	145	Kildare et al., 2007
Ruminant-associated *Bacteroidales* Rum-2-Bac	RumBacB2-590F: ACAGCCCGCGATTGATACTGGTAA RumBac708Rm: CAATCGGAGTTCTTCGTGAT RumBacB2-626P: 6-FAM-ATGAGGTGGATGGAATTCGTGGTGT-ZEN/IBFQ	TaqMan	60	99	Mieszkin et al., 2010
Swine-specific *Bacteroidales* Pig-2-Bac	Pig-2-Bac41F: GCATGAATTTAGCTTGCTAAATTTGAT Pig-2-Bac163Rm: ACCTCATACGGTATTAATCCGC Pig-2Bac113MGB: 6-FAM-TCCACGGGATAGCC-MGB	TaqMan	60	116	Mieszkin et al., 2009
Gull-specific *Catellicoccus marimammalium* Gull4	qGull7F: CTTGCATCGACCTAAAGTTTTGAG qGull8R: GGTTCTCTGTATTATGCGGTATTAGCA qGull7P: 6-FAM-ACACGTGGGTAACCTGCCCATCAGA-ZEN/IBFQ	TaqMan	60	116	Ryu et al., 2012
Bird-specific *Helicobacter* spp. GFD	GFDF: TCGGCTGAGCACTCTAGGG GFDR: GCGTCTCTTTGTACATCCCA	SYBR Green	57	123	Green et al., 2012
Bird-specific *Lactobacillus* sp. Av4143	Av4143F: TGCAAGTCGAACGAGGATTTCT Av4143R: TCACCTTGGTAGGCCGTTACC Av4143P: 6-FAM-AGGTGGTTTTGCTATCGCTTT-BHQplus	TaqMan	60	244	Ohad et al., 2016
Dog mitochondrial gene NADH dehydrogenase subunit 5 DogND5	DogF: GGCATGCCTTTCCTTACAGGATTC DogR: GGGATGTGGCAACGAGTGTAATTATG DogP: 6-FAM-TCATCGAGTCCGCTAACACGTCGAAT-BHQ1	TaqMan	60	102	Caldwell and Levine, 2009
Sheep mitochondrial cytochrome B SheepCytB	SheepF: ACGCATTCATTGATCTCCCAGCTC SheepR: TCGGCAAATGTGGGTTACAGAGGA SheepP: 6-FAM-ACTTTGGCTCTCTCCTAGGCATTTGC-BHQ1	TaqMan	57	167*	Schill and Mathes, 2008
Horse mitochondrial cytochrome B HorseCytB	HorseF: AGGAGCAACAGTCATCACGAACCT HorseR: AAATGTACGACTACCAGGGCTGTG HorseP: 6-FAM-ATCGGTACTACCCTCGTCGAGTGAAT-BHQ1	TaqMan	57	168*	Schill and Mathes, 2008

*6-FAM = 6-carboxyfluorescein. ZEN/IBFQ = ZEN-Iowa Black FQ quencher. BHQ1 = Black Hole Quencher 1. BHQplus = Black Hole Quencher plus. *The original article by Schill and Mathes (2008) did not define the amplicon size. The length was evaluated according to the standard sequences generated by NCBI Nucleotide BLAST.*

not follow a normal distribution. The statistical difference between copy numbers detected with RNA-based and DNA-based approaches was compared with the Mann–Whitney *U*-test. The differences in copy numbers on various hosts were compared with the Kruskal–Wallis test. The detection rate between RNA-based and DNA-based approaches was compared with the McNemar test. When the sample number was less than 20, Fisher's exact test was used to confirm the result. The difference was considered statistically significant when $p < 0.05$. All the statistical tests were conducted in IBM (2020), and figures were made on Origin (Pro), 2017.

RESULTS

Performance of qPCR Amplification

The qPCR assay characteristics are summarized based on the amplification of the targeted assay on negative control, LOD, amplification efficiency, R^2-value of the amplification curve, and the range of quantification from the qPCR runs with fecal samples in **Supplementary Table 3**. Except for Av4143, the lower range of amplification efficiency of the assays was above 80%. The R^2-value of the amplification curve ranged between 0.946 and 1.000; the highest was with a SheepCytB marker, and the lowest was in the Av4143 marker. All the assays had the range of quantification from 10 to 10^5 GC per µl template, except BacCan, which had the range of 10^2 to 10^5 GC per µl template. GenBac3 and DogND5 assays with a DNA-based method, and GenBac3, GFD, and Rum2-Bac assays with an RNA-based method, showed some amplification in the blank samples (**Supplementary Table 3**).

Copy Numbers and Detection Frequency on Fecal Samples

All bacterial assays produced higher copy numbers with the RNA-based approach than the DNA-based approach ($p < 0.001$–0.002, Mann–Whitney *U*-test; **Figure 2**). On average, BacCan

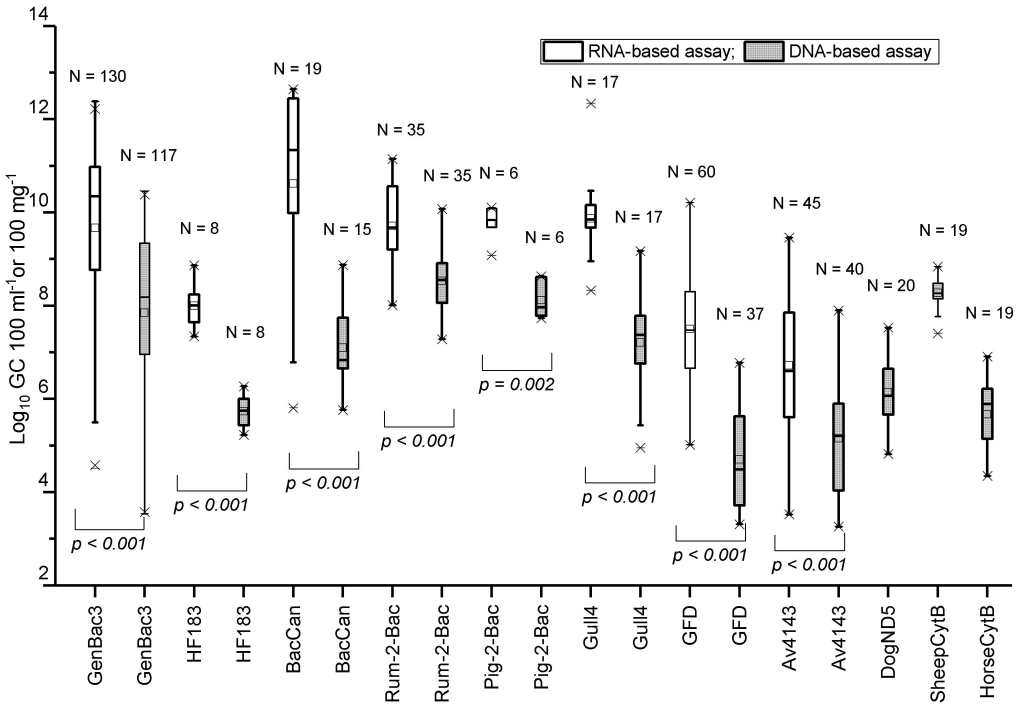


FIGURE 2 | The comparison between RNA-based and DNA-based copy numbers. The fecal and sewage samples were analyzed using the respective host-specific MST marker assays. The *p*-Value is based on the Mann–Whitney *U*-test. N: the number of samples included in the comparison (>LOQ).

assay produced the highest (10.6 $\log_{10}$) and the Av4143 assay produced the lowest (6.7 $\log_{10}$) RNA copy numbers per 100-mg feces, and the Rum-2-Bac assay produced the highest (8.5 $\log_{10}$) and GFD assay produced the lowest (4.7 $\log_{10}$) DNA copy number per 100-mg feces of targeted hosts (**Figure 2** and **Supplementary Table 4**).

Regarding detection frequency, the microbial targets were more frequently or at least equally detected from fecal samples with the RNA-based approach than with a DNA-based approach (**Supplementary Table 4**). The GFD assay was more frequently detected from targeted fecal samples with the RNA-based approach than with the DNA-based approach (p = 0.002, McNemar test, **Supplementary Table 4**). Assays HF183, Rum-2-Bac, Pig-2-Bac, and Gull4 had a 100% detection rate with both RNA-based and DNA-based approaches. Assays GenBac3, BacCan, and Av4143 had a higher detection percentage rate with the RNA-based approach, but the difference was not significant (**Supplementary Table 4**). Mostly, the detection frequency of assays in non-targeted fecal or sewage samples was increased with the RNA-based approach in comparison with the DNA-based approaches (**Table 3**).

Host Specificity and Cross-Reactivity With RNA-Based and DNA-Based Approaches

The GenBac3 assay targeting general Bacteroidales has been detected in 100% of mammal fecal samples with both RNA-based and DNA-based approaches but relatively less frequently (57% with RNA and 49% with DNA) from bird feces (**Table 3**, $p < 0.001$, McNemar test). There were also significantly higher GenBac3 copy numbers in the fecal material of mammals than in the fecal materials of birds, with both RNA-based and DNA-based approaches (**Supplementary Figure 1**; $p < 0.001$, Mann–Whitney *U*-test). However, there was no significant difference in the distribution of GenBac3 DNA ($p = 0.177$, Kruskal–Wallis) or the RNA ($p = 0.199$, Kruskal–Wallis) marker between bird (unknown bird species, gull, and goose) feces, but the distribution of the GenBac3 DNA marker was significantly different between mammal feces (horse, cow, sheep, dog, swine; $p < 0.001$, Kruskal–Wallis) (**Supplementary Figure 1**). Specifically, dog and cow feces expressed significantly lower GC numbers with the GenBac3 DNA-based approach than horse, swine, and sheep feces. Horse feces expressed significantly lower GC numbers with a GenBac3 DNA-based approach than sheep feces. As well, the distribution of the GenBac3 RNA marker was significantly different between mammal feces (horse, cow, sheep, dog, swine; $p < 0.001$, Kruskal–Wallis) (**Supplementary Figure 1**). Specifically, sheep feces expressed significantly lower GC numbers with the GenBac3 RNA-based approach than cow and dog feces. Horse feces expressed significantly lower GC numbers with the GenBac3 RNA-based approach than dog feces. The overall sensitivity of the GenBac3 assay was 81% with RNA-based and 77% with DNA-based templates (**Table 4**). The sample material in this study did not include true negative (non-fecal) samples for the GenBac3 assay; therefore, the specificity and accuracy of this assay were not calculated.

TABLE 3 | Sample-specific results of fecal samples.

Fecal source	Microbial source tracking markers																																					
	GenBac3				HF183				BacCan				Rum-2-Bac				Pig-2-Bac				Gull4				GFD				Av4143				Dog ND5				HorseCytB				SheepCytB	
	RNA		DNA		RNA		DNA		RNA		DNA		RNA		DNA		RNA		DNA		RNA		DNA		RNA		DNA		RNA		DNA		DNA		DNA		DNA	
	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T	NT
Effluent (n = 8)	8	–	8	–	8[#]	–	8[#]	–	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Dog (n = 21)	21	–	21	–	–	11	–	1	19	–	16	–	–	3	–	–	–	–	–	1	–	1	–	–	–	3	–	2	20	–	–	–	–	–	–	1		
Cattle (n = 16)	16	–	16	–	–	9	–	–	–	16	–	12	16	–	16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Swine (n = 6)	6	–	6	–	–	6	–	–	–	6	–	3	–	–	–	–	6	–	6	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Sheep (n = 19)	19	–	19	–	–	11	–	3	–	19	–	19	19	–	19	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	19	–	–	–	–
Horse (n = 19)	19	–	19	–	–	14	–	–	–	10	–	2	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	19	–	–	–	–	–	–	–
Hare (n = 2)	2	–	2	–	–	2	–	2	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Gull (n = 17)	13	–	8	–	–	15	–	1	–	–	–	–	–	–	–	–	2	–	–	17	–	17	–	15	–	8	–	16	–	14	–	–	–	–	–	–	–	–
Goose (n = 13)	11	–	9	–	–	2	–	–	–	4	–	1	–	–	–	–	–	–	–	–	–	2	13	–	8	–	1		1	–	–	–	–	–	–	–	–	–
Duck (n = 2)	1	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	1	–	1	–	1	–	–	–	–	–	–	–	–	–
Waterfowl (n = 2)	2	–	1	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	2	–	2	–	2	–	2	–	–	–	–	–	–	–	–	–
uBird (n = 34)	9	–	11	–	–	–	–	–	–	2	–	1	–	2	–	1	–	–	–	–	28	–	26	–	29	–	26	–	25	–	–	–	–	–	–	–	–	–	–

The different animal species and sewage effluent were analyzed using the tested MST assays with RNA-based and DNA-based templates. [#]Samples before wetland treatment, T, targeted; NT, not targeted; nc, not considered; –, none; uBird, fecal samples from birds with unidentified species.

The HF183 marker was detected in all targeted samples (sewage effluents before efficient tertiary treatment) from both RNA-based and DNA-based templates (**Table 3** and **Supplementary Table 4**). However, cross-reactions with non-targeted animal species (dog, cattle, swine, sheep, horse, hare, gull, and goose) happened more frequently with the RNA-based approach, compared with the DNA-based approach, which cross-reacted dog, sheep, hare, and gull (**Table 3**). The copy numbers from the RNA-based approach were significantly higher in targeted fecal samples than in non-targeted samples (**Figure 3**). The only exception was from the two hare fecal samples, from which the HF183 assay resulted in the highest recorded copy numbers from both RNA and DNA templates (**Figure 3**). The statistical test was not possible for DNA-based results due to a low number of samples exhibiting false-positive signals. The sensitivity of the HF183 assay was 100% with both RNA-based and DNA-based assays, but specificity was much lower, being 54% when RNA was the template and 95% when the template was DNA (**Table 4**). The accuracy of the HF183 assay was 56% with the RNA-based and 96% with the DNA-based approach.

From a total of 21 dog fecal samples, the dog-specific BacCan assay showed amplification in 19 with the RNA-based and 16 with the DNA-based approach. The sensitivity of the BacCan assay was 90% in the RNA-based and 76% in the DNA-based approach. The usability of this marker was questioned as the marker was amplified from fecal samples of mostly all animal species sampled in this study with both RNA-based and DNA-based approaches (**Table 3**). However, the BacCan copy numbers were significantly lower in the feces of non-targeted hosts as compared to the targeted canine feces (**Figure 4**). The specificity of the BacCan assay was 55% with an RNA-based and 71% with a DNA-based approach (**Table 4**). Further, the accuracy of this assay was 60% with RNA-based and 70% with RNA-based and DNA-based approaches.

TABLE 4 | Performance characteristics of the MST assays when using RNA and DNA as a template.

Assay	Sensitivity (%)		Specificity (%)		Accuracy (%)	
	RNA	DNA	RNA	DNA	RNA	DNA
GenBac3[#]	81	77	NA	NA	NA	NA
HF183[##]	100	100	54	95	56	96
BacCan	90	76	55	71	60	72
Rum-2-Bac	100	100	96	98	97	99
Pig-2-Bac	100	100	99	100	99	100
Gull4	100	100	97	97	97	97
GFD	88	66	99	100	94	85
Av4143	66	57	96	97	83	79
DogND5	NA	95	NA	100	NA	99
SheepCytB	NA	100	NA	99	NA	99
HorseCytB	NA	100	NA	100	NA	100

[#]The specificity and accuracy were not evaluated since the sample material did not include true negative samples for the GenBac3 assay. [##]HF183 assay values were determined by using sewage effluent samples before wetland treatment. NA, not applicable.

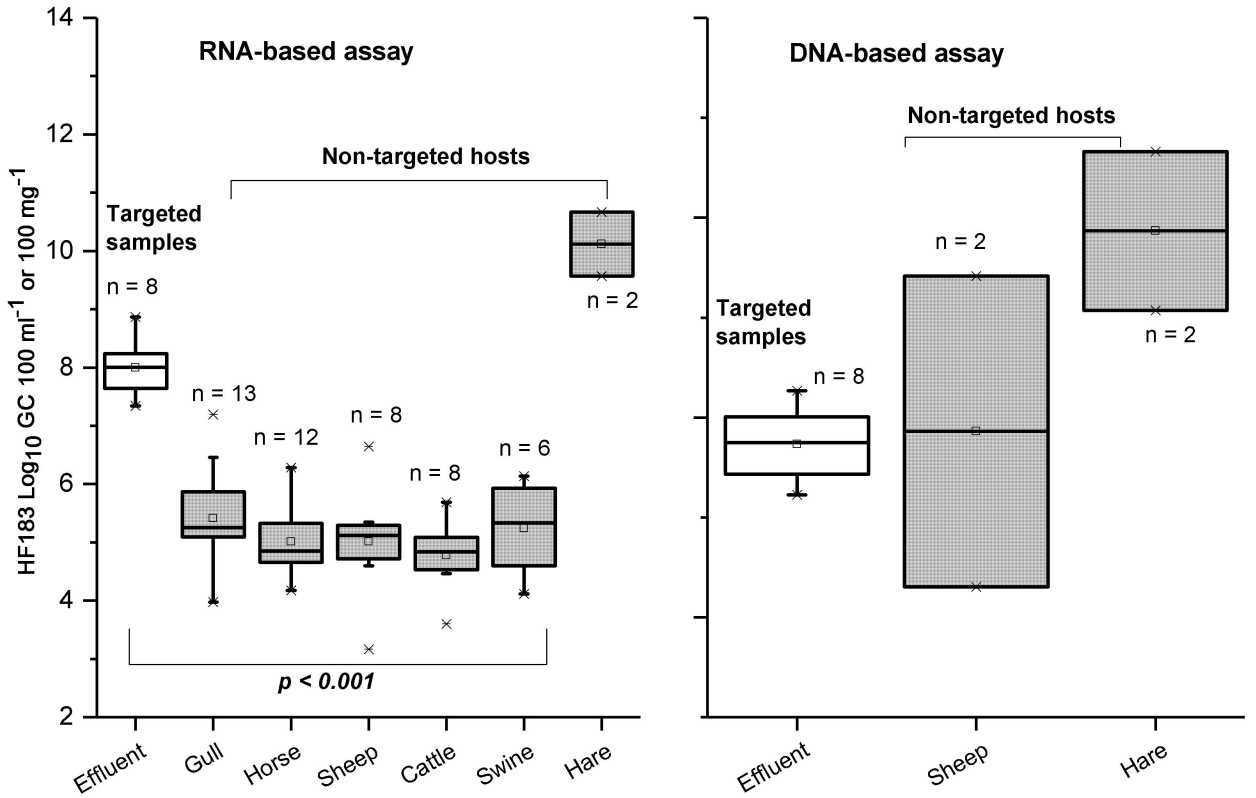


FIGURE 3 | Copy numbers generated with HF183 assay for targeted (sewage) and non-targeted (animal fecal) samples when using RNA and DNA as a template. Only the samples with results > LOQ included. The *p*-Value is based on the Kruskal–Wallis test.

The ruminant-specific Rum-2-Bac marker was 100% sensitive with cattle and sheep fecal samples with both RNA-based and DNA-based templates (**Table 4**). However, the detected GC was significantly higher in cattle fecal samples compared to sheep fecal samples ($p < 0.001$, Mann–Whitney *U*-test, **Supplementary Table 5**). The Rum-2-Bac marker cross-reacted with three dog and two bird fecal samples with the RNA-based approach and with one horse and one bird fecal sample with the DNA-based approach (**Table 3**). The marker had 96% specificity and 97% accuracy with the RNA-based approach and 98% specificity and 99% accuracy with the DNA-based approach.

Among bird-specific markers, the gull marker Gull4 was 100% sensitive with gull feces with both RNA-based and DNA-based assays (**Table 4**). There was a significant difference between the detected GC in the gull fecal samples compared to the fecal samples from unknown bird species with the RNA-based approach ($p < 0.001$, Kruskal–Wallis test), but the difference detected with the DNA-based approach was no longer significant ($p = 0.518$, Mann–Whitney *U*-test, **Supplementary Table 5**). The marker cross-reacted with one sheep fecal sample with the RNA-based approach and two goose and one dog fecal samples with the DNA-based approach (**Table 3**). The specificity and accuracy of this marker were 97% for both RNA-based and DNA-based approaches (**Table 4**). The sensitivity of the bird markers GFD and Av4143 were low, varying from 57 to 88% (**Table 4**). Instead, the specificity of the GFD assay was 99% with an RNA-based and 100% with a DNA-based approach, and specificity of the Av4143 assay was 96% with RNA-based and 97% with DNA-based approaches. Bird markers GFD and Av4143 cross-reacted with only a few (<3) canine fecal samples (**Table 3**). There were no significant differences in GC numbers between the different studied bird species (**Supplementary Table 5**).

In comparison to host-specific MST assays targeted to the 16S rRNA of bacteria, the performance of the mtDNA-based assays was generally better in terms of sensitivity, specificity, and accuracy calculated from the fecal sample test results (**Table 4**). The dog-specific assay DogND5 had the lowest (95%) sensitivity, whereas sheep-specific assay SheepCytB was the only one cross-reacting with one non-target (dog) sample (**Table 3**).

Probability of Target Detection

The probability of the tested MST markers, except the general fecal marker GenBac3, to correctly detect the presence of their targeted host feces in the water when using RNA and DNA as a template, was studied by the Bayesian statistical model. For the host-specific assays, as the prior probabilities were unknown, the range of prior probabilities from the worst-case scenario to the best-case scenario was used to visualize the performance of the markers to correctly detect their target. When the markers were compared by their ability to produce a positive result, in

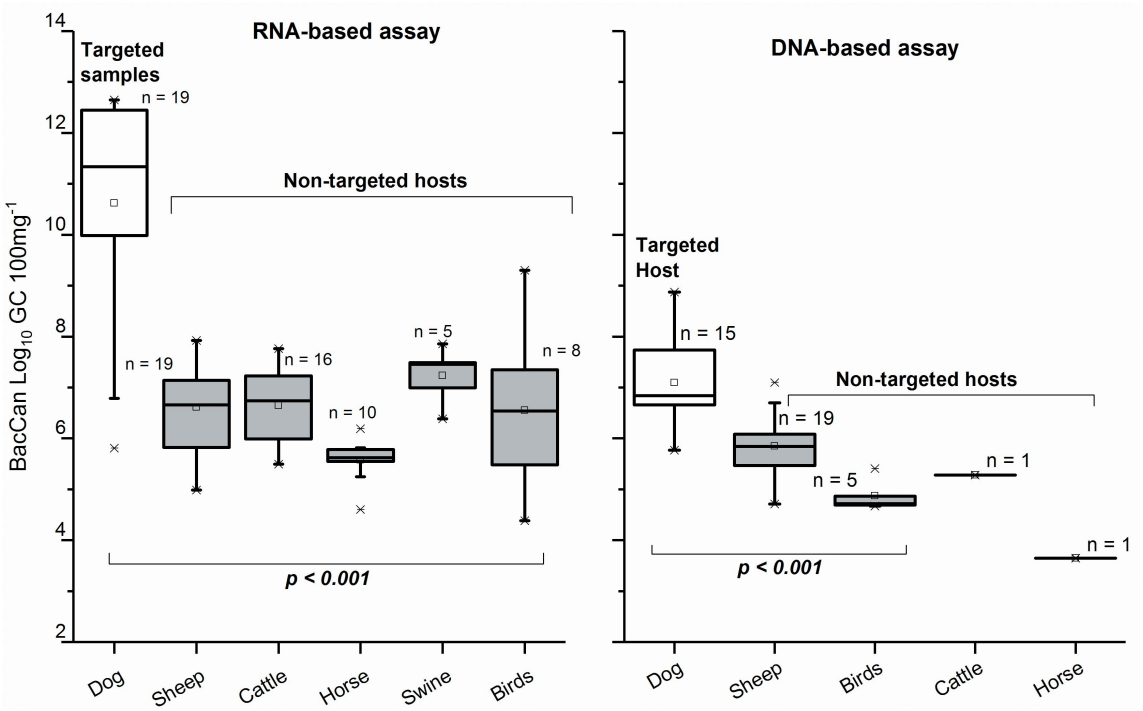


FIGURE 4 | Copy numbers generated with BacCan assay for targeted (dog) and non-targeted animal fecal samples when using RNA and DNA as a template. Only the samples with results > LOQ included. The *p*-Values are based on the Kruskal–Wallis test.

case the water matrix was contaminated with the feces of the target animal (**Figure 5**), the DogND5 assay exhibited a better capacity to correctly assign canine fecal contamination than the BacCan assay. Further, the capacity of the SheepCytB assay for detecting sheep feces, and the GFD assay for detecting bird feces was better than the capacity of the Rum-2-Bac and Av4143 assays, respectively. Nevertheless, the Rum-2-Bac and Av4143, as well as the HorseCytB, Pig-2-Bac, and Gull4 markers, showed a relatively good capacity to correctly detect their targets. By contrast, the capacity for the correct detection of BacCan and HF183 was relatively weak.

Detection of MST Markers in the Surface Water Samples

All bacterial markers (Rum-2-Bac, Gull4, GFD, and Av4143) were more frequently detected with the RNA-based approach than the DNA-based approach from the majority of the surface water sample types ($p < 0.001$, McNemar test, **Figure 6**), as well as all the samples together (**Supplementary Table 6**). In most of the sample groups, the copy number of rRNA was significantly higher than the rDNA copy number with all markers ($p < 0.001$, Mann-Whitney *U*-test, **Supplementary Figures 2–7**).

The RNA-based GenBac3 marker was detected in all sample groups in 98% of the total samples, and DNA-based GenBac3 markers were detected in all (100%) samples (**Figure 6**). The highest median of GC of 9.9 $\log_{10}$ GC/100 ml with an RNA-based approach and 7.07 $\log_{10}$ GC/100 ml with a DNA-based approach was detected in sewage samples (**Supplementary Figure 2**). Bathing water samples had the lowest GC values with both RNA-based (6.68 $\log_{10}$ GC/100 ml) and DNA-based (4.12 $\log_{10}$ GC/100 ml) approaches (**Supplementary Figure 2**).

The HF183 marker was detected from all sewage samples (100%) with a DNA-based approach. When an RNA-based approach was employed, the HF183 marker was detected more frequently from rural (fresh) surface water samples (93%, **Figure 6**). The RNA-based marker was detected only in 73% of sewage effluent samples, which was a lower detection rate than detected from urban surface water samples (80%). However, the highest median GC values were detected in sewage samples by using both RNA-based (7.73 $\log_{10}$ GC /100 ml) and DNA-based (5.45 $\log_{10}$ GC /100 ml) templates (**Supplementary Figure 3**). The detection frequency of this marker was less than 20% in bathing water, irrigation water, and rural (brackish) surface water samples.

The Rum-2-Bac marker was more frequently detected in rural brackish (60%) and fresh (100%) surface water when an RNA-based template was used. Overall, the detection frequency of the Rum-2-Bac marker was low when a DNA-based template was used. The highest detection frequency for a DNA-based template was 27% in rural (fresh) surface water samples. The highest median value of an RNA-based Rum-2-Bac assay was also detected from rural fresh surface water (5.38 $\log_{10}$ GC/100 ml) (**Supplementary Figure 4**). The Pig-2-Bac marker was not detected from any surface water samples in this study.

The Gull4 RNA markers were detected from all bathing water and irrigation water samples. However, the Gull4 DNA marker

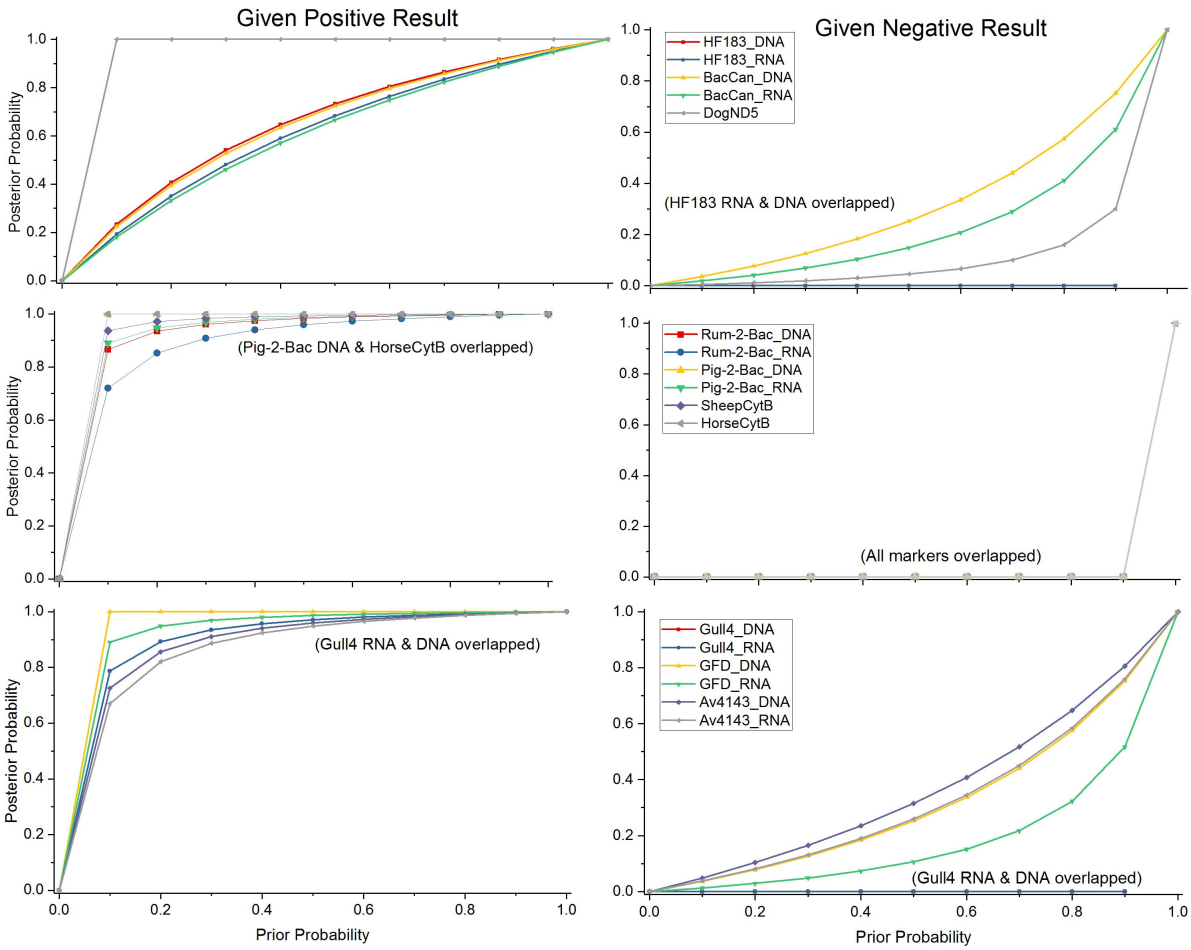


FIGURE 5 | The probabilities of true-positive and false-negative results in the case of animal fecal contamination by the Bayesian statistical model. Posterior probabilities of contamination given a positive qPCR result using the markers specific for dog, sheep, and bird, as well as human, gull, horse, and swine over a range of prior probabilities.

was detected in only about 50% of rural surface water (fresh and brackish), bathing water, and irrigation water samples (**Figure 6**). The highest median GC value with an RNA-based approach (5.89 $\log_{10}$ GC/100 ml) was detected in bathing water samples (**Supplementary Figure 5**).

The general bird-specific marker GFD was detected with the RNA-based approach in all sample groups. The highest detection frequency (96%) was noted in urban surface water and the lowest (60%) in irrigation water. By using the DNA-based approach, the GFD marker was detected only in urban surface water (32%) and bathing water (5%). The highest median value with RNA-based template (5.32 $\log_{10}$ GC/100 ml) was also from the urban surface water sample (**Supplementary Figure 6**). The other bird-specific marker Av4143 was less frequently detected in comparison to the corresponding GFD detection frequency when an RNA-based approach was used. On the contrary, Av4143 was detected more frequently than GFD when the DNA-based approach was used. The RNA-based Av4143 marker was not detected from sewage and irrigation water samples, although it was detected in 54% of rural surface water and 73% of urban surface water samples. The highest detection frequency with a DNA-based approach was only 20%, and it was reached from rural surface water (fresh and brackish). The highest median value (4.48 $\log_{10}$ GC/100 ml) of Av4143 RNA was also from the urban surface water sample (**Supplementary Figure 7**).

Mitochondrial DNA markers were less frequently detected in surface water samples than bacterial markers. A DogND5 marker was detected from two bathing water samples and one urban surface water sample with GC $\leq$ 3.43 $\log_{10}$ GC/100 mL. HorseCytB was detected only from one out of three samples from horse-farm runoff, and it was not detected at all from other environmental samples. SheepCytB was not detected from the environmental samples tested in this study.

DISCUSSION

This study tested the performance of previously reported MST assays targeting general fecal contamination, human, gull, ruminant, swine, dog, horse, sheep, and general birds. The goal

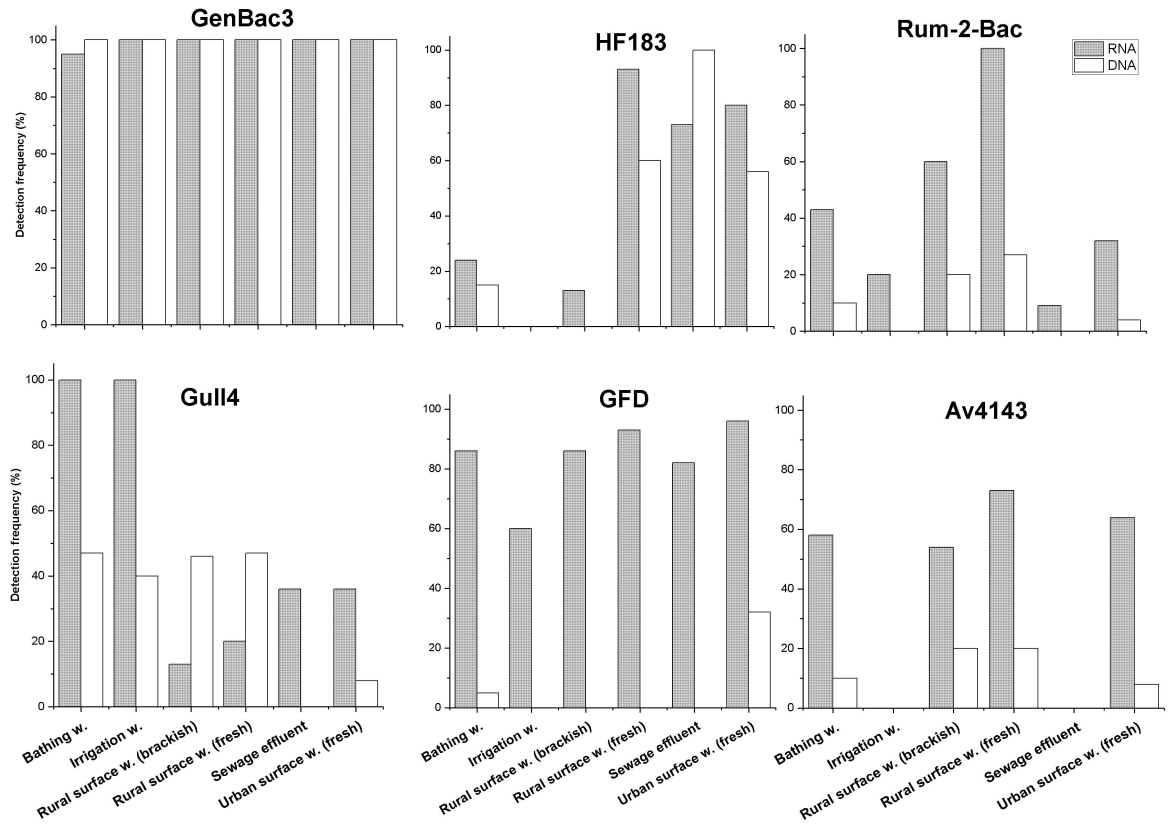


FIGURE 6 | Detection frequency (%) of the MST markers in the groups of surface water samples with both RNA-based and DNA-based approaches. Only markers with detection frequency > 20% in surface water groups are shown. In *x*-axis: sample group. W, water.

was to utilize the cDNA template produced from RNA, with a reverse transcriptase process, and compare its performance with a currently used rDNA-based approach and mtDNA-based approaches. To our knowledge, this is the first study to evaluate the performance of MST assays in Finland. Furthermore, to our knowledge, this is the first study to evaluate the performance of 16S rRNA-targeted MST assays with the use of the RNA-based template instead of the DNA template. As a main finding, the sensitivity, specificity, and accuracy of the assays targeting bird feces (Gull4, GFD, and Av4143) measured with the RNA-based approach were higher than, or at least similar to, the conventional DNA-based approach. In the case of mammal-specific markers (HF183, Rum-2-Bac, and Pig-2-Bac), the RNA-based approach resulted in a higher sensitivity, but the assay specificity and accuracy were lower than when using the DNA-based template.

A microbial assay with higher counts (CFU or GC) in fecal material has greater significance for water-quality monitoring; for example, such assay remains still detectable even after many folds of dilution in a surface water resource (Harwood et al., 2009, 2014; Layton et al., 2013). Sensitivity refers to the proportion of known positive controls that are correctly identified as positive. The higher sensitivity has practical significance; it better protects the public health than the methods with lower sensitivity (Harwood et al., 2014).

Laboratory methodologies developed in one geographical region mostly have global applicability. However, in the case of MST assays, mainly targeting host-specific bacteria, the assay performance can vary between the geographical locations, as gut bacterial communities are affected by animal feeding practices, herd size, and ages (Dick et al., 2005; Shanks et al., 2011; Ballesté et al., 2020). Such possible variation requires verification of the accuracy and reliability of MST markers before using them in a new geographical location (Roslev and Bukh, 2011). However, the probable fecal contamination sources in each watershed are different and each earlier study tested the marker in different animal fecal materials. In fact, multiple previous studies reported the cross-reaction of MST markers with the fecal materials from non-targeted species (Ryu et al., 2012; Boehm et al., 2013; Sinigalliano et al., 2013). Therefore, the performance characteristics, mainly specificity, related to the false-positive rate of the assays should be carefully evaluated in a new geographical location (Harwood et al., 2009; Stewart et al., 2013).

Comparing the Performance of the RNA-Based and DNA-Based Approaches

As reported in earlier studies (Pitkänen et al., 2013; Kapoor et al., 2014), also in our study, the RNA-based assays targeted to 16S rRNA had a higher detection frequency and higher

target copy numbers in fecal samples and also in surface water samples than the DNA-based assays. The explanation for the increased sensitivity is that an active cell contains ribosomes full of ribosomal RNA (Waters and McCuthan, 1990; Nogales et al., 2001; Péirez-Osorio et al., 2010). The rRNA target may indicate the activity and transcription rate status of bacterial cells, as metabolically active cells have greater amounts of rRNA per cell than non-viable cells (Gourse et al., 1996; Martinez et al., 2006; Péirez-Osorio et al., 2010).

In line with earlier studies using DNA as a template (Mieszkin et al., 2009, 2010; Ryu et al., 2012; Boehm et al., 2013; Raith et al., 2013; Harwood et al., 2014; Ohad et al., 2016), the sensitivity of HF183, Gull4, Rum-2-Bac, and Pig-2-Bac assays was more than 80% also in this study. In this study, this sensitivity was reached with both RNA-based and DNA-based approaches. Instead, the sensitivity of BacCan, GenBac3, and GFD assays remained below 80% when DNA was used as a template, while others have reported sensitivities of 63–100%, 100%, and 30–68%, respectively (Kildare et al., 2007; Boehm et al., 2013; Ahmed et al., 2015, Ahmed et al., 2016; Odagiri et al., 2015; Nshimyimana et al., 2017; Symonds et al., 2017). More than 80% sensitivity was achieved when RNA was used as a template. Of these assays, the GenBac3 assay was poorly amplified in the fecal materials of birds, with a lower detection rate and lower GC in comparison with mammal fecal materials. Earlier studies also reported the variation in the proportions of *Bacteroidetes* between different groups of birds, almost absent in waterfowls and broiler chickens, present in battery hens, and dominant in turkeys (Zhu et al., 2002; Scupham et al., 2008; Lu et al., 2009). However, such differences between the studied bird species were not noticed in this study.

The performance characteristics sensitivity, specificity, and accuracy of the human-specific marker (HF183) were good, being higher than 95% when a DNA-based template was used in this study. Surprisingly, the specificity and accuracy of this assay dropped to the levels of 54% and 56%, respectively, when we used rRNA as a template for the HF183 assay. Also, earlier studies with a DNA-based approach have reported cross-amplification of the HF183 assay with non-targeted species, such as dog, rabbit, chicken, swine, and cattle, with specificities between 80 and 100% (Boehm et al., 2013; Layton et al., 2013; Odagiri et al., 2015; Nshimyimana et al., 2017; Haramoto and Osada, 2018). Although the specificity of the RNA-based approach was lower than the specificity of the DNA-based approach in the present study, the GC difference in the fecal material between targeted and non-targeted hosts was many folds higher with the RNA-based approach than in the DNA-based approach. Thus, the cases where HF183 was detected in urban surface water simultaneously with both RNA-based and DNA-based approaches could tentatively be explained by human-derived fecal contamination such as accidental leakages of municipal sewage. However, the high cross-reactivity of the HF183 marker on a non-targeted host feces, especially with an RNA-based approach, calls for a need for method development toward more specific but still enough sensitive markers for human fecal contamination. For example, totally new targets could be found from the rapidly increasing metagenome data and also the further optimization of the PCR conditions of the current genetic targets might improve the assay performance as well.

In the case of the BacCan marker, many earlier studies have also reported the poor performance characteristics (Kildare et al., 2007; Boehm et al., 2013; Schriewer et al., 2013; Odagiri et al., 2015; Nshimyimana et al., 2017). For example, Schriewer et al. (2013) reported 100% sensitivity, but only 70% specificity of the BacCan assay. In our hands, the BacCan assay cross-reacted with feces of nearly all studied animal species. It is noteworthy that the GC counts were many folds lower in the feces of non-targeted hosts than the targeted hosts. This difference in the copy numbers was even greater when using rRNA as a template for the assay. However, acknowledging the poor specificity of this marker, it was discarded before the water sample analysis. Herein, the ruminant-specific Rum-2-Bac and Pig-2-Bac had similar performance characteristics with earlier studies (Mieszkin et al., 2009, 2010; Boehm et al., 2013; Raith et al., 2013). The highest detection rate of the Rum-2-Bac assay (RNA: 100% and DNA: 27%) fresh rural surface water confirms our hypothesis: the sensitivity of the RNA-based approach is crucial for the contamination source detection from watersheds. Regarding swine-specific Pig-2-Bac, Mieszkin et al. (2009) reported 98–100% sensitivity and 100% specificity while testing fecal materials from pig, cow, sheep, and horse. Haramoto and Osada (2018) reported 100% sensitivity, 66% specificity, and 77% accuracy of the Pig-2-Bac assay with a DNA-based assay. They reported the Pig-2-Bac marker amplification on cattle feces.

The performance of the Gull4 marker with the RNA-based and DNA-based approaches was in line with earlier studies (Ryu et al., 2012; Ohad et al., 2016; Ballesté et al., 2020). Ballesté et al. (2020) reported 85% sensitivity and 100% specificity of the Gull4 marker while testing the assay in human, ruminant, sheep, horse, pig, and gull feces. Ryu et al. (2012) reported 87% sensitivity and 91% specificity of the Gull4 assay with the DNA-based approach. The Gull4 marker had the highest detection frequency in irrigation water and bathing water (100% for both) among the six different surface water sample types studied (**Figure 6**). As already noted with other assays, the sensitivity of the GFD assay was also higher (88%) with the RNA-based approach than with the DNA-based approach (66%). In comparison, Green et al. (2012) reported sensitivity of 58%, whereas Ahmed et al. (2016) reposted sensitivities of 58% from a Brisbane, Australia sample and 30% from a Florida, United States sample (52% when combined). Symonds et al. (2017) reported 44% sensitivity (chicken and sea birds) and 56% specificity (cross-amplified with cow, dog, sewage, horse, and pig fecal materials) of this assay with a DNA-based approach. Therefore, the use of rRNA as a template seems to bring a needed boost for the assay sensitivity. The use of rRNA as a template seems feasible with the GFD assay, as the specificity of the assay was as good as 99% with an RNA-based approach (it was 100% when rDNA was used as a template). In the performance evaluation, the GFD assay was sensitive for the fecal material of multiple birds, including gulls, ducks, goose, and waterfowl, and when analyzing fecal samples from unspecified bird species. The GFD marker was the most frequently detected in urban surface water

(96%) and sewage effluent (82%), among six studied sample groups (**Figure 6**) with an RNA-based approach. This study recorded low sensitivity for the other bird marker tested, the Av4143 assay (RNA 66%, DNA 57%). This result deviates from an earlier study, where Ohad et al. (2016) reported a 95% sensitivity for this assay. The poor sensitivity of the Av4143 marker on different bird fecal materials may indicate that the bacterial group targeted with this marker may not be present in the gut of all bird populations. The potential geographical instability of this marker calls for further investigation. Due to the study outcome, we recommend the use of the GFD assay instead of Av4143 for use in Finnish surface water quality monitoring.

Although in general, the number of fecal samples per host used for this performance analysis was large, the HF183 and Pig-2-Bac assays targeting human and swine fecal contamination, respectively, were evaluated by using only eight sewage effluent and six swine fecal samples, which is below the recommended size of ten samples per each targeted host (US Environmental Protection Agency, 2005).

Mitochondrial DNA-Based Assays

This study demonstrated excellent performance characteristics of SheepCytB and HorseCytB assays. The high sensitivity and specificity (95–100%) of these mtDNA-based assays were consistent with earlier findings (Caldwell and Levine, 2009; Tambalo et al., 2012; He et al., 2016; Malla and Haramoto, 2020). The dog-specific DogND5 assay had higher sensitivity, specificity, and accuracy than the respective BacCan assay targeted to host-specific fecal bacteria. The central assumption of the host-specific, bacteria-based approach is that the targeted bacteria (or groups) have a strong relationship with a particular host. However, these fecal bacteria can be found from the feces of non-targeted hosts too, as noted in this study. In contrast, in the mtDNA-based MST approach, the detection of target DNA from exfoliated epithelial cells from the host alimentary canal has a much higher specificity than bacterial assays (Caldwell et al., 2011; Malla and Haramoto, 2020).

The usefulness of the highly specific mtDNA-based assays is somewhat hampered with the fact that the amplification efficiency of the assays could be relatively weak, as noted when using DogND5, SheepCytB, and HorseCytB assays for water sample testing in this study. The exact reason for low efficiency in our hands remains unclear, and some earlier studies have reported the higher amplification efficiency values of the DogND5 assay (Caldwell and Levine, 2009; Tambalo et al., 2012). Although the sensitivity and specificity of all three mtDNA-based assays were outstanding (∼100%) when fecal materials were tested, these targets were only seldom detected from the surface water samples. The mtDNA assays remained negative even when the feces of the targeted host animals were suspected to be present in the water, which creates uncertainty for the applicability of the mtDNA assays in real life. The dog-specific DogND5 marker was not detected from fresh rural and brackish rural surface water, where the detection was noted by the BacCan bacterial assay (RNA-based approach). However, due to the obvious specificity issues with RNA-based BacCan detection, the absence of the target feces from the samples tested cannot be out ruled either.

CONCLUSION

- The performance characteristics sensitivity, specificity, and accuracy of assays targeting birds with an RNA-based approach were higher than or equal to the DNA-based approach.
- The sensitivity of human and dog markers were higher with the RNA-based approach, but specificity and accuracy were higher with the DNA-based approach. The performance between using RNA and DNA as a template was similar to ruminant and swine markers.
- The performance of assays DogND5, HorseCytB, SheepCytB, GFD, Gull4, Rum-2-Bac, and Pig-2-Bac was shown as reliable for detecting dog, horse, sheep, bird, gull, ruminant, and pig fecal contamination sources, respectively, in Finnish watersheds. Still, all mtDNA targets and the Pig-2-Bac marker were not detected in surface water samples.
- The sensitivity of the human-specific marker HF183 was 100% with RNA-based and DNA-based approaches. However, the specificity and accuracy of the marker were higher with the DNA-based approach (95–96%) than with the RNA-based approach (54–56%). Despite the cross-reactivity, the GC values were many folds higher in targeted sewage samples than in non-targeted animal fecal samples. Therefore, the use of RNA as a template for the HF183 assay in the future could be justified when employed together with a DNA template.
- The general fecal marker GenBac3 had a higher detection rate and GC in studied mammal fecal materials than in bird fecal materials. It may indicate that measuring the marker targeted to general *Bacteroidales* may not cover the fecal contamination from bird species.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

AR, A-MH, TT, and TP contributed to conceptualization and design of the study. A-MH, SU, and TT organized the sampling. AR executed the laboratory study and the original data calculations and wrote the first draft of the manuscript. AT performed the statistical analysis under the supervision of AV and wrote major parts of the results, the first draft of the discussion, and visualized the data. TP supervised the work and was in charge of the funding acquisition. All authors contributed to manuscript editing, read, and approved the final version.

FUNDING

This research was supported in part by the cities of Kalajoki, Tampere, Kuopio, and the municipal wastewater treatment plants

Hämeenlinnan Seudun Vesi Ltd., Nokian Vesi Ltd., and Oulun Vesi Ltd., who provided samples for the study. Further, the work was partially funded by the Regional Council of Häme, grant number 518 HL/106/04.01.01/2018. We acknowledge all of the project partners and collaborators, especially the personnel of HAMK University of Applied Sciences, Hämeenlinnan Seudun Vesi Ltd., the Lammi Biological Station, and Ilkka Hirvonen of Led Future Ltd.

ACKNOWLEDGMENTS

We thank Tiina Heiskanen, Tarja Rahkonen, and Tarja Yli-Tuomi for their technical assistance and local health and environment authorities, and communal water supply personnel Jukka Meriluoto, Salla Leppänen, Anu Väänänen, Päivi Rissanen, Katja Ylönen, Hanna Jääskeläinen, and especially Ville Soininen and Inkeri Eronen, who received funding for the work from the North Ostrobothnia Centre for Economic Development, Transport, and the Environment (grant number POPELY/2687/2017), for organizing sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.673306/full#supplementary-material

REFERENCES

Ahmed, W., Harwood, V. J., Nguyen, K., Young, S., Hamilton, K., and Toze, S. (2016). Utility of *Helicobacter* spp. associated GFD markers for detecting avian fecal pollution in natural waters of two continents. *Water Res.* 88, 613–622. doi: 10.1016/j.watres.2015.10.050

Ahmed, W., Staley, C., Sadowsky, M. J., Gyawali, P., Sidhu, J. P., Palmer, A., et al. (2015). Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water. *Appl. Environ. Microbiol.* 81, 7067–7077. doi: 10.1128/AEM.02032-15

Anza, I., Vidal, D., Laguna, C., Diaz-Sanchez, S., Sanchez, S., Chicote, A., et al. (2014). Eutrophication and bacterial pathogens as risk factors for avian botulism outbreaks in wetlands receiving effluents from urban wastewater treatment plants. *Appl. Environ. Microbiol.* 80, 4251–4259. doi: 10.1128/AEM.00949-14

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

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Caldwell, J., Payment, P., and Villemur, R. (2011). "Mitochondrial DNA as source tracking markers of fecal contamination," in *Microbial Source Tracking: Methods, Applications, and Case Studies*, eds C. Hagedorn, A. R. Blanch, and V. J. Harwood (New York, NY: Springer New York), 229–250.

Caldwell, J. M., and Levine, J. F. (2009). Domestic wastewater influent profiling using mitochondrial real-time PCR for source tracking animal contamination. *J. Microbiol. Methods* 77, 17–22. doi: 10.1016/j.mimet.2008.11.007

Dick, L. K., Bernhard, A. E., Brodeur, T. J., Santo Domingo, J. W., Simpson, J. M., Walters, S. P., et al. (2005). Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71, 3184–3191. doi: 10.1128/AEM.71.6.3184-3191.2005

Gourse, R. L., Gaal, T., Bartlett, M. S., Appleman, J. A., and Ross, W. (1996). rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu. Rev. Microbiol.* 50, 645–677.

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

Guzman-Herrador, B., Carlander, A., Ethelberg, S., Freiesleben de Blasio, B., Kuusi, M., Lund, V., et al. (2015). Waterborne outbreaks in the Nordic countries, 1998 to 2012. *Euro Surveill.* 20:21160. doi: 10.2807/1560-7917.es2015.20.24.21160

Haramoto, E., and Osada, R. (2018). Assessment and application of host-specific Bacteroidales genetic markers for microbial source tracking of river water in Japan. *PLoS One* 13:e0207727. doi: 10.1371/journal.pone.0207727

Harwood, V. J., Brownell, M., Wang, S., Lepo, J., Ellender, R. D., Ajidahun, A., et al. (2009). Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. *Water Res.* 43, 4812–4819. doi: 10.1016/j.watres.2009.06.029

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. A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L., and Shanks, O. C. (2010). Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Syst. Appl. Microbiol.* 33, 348–357. doi: 10.1016/j.syapm.2010.06.001

He, X., Liu, P., Zheng, G., Chen, H., Shi, W., Cui, Y., et al. (2016). Evaluation of five microbial and four mitochondrial DNA markers for tracking human and pig fecal pollution in freshwater. *Sci. Rep.* 6:35311. doi: 10.1038/srep35311

Hokajärvi, A.-M., Pitkänen, T., Siljanen, H., Nakari, U., Torvinen, E., Siitonen, A., et al. (2013). Occurrence of thermotolerant *Campylobacter* spp. and adenoviruses in Finnish bathing waters and purified sewage effluents. *J. Water Health* 11, 120–134. doi: 10.2166/wh.2012.192

Hörman, A., Rimhanen-Finne, R., Maunula, L., von Bonsdorff, C. H., Torvela, N., Heikinheimo, A., et al. (2004). *Campylobacter* spp., *Giardia* spp., *Cryptosporidium* spp., noroviruses, and indicator organisms in surface water in southwestern Finland, 2000-2001. *Appl. Environ. Microbiol.* 70, 87–95. doi: 10.1128/aem.70.1.87-95.2004

IBM (2020). *IBM SPSS Statistics for Windows, Version 27.0*. Armonk, NY: IBM Corporation.

Inkinen, J., Jayaprakash, B., Siponen, S., Hokajärvi, A. M., Pursiainen, A., Ikonen, J., et al. (2019). Active eukaryotes in drinking water distribution systems of ground and surface waterworks. *Microbiome* 7:99. doi: 10.1186/s40168-019-0715-5

Kapoor, V., Pitkänen, T., Ryu, H., Elk, M., Wendell, D., and Santo Domingo, J. W. (2014). Distribution of human-specific Bacteroidales and fecal indicator bacteria in an urban watershed impacted by sewage pollution, determined using RNA- and DNA-based quantitative PCR assays. *Appl. Environ. Microbiol.* 81, 91–99. doi: 10.1128/AEM.02446-14

Kauppinen, A., Martikainen, K., Matikka, V., Veijalainen, A. M., Pitkänen, T., Heinonen-Tanski, H., et al. (2014). Sand filters for removal of microbes and nutrients from wastewater during a one-year pilot study in a cold temperate climate. *J. Environ. Manage.* 133, 206–213. doi: 10.1016/j.jenvman.2013.12.008

Kauppinen, A., Pitkänen, T., Al-Hello, H., Maunula, L., Hokajärvi, A.-M., Rimhanen-Finne, R., et al. (2019). Two Drinking Water Outbreaks Caused by Wastewater Intrusion Including Sapovirus in Finland. *Int. J. Environ. Res. Public Health* 16:4376.

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection

of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. *Water Res.* 41, 3701–3715. doi: 10.1016/j.watres.2007.06.037

Lamendella, R., Santo Domingo, J. W., Yannarell, A. C., Ghosh, S., Di Giovanni, G., Mackie, R. I., et al. (2009). Evaluation of swine-specific PCR assays used for fecal source tracking and analysis of molecular diversity of swine-specific "bacteroidales" populations. *Appl. Environ. Microbiol.* 75, 5787–5796. doi: 10.1128/AEM.00448-09

Layton, B. A., Cao, Y., Ebentier, D. L., Hanley, K., Balleste, E., Brandão, J., et al. (2013). Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res.* 47, 6897–6908. doi: 10.1016/j.watres.2013.05.060

Lu, J., Santo Domingo, J. W., Hill, S., and Edge, T. A. (2009). Microbial diversity and host-specific sequences of Canada goose feces. *Appl. Environ. Microbiol.* 75, 5919–5926. doi: 10.1128/AEM.00462-09

Malla, B., and Haramoto, E. (2020). Host-specific mitochondrial DNA markers for tracking the sources of fecal pollution. *Curr. Opin. Environ. Sustain.* 16, 34–46. doi: 10.1016/j.coesh.2020.02.006

Martinez, R. J., Mills, H. J., Story, S., and Sobecky, P. A. (2006). Prokaryotic diversity and metabolically active microbial populations in sediments from an active mud volcano in the Gulf of Mexico. *Environ. Microbiol.* 8, 1783–1796.

Matsuda, K., Tsuji, H., Asahara, T., Takahashi, T., Kubota, H., Nagata, S., et al. (2012). Sensitive quantification of Clostridium difficile cells by reverse transcription-quantitative PCR targeting rRNA molecules. *Appl. Environ. Microbiol.* 78, 5111–5118.

Mieszkin, S., Furet, J. P., Corthier, G., and Gourmelon, M. (2009). Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045–3054. doi: 10.1128/AEM.02343-08

Mieszkin, S., Yala, J. F., Joubrel, R., and Gourmelon, M. (2010). Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J. Appl. Microbiol.* 108, 974–984. doi: 10.1111/j.1365-2672.2009.04499.x

Nogales, B., Moore, E. R., Llobet-Brossa, E., Rossello-Mora, R., Amann, R., and Timmis, K. N. (2001). Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl. Environ. Microbiol.* 67, 1874–1884. doi: 10.1128/AEM.67.4.1874-1884.2001

Nshimyimana, J. P., Cruz, M. C., Thompson, R. J., and Wuertz, S. (2017). Bacteroidales markers for microbial source tracking in Southeast Asia. *Water Res.* 118, 239–248. doi: 10.1016/j.watres.2017.04.027

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., et al. (2015). Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* 502, 462–470. doi: 10.1016/j.scitotenv.2014.09.040

Ohad, S., Ben-Dor, S., Prilusky, J., Kravitz, V., Dassa, B., Chalifa-Caspi, V., et al. (2016). The development of a novel qPCR assay-set for identifying fecal contamination originating from domestic fowls and waterfowl in Israel. *Front. Microbiol.* 7:145. doi: 10.3389/fmicb.2016.00145

Origin (Pro), (2017). *Origin (Pro), Version*. Northampton, MA: OriginLab Corporation.

Péirez-Osorio, A. C., Williamson, K. S., and Franklin, M. J. (2010). Heterogeneous rpoS and rhlR mRNA levels and 16S rRNA/rDNA (rRNA gene) ratios within *Pseudomonas aeruginosa* biofilms, sampled by laser capture microdissection. *J. Bacteriol.* 192, 2991–3000.

Pitkänen, T., Ryu, H., Elk, M., Hokajärvi, A. M., Siponen, S., Vepsäläinen, A., et al. (2013). Detection of fecal bacteria and source tracking identifiers in environmental waters using rRNA-based RT-qPCR and rDNA-based qPCR assays. *Environ. Sci. Technol.* 47, 13611–13620. doi: 10.1021/es403489b

Pitkänen, T., Uusheimo, S., Kauppinen, A., Hokajärvi, A.-M., Tulonen, T., Hirvonen, I., et al. (2019). "Removal of fecal microbes from secondary treated sewage effluent by using constructed wetland and UV LED's," in *Proceedings of the Nordic Waste Water Conference 23-25 September 2019*, Helsinki.

Raith, M. R., Kelty, C. A., Griffith, J. F., Schriewer, A., Wuertz, S., Mieszkin, S., et al. (2013). Comparison of PCR and quantitative real-time PCR methods for the characterization of ruminant and cattle fecal pollution sources. *Water Res.* 47, 6921–6928. doi: 10.1016/j.watres.2013.03.061

Rankinen, K., Butterfield, D., Faneca Sanchez, M., Grizzetti, B., Whitehead, P., Pitkänen, T., et al. (2016). The INCA-Pathogens model: an application to the Loimijoki River basin in Finland. *Sci. Total Environ.* 572, 1611–1621. doi: 10.1016/j.scitotenv.2016.05.043

Roslev, P., and Bukh, A. S. (2011). State of the art molecular markers for fecal pollution source tracking in water. *Appl. Microbiol. Biotechnol.* 89, 1341–1355. doi: 10.1007/s00253-010-3080-7

Ryu, H., Griffith, J. F., Khan, I. U., Hill, S., Edge, T. A., Toledo-Hernandez, C., et al. (2012). Comparison of gull feces-specific assays targeting the 16S rRNA genes of *Catellicoccus marimammalium* and *Streptococcus* spp. *Appl. Environ. Microbiol.* 78, 1909–1916. doi: 10.1128/AEM.07192-11

Schill, W. B., and Mathes, M. V. (2008). Real-time PCR detection and quantification of nine potential sources of fecal contamination by analysis of mitochondrial cytochrome b targets. *Environ. Sci. Technol.* 42, 5229–5234.

Schriewer, A., Goodwin, K. D., Sinigalliano, C. D., Cox, A. M., Wanless, D., Bartkowiak, J., et al. (2013). Performance evaluation of canine-associated Bacteroidales assays in a multi-laboratory comparison study. *Water Res.* 47, 6909–6920. doi: 10.1016/j.watres.2013.03.062

Scupham, A. J., Patton, T. G., Bent, E., and Bayles, D. O. (2008). Comparison of the cecal microbiota of domestic and wild turkeys. *Microb. Ecol.* 56, 322–331. doi: 10.1007/s00248-007-9349-4

Shanks, O. C., McLellan, S., Huse, S. M., and Sogin, M. L. (2011). "Characterization of microbial population structures in recreational waters and primary sources of fecal pollution with a next-generation sequencing approach," in *Environmental Microbiology: Current Technology and Water Applications*, Chap. 9, eds K. Sen and N. J. Ashbolt (Norwich: Caister Academic Press).

Siefring, S., Varma, M., Atikovic, E., Wymer, L., and Haugland, R. A. (2008). Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instruments and reagent systems. *J. Water Health* 6, 225–237. doi: 10.2166/wh.2008.022

Sinigalliano, C. D., Ervin, J. S., Van De Werfhorst, L. C., Badgley, B. D., Ballesté, E., Bartkowiak, J., et al. (2013). Multi-laboratory evaluations of the performance of *Catellicoccus marimammalium* PCR assays developed to target gull fecal sources. *Water Res.* 47, 6883–6896. doi: 10.1016/j.watres.2013.02.059

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

Stewart, J. R., Boehm, A. B., Dubinsky, E. A., Fong, T. T., Goodwin, K. D., Griffith, J. F., et al. (2013). Recommendations following a multi-laboratory comparison of microbial source tracking methods. *Water Res.* 47, 6829–6838. doi: 10.1016/j.watres.2013.04.063

Symonds, E. M., Young, S., Verbyla, M. E., McQuaig-Ulrich, S. M., Ross, E., Jiménez, J. A., et al. (2017). Microbial source tracking in shellfish harvesting waters in the Gulf of Nicoya, Costa Rica. *Water Res.* 111, 177–184. doi: 10.1016/j.watres.2017.01.004

Tambalo, D. D., Boa, T., Liljebjelke, K., and Yost, C. K. (2012). Evaluation of two quantitative PCR assays using Bacteroidales and mitochondrial DNA markers for tracking dog fecal contamination in water bodies. *J. Microbiol. Methods* 91, 459–467. doi: 10.1016/j.mimet.2012.09.029

Tiwari, A., Hokajärvi, A. M., Santo Domingo, J. W., Kauppinen, A., Elk, M., Ryu, H., et al. (2018). Categorical performance characteristics of method ISO 7899-2 and indicator value of intestinal enterococci for bathing water quality monitoring. *J. Water Health* 16, 711–723. doi: 10.2166/wh.2018.293

US Environmental Protection Agency, (2005). *Microbial Source Tracking Guide Document EPA/600/R-05/064*. Washington, DC: U.S. Environmental Protection Agency.

Uusheimo, S. A., Huotari, J. T., Tulonen, T. V., Aalto, S. L., Rissanen, A. J., and Arvola, L. M. J. (2018). High nitrogen removal in a constructed wetland receiving treated wastewater in a cold climate. *Environ. Sci. Technol.* 52, 13343–13350. doi: 10.1021/acs.est.8b03032

Uusi-Kämppä, J., and Heinonen-Tanski, H. (2008). Evaluating slurry broadcasting and injection to ley for phosphorus losses and fecal microorganisms in surface runoff. *J. Environ. Qual.* 37, 2339–2350. doi: 10.2134/jeq2007.0428

Waters, A. P., and McCuthan, T. F. (1990). Ribosomal RNA: nature's own polymerase-amplified target for diagnosis. *Parasitol. Today* 6, 56–59. doi: 10.1016/0169-4758(90)90071-b

Zhu, X. Y., Zhong, T., Pandya, Y., and Joerger, R. D. (2002). 16S rRNA-based analysis of microbiota from the cecum of broiler chickens. *Appl. Environ. Microbiol.* 68, 124–137. doi: 10.1128/AEM.68.1.124-137.2002

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.

Copyright © 2021 Rytkönen, Tiwari, Hokajärvi, Uusheimo, Vepsäläinen, Tulonen and Pitkänen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 June 2021
doi: 10.3389/fmicb.2021.674047

Characterization of Stormwater Runoff Based on Microbial Source Tracking Methods

*Silvia Monteiro[1], Gaspar Queiroz[2], Filipa Ferreira[2] and Ricardo Santos[1]**

[1] Laboratório Análises, Tecnico Lisboa, University of Lisbon, Lisbon, Portugal, [2] Department of Civil Engineering, Tecnico Lisboa, University of Lisbon, Lisbon, Portugal

Rainfall and associated urban runoff have been linked to an increased deterioration of environmental waters, carrying several pollutants including pathogenic microorganisms. Such happens because fecal matter is washed into storm drainage pipes that are afterward released into environmental waters. Stormwater has not been extensively characterized as it is, because most studies are performed either on drainage pipes that are often impacted by sewage leakage or directly in environmental waters following a rain event. In this study, stormwater collected directly from the streets, was monitored for the presence of fecal indicator bacteria (FIB) and three potential important sources of fecal contamination in urban environments (human, cats, and dogs) in three distinct basins in Lisbon, Portugal. Stormwater was collected in sterilized plastic boxes inserted in the storm drains, therefore collecting only runoff. High concentration of fecal contamination was detected with a high percentage of the samples displayed at least one source of contamination. A strong relationship was found between the number of detected sources and the precipitation levels. Although no statistical correlation was found between the locations and the presence of FIB or source markers, the results show a trend in geographical information on the type of urban use in each basin. To the best of our knowledge, this is the first study analyzing the runoff collected directly from the streets. This study suggests that, in urban areas, stormwater runoff is highly impacted by fecal matter, not only from domestic animals but also from human origin, before any cross-contamination in the drainage system and may, by itself, pose a high risk to human health and the environment, particularly if water reuse of this water without further disinfection treatment is the final goal.

Keywords: stormwater, fecal contamination, mitochondrial DNA, microbial source tracking, urban environment

OPEN ACCESS

Edited by:
Georg H. Reischer, Vienna University of Technology, Austria

Reviewed by:
Brent Gilpin, Institute of Environmental Science and Research (ESR), New Zealand
Pradip Gyawali, Institute of Environmental Science and Research (ESR), New Zealand

***Correspondence:**
Ricardo Santos
ricardosantos@ist.utl.pt

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *28 February 2021*
Accepted: *13 May 2021*
Published: *10 June 2021*

Citation:
Monteiro S, Queiroz G, Ferreira F and Santos R (2021) Characterization of Stormwater Runoff Based on Microbial Source Tracking Methods. Front. Microbiol. 12:674047. doi: 10.3389/fmicb.2021.674047

INTRODUCTION

In the last quarter of the twentieth century, most countries developed an increasing concern over urban pollution, trying to limit the impacts on natural environments of both solid waste and wastewater.

The preservation of receiving waters depends strongly on the quality of wastewater and stormwater discharges. The latter is a direct result of precipitation over an urban basin, dragging all pollutants accumulated during dry weather: rooftop and traffic pollutants, single discharges of

waste and pollutants from several other sources. A number of different studies have found that pollutant loads are strongly related to factors of three different natures: geomorphology of the catchment (Butler and Memon, 1999) local climacteric conditions or precipitation regime over the basin (Gnecco et al., 2005); and the type and intensity of land use (Gray, 2004).

Wastewater discharges are, for instance, being monitored and controlled in the European Union countries and in the United States, under international legislation – European Union Directive 2006/7/EC (CEC Council Directive, 2006) and U.S. Clean Water Act 1972 (USEPA, 1972) – which tries to limit the impact on natural environments by these waters. Despite the growing concerns in the preservation of receiving waters, and the increasing demand in water quality standards, there is no legislation concerning stormwater in particular, and no mandatory control over stormwater runoff. Moreover, there is a growing need to find alternative water sources, especially in areas characterized by water scarcity. About four million people already life in locations with high water stress at least 1 month per year (Mekonnen and Arjen, 2016) and it is expected that approximately, one-third of the world's population will be impacted by physical and/or economic water scarcity by the year 2025 (UNCCD, 2014). Several water sources have been proposed, including stormwater since it is considered a relatively clean, and affordable alternative water source. A large number of studies have shown the presence of pathogenic organisms in stormwater (Ahmed et al., 2010, 2011, 2014; Jongman and Korsten, 2016; Waso et al., 2016; Bae et al., 2019).

Current fecal indicator bacteria (FIB) do not allow allocating the source of detected fecal contamination. Tracking the source of microorganisms allows for a better understanding of the water cycle in general and urban stormwater behavior in particular. Methods trying to identify the source of fecal contamination are called source tracking (ST) methods. Most ST methods are based on the premise that different intestinal systems select for different microorganism populations, due to diet and digestive differences of their hosts (Santo Domingo et al., 2007). There are, however, different approaches like the use of species-associated eukaryotic mitochondrial DNA markers (mtDNA) for the direct determination of sources of contamination.

Several studies tested a great variety of genetic markers, trying to assert the source of a given fecal contamination (Ballesté et al., 2010) in several different countries and hydrological contexts, namely: in rural hydrological catchments in Austria (Reischer et al., 2011) and California, United States (Kildare et al., 2007) in rivers in Canada (Martellini et al., 2005; Kortbaoui et al., 2009) and Spain (Ballesté and Blanch, 2010); and in surface and coastal waters of the United Kingdom (Baker-Austin et al., 2010). These methods have also been applied to more urban contexts to identify human fecal contamination in natural water courses during wet weather, in Australia (Ahmed et al., 2008), and to demonstrate the existence of illegal connections into stormwater collectors, that continuously shed a steady stream of human fecal contamination, even in dry weather periods, in California, United States (Sercu et al., 2009).

In this study, mtDNA markers were selected because most ST approaches used were not completely reliable (USEPA, 2005) and do not provide solutions for the discrimination of some of the animals targeted in this study. Therefore, the aim of this study was: (i) design and validate mtDNA markers targeting domestic animal fecal contamination (cat and dog); and (ii) analyze stormwater collected directly from the streets following rain events, for the presence of FIB and MST markers targeting human, dog, and cat fecal contamination. By using mtDNA markers, the nucleic acid of the animal species is targeted directly instead of microbial-associated species that may vary due to dietary and climate differences.

To the best of our knowledge, this is the first study conducted in the runoff collected directly from the streets, following a rain event seeking to: (i) evaluate the quality of stormwater runoff in the city of Lisbon, with special focus on fecal contamination levels; (ii) use mtDNA markers designed specifically for species commonly found in the urban environment (humans, cats and dogs) to assess the origin of registered fecal pollution in the city of Lisbon; and (iii) correlate the sources of pollution with the activity areas within the city of Lisbon. Such study will allow for a better understanding not only of the levels of fecal contamination of these waters but will also provide information on the potential sources and ultimately, the microbiological risks of discharging such waters directly in the environment.

MATERIALS AND METHODS

Experimental Basins

Three experimental basins were chosen in Lisbon city center due to their quite dissimilar topographic, morphologic and urban use characteristics: Alcântara (A), Bairro das Ilhas (I) and Madalena Street (M) (**Supplementary Figure 1**). Alcântara basin has residential areas as well as areas more dedicated to commerce and nightlife. There are streets with intense traffic, and both steep and flat streets. Most streets have trees in the sidewalk, intensifying the presence of vegetal debris. The basin Bairro das Ilhas is significantly smaller than Alcântara, and is mainly residential with no commercial activity, tight one-way streets and the average slope is generally smooth. The traffic is of low intensity with areas of exclusive pedestrian access, with only one relevant green space with 4,000 m^2. We also observed that some of the buildings' rooftops drain directly to the pavement. The third experimental basin include Madalena Street, right in the middle of the 18th century historical center of Lisbon. The street has an intense commercial activity and many residential areas. There are streets with intense traffic, with bus lines and tramlines.

Rain and Temperature Data

Rain data were supplied by Instituto Geofísico Infante Dom Luiz (IGIDL) and by Laboratório Nacional de Engenharia Civil (LNEC). Results obtained of instantaneous precipitation for 2011 and 2012, in each of the rain gages (LNEC-NEC, LNEC-WWTP and IGIDL) are shown in **Supplementary Figure 1**, in **Supplementary Material**. Also shown in the graphics are the dates of the campaigns. The LNEC-WWTP rain gage did not register any data from 04-11-2011 around 11 a.m. to 28-11-2011 around 10 a.m., possibly due to a malfunction in the

equipment. The differences between recorded rain were mainly due to relevant spatial variability.

Mean temperatures varied between 20°C on October 2011 and 8°C on February 2012 (**Supplementary Table 1** in **Supplementary Material**) (IPMA, 2012).

Sample Collection

The collection campaigns took place from October 2011 to May 2012. Immediately before each rain event, a previously decontaminated plastic box was inserted into the storm drains, where stormwater runoff was collected during that specific rain event (**Figure 1**). A total of 75 samples were collected during five rain events: six locations at Alcântara, six at Ilhas and three at Madalena per sampling event. The contents of the boxes were transferred to a sterile 1L sampling bottles (Deltalab, Spain), and taken to the laboratory at (5 ± 3) °C within 2 h of collection. Samples were processed within 8 h of collection for FIB and mtDNA markers.

Fecal and Raw Wastewater Sample Collection

Fecal and raw wastewater samples were used for the validation of the newly designed domestic animals mtDNA markers (dog and cat). Fecal samples were collected using sterile tools and placed in sterile 50-ml tubes. Cow (10 samples), pig (9 samples), poultry (7 samples), pigeon (12 samples), gull (10 samples), rat (9 samples), cat (15 samples), and dog (15 samples) fecal matter was collected from different locations in the Lisbon area or donated by a variety of people and stores. Human fecal matter (15 samples) were provided by a variety of individuals of different ethnicities and age. All samples from individual hosts were transported to the laboratory at (5 ± 3) °C. Raw wastewater (20 samples) was collected at distinct wastewater treatment plants (WWTP) in Portugal and send to the laboratory refrigerated at (5 ± 3) °C. Upon arrival to the laboratory, all samples (fecal matter and raw wastewater) were immediately stored at (−30 ± 5) °C until further processing.

Microbial Indicators

Escherichia coli (EC) and intestinal Enterococci (IE) were analyzed using Colilert and Enterolert (IDEXX, United States), respectively, according to the manufacturers' instructions.

Sample Concentration and DNA Extraction

One-liter of stormwater was centrifuged at 9,000 × *g* for 15 min (Sigma 3k-18, Sigma Germany) and the pellet resuspended in 5 mL of the supernatant. Mitochondrial DNA was extracted from 200 μL of concentrate using the QIAamp DNA Mini Kit (Qiagen, United States), following the manufacturer's instructions. DNA elution was performed in 200 μL final volume. A process control, consisting of sterile distilled water, was analyzed with each experiment. In addition, an extraction control, consisting of nuclease-free water, was also conducted.

Fecal matter (220 mg) from the different animals and raw wastewater samples (220 μL) were extracted using the QIAamp DNA Stool Mini kit (Qiagen, United States), according to the manufacturer's instructions, with DNA eluted in 200 μL final volume. Extracted samples were stored at (−30 ± 5) °C until further processing.

Design of Primers for Detection of Fecal Contamination From Cat and Dog Origin

Specific primers were designed for the detection of fecal contamination from cat and dog origin. Human, cat, and dog mitochondrial DNA (GenBank database accession no. J01415, NC_028310, and KF907307, respectively) were aligned using the ClustalW software (European Bioinformatics Institute, United Kingdom). After the alignment, locations in the cat and dog mtDNA with the greatest divergence were chosen and inserted in the Primer Express® 3 software (Applied Biosystems, United States) to obtain candidate primers. The chosen candidates were further analyzed, *in silico*, using BLAST to detect and eliminate those with interspecies reactivity (Altschul et al., 1990). Each set of primer pairs was analyzed together and separately to contemplate their use. For each species, a specific-primer pair was selected (**Table 1**) and used for nested PCR.

Single and Nested PCR Conditions

Single PCR was performed in a 25 μL final volume reaction using 0.4 pmol/μL of each primer, 5 μL of extracted DNA using the illustra™ puReTaq Ready-To-Go PCR Beads (GE HealthCare, United Kingdom) following the manufacturer's instructions. Nested PCR was performed in the same conditions except that 1 μL of the single PCR reaction was used as template DNA and the internal primers were used. With each PCR, 10-fold dilution of every DNA extract was also evaluated. Process and extraction controls and positive and negative controls were analyzed with each PCR run. Amplifications were carried out in a Veriti 96 well thermal cycler (Applied Biosciences, United States) using the following conditions: (i) human: 94°C for 5 min, 55°C for 5 min, 35 cycles of 72°C for 2 min, 94°C for 40 s and 55°C for 1 min, with a final elongation at 72°C for 10 min (Martellini et al., 2005); and (ii) cat and dog: 94°C for 5 min, 59°C for 5 min, 35 cycles of 72°C for 2 min, 94°C for 40 s and 59°C for 1 min, with a final elongation at 72°C for 10 min. Amplicons were visualized after electrophoresis on 2.5% agarose gels.

Validation of Primers

The validation of the assays consisted of two steps: (i) determining the sensitivity; and (ii) determining the specificity. First, targeted single-component fecal suspensions were analyzed by with the two designed assays in a nested PCR to test and determine the sensitivity. Sensitivity was determined by analyzing 10-fold dilutions of DNA, extracted from the feces of each animal and quantified using NanoDrop (Thermo Fischer Scientific). The lowest quantity detected in three separate experiments was chosen and determined as the limit of detection (LoD; sensitivity) for each assay. Secondly, each assay was surveyed for possible cross-reactivity (specificity) with other species such as human, cow, pig, poultry, pigeon, gulls, and rats. The cross-reactivity assay consisted of the analysis of mixed fecal suspensions

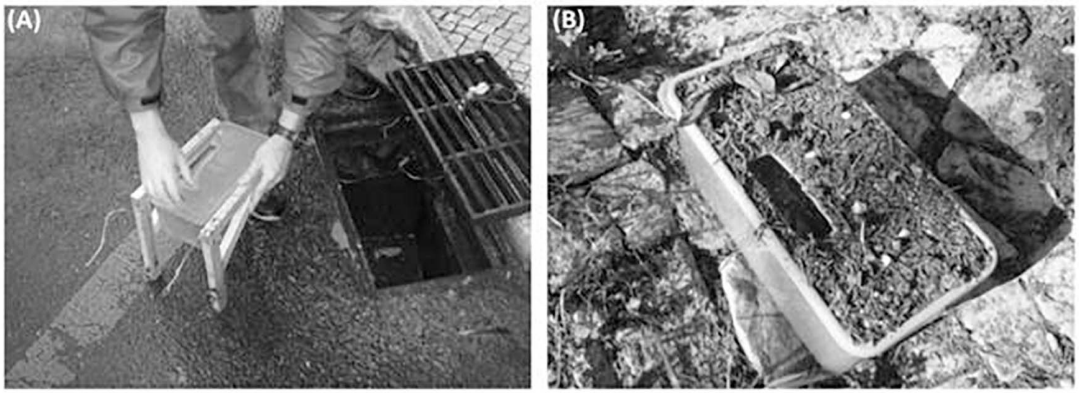


FIGURE 1 | (A) Simulation of experimental design with the placing of sample boxes in storm drains. **(B)** Sediment deposition after a rainfall event.

TABLE 1 | List of primers used in PCR and nested PCR in this study.

Source	Primers	Expected size	Reference
Human			Martellini et al., 2005
Single PCR			
Humito2-G	5′AGCCCTTCTAAACGCTAATCCAAGCCT-3′	659	
Humito2-D	5′-CTTGTCAGGGAGGTAGCGATGAGA-3′		
Nested PCR			
Humito11-G	5′-CCACTACTAGGCCTCCTCCTA-3′	612	
Humito11-D	5′-TAGCGATGAGAGTAATAGATAGGG-3′		
Cat			This study
Single PCR			
CatMito1-F	5′-CCTGTCCACACTACTTGTACTCATCGC-3′	539	
CatMito1-R	5′-AGATGGTTGTTTAGGATGGCTACG-3′		
Nested PCR			
CatMito2-F	5′-ATTTGATCCTATAGGGTCCGCC-3′	350	
CatMito2-R	5′-CCTATGAGCGACATGATGAAAGC-3′		
Dog			This study
Single PCR			
DogMito1-F	5′-ATGGCTCTAGCCGTTCGATTAAC-3′	638	
DogMito1-R	5′-GGCTAGGAGGACTGAGGTGTTGAG-3′		
Nested PCR			
DogMito2-F	5′-CATTAGGATTCACAACCAACCTGTTA-3	236	
DogMito2-R	5′-AATAATGCCGGTAGGAGGTCAG-3′		

composed of the non-target species mentioned above using the assays developed.

Statistical Analysis

All data analysis was done either with Microsoft Excel 2016 or IBM SPSS Statistics 25 (IBM, United States). Data were converted into a logarithmic format. Three factors – sampling area, FIB concentration and precipitation levels – were analyzed for significant difference in the variance of the occurrence of each marker in the runoff. One-way ANOVA was conducted to determine the association between sampling location and the presence of FIB (EC and IE) and MST markers (Human, dog, and cat). The association between the presence of FIB and MST markers was assessed using One-way ANOVA. Spearman rank order correlation was used for calculation of correlation coefficients between FIB and precipitation levels.

RESULTS

Microbial Indicators

The percentage of positive samples for the three basins was 100% for IE and 95% EC (results for EC ranging from 100% positive samples in Ilhas and Madalena basins and 89% in Alcântara). The distribution of EC and IE concentrations for the three basins is presented in **Figure 2**. The highest EC concentration was found in Alcântara (mean concentration: 4.29 ± 0.66 log MPN/100 mL) followed by Madalena (mean concentration: 3.38 ± 0.92 log

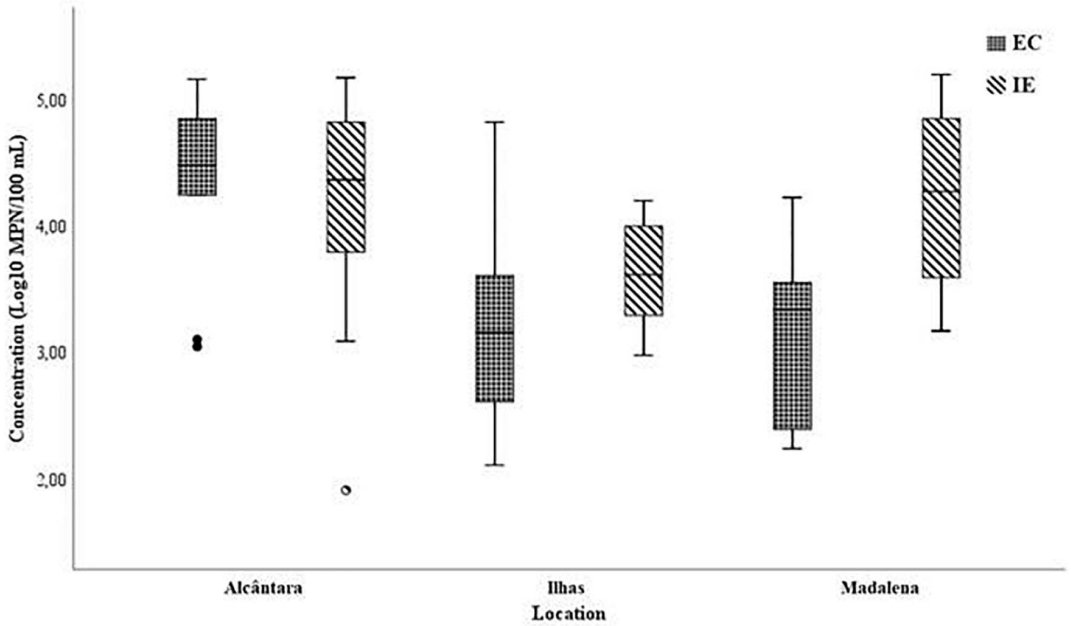


FIGURE 2 | Concentration of EC (A) and IE (B) in the three locations chosen, Boxes, 25th and 75th percentile; Whiskers, 10th and 90th percentile; inside the boxes, median.

MPN/100 mL) and Ilhas (mean concentration: 3.22 ± 0.71 log MPN/100 mL). Concentration of IE was highest in Madalena (mean concentration: 4.28 ± 0.78 log MPN/100 mL), followed by Alcântara and Ilhas (mean concentrations: 4.07 ± 0.92 log MPN/100 mL and 3.63 ± 0.43 log MPN/100 mL, respectively). The concentration of IE was greater than that of EC in two of the basins, Ilhas and Madalena. For Alcântara, the concentration of EC was higher than the concentration of IE, although being undetected in some of the samples.

Crossing the results for EC and IE with the rain data, there is no perceivable impact on the concentration of IE due to an increase in precipitation (**Table 2**). For EC, only for the Ilhas basin, a negative impact with the precipitation intensity was observed ($r = -0.528$, $\rho < 0.05$; **Table 2**).

Due to the impact of cumulative rain in the previous days of sampling, the mean concentration of EC and IE was calculated and compared with the cumulative values of rain in the day of the sampling and on the two previous days (**Figure 3**).

The mean concentration for both microorganisms was mostly similar throughout the sampling campaign showing little to no effect of the levels of rain in the number of microorganisms, with EC displaying generally lower mean concentration than IE (exception of sampling date 2). It is also important to note that no significant statistical difference was found for the presence of EC or IE with the location settings (One-way Anova, $p > 0.05$).

TABLE 2 | Spearman rank order correlation between the levels of precipitation and the concentration of FIB per basin area.

	Microorganism (spearman rank order correlation, *r*)	
Location	**EC**	**IE**
Alcântara	0.114	0.146
Ilhas	−0.528	−0.297
Madalena	0.103	0.128

Validation of Designed Primers

To determine the specificity of the designed primers for cat and dog sources, the primers were tested against fecal matter from different species, including humans, cow, pig, poultry, pigeons and gulls. These animals were chosen specifically since they could be the main sources of fecal contamination in urban and river environments found in the River Tagus catchment.

No cross-reactivity was observed for each species-specific primer, with the exception of a single raw wastewater sample that showed positive results for the dog targeting assay (**Table 3**). The assays developed for the specific detection of contamination from domestic animals produced positive results only when the corresponding species was analyzed representing the absence of non-specific amplifications in the fecal samples. The results are shown in **Supplementary Figure 2**.

Sensitivity tests were made for each set of nested-PCR by decreasing the quantity of the corresponding targeted animal DNA in the PCR reaction (**Supplementary Figure 3**). The sensitivity limit was determined to be around 1 pg for cat mitochondrial marker and 0.01 pg for dog mitochondrial marker following nested PCR.

Evaluation of Sources of Fecal Contamination in the Catchment Area

To determine the origin of fecal contamination responsible for the high concentration of EC and IE in the basins, three sources

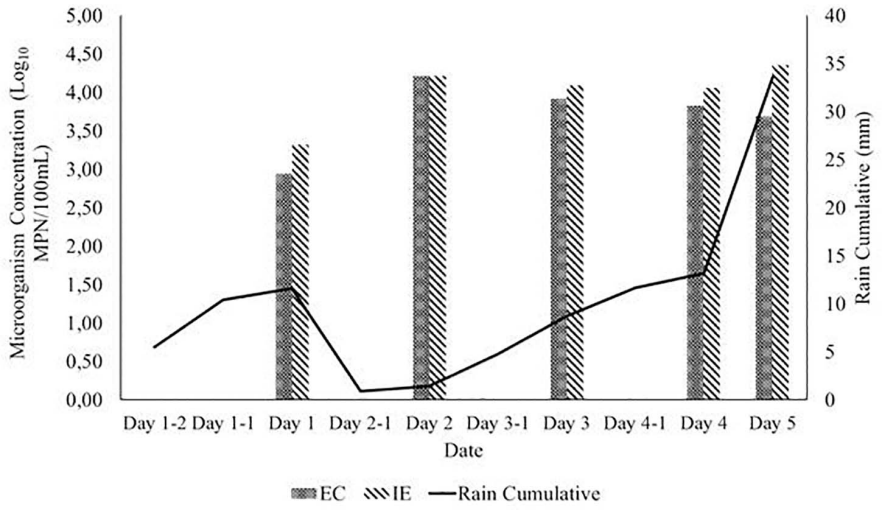


FIGURE 3 | Mean concentration of EC and IE per rain event: Day X-2 represents the rain intensity.

TABLE 3 | Specificity of the cat and dog mtDNA markers in samples from varying fecal origin.

Source	Percentage of positive samples with targeted assays (no. positive samples/total no. of tested samples)	
Fecal matter	**CatMito**	**DogMito**
Cat	100% (15/15)	0% (0/15)
Dog	0% (0/15)	100% (15/15)
Human	0% (0/15)	0% (0/15)
Cow	0% (0/10)	0% (0/10)
Pig	0% (0/9)	0% (0/9)
Poultry	0% (0/7)	0% (0/7)
Pigeon	0% (0/12)	0% (0/12)
Gulls	0% (0/10)	0% (0/10)
Rats	0% (0/9)	0% (0/9)
Raw Wastewater	0% (0/20)	5% (1/20)

were evaluated due to their potential impact and presence in urban environments: human, cats and dogs. Data showed that 68% of the samples were positive for at least one of the targeted sources, with 9% showing positive signals for all sources tested. Human was the most relevant source, followed by dog and cat at 49, 38, and 35%, respectively. **Figure 4** shows the percentage of positive samples for the three basins by source.

The prevalence of fecal markers in Alcântara was 55, 34, and 30%, for human, dog, and cat origin, respectively. In the remaining two basins, the prevalence of each marker was similar and closer. In Ilhas, the prevalence was 44, 41, and 36% for dog, cat and human assays, respectively. Finally, in the Madalena basin the results were 44, 38, and 38% for human, dog, and cat fecal contamination, respectively. In spite of the higher percentage of positive samples in Alcântara for human fecal contamination marker, there was no statistical significant difference between location and the presence of a particular fecal contamination (One-way Anova, $p > 0.05$). Additionally, no statistically significant difference was found when analyzing each fecal indicator bacteria (FIB) and the presence of targeted marker (One-way Anova, $p > 0.05$).

There was statistically significant difference between the precipitation levels and the assays targeting dog and cat fecal contamination ($p = 8.1 \times 10^{-15}$, and $p = 6.0 \times 10^{-6}$, respectively; $p < 0.001$). Additionally, there was strong association between the number of fecal contamination targets detected (zero markers, one marker, two markers, and all markers detected) and the levels of fecal contamination ($p = 1.1 \times 10^{-10}$; $p < 0.001$). Sampling dates with greater precipitation levels also showed a larger number of detected sources whereas days with low precipitation levels had lower numbers of detected markers. Comparison between the presence of different fecal contamination markers and precipitation is showed in **Figure 5**.

Observing data from **Figure 5** it is noticeable that sampling days with higher levels of precipitation (days 5 and 6) also displayed higher number of combined detected MST targeting assays, with detection of at least one MST marker in these days. For sampling date 2, displaying the lowest precipitation intensity, the combined number of MST markers varied between 0 and 1, at a mean marker detection level of 0.25. The highest percentage of positive samples was obtained during the month of May (80%) whereas the lowest was determined for the months of March and April (28 and 21%, respectively).

DISCUSSION

Little or no impact of the precipitation intensity on the concentration of FIB was detected throughout this study, with the exception of a moderate negative correlation between Ilhas basin and EC concentration. Such result may be a consequence of a "first flush" event caused by a

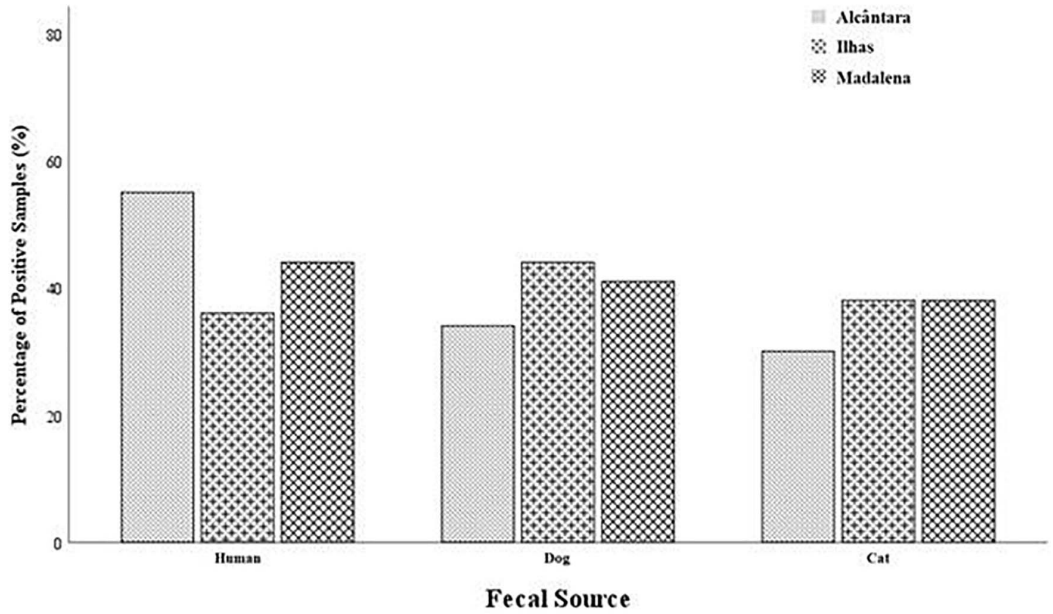


FIGURE 4 | Percentage of positive samples in each basin per source of fecal contamination.

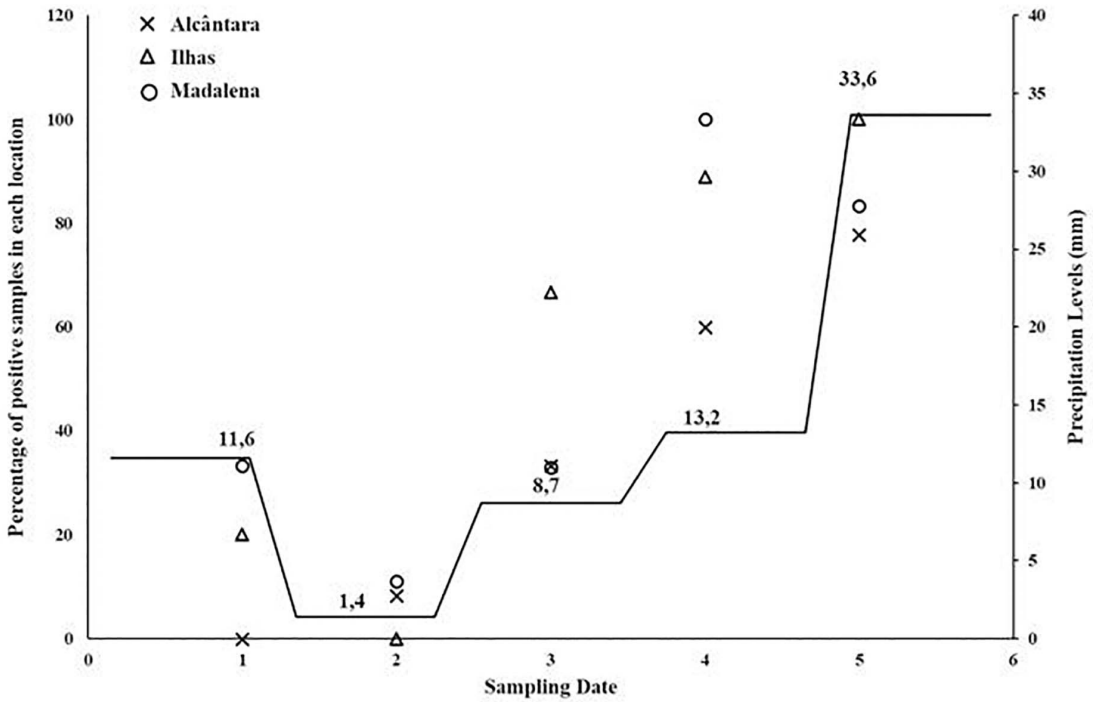


FIGURE 5 | Relation between the levels of precipitation and the percentage of detected MST assays per sampling date.

possible increase in the levels of EC followed by dilution effect with continuing rain and from the fact that Ilhas is a smaller residential area, with no commercial or nightlife areas. Similar results have been previously shown, where different sampling locations also showed different correlation for EC during rainfall events (Kleinheinz et al., 2009). Billian et al. (2018) studied the effect of precipitation in the movement, and therefore in the levels of FIB, through septic drain fields and found no relationship between precipitation and FIB.

To determine the sources responsible for to the high levels of fecal contamination found in the runoff, mtDNA markers were chosen to contemplate three major potential sources in urban areas: human and domestic animals (dogs and cats). Data from this study has shown that the existence of a given source of contamination is not related to the presence of FIB. Moreover, EC and IE were detected in samples in the absence of the targeted sources of fecal contamination. Such results are possibly due to FIB inputs from other sources in urban areas such as pigeons and gulls. These results are in accordance with previous studies (Waso et al., 2018a; Hajj-Mohamad et al., 2019) that tested the use of different markers, including human mitochondrial DNA marker, to determine correlation between MST markers, FIB and land use and have determined that no correlation existed between the markers and EC in residential areas. A study by Staley et al. (2016) determined that human and gull were the most important sources of fecal contamination in stormwater outfalls, indicating a large input from such sources in the levels of contamination (Staley et al., 2016). Similarly, a study by Waso et al. (2018b) concluded that fecal contamination from pigeon origin had a high impact on the quality of harvested rainwater.

In the present study, although no direct statistical correlation was found between location and the presence of a certain source of fecal contamination, data provided trends on the different activities occurring in the chosen basins. The city of Lisbon does not have open sewers, therefore the mtDNA results can be directly imputed solely to urban runoff. Alcântara basin is a residential area with dense commercial areas, heavy traffic and packed with nightlife clubs, resulting in the highest percentage of positive results for human-associated marker and the lowest for domestic animals. Alcântara basin, with a high number of homeless individuals and with a vibrant nightlife is the area where the impact of mtDNA is more noticeable. Ilhas basin is a small and quiet residential neighborhood in Lisbon, where most of the households have their own private garage minimizing the time that the residents are on the sidewalks. Usually the domestic animals are taken for walks in the surrounding streets. This impact is markedly visible due to the lowest human fecal contamination and the highest contamination from domestic animals detected. On the other hand, data obtained for the Madalena basin is a reflection of the urban use of this area, a mix between Alcântara and Ilhas basin with a mix of residential and commercial areas. Interestingly, the concentration of EC was also higher than that of IE in Alcântara, but lower for the remaining chosen sites. The relation between these two indicators is commonly accepted as an indication of human/animal source identification, further corroborating the results obtained by this study using mtDNA.

Data for MST markers are in agreement with previously published work showing an increase in the detection of FIB and other organisms from fecal contamination following a precipitation event (Lipp et al., 2001; Kelsey et al., 2004; Mallin et al., 2009). Billian et al. (2018) have shown a correlation between precipitation and the presence of *Bacteroides* HF183 marker when analyzing the movement of MST markers and conventional fecal indicators through septic drain fields. Human and animal *Bacteroides* markers have been consistently identified as the "gold standard" in determining the sources of fecal contamination (Harwood et al., 2014). However, they present several challenges, including dietary and climate dependence which may diminish the sensitivity of these markers. The choice of targeting mtDNA for source tracking in this particular study resulted from several premises: (i) mtDNA markers target specifically eukaryote cell DNA from the target animals rather than targeting specific bacteria or viruses associated to a specific animal; (ii) mtDNA evolves faster than nuclear DNA therefore providing an adequate number of sequence variations for the design of species-associated PCR assays; (iii) a large number of exfoliated cells are released in the feces of animals each containing a high number of mitochondrion and respective DNA sequences, which increases largely the sensitivity of the assays; (iv) specific strains of *Bacteroides* spp. may not be present in all of their specific hosts, as is the case for instance of the HF183 marker which is not present in the intestinal microbiome of all humans (Harwood et al., 2009); and (v) *Bacteroides* spp. are obligatory anaerobes and in our particular setting this could conduct rather immediately to their lysis and subsequent DNA degradation. In addition, the HF183 marker has been detected in gull waste samples, not only from the feces of gulls but also in the cloacae, indicating that gulls may transport human fecal pollution to places not impacted directly by human action (Alm et al., 2018).

The results from this study highlight the need for a better control of urban runoff, that may be a carrier of pathogenic organisms with the potential to cause risk to human health, indicating also the extreme importance of sources of fecal contamination other than human and domestic animals in the quality of the runoff.

CONCLUSION

Due to the experimental design, the present study, shows the direct analysis of urban runoff and the potential effects of this phenomenon in environmental waters. The majority of the studies targeting stormwater quality have been conducted either in the drainage pipes, many times cross-contaminated with sewage, or in environmental waters following a rain event. Such sampling approach change the load and type of fecal contamination that is being determined possibly masking largely the real input of urban runoff on the environment. This study shows, therefore, that stormwater runoff is one of the most important sources of biological pollutants that can potentially impair water quality and pose risks to the ecosystem and human health. Additionally, urban stormwater is being used in multiple situations to increase non-potable and potable water supplies within cities and other urban areas (Sidhu et al., 2012), with the potential to pose serious health problems. Based on the data presented in this study, it is important to identify the sources of fecal contamination in stormwater runoff for a better and more targeted remediation action to reduce risks to public health. In addition, disinfection treatment of stormwater must be performed if reuse is considered, even for non-potable purposes.

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

SM, GQ, FF, and RS contributed to conception and design of the study. GQ performed the sampling collection. SM performed the laboratory treatment of the samples. GQ and SM performed the statistical analysis. All authors contributed to manuscript revision, read, 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/fmicb.2021.674047/full#supplementary-material

REFERENCES

Ahmed, W., Gardner, T., and Toze, S. (2011). Microbiological quality of roof-harvested rainwater and health risks: a review. *J. Environ. Qual.* 40, 13–21. doi: 10.2134/jeq2010.0345

Ahmed, W., Goonetilleke, A., and Gardner, T. (2010). Implications of faecal indicator bacteria for the microbiological assessment of roof-harvested rainwater quality in southeast queensland. *Australia. Can. J. Microbiol.* 56, 471–479. doi: 10.1139/w10-037

Ahmed, W., Richardson, K., Sidhu, J. P. S., Jagals, P., and Toze, S. (2014). Inactivation of faecal indicator bacteria in a roof-captured rainwater system under ambient meteorological conditions. *J. Appl. Microbiol.* 116, 199–207. doi: 10.1111/jam.12342

Ahmed, W., Stewart, J., Gardner, T., and Powell, D. (2008). A real-time polymerase chain reaction assay for quantitative detection of human-specific enterococci surface protein marker in sewage and environmental waters. *Environ. Microbiol.* 10, 3255–3264. doi: 10.1111/j.1462-2920.2008.01715.x

Alm, E. W., Daniels-Witt, Q. R., Learman, D. R., Ryu, H., Jordan, D. W., Gehring, T. M., et al. (2018). Potential for gulls to transport bacteria from human waste sites to beaches. *Sci. T. Environ.* 615, 123–130. doi: 10.1016/j.scitotenv.2017.09.232

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bae, S., Maestre, J. P., Kinney, K. A., and Kirisits, M. J. (2019). An examination of the microbial community and occurrence of potential human pathogens in rainwater harvested from different roofing materials. *Water Res.* 159, 406–413. doi: 10.1016/j.watres.2019.05.029

Baker-Austin, C., Rangdale, R., Lowther, J., and Lees, D. N. (2010). Application of mitochondrial DNA analysis for microbial source tracking purposes in shellfish harvesting waters. *Water Sci. Technol.* 61, 1–7. doi: 10.2166/wst.2010.767

Ballesté, E., and Blanch, A. R. (2010). Persistence of *Bacteroides* species populations in a river as measured by molecular and culture techniques. *Appl. Environ. Microbiol.* 76, 7608–7616. doi: 10.1128/aem.00883-10

Ballesté, E., Bonjoch, X., Belanche, L. A., and Blanch, A. R. (2010). Molecular indicators used in the development of predictive models for microbial source tracking. *Appl. Environ. Microbiol.* 76, 1789–1795. doi: 10.1128/aem.02350-09

Billian, H., Krometis, L., Thompson, T., and Hagedorn, C. (2018). Movement of traditional fecal indicator bacteria and source-tracking targets through septic drainfields. *Sci. Tot. Environ.* 61, 1467–1475. doi: 10.1016/j.scitotenv.2017.08.131

Butler, D., and Memon, F. A. (1999). Dynamic modelling of roadside gully pots during wet weather. *Water Res.* 33, 3364–3372. doi: 10.1016/s0043-1354(99)00050-0

Council, C. E. C., and Directive. (2006). Council directive 2006/7/EC of 15 February 2006 concerning the management of bathing water quality. *OJ* 67, 37–51.

Gnecco, I., Berreta, C., Lanza, L., and La Barbera, P. (2005). Storm water pollution in the urban environment of Genoa. Italy. *Atmospheric Res.* 77, 60–73. doi: 10.1016/j.atmosres.2004.10.017

Gray, N. (2004). *Biology and Wastewater Treatment*. London: Imperial College Press.

Hajj-Mohamad, M., Hachad, M., Deschamps, G., Sauvé, S., Villemur, R., Blais, M.-A., et al. (2019). Fecal contamination of storm sewers: valuating wastewater micropollutants, human-specific *Bacteroides* 16S rRNA, and mitochondrial DNA genetic markers as alternative indicators of sewer cross connections. *Sci. T. Environ.* 659, 548–560. doi: 10.1016/j.scitotenv.2018.12.378

Harwood, V. J., Brownell, M., Wang, S., Lepo, J., Ellender, R. D., Ajidahun, A., et al. (2009). Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. *Water Res.* 43, 4812–4819. doi: 10.1016/j.watres.2009.06.029

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

Jongman, M., and Korsten, L. (2016). Microbial quality and suitability of roof-harvested rainwater in rural villages for crop irrigation and domestic use. *J. Water Health* 14, 961–971. doi: 10.2166/wh.2016.058

Kelsey, H., Porter, D. E., Scott, G., Neet, M., and White, D. (2004). Using geographic information systems and regression analysis to evaluate relationships between land use and fecal coliform bacterial pollution. *J. Exp. Mar. Bio. Ecol.* 298, 197–209. doi: 10.1016/s0022-0981(03)00359-9

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog fecal bacteroidales: a Bayesian approach. *Water Res.* 41, 3701–3715. doi: 10.1016/j.watres.2007.06.037

Kleinheinz, G. T., McDermott, C. M., Hughes, S., and Brown, A. (2009). Effects of rainfall on *E. coli* concentration at Door County. *Wisconsin Beaches. Int. J. Microbiol.* 2009:876050.

Kortbaoui, R., Locas, A., Imbeau, M., Payment, P., and Villemur, R. (2009). Universal mitochondrial PCR combined with species-specific dot-blot assay as a source-tracking method of human, chicken, ovine, and porcine in fecal-contaminated surface water. *Water Res.* 43, 2002–2010. doi: 10.1016/j.watres.2009.01.030

Lipp, E. K., Kurz, R., Vincent, R., Rodriguez-Palacios, C., Farrah, S. R., and Rose, J. B. (2001). The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. *Estuaries* 24, 266–276. doi: 10.2307/1352950

Mallin, M. A., Johnson, V. L., and Ensign, S. H. (2009). Comparative impacts of storm water runoff on water quality of an urban, suburban, and a rural system. *Environ. Monit. Assess.* 159, 475–491. doi: 10.1007/s10661-008-0644-4

Martellini, A., Payment, P., and Villemur, R. (2005). Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water Res.* 39, 541–548. doi: 10.1016/j.watres.2004.11.012

Mekonnen, M. M., and Arjen, Y. H. (2016). Four billion people facing severe water scarcity. *Sci. Adv.* 2:e1500323. doi: 10.1126/sciadv.1500323

Reischer, G. H., Kollanur, D., Vierheilig, J., Wehrspaun, C., Mach, R. L., Sommer, R., et al. (2011). Hypothesis-Driven approach for the identification of fecal pollution sources in water resources. *Environ. Sci. Technol.* 45, 4038–4045. doi: 10.1021/es103659s

Santo Domingo, J. W., Bambic, D. G., Edge, T. A., and Wuertz, S. (2007). Quo vadis source tracking? towards a strategic framework for environmental monitoring of fecal pollution. *Water Res.* 41, 3539–3552. doi: 10.1016/j.watres.2007.06.001

Sercu, B., Werfhorst, L. C. V. D., Murray, J., and Holden, P. A. (2009). Storm drains are sources of human fecal pollution during dry weather in three urban

southern California watersheds. *Environ. Sci. Technol.* 43, 293–298. doi: 10.1021/es801505p

Sidhu, J. P., Hodgers, L., Ahmed, W., Chong, M. N., and Toze, S. (2012). Prevalence of human pathogens and indicators in storm water runoff in Brisbane. Australia. *Water Res.* 46, 6652–6660. doi: 10.1016/j.watres.2012.03.012

Staley, Z. R., Grabuski, J., Sverko, E., and Edge, T. A. (2016). Comparison of microbial and chemical source tracking markers to identify fecal contamination sources in the Humber River (Toronto. Ontario, Canada) and associated storm water outfall. *Appl. Environ. Microbiol.* 82, 6357–6366. doi: 10.1128/aem.01675-16

UNCCD (2014). *Desertification: The Invisible Frontline*. Bonn: UNCCD.

USEPA (1972). *U.S. Clean Water Act*. Washington, DC: USEPA.

USEPA (2005). *Microbial Source Tracking guide document*. Cincinnati, OH: Office of Research and Development, United States Environmental Protection Agency. Document number EPA 600/R–05/064.

Waso, M., Khan, S., and Khan, W. (2018a). Development and small-scale validation of a novel pigeon-associated mitochondrial DNA source tracking marker for the detection of fecal contamination in harvested rainwater. *Sci. Environ.* 615, 99–106. doi: 10.1016/j.scitotenv.2017.09.229

Waso, M., Khan, S., and Khan, W. (2018b). Microbial source tracking markers associated with domestic rainwater harvesting systems: correlation to indicator organisms. *Environ. Res.* 161, 446–455. doi: 10.1016/j.envres.2017.11.043

Waso, M., Ndlovu, T., Dobrowsky, P. H., Khan, S., and Khan, W. (2016). Presence of microbial and chemical source tracking markers in roof-harvested rainwater and catchment systems for the detection of fecal contamination. *Environ. Sci. Poll. Res. Intern.* 23, 16987–16981. doi: 10.1007/s11356-016-6895-7

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.

Copyright © 2021 Monteiro, Queiroz, Ferreira and Santos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fmicb.2021.632850

Antibiotic Resistance and Sewage-Associated Marker Genes in Untreated Sewage and a River Characterized During Baseflow and Stormflow

Warish Ahmed[1], Pradip Gyawali[2], Kerry A. Hamilton[3,4], Sayalee Joshi[3,4], David Aster[5], Erica Donner[6], Stuart L. Simpson[7] and Erin M. Symonds[8]*

[1] CSIRO Land and Water, Ecosciences Precinct, Dutton Park, QLD, Australia, [2] Institute of Environmental Science and Research Ltd. (ESR), Porirua, New Zealand, [3] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, United States, [4] Biodesign Center for Environmental Health Engineering, The Biodesign Institute, Arizona State University, Tempe, AZ, United States, [5] Department of Agriculture and Fisheries, Ecosciences Precinct, Dutton Park, QLD, Australia, [6] Future Industries Institute, University of South Australia, University Boulevard, Mawson Lakes, SA, Australia, [7] CSIRO Land and Water, Lucas Heights, NSW, Australia, [8] College of Marine Science, University of South Florida, St. Petersburg, St. Petersburg, FL, United States

OPEN ACCESS

Edited by:
Georg H. Reischer, Vienna University of Technology, Austria

Reviewed by:
Anyi Hu, Institute of Urban Environment (CAS), China
Krassimira Hristova, Marquette University, United States
Julia Vierheilig, Karl Landsteiner University of Health Sciences, Austria

***Correspondence:**
Warish Ahmed
Warish.Ahmed@csiro.au

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *24 November 2020*
Accepted: *29 April 2021*
Published: *11 June 2021*

Citation:
Ahmed W, Gyawali P, Hamilton KA, Joshi S, Aster D, Donner E, Simpson SL and Symonds EM (2021) Antibiotic Resistance and Sewage-Associated Marker Genes in Untreated Sewage and a River Characterized During Baseflow and Stormflow. Front. Microbiol. 12:632850. doi: 10.3389/fmicb.2021.632850

Since sewage is a hotspot for antibiotic resistance genes (ARGs), the identification of ARGs in environmental waters impacted by sewage, and their correlation to fecal indicators, is necessary to implement management strategies. In this study, sewage treatment plant (STP) influent samples were collected and analyzed using quantitative polymerase chain reaction (qPCR) to investigate the abundance and correlations between sewage-associated markers (i.e., *Bacteroides* HF183, *Lachnospiraceae* Lachno3, crAssphage) and ARGs indicating resistance to nine antibiotics (belonging to aminoglycosides, beta-lactams, sulfonamides, macrolides, and tetracyclines). All ARGs, except bla_{VIM}, and sewage-associated marker genes were always detected in untreated sewage, and *ermF* and *sul1* were detected in the greatest abundances. *intI1* was also highly abundant in untreated sewage samples. Significant correlations were identified between sewage-associated marker genes, ARGs and the *intI1* in untreated sewage ($\tau = 0.488$, $p = 0.0125$). Of the three sewage-associated marker genes, the BIO-ENV procedure identified that HF183 alone best maximized correlations to ARGs and *intI1* ($\tau = 0.590$). Additionally, grab samples were collected from peri-urban and urban sites along the Brisbane River system during base and stormflow conditions, and analyzed for *Escherichia coli*, ARGs, the *intI1*, and sewage-associated marker genes using quantitative polymerase chain reaction (qPCR). Significant correlations were identified between *E. coli*, ARGs, and *intI1* ($\tau = 0.0893$, $p = 0.0032$), as well as with sewage-associated marker genes in water samples from the Brisbane River system ($\tau = 0.3229$, $p = 0.0001$). Of the sewage-associated marker genes and *E. coli*, the BIO-ENV procedure identified that crAssphage alone maximized correlations with ARGs

and *intI1* in river samples (τ = 0.4148). Significant differences in *E. coli*, ARGs, *intI1*, and sewage-associated marker genes, and by flow condition (i.e., base vs. storm), and site types (peri-urban vs. urban) combined were identified (R = 0.3668, p = 0.0001), where percent dissimilarities between the multi-factorial groups ranged between 20.8 and 11.2%. Results from this study suggest increased levels of certain ARGs and sewage-associated marker genes in stormflow river water samples compared to base flow conditions. *E. coli*, HF183 and crAssphage may serve as potential indicators of sewage-derived ARGs under stormflow conditions, and this merits further investigation. Data presented in this study will be valuable to water quality managers to understand the links between sewage pollution and ARGs in urban environments.

Keywords: microbial source tracking, antibiotic resistance, stormwater, human health risks, sewage pollution

INTRODUCTION

Sewage treatment plants (STPs) collect sewage from various sources, including households, hospitals, commercial and industrial sites followed by treatment processes to remove biological and chemical contaminants before the treated water is discharged back into the environment or recycled for commercial enterprises. Human and animal fecal waste contamination is a global problem and can occur from wet/dry weather overflows, septic tanks, faulty sewer lines, illicit sewer connections, lift stations, and in the event of natural disasters such as earthquakes and flooding (Olds et al., 2018; Ahmed et al., 2019a). Sewage contamination results in the dissemination of pathogens, nutrients, toxicants, endocrine disruptors, antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARGs) into the environment (Rodriguez-Manzano et al., 2010). In addition to sewage, defecation from wildlife, livestock, and pets contribute significant loads of pathogens, ARB, and ARGs to waterways by direct deposition or by stormwater runoff (Cox et al., 2005; Ahmed et al., 2019c).

Antibiotic resistant bacteria pose significant human health risks. For example, at least 2.8 million people contract an antibiotic-resistant infection, and more than 35,000 deaths occur each year in the United States as per CDC's 2019 AR Threats Report (CDC, 2019). It has been estimated that antibiotic-resistant infections would be responsible for 10 million excess deaths globally, and a cumulative cost of US$ 100 trillion by 2050 if the current rapid evolution and spread of antimicrobial resistance are not abated (The Review on Antimicrobial Resistance, 2014). In Australia, 290–1,600 deaths/year are due to antibiotic resistance (Australian Commission on Safety and Quality in Health Care (ACSQHC), 2019). ARGs are considered emerging contaminants in the aquatic environments since they can be transferred to pathogenic bacteria in ecosystems via horizontal gene transfer (HGT) (Nguyen et al., 2019). Therefore, mitigation strategies are needed to prevent their widespread dissemination. The World Health Organization (WHO) identified water safety and improved sanitation as critical components in preventing the spread of ARB (WHO, 2014).

The presence of fecal contamination is typically determined by monitoring fecal indicator bacteria (FIB), such as *Escherichia coli* for freshwater and enterococci for marine and estuarine waters (Rochelle-Newall et al., 2015; Ahmed et al., 2019a). However, limitations of FIB monitoring include factors such as differential decay rates and poor correlations to pathogens in aquatic environments (Harwood et al., 2005; Signor et al., 2005; Wade et al., 2008; Ferguson et al., 2012; Korajkic et al., 2019), potential environmental, non-fecal sources (Badgley et al., 2011; Byappanahalli et al., 2012), and most importantly, their inability to identify the source of contamination (Harwood et al., 2014). Without knowing the sources, relative pathogen contributions cannot be assigned, thus inhibiting the ability to manage human health risks (Soller et al., 2015).

The application of microbial source tracking (MST) tools enabled researchers and regulators to differentiate between sources of fecal contamination (Harwood et al., 2014; Ahmed et al., 2016). The most widely used MST tools involve the analysis of host-associated marker genes using qPCR (Harwood et al., 2014). The benefit of host-associated marker genes is that they occur in far greater abundances compared to pathogens and are highly specific to the type of contamination source (e.g., human vs. wildlife) (Stoeckel and Harwood, 2008; Ahmed et al., 2016). These tools are currently being applied throughout the world to gain insight into the sources of fecal contamination in water bodies and to help guide risk management solutions (Harwood et al., 2014; Derx et al., 2016; Ahmed et al., 2018).

Sewage treatment plants are considered hot spots for ARGs (Rizzo et al., 2013; Pazda et al., 2019). Sewage-associated marker genes such as *Bacteroides* HF183, crAssphage, *Lachnospiraceae* Lachno3, and others are highly abundant in untreated sewage (Ahmed et al., 2016; Feng et al., 2018; Farkas et al., 2019). Therefore, in the event of recent sewage pollution, it is highly likely that both sewage-associated markers and certain ARGs will be present in environmental waters. A recent study analyzed relative ARG abundance and accompanying extent of fecal contamination in publicly available metagenomic data, using crAssphage sequences as a marker of sewage contamination (Karkman et al., 2019). The authors were investigating whether an increased abundance of ARGs in sewage and sewage-impacted environments was due to on-site selection pressure by residual antibiotics or was simply a result

of fecal pollution with ARB. The analysis suggested that the presence of ARGs in the environments can largely be explained by fecal pollution, with no clear signs of on-site selection in the environment.

Patterns of increased human fecal pollution and ARGs were previously identified in environmental waters (Ahmed et al., 2018; Reynolds et al., 2020; Stange and Tiehm, 2020; Moretto et al., 2022). Abundances of 47 ARGs in several storm drain outfalls during dry and wet weather in Tampa Bay, FL, United States along with sewage-associated *Bacteroides* HF183 and crAssphage markers were monitored (Ahmed et al., 2018). The study found that the abundances of sewage-associated markers and many ARGs in were relatively high, in water samples collected during wet weather compared to dry weather, and that storm drain outfalls may be potential hot spots for microbial contamination in Tampa Bay. CrAssphage and HF183 were significantly correlated with *intl1*, and several ARGs such as *sul1*, tet*(M)*, *ampC*, *mexB*, and tet*(W)*. Similarly, greater abundances of human MST markers and ARGs were measured after heavy rains in a German spring (Stange and Tiehm, 2020). While HF183 was consistently and significantly correlated to several ARGs in the Liffey Estuary, Ireland, HF183 was not always significantly correlated with ARGs in the urban streams nearby (Reynolds et al., 2020).

Information on the positive correlation between the presence of sewage-associated marker genes and ARGs in environmental waters may aid in the management of fecal pollution and ARG dissemination. Yet it is also possible that ARG abundance may be linked to animal fecal (Lee et al., 2020) and non-fecal sources. Sediments can potentially act as reservoirs for ARB and ARG, and resuspension from sediments during storm events or due to other disturbances, is likely to occur. Furthermore, ARGs have also been detected in Lake Tai, China, despite low levels of human, ruminant, and pig fecal pollution (Stange et al., 2019).

The main objective of this study was to better understand the abundance and correlation between sewage-associated markers and ARGs in sewage and river water exposed to variable amounts of fecal pollution. To achieve this objective, first we investigated the abundance and correlations between ARGs and sewage-associated marker genes in untreated sewage. Subsequently we characterized the occurrence and abundance of *E. coli*, ARGs, and sewage-associated marker genes at a range of peri-urban to urban sites along the Brisbane River system, which were exposed to different amounts of fecal pollution. Correlations between *E. coli*, ARGs, and sewage-associated marker genes in a river system were analyzed to identify if any sewage indicator could be used to imply the presence of ARGs in river water. Finally, we compared the occurrence and abundance of *E. coli*, ARGs, and sewage-associated marker genes in water samples collected during baseflow and under stormflow conditions to better understand the dynamics between them. This study is the first of its kind to elucidate the relation between sewage-associated marker genes and ARGs in an Australian urban river system.

MATERIALS AND METHODS

Untreated Sewage Sampling

Untreated sewage samples were collected from a municipal STP in Southeast Queensland (SEQ), Australia. The STP treats sewage from approximately 250,000 people and a hospital. The treatment process consists of primary treatment, a secondary treatment (activated sludge), and disinfection with chlorine and UV. Approximately 100 mL of untreated sewage (grab) samples were collected in sterile bottles from the influent of the STP. Weekly samples were collected in triplicate over a period of 6 weeks yielding 18 samples in total for monitoring of ARGs, and sewage-associated marker genes. Samples were transported on ice to the laboratory and stored at 4°C for up to 24 h before analysis.

Environmental Water Sampling

Water samples (i.e., single grab) were collected from Brisbane River and associated creeks (Oxley Creek and Boggy Creek) in Brisbane, Australia (**Figure 1**). Oxley Creek is a tributary of Brisbane River and is tidally influenced. Boggy Creek is a tributary of lower Brisbane River and drains into the mouth of Brisbane River. Twelve sampling sites (BR1-BR12) were chosen along the entire length (i.e., 344 km) of the river. In addition, one sampling site (OX1) was chosen in Oxley Creek and one sampling site (BC1) was chosen in Boggy Creek. Sampling sites BR1-BR5 are sparsely populated with forested hills and grazing land. Based on land-uses and population, these sites are considered peri-urban sites. The middle and lower catchments (sites BR6-BR12, OX1, and BC1) are highly populated and characterized by industrial, residential, and urban areas. From each site, samples were collected on four separate occasions between 15 and 26th May [15th May (Event 1), 19th May (Event 2), and 26th May (Event 3)], 2019 and 13th February (Event 4) 2020. Samples collected between 15 and 26th May represented baseflow samples when the study area did not receive any precipitation 10 days before the respective water sampling. Samples collected on the 13th of February 2020 represented a storm weather event when the study area received 78.8 mm rainfall during the 3 days prior to sampling. The study area did not receive any additional rainfall to allow more sampling to be undertaken during stormflow conditions. Baseflow samples were collected during low tides, while stormflow samples were collected during a spring tide. A total of 56 water samples were collected for monitoring of *E. coli*, ARGs, and sewage-associated marker genes. For river water samples, *E. coli* was included as a general FIB to indicate fecal pollution from both humans and animals. Water samples were collected in 500 mL sterilized PET bottles at approximately 30 cm below the water surface and transported on ice to the laboratory and analyzed within 16 h.

Concentration of River Water Samples

For qPCR analysis of *E. coli*, ARGs, and sewage-associated marker genes, 500 mL of each water sample was filtered through a 90-mm, 0.45-μm pore size HAWP membrane (Millipore, Tokyo,

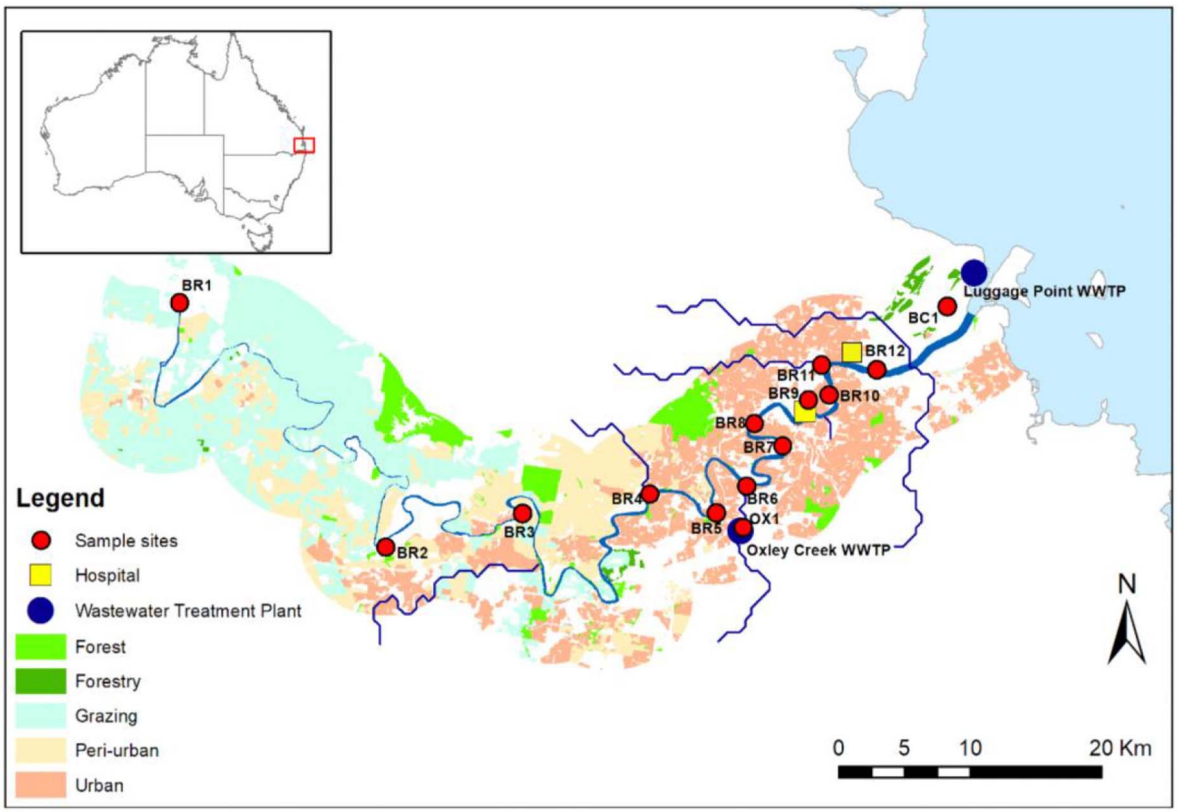


FIGURE 1 | Map of the study sites along the Brisbane River system and its tributaries Oxley Creek, and Boggy Creek, located in Brisbane, Australia. Site OX1 is located downstream of a WWTP, site BR8 is located in proximity to a storm water drain, and sites BR9, BR10, BR11, and BR12 are located downstream of hospitals.

Japan). Before filtration, the pH in all water samples was adjusted to 3.5, using 2.0 N HCl to capture both bacteria and viruses simultaneously (Ahmed et al., 2015).

DNA Extraction

DNA was extracted from an aliquot of 250 μL of untreated sewage sample using the MO Bio PowerSoil DNA isolation kit (Mo Bio Laboratories, Carlsbad, CA, United States) with minor modifications as described elsewhere (Ahmed et al., 2015). For the river water samples, a DNeasy PowerWater Kit (Qiagen, Valencia, CA, United States) was used to extract DNA directly from the membrane. DNA concentrations were measured with a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, Wilmington, DE, United States). All DNA samples were stored at −80°C until further analysis.

PCR Inhibition

An experiment was conducted to determine the presence of PCR inhibitors in DNA samples from untreated sewage and water samples collected from the Brisbane River system using a Sketa22 qPCR assay (Haugland et al., 2005). DNA samples with a 2-quantification cycle (Cq) delay were considered to have potential PCR inhibitors (Ahmed et al., 2019a). Samples with PCR inhibitors were subjected to a 10-fold dilution with TE buffer and reanalyzed with the Sketa22 assay. PCR-uninhibited and 10-fold diluted (inhibition relieved) samples were used for qPCR analysis.

qPCR Assays

Previously published qPCR assays were used for the analysis of *E. coli* 23S rRNA gene (Chern et al., 2011), *aacA* (Szczepanowski et al., 2009), $bla_{\text{ctx-m-32}}$ (Szczepanowski et al., 2009), bla_{KPC} (Hindiyeh et al., 2008), bla_{VIM} (Szczepanowski et al., 2009), *ermF* (Ma et al., 2011), *intl1* (Goldstein et al., 2001), *sul1* (Heuer and Smalla, 2007), *sul2* (Szczepanowski et al., 2009), *tet(M)* (Ng et al., 2001), and *vanA* (Bell et al., 1998), and sewage-associated marker genes HF183 (Green et al., 2014), crAssphage CPQ_056 (Stachler et al., 2017), and Lachno3 (Feng et al., 2018). For this study, we chose *E. coli* as it has been recommended that *E. coli* is suitable for freshwater monitoring by the National Health and Medical Research Council (NHMRC, 2008). For ARGs, we chose aminoglycosides, beta-lactams, sulfonamides, macrolides, and tetracycline resistance genes because of their widespread usage in animal farming worldwide. More than half of the antibiotics prescribed to humans are beta-lactams (Lachmayr et al., 2009; Al Salah et al., 2019). We chose *aacA*, $bla_{\text{ctx-m-32}}$, bla_{KPC}, bla_{VIM}, *ermF*, *sul1*, *sul2*, *tet(M)*, *vanA*, and the *intl1* because these were suggested as

possible indicators to assess the antibiotic resistance status in environmental settings and some of them are highly prevalent in contaminated environments (Berendonk et al., 2015; Gillings, 2018). We also chose three sewage-associated marker genes namely HF183, Lachno3, and crAssphage to provide evidence of sewage contamination in the studied river system. These markers were selected based on their high host-specificity and sensitivity in Brisbane, Australia as determined in our previous studies (Hughes et al., 2017: Ahmed et al., 2019b). The primers and probes for each assay are shown in **Supplementary Table 1** along with qPCR cycling parameters. All qPCR amplifications were performed in 20 μL reaction mixtures using SsoAdvanced Universal Probes Supermix or SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Richmond, CA, United States).

For HF183, Lachno3, and CPQ_056 assays, qPCR mixtures contained 10 μL of SsoAdvanced Universal Probes Supermix, 1000 nM forward primer, 1000 nM reverse primer, and 100 nM of probe (for the HF183, Lachno3, and CPQ_056). For ARGs, qPCR mixtures contained 10 μL of SsoAdvanced Universal SYBR Green Supermix, 300 nM of the forward primer, and 300 nM of the reverse primer. To separate the specific product from non-specific products, including primer dimers, a melting curve analysis was performed for each qPCR run. During the melt curve analysis, the temperature was increased from 65 to 95°C at 0.5°C increment. Samples were considered positive when the melting points were matched with the melting point of the standard curve amplification within a tolerance of 0.5°C (Nutz et al., 2011).

The qPCR assays were performed using a Bio-Rad CFX96 thermal cycler. All qPCR reactions were performed in triplicate. gBlocks gene fragments were used to prepare qPCR standards, ranging from 10^6 to 1 gene copies (GC)/μL of DNA (Integrated DNA Technology, Coralville, IA, United States). For each qPCR run, a series of standards (3×10^6 to 3 GC/reaction), and no template controls (n = 3) were included. A standard curve was generated for each assay and instrument run.

Quantitative polymerase chain reaction performance characteristics such as amplification efficiencies (E), correlation coefficient (r^2), slopes and Y-intercepts were determined from the standard curves for each assay and were within the prescribed limits (Bustin et al., 2009). The assay limit of detection (ALOD) and quantification (ALOQ) for the different assays used were defined as the minimum copy number detected (e.g., lowest copy number detected 95% of the time) and quantifiable (e.g., lowest copy number detected 100% of the time), respectively, as previously described (Verbyla et al., 2016). The abundances in the original sample were back-calculated to take into account each step of the methods. Any sample with triplicate Cq measurements below the ALOQ was considered detected but not quantifiable.

Quality Control

A reagent blank was included for each batch of DNA samples to ensure no carryover contamination occurred from DNA extraction reagents. No carryover contamination was observed in extracted DNA samples. To minimize qPCR contamination, DNA extraction and qPCR setup were performed in separate laboratories.

Data Analysis

All descriptive and multivariate statistical analyses were executed in R version 4.0.2 using the NADA and Vegan packages (R Core Team, 2013; Oksanen et al., 2016; Lee, 2017). All data were interval censored, differentiating between non-censored (i.e., quantifiable) measurements, and left-censored (i.e., concentrations less than the ALOD, and those positive, but not quantifiable), following the recommendations for left-censored data (Helsel, 2011). To execute multivariate statistical analyses that accommodate left-censored data, the u-Score ranks were calculated from the $\log_{10}$ transformed data (https://practicalstats.teachable.com/ on June 20, 2020) (Helsel, 2011, 2019). Then Euclidean distance matrices were calculated from the u-Score ranks (Helsel, 2011) prior to analyses.

Descriptive Statistics

With respect to the sewage samples, the mean concentration of each microbial target was calculated from the triplicate samples collected during each sampling event (n = 6) and used in subsequent analyses. The mean, standard deviation, and frequency of detection were determined for each microbial target in sewage. For all river water samples, the median, minimum, and maximum were reported. For microbial targets with censored observations (up to 80% censored), the median was estimated using Robust Regression on Order Statistics (rROS) (lognormal distribution assumed for modeling the left-censored portion of the distribution). If more than 80% of observations were censored, then only the minimum and/or maximum values measured were reported taking into account the process limit of detection (PLOD) and process limit of quantification (PLOQ). PLOD and PLOQ were calculated from ALOD and ALOQ values.

Multivariate Correlation Analyses

To understand correlations between sewage-associated marker genes and ARGs in sewage at the community-composition-level, the Mantel test (multivariate approach) was executed using the Kendall method and 9,999 permutations as recommended (n = 6 per microbial target; Legendre and Legendre, 2012). Subsequently, the best subset of sewage-associated markers that maximized the correlations with the ARGs was identified using the BIO-ENV procedure [iterative Mantel tests with the Kendall method designed to identify the best subset of a group of variables, such that they maximize (rank) correlations with a dissimilarity matrix of other variables; Clarke and Ainsworth (1993)]. Similarly, the Mantel test and BIO-ENV procedure were also used to identify correlations between *E. coli*, sewage-associated marker genes, and ARGs at the community-composition-level in the river water (n = 56 per microbial target).

Multivariate Hypothesis Testing

Analysis of similarities (ANOSIM) with 9,999 permutations was executed to test whether there was a significant difference in

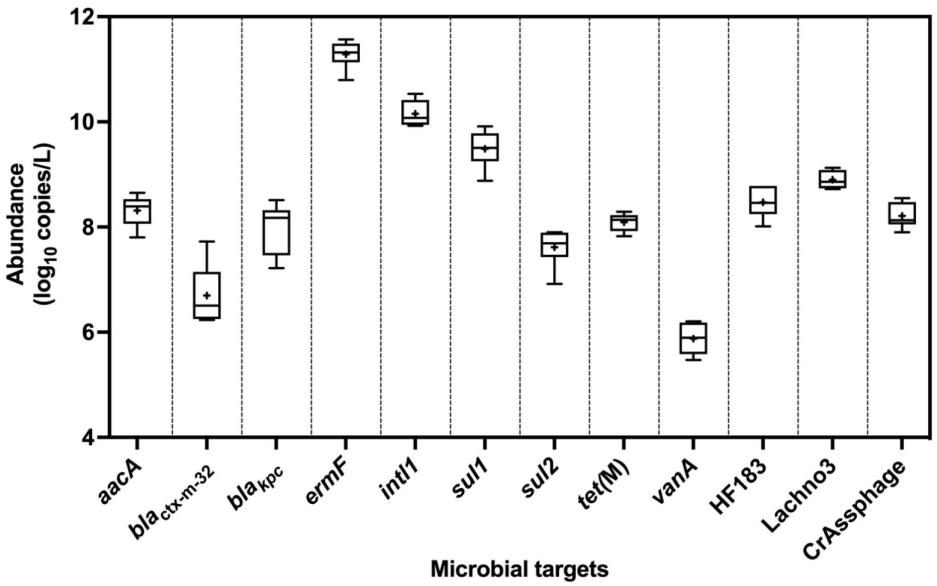


FIGURE 2 | Abundance (log_{10} copies/L) of ARGs (*aacA*, $bla_{ctx-m-32}$, bla_{KPC}, *ermF*, *sul1*, *sul2*, *tet(M)*, and *vanA*), *intI1*, and sewage-associated marker genes (HF183, crAssphage CPQ_056, and Lachno3) in untreated sewage samples, which were always detected in quantifiable abundances. ARG bla_{VIM} was positively detected, but not in quantifiable abundances (between 2.42- and 2.59-log_{10} copies/L), in only two of the six untreated sewage samples. +denotes mean while the outer box lines represent 25th and 75th percentiles and the whiskers extended to the range, and lines inside the boxes represent median values.

E. coli, sewage-associated markers, and ARG abundances in the river between samples collected from different site types [urban (n = 9 sites) vs. peri-urban (n = 5 sites)] and flow conditions [baseflow (n = 42 samples) vs. stormflow (n = 14 samples)], separately, and combined [urban baseflow (n = 27 samples), urban stormflow (n = 9 samples), peri-urban baseflow (n = 15 samples), peri-urban-stormflow (n = 5 samples)] (Clarke, 1993). The Similarity Percentages (SIMPER) procedure was executed with 9,999 permutations to determine which microbial targets differed between groupings of site type and stormflow. Non-metric multidimensional scaling (NMDS) was also used to visualize differences between site types and flow conditions. A p-value of <0.05 was considered significant for all multivariate analyses.

RESULTS

Prevalence and Abundance (log_{10} Copies/L) of ARGs and Sewage-Associated Marker Genes in Untreated Sewage Samples

All ARGs and sewage-associated marker genes were detected in samples collected from six out of the six sampling events except bla_{VIM}, which was only detected twice out of the six sampling events but in abundances less than the ALOQ (**Figure 2**). Among the ARGs tested, the mean abundances of *ermF* (11 ± 0.27 log_{10} copies/L) and *sul1* (9.50 ± 0.25 log_{10} copies/L) were ∼1.10 to 5.50 log_{10} copies greater than *aacA*, bla_{KPC}, $bla_{ctx-m-32}$, *sul2*, *tet(M)*, and *vanA*. The mean abundance of *intI1* was 10 ± 0.25 log_{10} copies/L of sewage. Among the three sewage-associated marker genes, the mean abundance (8.90 ± 0.17 log_{10} copies/L) of the Lachno3 was ∼0.40 to 0.70 log_{10} greater than HF183 (8.50 ± 0.30 log_{10} copies/L) and crAssphage (8.20 ± 0.24 log_{10} copies/L).

Occurrence and Abundance of *E. coli*, ARGs, and Sewage-Associated Marker Genes in Water Samples From the Brisbane River System

A total of 56 river water samples were collected from sampling 14 sites on four occasions, in which three occurred during baseflow conditions and one occurred during stormflow conditions. All water samples collected during the baseflow and stormflow from peri-urban and urban sites were positive for *E. coli*. The ARGs $bla_{ctx-m-32}$, *sul2*, and *vanA* were not detected in any water sample collected during baseflow from these peri-urban and urban sites. *intI1* and ARG *sul1* were most frequently detected in both baseflow and stormflow, however, the frequency of detection was greater in peri-urban sites compared to urban sites. The occurrence of *aacA*, bla_{VIM}, *ermF*, *sul1*, *sul2*, *tet(M)*, and *intI1* and all three sewage-associated marker genes was greater in stormflow than baseflow samples for both peri-urban and urban sites.

Among the 14 sites [five peri-urban (BR1-BR5) and nine urban sites (BR6-BR12, OX1, and BC1)] sampled in baseflow at Event 1, water samples from upstream peri-urban sites (i.e., BR1, BR2, and BR3) and sites located near the STP or stormwater drains (OX1, BR8, and BR12) were positive for one to three ARGs and *intI1* but negative (i.e., <ALOD) for sewage-associated marker genes (**Supplementary Table 2**). Similar patterns were

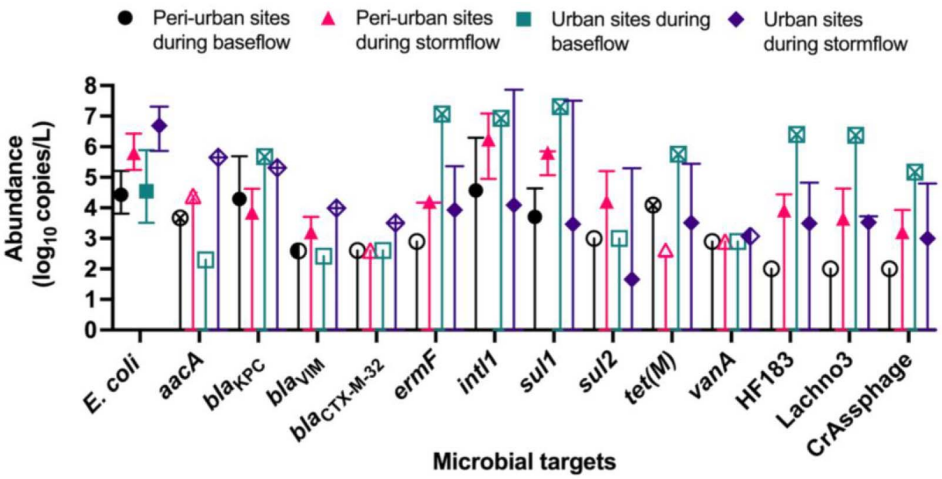


FIGURE 3 | Median abundance (log_{10} copies/L) with maximum and minimum abundances (error bars) reported for *E. coli* (PLOD = 2.00 log_{10} copies/L), *aacA* (PLOD = 2.30 log_{10} copies/L), $bla_{ctx-m-32}$ (PLOD = 2.60 log_{10} copies/L), bla_{KPC} (PLOD = 2.30 log_{10} copies/L), bla_{VIM} (PLOD = 2.42 log_{10} copies/L), *ermF* (PLOD = 2.90 log_{10} copies/L), *intI1* (PLOD = 2.30 log_{10} copies/L), *sul1* (PLOD = 2.00 log_{10} copies/L), *sul2* (PLOD = 3.00 log_{10} copies/L), *tet*(M) (PLOD = 2.60 log_{10} copies/L), and *vanA* (PLOD = 2.90 log_{10} copies/L), and sewage-associated marker genes (HF183, crAssphage CPQ_056, and Lachno3; all PLOD = 2.00 log_{10} copies/L) in water samples collected from peri-urban and urban sites from the Brisbane River system during the base and stormflow (solid-filled shape). If a particular microbial target had left-censored values, the uncertainty of the minimum abundance was depicted with a straight, vertical line. If it was not possible to calculate a median abundance (>80% left-censored), then the maximum abundance measured was depicted with an "X" inside the site/flow shape. When a microbial target was only detected below the limit of quantification, then a half-filled shape depicted the limit of quantification. Finally, an un-filled shape represents 100% left-censored values, and the process limit of detection is depicted.

also observed for the samples collected during the Event 2, however, the sample from site BR8 was positive for four ARGs and three sewage-associated marker genes. Sewage-associated marker genes could not be detected in all water samples collected during baseflow at Event 3; however, water samples from upstream sites (i.e., BR1, BR2, and BR3) and one urban site (i.e., OX1) located downstream of the STP were positive for two to four ARGs. Most of the samples collected during stormflow at Event 4 were positive for multiple ARGs and sewage-associated marker genes except for a few urban sites (i.e., BR7, BR9, BR10, and BR11). Notably, the sample collected from site BC1 was heavily polluted (i.e., positive for ten ARGs and three sewage-associated marker genes).

The abundances of *E. coli*, ARGs, and sewage-associated marker genes are shown in **Figure 3**. The levels of *E. coli* in water samples collected during the baseflow from peri-urban sites ranged from 3.80- to 5.20-log_{10} copies/L with a median value of 4.43-log_{10} copies/L. The levels of *E. coli* in urban sites baseflow samples were slightly greater than in peri-urban sites. We observed increased levels (i.e., ∼2–3 orders of magnitude greater) of *E. coli* in stormflow samples collected from both urban (range = 5.86- to 7.30-log_{10} copies/L with a median value of 6.60-log_{10} copies/L) and peri-urban (range = 5.24- to 6.42-log_{10} copies/L with a median value of 5.79-log_{10} copies/L) sites compared to baseflow samples.

The abundance of bla_{Kpc} in baseflow water samples ranged from 2.47- to 5.69-log_{10} copies/L with a median value of 4.29-log_{10} copies/L for peri-urban sites; however, bla_{Kpc} was rarely detected in urban sites and when detected, abundance ranged from 2.47- to 5.69-log_{10} copies/L. Other ARGs such as *aacA* (range 2.47- to 5.65-log_{10} copies/L), bla_{VIM}, (range 2.59- to 3.99-log_{10} copies/L) $bla_{CTX-M-32}$ (2.77- to 3.50-log_{10} copies/L), and *ermF* (range 2.90- to 7.07-log_{10} copies/L) were sporadically detected in baseflow and stormflow samples in both urban and peri-urban sites. The abundance of *intI1* (range 4.95- to 7.86-log_{10} copies/L) and *sul1* (5.85 ± 7.50-log_{10} copies/L) were generally greater in stormflow samples collected from both urban and peri-urban sites compared to the baseflow events (2.30- to 6.93-log_{10} copies/L for *intI1* and 2.00- to 7.31-log_{10} copies/L for *sul1*). The abundances of *sul2* in water samples collected from urban and peri-urban sites during the stormflow ranged from 3.00- to 5.30-log_{10} copies/L, while *sul2* could not be detected (i.e., <PLOD) in any water samples collected during the baseflow conditions. Among the three sewage associated marker genes, the abundance of Lachno3, HF183, and crAssphage in stormflow samples ranged from 2.17- to 4. 69-, 2.17- to 4. 83-, and 2.17- to 4.80-log_{10} copies/L, respectively, and were detected in both urban and peri-urban. In contrast, these marker genes were not detected in baseflow water samples collected from peri-urban sites but were present in an urban site (BR8; Event 2).

Correlations at the Community-Composition-Level Between Microbial Targets in Untreated Sewage and Water Samples From the Brisbane River System

Significant correlations at the community-composition-level were identified between sewage-associated markers and ARGs in untreated sewage samples (n = 6) using the Mantel test

(τ = 0.488, p = 0.0125). Among all of the sewage-associated markers, the BIO-ENV procedure identified that HF183 alone best maximized correlations to ARGs in untreated sewage samples (τ = 0.5898). Significant correlations at the community-composition-level were also identified between ARGs and *E. coli* (τ = 0.0893, p = 0.0032), as well as with sewage-associated markers in the 56 river water samples (τ = 0.3229, p = 0.0001), separately. Of *E. coli* and the sewage-associated markers combined, the BIO-ENV procedure identified that crAssphage alone maximized correlations with patterns of ARGs in the river water samples (τ = 0.4148).

Differences in *E. coli*, Sewage-Associated Markers, and ARGs Observed by Flow Conditions and Site Types in Water Samples From the Brisbane River System

Significant differences in *E. coli*, ARGs, *intl1*, and sewage-associated markers occurred under different flow conditions [baseflow (n = 42 samples) vs. stormflow (n = 14 samples); R = 0.5340, p = 0.0001], while non-significant weak differences were identified by site types [urban (n = 36 samples) vs. peri-urban (n = 20 samples); R = 0.09823, p = 0.0529]. Subsequently, differences in *E. coli*, ARGs, and sewage-associated markers in river water samples were tested using a multi-factorial ANOSIM analysis and significant differences were identified by stormflow and site type combined [urban baseflow (n = 27 samples), urban stormflow (n = 9 samples), peri-urban baseflow (n = 15 samples), peri-urban-stormflow (n = 5 samples); R = 0.3668, p = 0.0001], which were also visualized in the NMDS plot (**Figure 4**). SIMPER analysis revealed that average percent dissimilarities were greatest between urban-baseflow vs. peri-urban-stormflow (20.8%), followed by peri-urban-stormflow vs. baseflow (19.5%), urban-stormflow and peri-urban baseflow (19.3%), urban- vs. peri-urban during stormflow (18.9%), urban-baseflow vs. -stormflow (17.6%), and urban vs. peri-urban during baseflow (11.2%) (**Table 1**). Depending upon site type-flow condition groupings, four to eight microbial targets were identified as the most influential in contributing to the differences observed between multi-factorial groupings. Only bla_{KPC} and *intl1* significantly differed when comparing peri-urban and urban sites during baseflow conditions. The microbial targets that most frequently were identified as most influential with respect to the differences observed by site type and flow condition included *E. coli*, two sewage-associated markers (HF183 and crAssphage), as well as *sul1* and *intl1*.

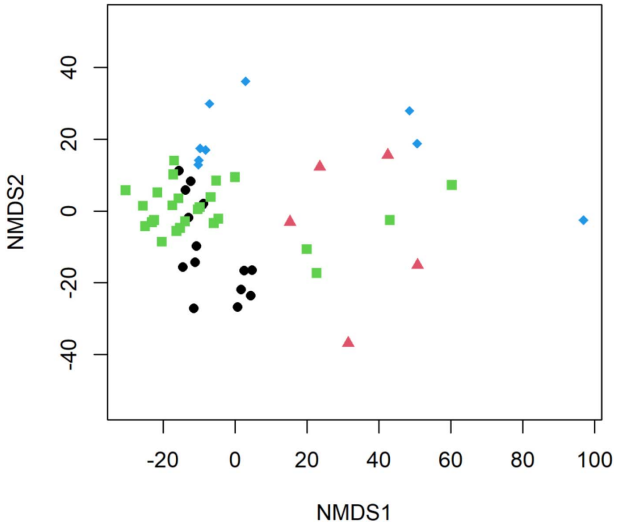


FIGURE 4 | Non-metric multidimensional scaling (NMDS) plot of *E. coli*, ARGs, and sewage-associated markers, measured in water samples (n = 56) collected from peri-urban sites during baseflow (black circle) and stormflow (red triangle), as well as and urban sites during baseflow (green square) and stormflow (blue diamond) along the Brisbane River system, Brisbane, Australia.

DISCUSSION

Antibiotic resistance genes are considered emerging contaminants (Pruden et al., 2006). Additionally, they are considered causes of a potential hazard whether contained within an ARB or as naked DNA, due to the fact that genetic material can be transferred between bacteria through mechanisms of HGT (Thomas and Nielsen, 2005). Although a previous study reported the presence of class 1 integron integrase genes in *E. coli* isolates in water samples collected from the Brisbane River system (Sidhu et al., 2017), little is known regarding the occurrence of ARGs and sewage-associated marker genes in baseflow and stormflow water samples in Brisbane, Australia. We determined the abundances of several ARGs, *intl1*, and sewage-associated marker genes in untreated sewage samples and then determined their abundances in baseflow and stormflow samples collected from the Brisbane River system characterized by peri-urban and urban areas.

The abundance of *ermF* was the greatest in sewage samples followed by *intl1* and *sul1*. These genes are also reported to be highly prevalent in sewage in Europe (Cacace et al., 2019; Osińska et al., 2020; Pärnänen et al., 2019). Schmitz et al. (2019) reported the co-occurrence and high correlation between *intl1* and *sul1* throughout sewage treatment processes in the United States. The high abundances of *intl1* in untreated sewage samples are not surprising as human feces often carry up to 10^{11} copies of *intl1*/g of feces (Gillings, 2018).

Sewage samples were also positive for the carbapenemase encoding gene bla_{VIM}, which is an emerging resistance in Australian human infections and has been implicated in a small number of outbreaks in Queensland (Australian Commission on Safety and Quality in Health Care (ACSQHC), 2019). Human infections associated with the ARG bla_{KPC} in Queensland are considered to be rare but the abundance was greater in sewage samples compared to bla_{VIM}. The prevalence of emerging ARGs, such as bla_{KPC}, bla_{VIM}, and $bla_{ctx-m-32}$, provides baseline data but are not yet prevalent in human infection monitoring data from Australia. It is likely that the abundance of these genes will

TABLE 1 | The average percent dissimilarities and the corresponding percent contributions of the most influential (explaining >70% of the differences) microbial targets (*E. coli*, sewage-associated markers, and ARGs) observed between the multi-factorial groupings (flow condition and site type) for river water samples, as identified by the Similarity Percentages (SIMPER) procedure.

Multifactorial Grouping	Urban Sites in Baseflow conditions		Urban Sites in Stormflow conditions		Peri-urban Sites in Baseflow conditions	
	Average dissimilarity (%)	Contributions of the most influential microbial targets (%)	Average dissimilarity (%)	Contributions of the most influential microbial targets (%)	Average dissimilarity (%)	Contributions of the most influential microbial targets (%)
Urban Sites in Stormflow conditions	17.6%	*E. coli* (20.5%) *sul1* (10.4%)* *ermF* (9.6%) *intI1* (9.4%)* *tet(M)* (9.0%) HF183 (8.5%) CrAssphage (8.2%)				
Peri-urban Sites in Baseflow conditions	11.2%	*intI1* (19.4%) *sul1* (18.0%)* *E. coli* (16.9%)* bla_{KPC} (16.2%)	19.3%	*E. coli* (19.6%) *intI1* (11%)* *sul1* (10.9%)* bla_{KPC} (8.3%)* *ermF* (8.4%) *tet(M)* (7.7%)* HF183 (7.2%)*		
Peri-urban Sites in Stormflow conditions	20.8%	*sul1* (14.2%) *intI1* (13.6%) *E. coli* (12.4%)* *sul2* (12.3%)* HF183 (8.7%) CrAssphage (8.4%) Lachno3 (7.0%)*	18.9%	*sul1* (14.5%) *intI1* (13.7%) *sul2* (10.0%) HF183 (8.5%)* CrAssphage (8.1%) *ermF* (7.7%)* Lachno3 (7.4%)* *tet(M)* (6.4%)*	19.5%	*E. coli* (14.2%)* *sul2* (13.0%) *sul1* (11.5%)* *intI1* (9.6%)* HF183 (9.5%) CrAssphage (9.0%) bla_{VIM} (7.9%)

For each microbial target identified among the most influential, all percent contributions had p values ($p < 0.05$, unless indicated by an asterisk (*).

be increased in sewage and also their abundance is expected to potentially increase in environmental waters.

The abundances of sewage-associated markers HF183, Lachno3, and crAssphage under stormflow conditions corroborates with previous studies and shows them to be highly sensitive marker genes for tracking sewage pollution in environmental waters (Hughes et al., 2017; Ahmed et al., 2019b; Korajkic et al., 2020). The host-specificity of sewage-associated markers were not determined in this study as in our previous studies, we extensively evaluated the host-specificity of the HF183, Lachno3, and crAssphage marker genes (Hughes et al., 2017; Ahmed et al., 2019b). Abundances of *ermF*, *intl1*, and *sul1* were greater in sewage compared to sewage-associated marker genes suggesting that in the event of sewage pollution in environmental waters, these ARGs are likely to be present. The concentrations of these three genes were greater in sewage compared to HF183 and Lachno3 probably due to the high antibiotic usage and associated gene transfer from bacteria to bacteria. *intl1* data presented in this study corroborate with a previous report that *intl1* may be a useful proxy for anthropogenic pollution in catchment waters (Borruso et al., 2016; Gillings, 2018).

Overall, significant differences in *E. coli*, sewage-associated markers, and ARGs were identified by site type and flow conditions (**Figure 4**). All water samples collected from the Brisbane River system in this study were positive for *E. coli* regardless of flow conditions and site types. We did not see a stark difference between the levels of HF183 and crAssphage in water samples collected between urban and peri-urban sites during the baseflow, suggesting human fecal pollution may not be occurring in many sites in the studied river system during baseflow conditions except an urban site (i.e., BR8) at event 2 (one of three base flow events), which was in close proximity to a storm drain. However, the abundance of *E. coli*, HF183, and crAssphage increased significantly in stormflow samples collected from both urban and peri-urban sites compared to baseflow samples. A previous study reported a greater abundance of *E. coli* and enteric pathogens in water samples collected from Brisbane River and its tributaries in stormflow compared to baseflow (Sidhu et al., 2012). Similarly, in North Queensland, higher abundances of *E. coli* were found near sewage outfalls, with a change in the microbial community noted between baseflow and stormflow water samples (Neave et al., 2014). For *E. coli* monitoring, we used qPCR, however, the guideline values are based on culturable FIB with the most stringent category "A" setting a limit at less than or equal to 40 CFU/100 mL. Therefore, *E. coli* was not comparable to this guideline. Nevertheless, the combined presence of *E. coli* and two sewage-associated markers

indicated an increase in fecal pollution during storm conditions. *E. coli* was also detected in water samples that were negative for sewage-associated marker genes, suggesting that animal fecal pollution is also occurring in the Brisbane River system. The extraintestinal growth of *E. coli* in the Brisbane River system also cannot be ruled out and requires further investigation by characterizing *E. coli* or applying animal fecal markers to determine the sources of *E. coli*.

When comparing the different combinations of site types and flow conditions, different ARGs were identified as the most influential depending upon the site type and flow condition. *intl1* and *sul1* were always identified as contributing most to differences observed regardless of the groupings, followed by *sul2*, *tet(M)*, and *ermF*. The abundance of *intl1* and *sul1* were generally greater in stormflow samples collected from both urban and peri-urban sites, compared to the baseflow samples from the same sites suggesting surface run-off from point and non-point sources contribute ARGs to the Brisbane River system.

While HF183 and crAssphage significantly contributed to the differences observed for each site type when comparing flow conditions, the occurrence and abundance of all three sewage-associated marker genes were greater in stormflow samples compared to baseflow water samples for both urban and peri-urban sites. With differential persistence of bacteria and the potential for HGT, stormflow conditions may result in increased proliferation of ARGs. The presence of sewage-associated marker genes also indicates the potential presence of enteric viruses in the Brisbane River system (Sidhu et al., 2012). These factors collectively may increase health risks to recreational users.

As there are numerous ARGs that can be present in the environment with relevance to public health (Gatica et al., 2016), at sometimes high background levels can be detected even in "pristine" environments, there is not yet a risk-based consensus regarding which ARGs would be "high priority" for monitoring purposes (Hamilton et al., 2020). Therefore, assessing the correlation of ARGs with more commonly monitored factors provides value if surrogate monitoring targets could reasonably be used for assessing the antibiotic resistance impacts of stormwater on receiving water bodies. ARGs were significantly correlated with sewage-associated markers and *E. coli*. CrAssphage alone maximized correlations to the group of ARGs analyzed in river water samples collected from peri-urban and urban areas. Interestingly, HF183 alone best maximized correlations to ARGs in untreated sewage.

Different sewage-associated markers were identified to maximize correlations to ARGs depending on the matrix, thus emphasizing the importance of assessing correlations in the matrix of interest as ARG and sewage-associated marker persistence likely varies. Previous studies noted positive correlations between FIB and antimicrobial resistance in impacted surface waters in Southeast Queensland, indicating their role as a potential reservoir of resistance (Watkinson et al., 2007). Stachler et al. (2019) reported greater abundances of ARGs and crAssphage on wet weather days than on dry weather days. They also demonstrated a strong correlation of various ARGs with culturable *E. coli*, culturable enterococci, and HF183 qPCR assay, suggesting that the bacterial-based markers could be an indicator for the presence of ARGs in environmental waters due to human fecal pollution. Nevertheless, the results presented here demonstrate the application of sewage-associated marker genes and *E. coli* as indicators of probable ARG prevalence under stormflow conditions but requires further validation. The other benefits of the sewage-associated marker genes are that their presence can be used to identify hot spots of sewage pollution or to detect broken pipes. This could be a useful source control option to decrease environmental spread of ARGs.

In this study, sewage-associated marker genes were detected near the storm drains such as site BR8 (Event 2). Interestingly, most of the samples collected during stormflow (Event 4) were positive for multiple ARGs and sewage-associated marker genes, suggesting various point and non-point sources are contributing to these microbial targets. Notably, the sample collected from site BC1 (urban) was positive for ten ARGs and three sewage-associated marker genes. Site BC1 is located downstream of an urban WWTP. The water samples from this site and others in the upstream sites were collected during the high tide. It is highly likely that high tide pushed ARGs and sewage-associated marker genes upstream to site BR12. We also noted increased levels of *ermF*, *intl1*, and *sul1* at sites BR5, OX1, BR12, and BC1 in the stormflow samples, and samples from these sites were also positive for one or more sewage-associated marker genes. The results presented in this study could provide valuable information to the water quality managers for mitigation of emerging contaminants and sewage pollution in urban rivers or similar aquatic environments.

There are several limitations of this study. For example, the grab sampling method used in this study only provided a snapshot of microbial targets in the studied river system. Future studies would also benefit from using flow-weighted composite samples, particularly during stormflow conditions, to characterize *E. coli*, ARGs, and sewage-associated marker abundances. We used a moderate volume (i.e., 500 mL) of water samples for the concentration of microbial targets, however, during the baseflow, processing a large volume of water samples may increase the detection sensitivity. In this study, for river water samples, left-censored data analysis techniques were used to assess differences in microbial target abundance by areas and flow conditions, as well as to characterize the correlation between ARGs, *E. coli*, and sewage-markers because several microbial targets were not always detected, and/or detected in abundances too low to quantify (Helsel, 2011). While baseflow conditions were sampled on three occasions, stormflow conditions (i.e., 57 mm rainfall) were sampled on only one occasion because the study area did not receive any rainfall except for only once during the study period. Since the power of statistical analyses decreases with increased data censoring and unbalanced sampling, future studies are needed to confirm the findings presented here.

Multivariate statistical analyses appropriate for left-censored data were used in this study to understand the correlation between fecal indicator microorganisms and ARGs in untreated sewage and river water, as well as to test hypotheses about land use and flow conditions with respect to ARGs and fecal indicator microorganisms in samples collected from the Brisbane River system. The abundance of ARGs and sewage-associated

markers increased during stormflow conditions. Specifically, HF183 and crAssphage may serve as potential indicators of ARGs of sewage origin that merit further investigation. It can be difficult to identify cause-effect relationships between ARGs and sewage-associated markers in an uncontrolled, monitoring-based study. However, this study demonstrated that sewage-associated marker genes can be used to indicate the presence of the group of sewage-associated ARGs analyzed in this study under stormflow conditions. Data presented in this study may be valuable to water quality managers to manage sewage pollution and ARGs or pathogens in environmental waters. Future work, particularly mesocosm experiments, can build on this study to better understand the relationship between ARGs, HF183, and crAssphage genome.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

WA: investigation, resources, and writing—original draft. PG: sampling and study design. KH: sample analysis and writing—original draft. SJ: sample analysis and writing—original draft. DA: sampling and data collection. ED: writing—original draft. SS: resources and writing—original draft. ES: statistical analysis and writing—original draft. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

ES was supported by the US National Science Foundation grant OCE-1566562.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.632850/full#supplementary-material

REFERENCES

Ahmed, W., Harwood, V. J., Gyawali, P., Sidhu, J. P. S., and Toze, S. (2015). Comparison of concentration methods for quantitative detection of sewage-associated viral markers in environmental waters. *Appl. Environ. Microbiol.* 81, 2042–2049. doi: 10.1128/aem.03851-14

Ahmed, W., Hughes, B., and Harwood, V. J. (2016). Current status of marker genes of *Bacteroides* and related taxa for identifying sewage pollution in environmental waters. *Water* 8:231. doi: 10.3390/w8060231

Ahmed, W., Hamilton, K. A., Lobos, A., Hughes, B., Staley, C., Sadowsky, M. J., et al. (2018). Quantitative microbial risk assessment of microbial source tracking markers in recreational water contaminated with fresh untreated and secondary treated sewage. *Environ. Int.* 117, 243–249. doi: 10.1016/j.envint.2018.05.012

Ahmed, W., Payyappat, S., Cassidy, M., and Besley, C. (2019a). Enhanced insights from human and animal host-associated molecular marker genes in a freshwater lake receiving wet weather overflows. *Sci. Rep.* 9:12503.

Ahmed, W., Gyawali, P., Feng, S., and McLellan, S. L. (2019b). Host specificity and sensitivity of established and novel sewage-associated marker genes in human and nonhuman fecal samples. *Appl. Environ. Microbiol.* 85:e00641-19.

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

Al Salah, D. M. M., Laffite, A., and Pote, J. (2019). Occurrence of bacterial markers and antibiotic resistance genes in Sub-Saharan Rivers receiving animal farm wastewater. *Sci. Rep.* 9:14847.

Australian Commission on Safety and Quality in Health Care (ACSQHC) (2019). *AURA 2019: Third Australian Report on Antimicrobial Use and Resistance in Human Health*. Sydney: Australian Commission on Safety and Quality in Health Care (ACSQHC).

Badgley, B. D., Thomas, F. I., 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

Bell, J. M., Paton, J. C., and Turnidge, J. (1998). Emergence of vancomycin-resistant enterococci in Australia: phenotypic and genotypic characteristics of isolates. *J. Clin. Microbiol.* 36, 2187–2190. doi: 10.1128/jcm.36.8.2187-2190.1998

Berendonk, T. U., Manaia, C. M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., et al. (2015). Tackling antibiotic resistance: the environmental framework. *Nat. Rev. Microbiol.* 13, 310–317.

Borruso, L., Harms, K., Johnsen, P. J., Nielsen, K. M., and Brusetti, L. (2016). Distribution of class 1 integrons in a highly impacted catchment. *Sci. Total Environ.* 566, 1588–1594. doi: 10.1016/j.scitotenv.2016.06.054

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

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.

Cacace, D., Fatta-Kassinos, D., Manaia, C. M., Cytryn, E., Kreuzinger, N., Rizzo, L., et al. (2019). Antibiotic resistance genes in treated wastewater and in the receiving water bodies: A pan-European survey of urban settings. *Water Res.* 162, 320–330. doi: 10.1016/j.watres.2019.06.039

CDC (2019). *Antibiotic Resistance Threats in the United States*. Atlanta, GA: U.S. Department of Health and Human Services.

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. *Lett. Appl. Microbiol.* 52, 298–306. doi: 10.1111/j.1472-765x.2010.03001.x

Clarke, K. R., and Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92, 205–205. doi: 10.3354/meps092205

Clarke, K. R. (1993). Non−parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Cox, P., Griffith, M., Angles, M., Deere, D., and Ferguson, C. (2005). Concentrations of pathogens and indicators in animal feces in the Sydney Watershed. *Appl. Environ. Microbiol.* 71, 5929–5934. doi: 10.1128/aem.71.10.5929-5934.2005

Derx, J., Schijven, J., Sommer, R., Zoufal-Hruza, C. M., van Driezum, I. H., Reischer, G., et al. (2016). QMRAcatch: human-associated fecal pollution and infection risk modeling for a river/floodplain environment. *J. Environ. Qual.* 45, 1205–1214. doi: 10.2134/jeq2015.11.0560

Farkas, K., Adriaenssens, E. M., Walker, D. I., McDonald, J. E., Malham, S. K., and Jones, D. L. (2019). Critical evaluation of crAssphage as a molecular marker

for human-derived wastewater contamination in the aquatic environment. *Food Environ. Virol.* 11, 113–119. doi: 10.1007/s12560-019-09369-1

Feng, S., Bootsma, M., and McLellan, S. L. (2018). Novel human-associated *Lachnospiraceae* genetic markers improve detection of fecal pollution sources in urban waters. *Appl. Environ. Microbiol.* 84:e00309-18.

Ferguson, A. S., Layton, A. C., Mailloux, B. J., Culligan, P. J., Williams, D. E., Smartt, A. E., et al. (2012). Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. *Sci. Total Environ.* 431, 314–322. doi: 10.1016/j.scitotenv.2012.05.060

Gatica, J., Tripathi, V., Green, S., Manaia, C. M., Berendonk, T., Cacace, D., et al. (2016). High throughput analysis of integron gene cassettes in wastewater environments. *Environ. Sci. Technol.* 50, 11825–11836. doi: 10.1021/acs.est.6b03188

Gillings, M. R. (2018). DNA as a pollutant: the clinical class 1 integron. *Curr. Pollut. Rep.* 4, 49–55. doi: 10.1007/s40726-018-0076-x

Goldstein, C., Lee, M. D., Sanchez, S., Hudson, C., Phillips, B., Register, B., et al. (2001). Incidence of class 1 and 2 integrases in clinical and commensal bacteria from livestock, companion animals, and exotics. *Antimicrob. Agents Chemo.* 45, 723–726. doi: 10.1128/aac.45.3.723-726.2001

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

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

Harwood, V. J., Staley, C., Badgely, 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. A., Siefring, S. C., Wymer, L. J., Brenner, K. P., and Dufour, A. P. (2005). Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Res.* 39, 559–568. doi: 10.1016/j.watres.2004.11.011

Hamilton, K. A., Garner, E., Joshi, S., Ahmed, W., Ashbolt, N., Medema, G., et al. (2020). Antimicrobial-resistant microorganisms and their genetic determinants in stormwater: A systematic review. *Curr Opin Environ. Sci. Health* 16, 101–112. doi: 10.1016/j.coesh.2020.02.012

Helsel, D. R. (2011). *Statistics for Censored Environmental Data using Minitab® and R: Second Edition, Statistics for Censored Environmental Data Using Minitab® and R: Second Edition*. Hoboken, NJ: John Wiley and Sons. doi: 10.1002/9781118162729

Helsel, D. (2019). *Statistical Analysis of Censored Data (Data with Nondetects). [WWW Document.]*. Available online at: https://practicalstats.teachable.com/p/nada (accessed December 21, 2019).

Heuer, H., and Smalla, K. (2007). Manure and sulfadiazine synergistically increased bacterial antibiotic resistance in soil over at least two months. *Environ. Microbiol.* 9, 657–666. doi: 10.1111/j.1462-2920.2006.01185.x

Hindiyeh, M., Smollen, G., and Grossman, Z. (2008). Rapid detection of bla_{KPC} carbapenemase genes by real-time PCR. *J. Clin. Microbiol.* 46, 2879–2883.

Hughes, B., Beale, D. J., Dennis, P. G., Cook, S., and Ahmed, W. (2017). Cross-comparison of human wastewater-associated molecular markers in relation to fecal indicator bacteria and enteric viruses in recreational beach waters. *Appl. Environ. Microbiol.* 83:e00028-17.

Karkman, A., Parnanen, K., and Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropology impacted environments. *Nat. Comm.* 10:80.

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

Korajkic, A., McMinn, B., Herrmann, M. P., Sivaganesan, M., Kelty, C. A., Clinton, P., et al. (2020). Viral and bacterial fecal indicators in untreated wastewater across the contiguous United States exhibit geospatial trends. *Appl. Environ. Microbiol.* 86:e02967-19.

Lachmayr, K. L., Kerhof, L. J., Dirienzi, A. G., Cavanaugh, C. M., and Ford, T. E. (2009). Quantifying nonspecific TEM beta-lactamase (bla_{TEM}) genes in a wastewater streams. *Appl. Environ. Microbiol.* 75, 203–211. doi: 10.1128/aem.01254-08

Legendre, P., and Legendre, L. (2012). *Numerical Ecology*, 3rd Edn. Amsterdam: Elsevier.

Lee, L. (2017). *CRAN - Package NADA [WWW Document]*. Available online at: https://cran.r-project.org/web/packages/NADA/index.html (accessed October 13, 2019).

Lee, S., Suits, M., Wituszynski, D., Winston, R., Martin, J., and Lee, J. (2020). Residential urban stormwater runoff: a comprehensive profile of microbiome and antibiotic resistance. *Sci. Total Environ.* 723:138033. doi: 10.1016/j.scitotenv.2020.138033

Ma, Y., Wilson, C. A., Novak, J. T., Riffat, R., Aynur, S., Murthy, S., et al. (2011). Effect of various sludge digestion conditions on sulfonamide, macrolide, and tetracycline resistance genes and class I integrons. *Environ. Sci. Technol.* 45, 7855–7861. doi: 10.1021/es200827t

Moretto, V. T., Bartley, P. S., Ferreira, V. M., Santos, C. S., Silva, L. K., Ponce-Terashima, R. A., et al. (2022). Microbial source tracking and antimicrobial resistance in one river system of a rural community in Bahia, Brazil. *Braz. J. Biol.* 82:e231838.

Neave, M., Luter, H., Padovan, A., Townsend, S., Schobben, X., and Gibb, K. (2014). Multiple approaches to microbial source tracking in tropical northern Australia. *MicrobiologyOpen* 3, 860–874. doi: 10.1002/mbo3.209

Ng, L.-K., Martin, I., Alfa, M., and Mulvey, M. (2001). Multiplex PCR for the detection of tetracycline resistant genes. *Mol. Cell Prob.* 15, 209–215. doi: 10.1006/mcpr.2001.0363

Nguyen, T. N., Kasuga, I., Liu, M., and Katayama, H. (2019). "Occurrence of antibiotic resistance genes as emerging contaminants in watersheds of Tama River and Lake Kasumigaura in Japan," in *IOP Conference Series: Earth and Environmental Science, Volume 266, 4th International Forum on Sustainable Future in Asia/4th NIES International Forum. 23-24 January*, Pan Pacific Hanoi.

NHMRC (2008). *National Health and Medical Research Council*. Available online at: www.nhmrc.gov.au/publications/synopses/eh38.htm

Nutz, S., Doll, K., and Karlovsky, P. (2011). Determination of the LOQ in real-time PCR by receiver operating characteristics curve analysis: application to qPCR assays for *Fusarium verticillioides* and *F. proliferatus*. *Anal. Bioanal. Chem.* 401, 717–726. doi: 10.1007/s00216-011-5089-x

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016). *Vegan: Community Ecology Package. R Package Version 2.3-5*.

Olds, H. T., Corsi, S. R., Dila, D. K., Halmo, K. M., Bootsma, M. J., and McLellan, S. L. (2018). High levels of sewage contamination released from urban areas after storm events: a quantitative survey with sewage specific bacterial indicators. *PLoS Med.* 15:e1002614. doi: 10.1371/journal.pmed.1002614

Osiñska, A., Korzeniewska, E., Harnisz, M., Felis, E., Bajkacz, S., Jachimowicz, P., et al. (2020). Small-scale wastewater treatment plants as a source of the dissemination of antibiotic resistance genes in the aquatic environments. *J. Haz. Mat.* 381:121221. doi: 10.1016/j.jhazmat.2019.121221

Pazda, M., Kumirska, J., Stepnowski, P., and Mulkiewicz, E. (2019). Antibiotic resistance genes identified in wastewater treatment plant systems – a review. *Sci. Total Environ.* 697:134023. doi: 10.1016/j.scitotenv.2019.134023

Pärnänen, K. M. M., Narciso-da-Rocha, C., Kneis, D., Berendonk, T. U., Cacace, D., Do, T. T., et al. (2019). Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Sci. Adv.* 5:eaau9124.

Pruden, A., Pei, R., Storteboom, H., and Carlson, K. H. (2006). Antibiotic resistance genes as emerging contaminants: studies in Northern Colorado. *Environ. Sci. Technol.* 40, 7445–7450. doi: 10.1021/es060413l

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

Reynolds, L. J., Sala-Comorera, L., Martin, N. A., Nolan, T. M., Stephens, J. H., Gitto, A., et al. (2020). Correlation between antimicrobial resistance and faecal contamination in small urban streams and bathing waters. *Sci. Total Environ.* 739:140242. doi: 10.1016/j.scitotenv.2020.140242

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M. C., et al. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Sci. Total Environ.* 447, 345–360. doi: 10.1016/j.scitotenv.2013.01.032

Rochelle-Newall, E., Nguyen, T. M. H., Le, T. P. Q., Sengtaheuanghoung, O., and Ribolzi, O. (2015). A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions. *Front. Microbiol.* 6:308.

Rodriguez-Manzano, J., Miagostovich, M. P., Hundesa, A., Clemente-Casares, P., Carratala, A., Buti, M., et al. (2010). Analysis of the evolution in the circulation of HAV and HEV in Eastern Spain by testing urban sewage samples. *J. Water Health* 8, 346–354. doi: 10.2166/wh.2009.042

Schmitz, B. W., Innes, G. K., Xue, J., Gerba, C. P., Pepper, I. L., and Sherchan, S. (2019). Reduction of erythromycine resistance gene erm(F) and class 1 integron-integrase genes in wastewater by Bardenpho treatment. *Sci. Adv.* 5:eaau9124.

Sidhu, J. P. S., Hodgers, L., Ahmed, W., Chong, M. N., and Toze, S. (2012). Prevalence of human pathogens and indicators in stormwater runoff in Brisbane, Australia. *Water Res.* 46, 6652–6660. doi: 10.1016/j.watres.2012.03.012

Sidhu, J. P. S., Jagals, P., Smith, A., and Toze, S. (2017). Comparative prevalence of *Escherichia coli* carrying virulence genes and class 1 and 2 integrons in sub-tropical and cool temperate freshwater. *Environ. Sci. Pollut. Res.* 24, 18263–18272. doi: 10.1007/s11356-017-9497-0

Signor, R. S., Roser, D. J., Ashbolt, N. J., and Ball, J. E. (2005). Quantifying the impact of runoff events on microbiological contaminant concentrations entering surface drinking source waters. *J. Water Health* 3, 453–468. doi: 10.2166/wh.2005.052

Soller, J., Bartrand, T., Ravenscroft, J., Molina, M., Whelan, G., Schoen, M., et al. (2015). Estimated human health risks from recreational exposures to stormwater runoff containing animal faecal material. *Environ. Model. Soft.* 72, 21–32. doi: 10.1016/j.envsoft.2015.05.018

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Stachler, E., Crank, K., and Bibby, K. (2019). Co-occurrence of crAssphage with antibiotic resistance genes in an impacted urban watershed. *Environ. Sci. Technol. 6*, 216–221. doi: 10.1021/acs.estlett.9b00130

Stoeckel, D. M., and Harwood, V. J. (2008). Performance, design and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* 73, 2405–2415. doi: 10.1128/aem.02473-06

Stange, C., Yin, D., Xu, T., Guo, X., Schäfer, C., and Tiehm, A. (2019). Distribution of clinically relevant antibiotic resistance genes in Lake Tai, China. *Sci. Total Environ.* 655, 337–346. doi: 10.1016/j.scitotenv.2018.11.211

Stange, C., and Tiehm, A. (2020). Occurrence of antibiotic resistance genes and microbial source tracking markers in the water of a karst spring in Germany. *Sci. Total Environ.* 742:140529. doi: 10.1016/j.scitotenv.2020.140529

Szczepanowski, R., Linke, B., and Krahn, I. (2009). Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology* 155(Pt 7), 2306–2319. doi: 10.1099/mic.0.028233-0

The Review on Antimicrobial Resistance (2014). *Tackling a Crisis for the Health and Wealth of Nations*, London.

Thomas, C., and Nielsen, K. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234

Verbyla, M. E., Symonds, E. M., Kafle, R. C., Cairns, M. R., Iriarte, M., Mercado Guzmán, A., et al. (2016). Managing microbial risks from indirect wastewater reuse for irrigation in urbanizing watersheds. *Environ. Sci. Technol.* 50, 6803–6813. doi: 10.1021/acs.est.5b05398

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

Watkinson, A. J., Micalizzi, G. R., Bates, J. R., and Costanzo, S. D. (2007). Novel method for rapid assessment of antibiotic resistance in *Escherichia coli* isolates from environmental waters by use of a modified chromogenic agar. *Appl. Environ. Microbiol.* 73, 2224222–2224229.

WHO (2014). *Antimicrobial Resistance: Global Report on Surveillance*. Geneva: World Health Organization.

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.

Copyright © 2021 Ahmed, Gyawali, Hamilton, Joshi, Aster, Donner, Simpson and Symonds. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fmicb.2021.660368

Synergistic Application of Molecular Markers and Community-Based Microbial Source Tracking Methods for Identification of Fecal Pollution in River Water During Dry and Wet Seasons

Hongxia Liang[1], Zhisheng Yu[1,2], Bobo Wang[1], Fabrice Ndayisenga[1], Ruyin Liu[1], Hongxun Zhang[1] and Gang Wu[3,4]*

[1] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, [2] RCEES-IMCAS-UCAS Joint-Lab of Microbial Technology for Environmental Science, Beijing, China, [3] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, University of Chinese Academy of Sciences, Beijing, China, [4] University of Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:
Anicet Ramon Blanch, University of Barcelona, Spain

Reviewed by:
Jill Stewart, University of North Carolina at Chapel Hill, United States
Amity Zimmer-Faust, Southern California Coastal Water Research Project, United States

***Correspondence:**
Zhisheng Yu
yuzs@ucas.ac.cn

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *29 January 2021*
Accepted: *12 May 2021*
Published: *14 June 2021*

Citation:
Liang HX, Yu ZS, Wang BB, Ndayisenga F, Liu RY, Zhang HX and Wu G (2021) Synergistic Application of Molecular Markers and Community-Based Microbial Source Tracking Methods for Identification of Fecal Pollution in River Water During Dry and Wet Seasons. Front. Microbiol. 12:660368. doi: 10.3389/fmicb.2021.660368

It is important to track fecal sources from humans and animals that negatively influence the water quality of rural rivers and human health. In this study, microbial source tracking (MST) methods using molecular markers and the community-based FEAST (fast expectation–maximization microbial source tracking) program were synergistically applied to distinguish the fecal contributions of multiple sources in a rural river located in Beijing, China. The performance of eight markers were evaluated using 133 fecal samples based on real-time quantitative (qPCR) technique. Among them, six markers, including universal (BacUni), human-associated (HF183-1 and BacH), swine-associated (Pig-2-Bac), ruminant-associated (Rum-2-Bac), and avian-associated (AV4143) markers, performed well in the study. A total of 96 water samples from the river and outfalls showed a coordinated composition of fecal pollution, which revealed that outfall water might be a potential input of the Fsq River. In the FEAST program, bacterial 16S rRNA genes of 58 fecal and 12 water samples were sequenced to build the "source" library and "sink," respectively. The relative contribution (<4.01% of sequence reads) of each source (i.e., human, swine, bovine, or sheep) was calculated based on simultaneous screening of the operational taxonomic units (OTUs) of sources and sinks, which indicated that community-based MST methods could be promising tools for identifying fecal sources from a more comprehensive perspective. Results of the qPCR assays indicated that fecal contamination from human was dominant during dry weather and that fecal sources from swine and ruminant were more prevalent in samples during the wet season than in those during the dry season, which were consistent with the findings predicted by the FEAST program using a very small sample size. Information from the study could be valuable for the development of improved regulation policies to reduce the levels of fecal contamination in rural rivers.

Keywords: microbial source tracking, fecal pollution, molecular markers, FEAST program, real-time quantitative PCR, high-throughput sequencing, dry and wet seasons

INTRODUCTION

Fecal contamination of surface water has been recognized as one of the leading causes of the decline of water quality worldwide (Blanch et al., 2006; Reischer et al., 2013). Human and animal feces could be inputted directly or indirectly into freshwater by multiple-point or nonpoint source pathways such as wastewater discharge, uncontrolled discard of feces from humans and animals, and rainwater runoff (Converse et al., 2012; De Kwaadsteniet et al., 2013; Dubinsky et al., 2016; Ahmed et al., 2017). In particular, rainwater runoff was considered one of the major sources of fecal matter transported into the freshwater environment, especially during the wet season in rural regions (Chidamba and Korsten, 2015; Kostyla et al., 2015; Ahmed et al., 2016). Microbial contaminations from human and animal fecal matter are a leading cause of the deterioration of water quality, raising public health concerns (Lee et al., 2013; Reischer et al., 2013). It is important to identify the sources of microbial contamination for the development of regulatory policies to protect water quality and avoid threats to human health caused by potential pathogenic bacteria from feces.

Microbial source tracking (MST) methods can distinguish fecal matter from different host species (Kildare et al., 2007; Staley et al., 2018; Balleste et al., 2020). The initially developed library-dependent MST methods relied on microbial reference libraries such as the routine collecting, monitoring, and typing of many isolates; therefore, it was time-consuming and costly to identify the potential fecal sources (Moore et al., 2005; Stoeckel and Harwood, 2007). Conversely, library-independent MST methods, targeting specific gene fragments (e.g., 16S rRNA gene) or taxonomic groups of the fecal sources associated with specific hosts, are considered accurate tools for the detection of fecal contamination and could be applied in diverse geographic settings with less time and effort (Fong et al., 2005; Kildare et al., 2007; Harwood et al., 2014).

In particular, MST methods based on molecular markers using real-time quantitative PCR (qPCR) and community-based programs using high-throughput sequencing (HTS) data have recently emerged. Up to now, MST molecular markers have been developed to distinguish fecal contamination from various host species, including human (Kildare et al., 2007; Reischer et al., 2007), swine (Mieszkin et al., 2009), ruminant (Kildare et al., 2007), avian (Ohad et al., 2016), and others (Marti et al., 2013). Although qPCR assays using MST markers with high sensitivity and specificity can distinguish fecal contamination from different host sources, each marker can only detect a specific source of contamination. In recent years, community-based MST methods, including random forest classifier (Smith et al., 2010; Roguet et al., 2018), the SourceTracker program (Knights et al., 2011), and the FEAST (fast expectation–maximization microbial source tracking) program (Shenhav et al., 2019), have been utilized to estimate the relative contribution of microbial contamination of each source to freshwater. These tools can match the microbial community profiles between potential sources and sink samples using the operational taxonomic unit (OTU) dataset of the 16S ribosomal RNA (rRNA) gene. More specifically, FEAST exhibited stronger robustness and higher running speed than the random forest classifier and SourceTracker under the same conditions (Shenhav et al., 2019). It has also been reported that the program could simultaneously determine the relative contribution of different sources in the same sample (Chen et al., 2020; Kim et al., 2020). However, the accuracy of the FEAST program in tracking fecal pollution requires further verification because its research on the fecal inputs of river waters has been minimal.

The objectives of this study were to distinguish the sources of fecal microbes in a rural river and quantify their relative contributions to the microbial community of the river water. Given the high accuracy of the molecular markers and the ability of FEAST to identify contributions from multiple sources, both MST methods were concomitantly applied to distinguish the sources of fecal pollution in the Fsq River, located in Beijing, China, during the dry and wet seasons.

MATERIALS AND METHODS

Study Area

The mainstream of the Fsq River, located in Beijing, China, spans 18.2 km and covers 58.8 km^2. Fsq River, one of the main tributaries of the Wenyu River, plays an important role in providing recreational and landscape water for surrounding residents. The Fsq River flows through many villages in the studied area, where the villagers breed swine or bovine on a small or a large scale. The Fsq River has been reported to be severely polluted by various sources such as human and animal feces, industrial wastewater, and farmland and woodland deposits. The sewage from many villages on both sides of the river is almost discharged directly into the Fsq River, resulting in poor water quality and subsequently affecting the water quality of the Wenyu River (Qi and Chen, 2012). Feces from villagers are stored in septic tanks and regularly removed by professional companies. There are sewage treatment stations in villages near the Fsq River, but only a small number of stations are in operation. Therefore, domestic sewage without treatment is usually discharged into the river through drainage ditches. In addition, farms near the river are not equipped with treatment systems for animal waste. The piled-up manure is likely to be washed into the river by rainwater runoff during the wet season.

The wet season with frequent rainfall events in the city of Beijing where the Fsq River is located mainly occurs from July to September, whereas the dry season covers the remaining months with little or no rainfall. All outlets on both sides of the Fsq River are open drainage ditches for the discharge of rainwater and domestic sewage. Only domestic sewage flows into the river through the drainage ditches in the dry season, whereas mixtures of rainwater, domestic sewage, and animal feces are discharged into the Fsq River in the wet season.

Sampling

Fecal Samples

A total of 184 individual fecal samples were collected from each of human (n = 28) and 12 animal host species, including swine (n = 20), canine (n = 6), equine (n = 11), donkey (n = 4), bovine (n = 14), sheep (n = 18), goat (n = 5), chicken

(n = 12), duck (n = 10), goose (n = 10), pigeon (n = 9), and fish (n = 37) (**Supplementary Table 1**). All fecal samples were collected in sterile 50-ml polypropylene tubes, which were transported directly to the lab on ice and immediately kept in the laboratory at −80°C until DNA extraction.

Water Samples

A total of 96 water samples were collected along the Fsq River in a downstream to upstream manner. River water samples (n = 80) were collected from 17 sites, R1 to R17, and outfall water samples (n = 16) were obtained from five sites, including FR2, FR7, FR9, FR14, and FR15 (**Figure 1** and **Supplementary Table 2**). All water samples were collected from five sampling events, named BF, AF1, AF2, AF3, and AF4, between July and November 2019. There was no rainfall within 1 week before the occurrence of sampling event BF during the dry season, thus reflecting the pollution status of the river water without rain interference. Samples for sampling events AF1 to AF4 were obtained as early as possible after the start of rain events during the wet season. Water samples were collected in 1-L bottles using either rope-suspended samplers or sampling poles. All water samples were transported directly to the lab on ice and immediately filtered. Briefly, 1 L of each water sample was filtered through 47 mm × 0.22 μm nitrocellulose membrane (Millipore, Billerica, MA, United States) to retain microbial cells, and then filter papers were stored in the laboratory at −80°C as fecal samples.

DNA Extraction and Sequencing

Genomic DNA was extracted from fecal (180–220 mg) and water (filter papers) samples using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) used in the Earth Microbiome Project and then purified using the OneStep™ PCR Inhibitor Removal Kit (Zymo Research Corp., Irvine, CA, United States) following the manufacturer's instructions. The purity and concentration of genomic DNA were measured with a NaNodrop 2000 spectrophotometer (Thermo Fisher Technologies, Foster City, CA, United States). Purified DNA extracts were sent to Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China) for sequencing of the bacterial V3–V4 region of the 16S rRNA gene on an Illumina MiSeq platform using forward primer 338F (5′-ACTCCTACGGGAGGCAGCAG-3′) and reverse primer 806R (5′-GGACTACHVGGGTWTCTAAT-3′) (Mori et al., 2014). The reads obtained in the current study have been deposited in the SRA database (NCBI) under the BioProject accession number PRJNA713417, with Biosample accession numbers SAMN18274560–SAMN18274616 and SAMN18252387–SAMN18252398. Besides, the sequence data of one bovine feces have been deposited in the SRA database (NCBI) under the BioProject accession number PRJNA392724, with biosample number SAMN07305415.

Selection of Markers and Community-Based Programs

For qPCR assays, eight molecular markers that performed well in previous reports were selected in the study, including universal marker BacUni, four human-associated markers (HF183-1, HF183-2, BacH, and BacHum), swine-associated marker Pig-2-Bac, ruminant-associated marker Rum-2-Bac, and avian-associated marker AV4143 (**Supplementary Table 3**). The

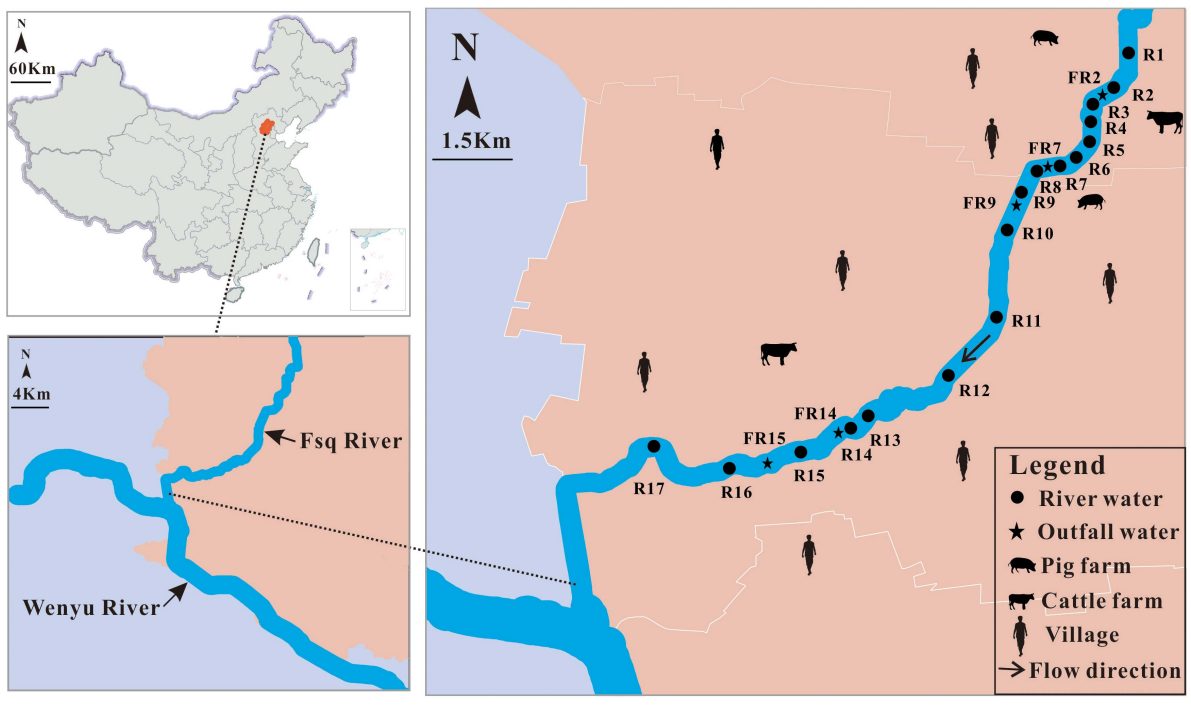


FIGURE 1 | Overview of the sampling sites in Fsq River located in Beijing, China.

universal marker BacUni has demonstrated a 100% positive rate against human and animal fecal samples in several countries (Kildare et al., 2007; Nshimyimana et al., 2017), including China (Liang et al., 2020). Four human-associated markers—HF183-1, BacH, HF183-2, and BacHum—were recommended in the studies conducted in China (Vadde et al., 2019; Zhang et al., 2020) and other countries (Reischer et al., 2013; Odagiri et al., 2015; Haramoto and Osada, 2018). The high prevalence rates of Pig-2-Bac and Rum-2-Bac in target host species were also reported in China (He et al., 2016; Xu et al., 2020; Zhang et al., 2020) and in an evaluation study of 27 labs (Boehm et al., 2013). The avian-associated marker AV4143 was modified in our previous study and showed high sensitivity and specificity against fecal samples from China (Liang et al., 2020).

The relative contributions of microbes from potential fecal sources to sink samples were estimated using FEAST, a highly efficient expectation/maximization-based program (Shenhav et al., 2019). A closed OTU dataset was generated consisting of 70 regionally specific source and sink samples. The OTU dataset of individual fecal samples (n = 58) was obtained to develop the "source" library (**Supplementary Tables 1, 4**). The OTU dataset of river water samples (n = 12) collected from locations R1 to R6 during the dry (BF) and wet (AF3) seasons was designated "sink." The FEAST program was run under the "FEAST example Multiple sinks" script to screen fecal sources to sink samples (**Supplementary Table 4**). For each sink, five independent runs were carried out in R version 3.6.3 (Vienna, Austria) to reduce the effect of false predictions, as in a previous study (Henry et al., 2016).

In the qPCR assays, the performances of eight selected markers were first evaluated in 133 fecal samples from 13 host species (**Supplementary Table 1**). Subsequently, only markers that performed well were used to track the sources of all 96 water samples (**Supplementary Table 2**). In the FEAST program, 16S rRNA genes of fecal (n = 58) and water (n = 12) samples (**Supplementary Tables 1, 4**) were sequenced to build the "source" library and "sink." Compared with a foreign source library, a local source library could efficiently distinguish fecal sources in the sink samples (Staley et al., 2018), and thus the fecal samples used to build the "source" library were collected from China in this study. In addition, fecal sources of the water samples were identified with MST markers, and then only the identified fecal sources were contained in the "source" library to obtain more reliable prediction results by the FEAST program (Brown et al., 2019).

qPCR Assays

All the MST markers in this study were monitored on an ABI 7500 real-time qPCR system (Applied Biosystems, Foster City, CA, United States). The optimized reaction mixture (20 μl) was composed of 10 μl 2 × Premix Ex Taq (Probe qPCR; Takara Bio, Otsu, Japan), 0.2 μl Rox Reference Dye II (50×; Takara Bio, Otsu, Japan), primers and probes at the final concentrations in the mix shown in **Supplementary Table 3**, 2 μl template DNA, and nuclease-free water to a final volume of 20 μl. The TaqMan PCR program was initiated at 50°C for 2 min and 95°C for 30 s, followed by 40 cycles of denaturation at 95°C for 5 s and annealing/extension at 60°C for 1 min. All reactions, including those of the samples tested, standards (**Supplementary Table 5**), and no-template controls, were performed in triplicate using MicroAmp Optical 96-well reaction plates.

The standard curve for each of eight TaqMan qPCR assays (**Supplementary Table 3**) was established using six 10-fold serial dilutions (10^3–10^8 gene copies per reaction) of plasmid standards (**Supplementary Table 5**) containing the target gene sequences. The amplification efficiency, limit of detection (LOD), and limit of quantification (LOQ) were measured based on the generated standard curves (**Supplementary Table 6**). The amplification efficiency (E) was calculated according to the formula: $E = 10^{(1/-\text{slope})} - 1$ (Bustin et al., 2009). LOD was the lowest number of gene copies detected in the target host samples. LOQ was considered the lowest concentration within the linear range of quantification. Once the LOD and LOQ were confirmed, the absence of markers in samples can be divided into ND (not detected, no amplification), DNQ (detected but not quantifiable, LOQ $< C_t <$ LOD), and ROQ (detected within the range of quantification, $C_t <$ LOQ). To ensure reproducibility, the standards of each marker were tested in different plates (**Supplementary Table 7**), as described in previous studies (Nshimyimana et al., 2017; Zhang et al., 2020). The performances of the universal and specific markers were evaluated according to sensitivity (R), specificity (S), and accuracy (A) (Kildare et al., 2007; Odagiri et al., 2015). They were defined as follows: R = TP/(TP + FN) × 100%, S = TN/(TN + FP) × 100%, and A = (TP + TN)/(TP + FP + TN + FN) × 100%, where TP and FN represent true positive and false negative, respectively, in the target host samples and TN and FP are true negative and false positive, respectively, in the non-target host samples.

Cross-reactivity was defined as the condition at which the marker was found to be positive in the non-target samples. A "25th/75th" metric was applied to determine the classification of cross-reactivity for each marker in the non-target host samples (i.e., 25th/75th metric = 25th percentile$_{target}$ - 75th percentile$_{non-target}$) (Reischer et al., 2013). Accordingly, the samples were classified into four groups (Zhang et al., 2020) as follows: no cross-reactivity (NCR), when the marker did not show any positive signal in the non-target samples; weak cross-reactivity (WCR), when the "25th/75th metric" > 0; moderate cross-reactivity (MCR), when the "25th/75th metric" <0; and strong cross-reactivity (SCR), when either the disparity between the mean gene copies of the target and non-target samples was below one order of magnitude or the number of mean gene copies of the non-target samples was higher than that in the target samples.

Bioinformatics and Statistical Analyses

The raw 16S rRNA gene sequences were quality filtered using Fastp software (v.0.20.0) (Chen et al., 2018). Reads were filtered to remove adapters and trimmed to remove any terminal stretches of bases at or below Q30. FLASH software (v.1.2.11) was used to merge into a single read of sequences that passed the quality control (Magoc and Salzberg, 2011). Chimeric sequences were checked and removed using USEARCH v.7.0 (Edgar, 2010). Sequences with ≥97% similarity were assigned to the same OTU

and the representative sequence was screened for each OTU using UPARSE v.7.0 (Edgar, 2013). The sequences were further filtered to remove OTUs that accounted for less than 0.005% of the total sequence counts (Bokulich et al., 2013). Representative sequences were used to annotate taxonomic information against the Silva database (release 138) using the RDP Classifier (v.2.2) on the QIIME platform 1.9.0 based on the Bayesian algorithm (Caporaso et al., 2010). Statistical analyses and data visualization were carried out using SPSS version 25.0 (SPSS Inc., Chicago, IL, United States) and R version 3.6.3 (Vienna, Austria), respectively.

RESULTS

Selection of MST Markers Used for Water Samples

Prevalence of MST Markers in Human and Animal Fecal Samples

We used 133 individual fecal samples to evaluate the sensitivity, specificity, and accuracy of eight MST markers using TaqMan qPCR. Based on this, we only selected the markers effective at detecting the fecal pollution of water samples. Samples identified as ND were considered negative, whereas those identified as DNQ and ROQ were treated as positive in this study.

We found that the universal marker BacUni exhibited 100% sensitivity because it was detected in all fecal samples from humans (n = 13) and animals (n = 120) (**Figure 2** and **Supplementary Table 8**). In the case of the four human-associated markers, we observed that both HF183-1 and BacH performed well, with ≥92% accuracy, followed by HF183-2 and BacHum, with ≤77% accuracy. The sensitivity of each of HF183-1, BacH, HF183-2, and BacHum was calculated as 100%, but their specificity values were 93%, 92%, 76%, and 77%, respectively. We also found that the markers Pig-2-Bac, Rum-2-Bac, and AV4143 displayed ≥91% accuracy. More specifically, we noticed that the three markers were not only present in the fecal samples of 100% swine, 100% ruminant (i.e., bovine, sheep, and goat), and 86% avian (i.e., chicken, duck, goose, and pigeon), respectively, but also demonstrated ≥97% specificity against the non-target host samples (**Figure 2** and **Supplementary Table 8**).

Concentrations of MST Markers in Human and Animal Fecal Samples

Using standards in each qPCR assay, we observed that the amplification efficiencies ranged from 87.1% to 101.9% (**Supplementary Table 6**), meeting the requirements of the MIQE (minimum information for the publication of qPCR) guidelines (Bustin et al., 2009). The LOD and LOQ for each qPCR assay are also exhibited in **Supplementary Table 6**. Moreover, the inhibitors in the samples used in this study were found to have little effect on the qPCR assays, in accordance with our previous research (Liang et al., 2020). The credibility of the qPCR results was further proven by the ideal reproducibility of the qPCR assays on different reaction plates. We found that the mean coefficient of variability (CV) was less than 4.0% ± 0.1% for each standard between 10^3 and 10^8 gene copies/μl (**Supplementary Table 7**). Marker abundance was characterized as log10 (gene copies) per gram wet feces or log10 (gene copies) per 100 ml water, as discussed in previous studies (Kildare et al., 2007; Nshimyimana et al., 2017).

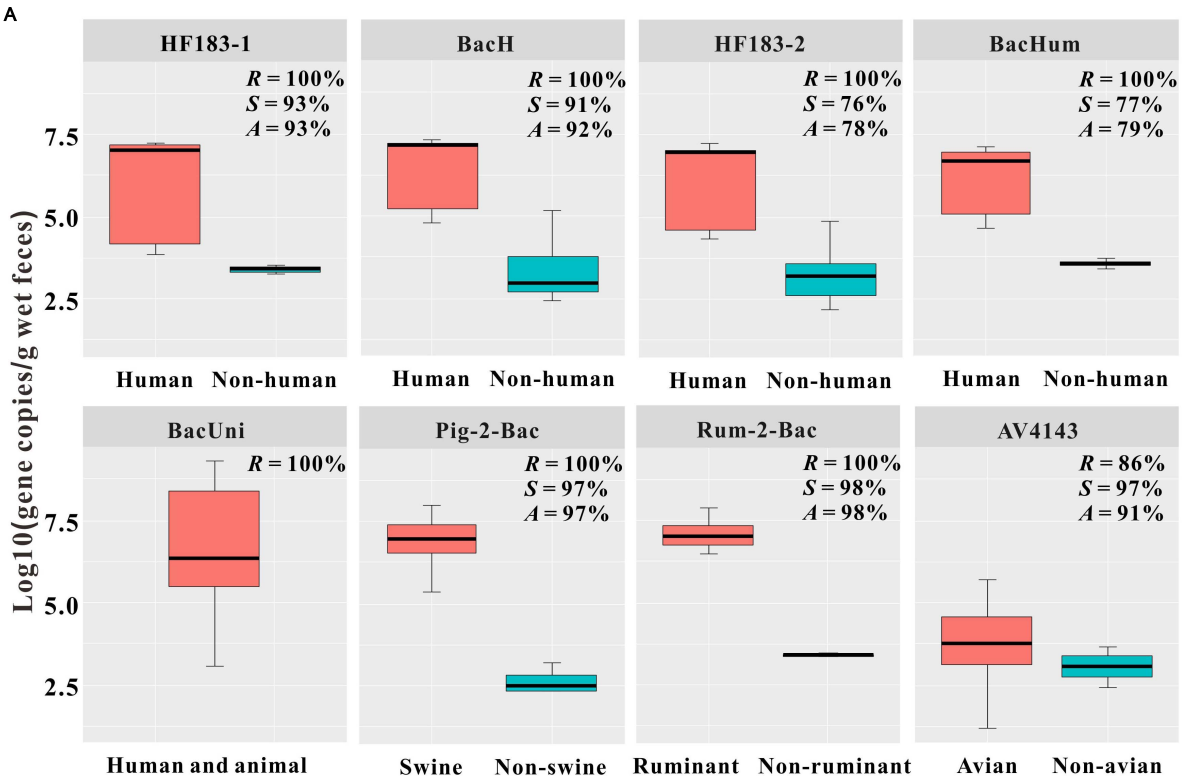
In general, we observed that the universal marker BacUni showed higher gene copy numbers than the specific markers in the target hosts. In particular, we found that the concentrations of BacUni were 6.70 ± 1.71 (mean ± standard deviation) in all the fecal samples tested from humans and animals (**Figure 2A** and **Supplementary Table 8**). Despite the presence of false-positive samples in each human-associated marker tested in this study, we noticed significant differences in the concentration of each marker between the fecal samples from humans (mean ± SD > 6.16 ± 1.4) and non-human hosts (mean ± SD < 3.59 ± 0.33; one-way ANOVA: $P < 0.01$) (**Figure 2A** and **Supplementary Table 8**). Likewise, approximately four orders of magnitude and significant differences were also observed for Pig-2-Bac and Rum-2-Bac between the target host (mean ± SD = 6.84 ± 0.84 and 7.02 ± 0.50) and the non-target host (mean ± SD = 2.66 ± 0.41 and 3.41 ± 0.06) fecal samples (one-way ANOVA: $P < 0.01$) (**Figure 2A** and **Supplementary Table 8**). In contrast, we noticed that the avian-associated marker AV4143 exhibited low concentrations (mean ± SD = 3.81 ± 1.09) in the target samples compared with the other host-associated markers tested in this study. The performances of HF183-2 and BacHum were poor (**Figure 2A** and **Supplementary Table 8**); thus, we did not further classify the cross-reactivity of these two markers. We consecutively divided the cross-reactivities of HF183-1, BacH, Pig-2-Bac, and Rum-2-Bac in the non-target host fecal samples into NCRs or WCRs (**Figure 2B** and **Supplementary Table 9**). In the AV4143 assay, we found that only one positive sample from equine was classified as MCR, whereas the others were NCRs or WCRs (**Figure 2B** and **Supplementary Table 9**).

Screening of MST Markers Used for Water Samples

Based on the results of the qPCR assays, we selected six markers, including the universal marker BacUni, the human-associated markers HF183-1 and BacH, swine-associated marker Pig-2-Bac, the ruminant-associated marker Rum-2-Bac, and the avian-associated marker AV4143, to distinguish fecal pollution in water samples. Each of these markers met the 80% specificity and sensitivity benchmarks used in a previous report (Boehm et al., 2013) and the guideline document of the United States Environmental Protection Agency (USEPA) (U.S. Environmental Protection Agency, 2005) when DNQ was considered positive (**Figure 2A** and **Supplementary Table 8**). In addition, we found that the cross-reactivity for each selected marker ranged only from 2% to 9% in the non-target hosts, with almost all of them being grouped into WCR or NCR, which indicated that there was no or little effect on the qPCR assays (**Figure 2B** and **Supplementary Table 9**).

Application of MST Markers Selected in Water Samples

We analyzed all 96 water samples, including river water (n = 80) and outfall water (n = 16), using the six aforementioned markers selected from the previous step. We accordingly found that all



B

	BacUni	HF183-1	BacH	Pig-2-Bac	Rum-2-Bac	AV4143
Human	TP	TP	TP	WCR	NCR	NCR
Swine	TP	NCR	NCR	TP	NCR	WCR
Canine	TP	WCR	WCR	NCR	NCR	WCR
Equine	TP	WCR	NCR	NCR	WCR	MCR
Donkey	TP	NCR	NCR	NCR	NCR	NCR
Bovine	TP	NCR	NCR	NCR	TP	NCR
Sheep	TP	NCR	NCR	NCR	TP	NCR
Goat	TP	NCR	NCR	NCR	TP	NCR
Chicken	TP	NCR	NCR	NCR	NCR	TP
Duck	TP	WCR	WCR	NCR	NCR	TP
Goose	TP	NCR	NCR	NCR	WCR	TP
Pigeon	TP	NCR	WCR	NCR	NCR	TP
Fish	TP	WCR	WCR	NCR	NCR	NCR

FIGURE 2 | Concentrations **(A)** and classification of cross-reactivity. **(B)** for each tested marker. **(A)** Prevalence and concentrations of the universal marker BacUni, human-associated markers (HF183-1, BacH, HF183-2, and BacHum), pig-associated marker (Pig-2-Bac), ruminant-associated marker (Rum-2-Bac), and avian-associated marker (AV4143) in fecal samples from human, swine, canine, equine, donkey, ruminant (bovine, sheep, and goat), avian (chicken, duck, goose, and pigeon), and fish. Letters *R*, *S*, and *A* represent the sensitivity, specificity, and accuracy of each marker, respectively. Each box plot shows the median, upper, and lower quartiles spanning the maximum and minimum observations. Log-transformed gene copies of negative results were treated as 0 value. Only positive results within the limit of detection (LOD) were displayed in the graph. **(B)** Classification of cross-reactivity in the non-target samples for each microbial source tracking (MST) marker used to detect fecal pollution from environmental samples. The results were colored based on the following criteria: TP (true positive); NCR (no cross-reactivity), no false-positive signal was amplified; WCR (weak cross-reactivity), the "25th/75th metric" >0; MCR (moderate cross-reactivity), the "25th/75th metric" <0; SCR (strong cross-reactivity), the disparity of the mean gene copies between the non-target and target hosts is less than one order of magnitude or the mean gene copies of the non-target samples are higher than those in the target samples.

the water samples from the river and outfalls consistently showed 100% positive signals for total Bacteroidales (BacUni; **Table 1**). In particular, we detected that the human-associated markers (60% in HF183-1 and 69% in BacH) exhibited the highest prevalence in all the river water samples tested, followed by the ruminant-associated marker Rum-2-Bac (29%) and finally the swine-associated marker Pig-2-Bac (3%) and the avian-associated marker AV4143 (0%; **Table 1**). The positive rates of markers HF183-1 and BacH were demonstrated to be higher in river samples during the dry season (87% and 93%, respectively) than those from the wet season (60% and 69%, respectively). However, we observed the opposite trend in the detection rates for markers Pig-2-Bac (0% and 3%) and Rum-2-Bac (7% and 34%) in the samples collected from the dry and wet seasons, respectively (**Table 1**). A similar pollution trend was also observed in the outfall water samples. More specifically, we found that markers HF183-1 and BacH showed the highest positive rates (38 and 56%), followed by Rum-2-Bac (13%) and finally Pig-2-Bac (6%) and AV4143 (0%) in the outfall water samples (**Table 1** and **Supplementary Table 10**).

We further observed a strong correlation of the mean concentrations of the tested markers between the positive river and outfall water samples (Spearman: $R = 0.85$, $P < 0.01$). Signals of HF183-1 and BacH were detected in both the river (mean ± SD = 3.76 ± 0.72 and 3.61 ± 0.75) and outfall (mean ± SD = 4.03 ± 0.78 and 3.56 ± 1.16) water samples (**Figure 3A** and **Supplementary Table 10**). Likewise, Bacteroidales from ruminant feces (Rum-2-Bac) were quantified at the concentrations of 2.97 ± 0.65 in 23 of the 80 river water samples and 2.93 ± 1.16 in two of the 16 outfall water samples (**Figure 3A** and **Supplementary Table 10**). Signals of Pig-2-Bac were detected only in the river water samples (mean ± SD = 2.50 ± 0.05) collected at locations R2 (2.53) and R3 (2.47) and in the outfall water sample collected at location FR9 (2.47) during the wet season (**Figure 1**), which were adjacent to a large-scale pig farm in the village nearby (**Figures 1, 3A** and **Supplementary Table 10**). No amplification of AV4143 was found in the outlet water at each sampling event (**Figure 3A**).

There were significant statistical differences in the concentrations of the markers BacUni, HF183-1, BacH, Pig-2-Bac, and Rum-2-Bac between the river water samples collected during the dry and wet seasons (independent-samples *t*-test: $P < 0.05$; **Figure 3B**). In particular, we found that the concentrations of total and human-specific Bacteroidales (i.e., BacUni, HF183-1, and BacH) in the river water samples were higher in dry weather (mean ± SD = 6.53 ± 0.84, 4.34 ± 0.82, and 4.25 ± 0.90, respectively) than those in the wet season (mean ± SD = 6.13 ± 0.60, 3.55 ± 0.56, and 3.40 ± 0.56, respectively) (**Figure 3B** and **Supplementary Table 10**). On the contrary, higher levels of Pig-2-Bac and Rum-2-Bac were observed in the river water samples collected during the wet season (mean ± SD = 2.50 ± 0.05 and 2.97 ± 0.65) than during the dry season (0 and 1.54) (**Figure 3B** and **Supplementary Table 10**).

High-Throughput Sequencing Analyses

A total of 58 fecal samples and 12 river water samples were sequenced in this study (**Supplementary Table 4**). The average sequence numbers from the water and fecal samples were 48,288 and 36,515, respectively (**Supplementary Table 11**). We noticed that the microbial community showed a lower average α diversity in the fecal samples than in the river water samples (**Supplementary able 11**). Feces had an average Shannon index of 4.30 ± 0.93 and 3,666 OTUs that clustered at 97% similarity, whereas freshwater samples had an average Shannon index of 4.92 ± 1.18 and 5,561 OTUs clustering at 97% similarity (**Supplementary Table 11**).

Fecal and river water samples were clustered separately based on the analysis of the hierarchical clustering tree of the Bray–Curtis dissimilarity matrix (**Figure 4** and **Supplementary Figure 1**). Differences in the microbial communities between the fecal and water samples were demonstrated according to the presence and abundance of specific bacteria. In the bacterial community of the fecal samples, four of 17 orders accounted for 56.77%–72.61% of the sequence data across human and animal fecal samples. The four orders were Bacteroidales (25.92%–39.12%), Oscillospirales (12.23%–30.72%), Lachnospirales (5.42%–28.71%), and Peptostreptococcales–Tissierellales (2.10%–8.83%) (**Figure 4** and **Supplementary Table 12**). The orders Flavobacteriales (41.06% and 5.53%) and Burkholderiales (27.34% and 15.74%) were found to be the most common taxa in the water samples during the dry and wet seasons, respectively (**Figure 4** and **Supplementary Table 12**). In addition, we observed that each of the nine orders showed

TABLE 1 | Prevalence of the microbial source tracking (MST) markers in the water samples collected from the river and outfalls.

Sample	Season	Sampling events	No. of samples	Positive ratios					
				Total (%)	Human (%)		Swine (%)	Ruminant (%)	Avian (%)
				BacUni	HF183-1	BacH	Pig-2-Bac	Rum-2-Bac	AV4143
River	Dry	BF	15	100	87	93	0	7	0
	Wet	AF	65	100	60	69	3	34	0
	Total		80	100	60	69	3	29	0
Outfall	Dry	BF	1	100	100	100	0	0	0
	Wet	AF	15	100	33	53	7	13	0
	Total		16	100	38	56	6	13	0

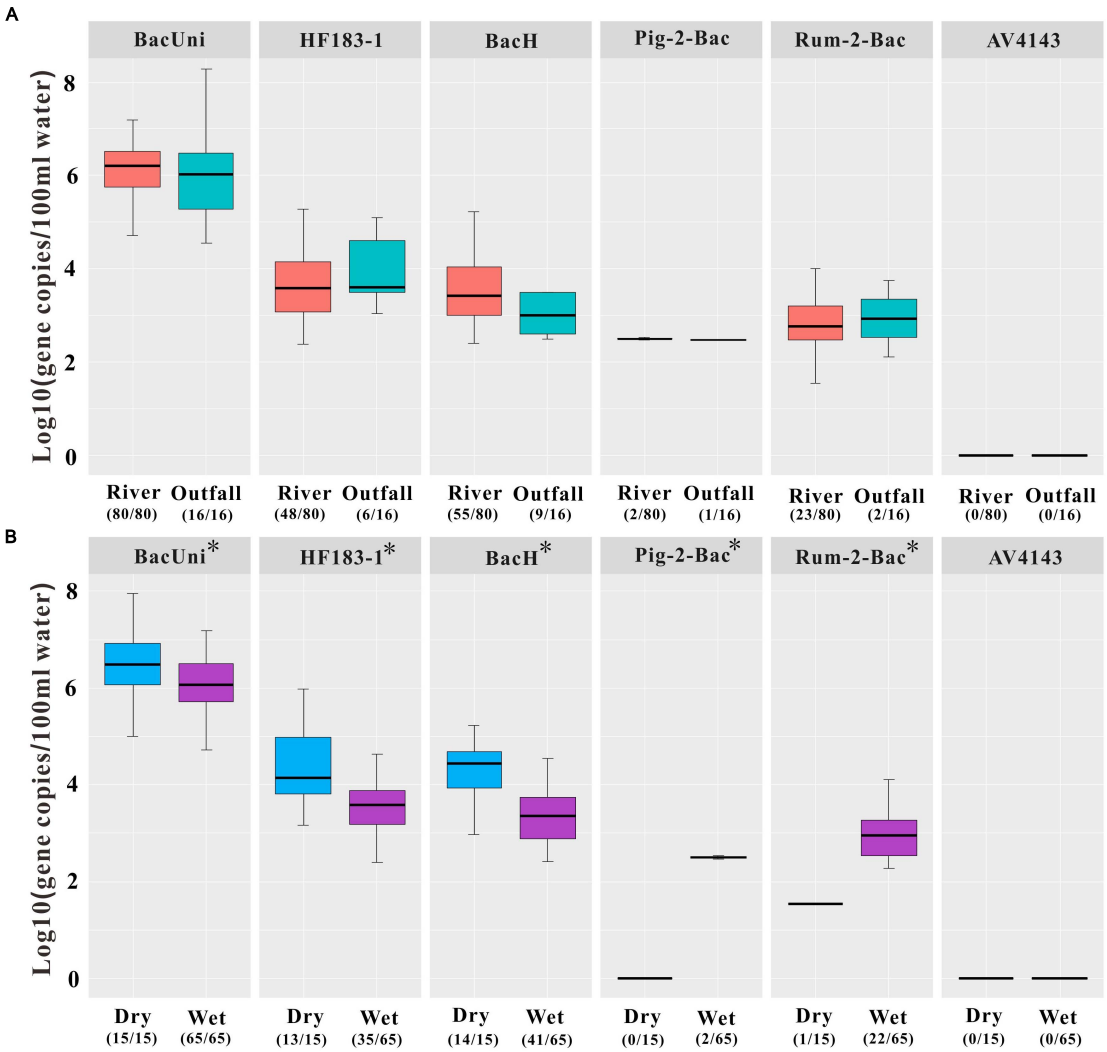


FIGURE 3 | Concentrations of the universal marker BacUni, the human-associated markers HF183-1 and BacH, the swine-associated marker Pig-2-Bac, the ruminant-associated marker Rum-2-Bac, and the avian-associated marker AV4143 in environmental samples, including **(A)** all river water and outfall water samples and **(B)** river water samples during the dry and wet seasons. *Asterisk* indicates that statistically significant differences were observed in the concentrations of the markers in the river water samples between the dry and wet seasons by independent-samples *t*-test at the 0.05 level of significance. *Numbers in parentheses* represent the sample number within the limit of detection (LOD)/total samples number tested. Each box plot shows the median, upper, and lower quartiles spanning the maximum and minimum observations. Log-transformed gene copies of negative results were treated as 0 value. Only positive results within the limit of quantification (LOQ) were displayed in the graph.

different relative abundance among water and each kind of host fecal sample, including Bacteroidales (0.54%–39.12%), Oscillospirales (0.03%–30.72%), Lachnospirales (0.08%–28.71%), Burkholderiales (0.01%–27.34%), Peptostreptococcales–Tissierellales (0.09%–8.83%), Christensenellales (0.01%–7.53%), Lactobacillales (0.02%–11.73%), Clostridiales (0.20%–5.77%), and Enterobacterales (0.02%–7.79%).

We also performed non-metric multidimensional scaling (NMDS) analyses based on the Bray–Curtis dissimilarities between the fecal (i.e., human, swine, bovine, and sheep) and river water samples collected during the dry and wet seasons (**Figure 5**). Similar to the results of hierarchical clustering, the fecal and river water samples were clustered independently with a statistically significant difference (stress = 0.123, R = 0.91, $P < 0.01$). Moreover, we found that the fecal samples from different host species were also clustered independently, whereas those of bovine and sheep belonging to ruminants were classified as the same cluster.

Fecal Sources Determined by FEAST Analyses

As fecal pollution from hosts of humans, swine, and ruminants (i.e., bovine and sheep) has been identified as a source in the Fsq River based on the results of the qPCR assays, we sequenced the fecal samples (n = 58) from these host species to create a "source"

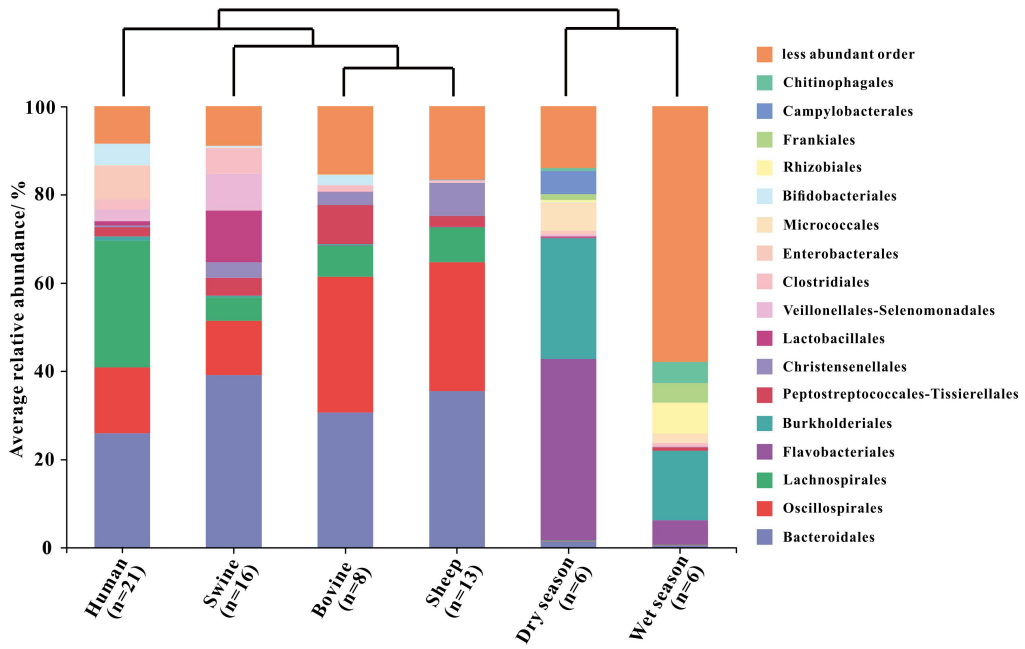


FIGURE 4 | Hierarchical clustering tree based on Bray–Curtis and the relative abundance of the bacterial community on the order level in the fecal samples (human, swine, bovine, and sheep) and river water samples collected from the dry and wet seasons. Only orders with a relative abundance higher than 4% were represented in the legend.

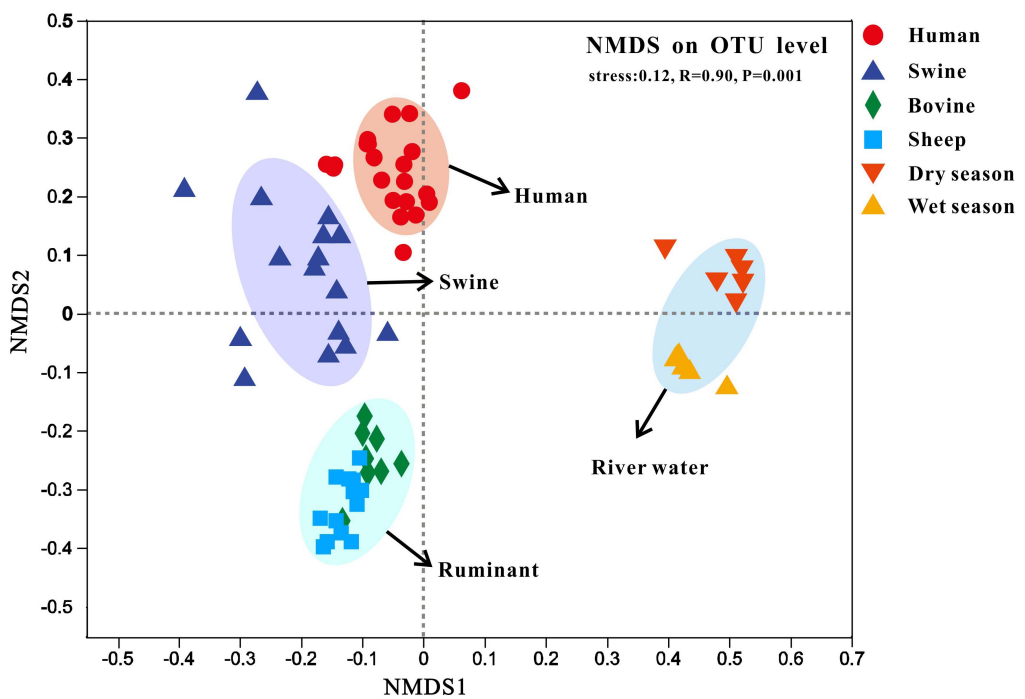


FIGURE 5 | Non-metric multidimensional scaling (NMDS) analysis of the Bray-Curtis dissimilarities among sample bacterial communities on the operational taxonomic unit (OTU) level. Fecal samples were collected from human, swine, bovine, and sheep. River samples were collected from the Fsq River during the dry and wet seasons.

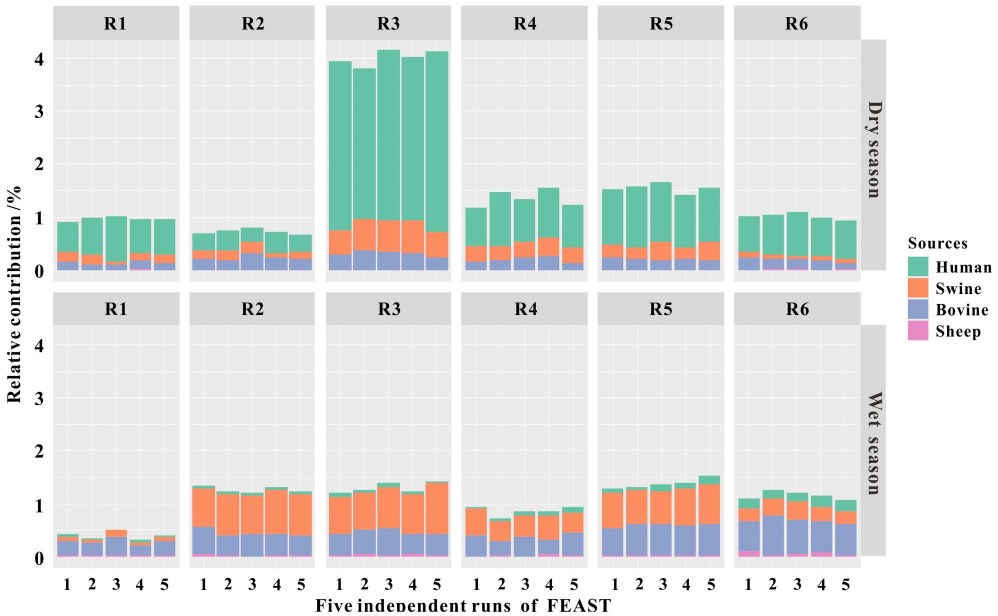


FIGURE 6 | Fast expectation–maximization microbial source tracking (FEAST) analyses showing the relative contributions of fecal pollution sources in river water collected from R1 to R6 during the dry and wet seasons. A *single bar* represents the relative contribution in an independent FEAST run using the same script. FEAST was run in five independent runs on the same script.

library. The sink was built using the 16S rRNA gene sequences of the river water samples (n = 12) collected from R1 to R6, six different locations in the Fsq River, during the dry and wet seasons. Subsequently, we ran a total of five independent analyses on the OTU level taxa tables by the FEAST program to predict the proportions of fecal sources in the river water samples. We accordingly observed that bacterial signatures resembling fecal sources represented low levels of water contamination (<4.01% of sequence reads) in the bacterial communities of the river samples (**Figure 6** and **Table 2**). Potentially unknown sources accounted for a high level (>95.99% of sequence reads), owing to the complex community composition (**Table 2**).

To evaluate the accuracy of FEAST for predicting the relative contributions of sources, Spearman's rank correlations were performed to relate sink predictions with the relative standard deviation (RSD) values obtained from five individual FEAST runs. We observed strong and significant negative correlations between the FEAST predictions and the RSD values ($\rho = -0.82$, $P < 0.001$). Larger relative contributions such as human feces showed lower RSD values (7–16%), whereas smaller source proportions such as sheep had high RSD values (>75%) during the dry season (**Supplementary Table 13**). Importantly, the source contributing the largest fecal pollution in each sink showed low RSD values (7%–19%; **Supplementary Table 13**).

DISCUSSION

In this study, MST methods based on molecular markers and machine learning programs were applied together to distinguish the fecal inputs from multiple sources in a rural river located in Beijing, China. Previous studies have shown the credibility of qPCR assays using markers with high sensitivity and specificity in distinguishing the different fecal sources in the river water (Layton et al., 2013; Zhang et al., 2020). However, the MST method using markers has some limitations, such as detecting only a single specific source of pollution rather than

TABLE 2 | Relative contributions of potential fecal sources based on FEAST analyses in river water samples.

Season	Sink	Relative contribution of source (%)				
		Human	Swine	Bovine	Sheep	Unknown
Dry	R1	**0.69**[a]	0.15	0.13	0.01	99.02
	R2	**0.34**	0.15	0.24	0	99.27
	R3	**3.15**	0.55	0.31	0	95.99
	R4	**0.85**	0.3	0.2	0	98.65
	R5	**1.08**	0.27	0.21	0	98.44
	R6	**0.74**	0.09	0.18	0.01	98.98
Wet	R1	0.03	0.08	**0.26**	0.02	99.61
	R2	0.06	**0.77**	0.42	0.02	98.73
	R3	0.06	**0.78**	0.43	0.03	98.7
	R4	0.07	**0.43**	0.34	0.02	99.14
	R5	0.11	**0.67**	0.57	0.03	98.62
	R6	0.19	0.28	**0.62**	0.06	98.85

FEAST, fast expectation–maximization microbial source tracking.
[a]*Values in bold indicate that the fecal source was calculated as having the largest relative contribution in total fecal sources.*

multiple sources at the same time. FEAST, a newly emerging computational tool, could be applied to estimate multiple potential sources and the relative contributions of various fecal inputs at the same time. This study comprehensively combined these two advanced MST methods and found that the relative contributions of fecal pollution in the river were influenced by rainfall events. As such, more measures should be taken to prevent human and animal feces from flowing into rivers during the wet season.

Evaluation of MST qPCR Assays in Fecal Samples

The universal marker BacUni, the human-associated markers HF183-1 and BacH, swine-associated marker Pig-2-Bac, ruminant-associated Rum-2-Bac, and the avian-associated marker AV4143 performed well for the target and non-target host species in this study. The universal marker BacUni, targeting the 16S rRNA gene of Bacteroidales, was detected in all human and animal fecal samples (**Figure 2** and **Supplementary Table 8**), which was comparable to the 100% sensitivity observed in other regions, including the United States (Kildare et al., 2007), India (Odagiri et al., 2015), Thailand (Somnark et al., 2018), and Kenya (Jenkins et al., 2009). Both HF183-1 and BacH exhibited sensitivity and specificity greater than 91%, whereas the two remaining human-associated markers HF183-2 and BacHum demonstrated specificity less than 80% despite being amplified in all the tested human fecal samples. According to the guideline document of the USEPA (U.S. Environmental Protection Agency, 2005), a marker for MST can be creditable only when its specificity is 80% or higher (maximum value of 100%). Additionally, our results indicated that HF183-1 and BacH had good distinguishing effects, which was consistent with previous studies carried out in China (Vadde et al., 2019; Zhang et al., 2020). The swine-associated marker Pig-2-Bac, the ruminant-associated marker Rum-2-Bac, and the avian-associated marker AV4143 were found to be qPCR-positive in at least 86% of the target host samples tested in this study (**Figure 2** and **Supplementary Table 8**), which was also consistent with other investigations carried out on samples collected from China by He et al. (2016), Liang et al. (2020), Zhang et al. (2020).

The classification method of the 25th/75th metric based on the concentrations of the MST markers can effectively filter out false positives with low concentrations and then improve the reliability of the MST markers. Fecal samples from 13 host species were tested in this study, which met the requirements of the USEPA MST guidelines that more than 10 species of animals should be used for evaluations of host specificity (U.S. Environmental Protection Agency, 2005). Lower false-positive signals from WCRs in the fecal samples could not be detected in the environmental samples because fecal droppings were diluted after entering rivers (Balleste et al., 2020; Zhang et al., 2020). The cross-reactivities of the markers that performed well in sensitivity and specificity (i.e., BacUni, HF183-1, BacH, Pig-2-Bac, Rum-2-Bac, and AV4143) were almost classified as NCRs or WCRs (except an MCR in AV4143 with an equine sample), indicating that the signals from the non-target samples might not be detected in the river water samples. Therefore, the markers BacUni, HF183-1, BacH, Pig-2-Bac, Rum-2-Bac, and AV4143 were considered suitable for subsequent experiments for the detection of the presence or absence of human or animal fecal sources in environmental samples.

Application of MST Markers in Water Samples

The prevalence of human or animal feces in the Fsq River was related to the presence or absence of rainfall events. The detection rates of the human-associated markers were higher in the samples collected during the dry season than in those collected during the wet season (**Table 1**), indicating baseline human fecal pollution in the Fsq River. With rainfall flowing into the river, the original concentrations of the human markers in the river water were diluted, resulting in the lower detection rates of human feces, consistent with the results of previous studies (Newton et al., 2013; Balleste et al., 2020). Conversely, the amplified signals of swine feces were only found in the samples from the wet season, but not in the dry season. A similar trend was also observed for Rum-2-Bac, indicating the increased input of ruminant feces owing to runoffs, as previously reported (Sidhu et al., 2012). Therefore, the prevalence difference of the markers between the samples collected during dry and wet weathers demonstrated the influence of bacterial inputs from different fecal sources due to rainfall events. In addition, the detection rates of fecal pollution in the outfall water samples revealed the prevalence rates of human, ruminant, and swine sources from high to low, consistent with the observations in the river water samples, suggesting that outfall water might contribute to pollution of the river.

Human fecal inputs were the most frequently detected in the Fsq River, followed by swine and ruminant fecal inputs, whereas no avian feces were detected. The Fsq River flows through many villages in the studied area, where swine and cattle are bred on a small or a large scale. Outfall water is one of the ways in which rainwater runoffs enter the receiving water. Runoff, especially seasonal rainwater in the wet season, has been widely recognized as a major transport vector of pollutants into the receiving water; therefore, runoff was considered a significant contributor to the deterioration of the quality of rural receiving waters (Barbosa et al., 2012; Kostyla et al., 2015). A higher level of fecal contamination has been observed in research of rural areas compared with that in urban areas during the wet season (Kostyla et al., 2015). Therefore, it is imperative to impose more measures such as farm management or ditch flow control in rural settings during the wet season to improve water quality.

Application of FEAST Program in Water Samples

Only concentrations of the same marker, rather than different markers, can be used to compare the levels of fecal pollution of the targeting source among different samples. It should be noted that the concentrations of the different markers might not be comparable, especially when these markers were designed from different gene fragments of diverse bacteria, according to previous reports (Liang et al., 2020; Zhang et al., 2020).

Therefore, a higher gene copy number might not indicate a higher pollution level of the corresponding host species. In the case of the FEAST program, a "source" library including the potential sources was built, and each sample tested was treated as an individual "sink." The relative contribution of each source to each sink was calculated according to the matching ratio of the OTUs between the "sink" and the "source" library. Therefore, FEAST could calculate the relative contribution of each source in the tested samples, which cannot be achieved by the MST method based on the qPCR technique. Simultaneous screening of the OTUs of all sources was carried out using the FEAST program to match the OTUs of each sink, and thus the relative contribution of each source in each sink could be identified at the same time.

The construction of the "source" library included local fecal samples from humans, swine, and ruminants (i.e., bovine and sheep) detected in the Fsq River based on the results of the qPCR assays. Firstly, compared with a foreign source library, a local source library could efficiently distinguish fecal sources in the sink samples (Staley et al., 2018), and thus the fecal samples used to build the "source" library were collected from China in this study. Secondly, the prediction results were more reliable when only the known potential sources were included in the "source" library rather than more but random sources (Brown et al., 2019). In addition, highly clustered samples were selected using hierarchical clustering of individual fecal samples to constitute a representative "source" library (**Supplementary Figure 1**) because the low intragroup variability of the source profiles also enhanced the accuracy of prediction of community-based MST methods (Brown et al., 2019). Analyses of the fecal (n = 58) and river water (n = 12) samples indicated that the composition of bacterial community differed significantly between the fecal and water samples (**Figure 4**), ensuring the determination of the presence of fecal inputs and the identification of specific inputs.

fast expectation–maximization microbial source tracking predicted that the largest fecal input in sites R1 to R6 was from human source under dry weather conditions, whereas a bovine or swine source was dominant during the wet season, in agreement with the analysis results using host-associated molecular markers. FEAST is a promising tool to detect low-level bacterial signatures of freshwater, which were similar to those obtained using SourceTracker (O'Dea et al., 2019). FEAST analyses assigned the contamination of the river water samples collected during dry weather to human fecal signatures, comprising 0.34%–3.15% of the total bacterial community (**Figure 6** and **Table 2**). Similar results, in which <10% (Newton et al., 2013) and 1–13% (Ahmed et al., 2015) were assigned to human fecal inputs, were reported in previous studies that investigated the presence of human feces using SourceTracker in water samples. The non-human fecal inputs always constituted <1% of the total proportion of the sink community in the river water samples, even though the contributing percentage of bovine or swine was the highest in the total fecal inputs during wet weather periods. Previous reports have also indicated that non-human fecal inputs were predicted to be trivial when they were compared with environmental samples near the sampling regions (Newton et al., 2013; Baral et al., 2018).

The results using the FEAST program validated its use as a tool to distinguish fecal pollution of various sources and indicated that it could be a promising addition to the toolbox for identifying pollution sources and for suggesting appropriate migration strategies to improve water quality. Of course, this study also had some limitations, even though the predicted results of the FEAST program were promising. For example, only a very small sample size was analyzed by the FEAST program and only fecal samples were included to build the "source" library. In this study, most microbial taxa in the sink did not match the fecal signature in the "source" library, thus being classified as unknown, as in the case of published studies using SourceTracker (Newton et al., 2013; Staley et al., 2018; O'Dea et al., 2019). Therefore, more samples, especially potential sources near the sampling sites (e.g., soil and rainwater samples), need to be included in the "source" library if the composition of the unknown source is to be clarified. SourceTracker, a community-based program like FEAST, could accurately predict the fecal compositions of fecal sources, according to a previous double-blinded study (Staley et al., 2018). In previous studies using SourceTracker, unknown sources were mainly identified as treated effluent (Henry et al., 2016), wastewater effluent or influent (Ahmed et al., 2017; Brown et al., 2017), and embankment soil or streambed sediment (Baral et al., 2018; Li et al., 2020), which were collected in the studied area near the sink samples.

CONCLUSION

The results of the FEAST program were consistent with those of the qPCR assays. Both methods revealed that fecal contamination from humans was dominant during dry weather and that the baseline of human fecal pollution might exist in the Fsq River. Swine and ruminant fecal sources were more prevalent in the samples during the wet season than in those during the dry season, owing to their potential discharge into the river water *via* the runoff system. MST methods using molecular markers retain certain advantages because they can monitor the sources of pollutants quickly and accurately without building the "source" library, especially when the sample size is small. However, there is no doubt that FEAST based on machine learning could provide promising advantages over traditional culture-based fecal indicator approaches and single-target qPCR assays. The method only needs to characterize relevant source samples instead of building thousands of libraries, thus reducing the costs or labor. Of course, each MST approach has its limitations, and the FEAST program is no exception. The development of a representative library requires time, but efficient and rapid advantages of the method could be fully exploited in the long-term dynamic monitoring of environmental samples from multiple pollution sources once a representative library is created. Therefore, the future development of FEAST should not only focus on its technical limitations but should also make better use of the technology academically and practically so that it could be better applied to different avenues of microbial community research, such as identifying contamination, environment mixing, or microbial migration.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: SRA database (NCBI) under the BioProject accession number PRJNA713417, with Biosample accession numbers SAMN18274560–SAMN18274616 and SAMN18252387–SAMN18252398. Besides, sequence data of one bovine feces has been deposited in the SRA database (NCBI) under the BioProject accession number PRJNA392724, with biosample number SAMN07305415.

AUTHOR CONTRIBUTIONS

ZSY and HXL designed the experiment. HXL and BBW carried out the field sampling. HXL, BBW, and FN performed the laboratory work. HXL and ZSY analyzed the data and wrote the manuscript. RYL, HXZ, and GW contributed to the revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (grant number 2016YFC0503601) from the Ministry of Science and Technology of the People's Republic of China.

ACKNOWLEDGMENTS

The authors would like to thank Bowen Hu and Xiangyang Wang for their assistance in obtaining the fecal and water samples used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.660368/full#supplementary-material

REFERENCES

Ahmed, W., Hamilton, K. A., Gyawali, P., Toze, S., and Haas, C. N. (2016). Evidence of avian and possum fecal contamination in rainwater tanks as determined by microbial source tracking approaches. *Appl. Environ. Microbiol.* 82, 4379–4386. doi: 10.1128/aem.00892-16

Ahmed, W., Staley, C., Sadowsky, M. J., Gyawali, P., Sidhu, J. P. S., Palmer, A., et al. (2015). Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water. *Appl. Environ. Microbiol.* 81, 7067–7077. doi: 10.1128/aem.02032-15

Ahmed, W., Staley, C., Sidhu, J., Sadowsky, M., and Toze, S. (2017). Amplicon-based profiling of bacteria in raw and secondary treated wastewater from treatment plants across Australia. *Appl. Microbiol. Biotechnol.* 101, 1253–1266. doi: 10.1007/s00253-016-7959-9

Balleste, 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

Baral, D., Speicher, A., Dvorak, B., Admiraal, D., and Li, X. (2018). Quantifying the relative contributions of environmental sources to the microbial community in an urban stream under dry and wet weather conditions. *Appl. Environ. Microbiol.* 84:e00896-18. doi: 10.1128/aem.00896-18

Barbosa, A. E., Fernandes, J. N., and David, L. M. (2012). Key issues for sustainable urban stormwater management. *Water Res.* 46, 6787–6798. doi: 10.1016/j.watres.2012.05.029

Blanch, A. R., Belanche-Munoz, L., Bonjoch, X., Ebdon, J., Gantzer, C., Lucena, F., et al. (2006). Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Appl. Environ. Microbiol.* 72, 5915–5926. doi: 10.1128/aem.02453-05

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Brown, C. M., Mathai, P. P., Loesekann, T., Staley, C., and Sadowsky, M. J. (2019). Influence of library composition on source tracker predictions for community-based microbial source tracking. *Environ. Sci. Technol.* 53, 60–68. doi: 10.1021/acs.est.8b04707

Brown, C. M., Staley, C., Wang, P., Dalzell, B., Chun, C. L., and Sadowsky, M. J. (2017). A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary. *Environ. Sci. Technol.* 51, 8263–8271. doi: 10.1021/acs.est.7b01353

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

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

Chen, H. Y., Li, Y. Z., Sun, W. C., Song, L. T., Zuo, R., and Teng, Y. G. (2020). Characterization and source identification of antibiotic resistance genes in the sediments of an interconnected river-lake system. *Environ. Int.* 137:105538. doi: 10.1016/j.envint.2020.105538

Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 884–890. doi: 10.1093/bioinformatics/bty560

Chidamba, L., and Korsten, L. (2015). Antibiotic resistance in *Escherichia coli* isolates from roof-harvested rainwater tanks and urban pigeon faeces as the likely source of contamination. *Environ. Monit. Assess.* 187:405. doi: 10.1007/s10661-015-4636-x

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

De Kwaadsteniet, M., Dobrowsky, P. H., Van Deventer, A., Khan, W., and Cloete, T. E. (2013). Domestic rainwater harvesting: microbial and chemical water quality and point-of-use treatment systems. *Water Air Soil Pollut.* 224:1629. doi: 10.1007/s11270-013-1629-7

Dubinsky, E. A., Butkus, S. R., and Andersen, G. L. (2016). Microbial source tracking in impaired watersheds using PhyloChip and machine-learning classification. *Water Res.* 105, 56–64. doi: 10.1016/j.watres.2016.08.035

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

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

Fong, T. T., Griffin, D. W., and Lipp, E. K. (2005). Molecular assays for targeting human and bovine enteric viruses in coastal waters and their application for library-independent source tracking. *Appl. Environ. Microbiol.* 71, 2070–2078. doi: 10.1128/aem.71.4.2070-2078.2005

Haramoto, E., and Osada, R. (2018). Assessment and application of host-specific Bacteroidales genetic markers for microbial source tracking of river water in Japan. *PLoS One* 13:e0207727. doi: 10.1371/journal.pone.0207727

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

He, X. W., Liu, P., Zheng, G. L., Chen, H. M., Shi, W., Cui, Y. B., et al. (2016). Evaluation of five microbial and four mitochondrial DNA markers for tracking human and pig fecal pollution in freshwater. *Sci. Rep.* 6:35311. doi: 10.1038/srep35311

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., et al. (2016). Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Res.* 93, 242–253. doi: 10.1016/j.watres.2016.02.029

Jenkins, M. W., Tiwari, S., Lorente, M., Gichaba, C. M., and Wuertz, S. (2009). Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res.* 43, 4956–4966. doi: 10.1016/j.watres.2009.07.028

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. *Water Res.* 41, 3701–3715. doi: 10.1016/j.watres.2007.06.037

Kim, G., Bae, J., Kim, M. J., Kwon, H., Park, G., Kim, S.-J., et al. (2020). Delayed establishment of gut microbiota in infants delivered by cesarean section. *Front. Microbiol.* 11:2099. doi: 10.3389/fmicb.2020.02099

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., et al. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–765. doi: 10.1038/nmeth.1650

Kostyla, C., Bain, R., Cronk, R., and Bartram, J. (2015). Seasonal variation of fecal contamination in drinking water sources in developing countries: a systematic review. *Sci. Total Environ.* 514, 333–343. doi: 10.1016/j.scitotenv.2015.01.018

Layton, B. A., Cao, Y. P., Ebentier, D. L., Hanley, K., Balleste, E., Brandao, J., et al. (2013). Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res.* 47, 6897–6908. doi: 10.1016/j.watres.2013.05.060

Lee, C., Marion, J. W., and Lee, J. (2013). Development and application of a quantitative PCR assay targeting *Catellicoccus marimammalium* for assessing gull-associated fecal contamination at Lake Erie beaches. *Sci. Total Environ.* 454, 1–8. doi: 10.1016/j.scitotenv.2013.03.003

Li, D., Van De Werfhorst, L. C., Dunne, T., Devarajan, N., Ayala, T. G., and Holden, P. A. (2020). Surf zone microbiological water quality following emergency beach nourishment using sediments from a catastrophic debris flow. *Water Res.* 176:115733. doi: 10.1016/j.watres.2020.115733

Liang, H. X., Yu, Z. S., Ndayisenga, F., Liu, R. Y., Zhang, Y. M., Zhang, H. X., et al. (2020). A combination of mitochondrial DNA markers Ckmito and ND5-CD is recommended as the most reliable indicator for microbial source tracking to identify faecal pollution from poultry in China. *Ecol. Indic.* 115:106334. doi: 10.1016/j.ecolind.2020.106334

Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Marti, R., Yun, Z., Tien, Y. C., Lapen, D. R., and Topp, E. (2013). Assessment of a new Bacteroidales marker targeting North American beaver (*Castor canadensis*) fecal pollution by real-time PCR. *J. Microbiol. Methods* 95, 201–206. doi: 10.1016/j.mimet.2013.08.016

Mieszkin, S., Furet, J. P., Corthier, G., and Gourmelon, M. (2009). Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045–3054. doi: 10.1128/aem.02343-08

Moore, D. F., Harwood, V. J., Ferguson, D. M., Lukasik, J., Hannah, P., Getrich, M., et al. (2005). Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed. *J. Appl. Microbiol.* 99, 618–628. doi: 10.1111/j.1365-2672.2005.02612.x

Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., et al. (2014). Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res.* 21, 217–227. doi: 10.1093/dnares/dst052

Newton, R. J., Bootsma, M. J., Morrison, H. G., Sogin, M. L., and McLellan, S. L. (2013). A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microb. Ecol.* 65, 1011–1023. doi: 10.1007/s00248-013-0200-9

Nshimyimana, J. P., Cruz, M. C., Thompson, R. J., and Wuertz, S. (2017). Bacteroidales markers for microbial source tracking in Southeast Asia. *Water Res.* 118, 239–248. doi: 10.1016/j.watres.2017.04.027

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., et al. (2015). Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* 502, 462–470. doi: 10.1016/j.scitotenv.2014.09.040

O'Dea, C., Zhang, Q., Staley, C., Masters, N., Kuballa, A., Fisher, P., et al. (2019). Compositional and temporal stability of fecal taxon libraries for use with SourceTracker in sub-tropical catchments. *Water Res.* 165:114967. doi: 10.1016/j.watres.2019.114967

Ohad, S., Ben-Dor, S., Prilusky, J., Kravitz, V., Dassa, B., Chalifa-Caspi, T., et al. (2016). The development of a novel qPCR assay-set for identifying fecal contamination originating from domestic fowls and waterfowl in Israel. *Front. Microbiol.* 7:145. doi: 10.3389/fmicb.2016.00145

Qi, J., and Chen, L. (2012). Application of storm water utilization and water quality improvement. *Beijing Water* 3, 54–57.

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Astrom, J., et al. (2013). Performance characteristics of qPCR assays targeting human- and ruminant-associated Bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* 47, 8548–8556. doi: 10.1021/es304367t

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

Roguet, A., Eren, A. M., Newton, R. J., and McLellan, S. L. (2018). Fecal source identification using random forest. *Microbiome* 6:185. doi: 10.1186/s40168-018-0568-3

Shenhav, L., Thompson, M., Joseph, T. A., Briscoe, L., Furman, O., Bogumil, D., et al. (2019). FEAST: fast expectation-maximization for microbial source tracking. *Nat. Methods* 16, 627–632. doi: 10.1038/s41592-019-0431-x

Sidhu, J. P. S., Hodgers, L., Ahmed, W., Chong, M. N., and Toze, S. (2012). Prevalence of human pathogens and indicators in stormwater runoff in Brisbane. *Australia. Water Res.* 46, 6652–6660. doi: 10.1016/j.watres.2012.03.012

Smith, A., Sterba-Boatwright, B., and Mott, J. (2010). Novel application of a statistical technique, Random Forests, in a bacterial source tracking study. *Water Res.* 44, 4067–4076. doi: 10.1016/j.watres.2010.05.019

Somnark, P., Chyerochana, N., Mongkolsuk, S., and Sirikanchana, K. (2018). Performance evaluation of Bacteroidales genetic markers for human and animal microbial source tracking in tropical agricultural watersheds. *Environ. Pollut.* 236, 100–110. doi: 10.1016/j.envpol.2018.01.052

Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V. J., Brown, C. M., et al. (2018). Application of SourceTracker for accurate identification of fecal pollution in recreational freshwater: a double-blinded study. *Environ. Sci. Technol.* 52, 4207–4217. doi: 10.1021/acs.est.7b05401

Stoeckel, D. M., and Harwood, V. J. (2007). Performance, design, and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* 73, 2405–2415. doi: 10.1128/aem.02473-06

U.S. Environmental Protection Agency, (2005). *Microbial Source Tracking Guide Document EPA/600/R-05/064*. Washington DC: U.S. Environmental Protection Agency.

Vadde, K. K., McCarthy, A. J., Rong, R., and Sekar, R. (2019). Quantification of microbial source tracking and pathogenic bacterial markers in water and sediments of Tiaoxi River (Taihu Watershed). *Front. Microbiol.* 10:699. doi: 10.3389/fmicb.2019.00699

Xu, Y. F., Li, Z., Liu, R. Y., Liang, H. X., Yu, Z. S., and Zhang, H. X. (2020). Validation of Bacteroidales-based microbial source tracking markers for pig fecal pollution and their application in two rivers of North China. *Appl. Environ. Microbiol.* 14:67. doi: 10.1007/s11783-020-1246-9

Zhang, Y., Wu, R. R., Lin, K. R., Wang, Y. S., and Lu, J. Q. (2020). Performance of host-associated genetic markers for microbial source tracking in China. *Water Res.* 175:115670. doi: 10.1016/j.watres.2020.115670

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.

Copyright © 2021 Liang, Yu, Wang, Ndayisenga, Liu, Zhang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 June 2021
doi: 10.3389/fmicb.2021.673190

Bather Shedding as a Source of Human Fecal Markers to a Recreational Beach

*Dong Li[1], Laurie C. Van De Werfhorst[1], Brandon Steets[2], Jared Ervin[2], Jill L. S. Murray[3], Naresh Devarajan[1] and Patricia A. Holden[1]**

[1] Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA, United States, [2] Geosyntec Consultants, Santa Barbara, CA, United States, [3] Creeks Division, Department of Parks and Recreation, Santa Barbara, CA, United States

OPEN ACCESS

Edited by:
Michèle Gourmelon, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Christopher Sinigalliano, Atlantic Oceanographic and Meteorological Laboratory (NOAA), United States
Hyatt Green, Suny College of Environmental Science and Forestry, United States
Patrick C. Hanington, University of Alberta, Canada

***Correspondence:**
Patricia A. Holden
holden@bren.ucsb.edu

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *27 February 2021*
Accepted: *30 April 2021*
Published: *25 June 2021*

Citation:
Li D, Van De Werfhorst LC, Steets B, Ervin J, Murray JLS, Devarajan N and Holden PA (2021) Bather Shedding as a Source of Human Fecal Markers to a Recreational Beach. Front. Microbiol. 12:673190. doi: 10.3389/fmicb.2021.673190

Microbial source tracking (MST) can identify and locate surf zone fecal indicator bacteria (FIB) sources. However, DNA-based fecal marker results may raise new questions, since FIB and DNA marker sources can differ. Here, during 2 years of summertime (dry season) MST for a Goleta, California recreational beach, surf zone FIB were mainly from gulls, yet low level human-associated DNA-based fecal marker (HF183) was detected in 25 and 14% of surf zone water samples, respectively. Watershed sources were hypothesized because dry weather creek waters had elevated FIB, and runoff-generating rain events mobilized human (and dog) fecal markers and *Salmonella* spp. into creeks, with human marker HF183 detected in 40 and 50% of creek water samples, dog markers detected in 70 and 50% of samples, and *Salmonella* spp. in 40 and 33.3% of samples, respectively over 2 years. However, the dry weather estuary outlet was bermed in the first study year; simultaneously, creek fecal markers and pathogens were lower or similar to surf zone results. Although the berm breached in the second year, surf zone fecal markers stayed low. Watershed sediments, intertidal beach sands, and nearshore sediments were devoid of HF183 and dog-associated DNA markers. Based on dye tests and groundwater sampling, beach sanitary sewers were not leaking; groundwater was also devoid of HF183. Offshore sources appeared unlikely, since FIB and fecal markers decreased along a spatial gradient from the surf zone toward nearshore and offshore ocean waters. Further, like other regional beaches, surf zone HF183 corresponded significantly to bather counts, especially in the afternoons when there were more swimmers. However, morning detections of surf zone HF183 when there were few swimmers raised the possibility that the wastewater treatment plant (WWTP) offshore outfall discharged HF183 overnight which transported to the surf zone. These findings support that there may be lowest achievable limits of surf zone HF183 owing to several chronic and permanent, perhaps diurnal, low concentration sources.

Keywords: HF183 bacteroides, bathers, outfall, diurnal sources, chronic

INTRODUCTION

Fecal contamination of recreational coastal waters is a global public health concern. Worldwide there are annually more than 120 million gastrointestinal and 50 million respiratory illnesses estimated to be associated with activities in contaminated coastal waters (Shuval, 2003). To determine microbial water quality as related to public health, routine monitoring of fecal indicator bacteria (FIB) including total coliforms, fecal coliforms, *Escherichia coli*, and enterococci may be mandated, such as by Assembly Bill 411 (AB411) in California (Sikich et al., 2018), with swimming advisories and beach closures implemented accordingly. However, by epidemiological studies, FIB may not relate to human illness (Colford et al., 2007; Arnold et al., 2013), as FIB may originate from animal feces that are of low risk to human health due to host specificity of pathogens, particularly enteric viruses (Sinclair et al., 2009). Fecal contamination in recreational waters is in fact frequently attributed to gulls and canines, with management practices implemented and FIB successfully diminished (Wright et al., 2009; Converse et al., 2012; Goodwin et al., 2016). Further, FIB can survive and grow in diverse environments including soil, sediments, beach sand, wrack and aquatic vegetation (Field and Samadpour, 2007; Harwood et al., 2014). As such, microbial source tracking (MST) is needed to discern FIB sources that should be remedied for protecting public health.

Modern MST typically discovers FIB sources by analyzing field samples for molecular markers that target genes of host-coevolved or -associated bacteria, using quantitative PCR (qPCR) or droplet digital PCR (ddPCR) (Field and Samadpour, 2007; Boehm et al., 2013, 2015; Harwood et al., 2014; Mayer et al., 2018). Although host-associated microorganisms and related DNA markers decay in marine waters (Green et al., 2011; Ahmed et al., 2014; Mattioli et al., 2017), markers can sensitively and specifically reveal hosts' wastes, and allow for quantitative microbial risk assessment (QMRA) (Boehm et al., 2015; Brown et al., 2017b). As such, host-associated fecal markers such as the human fecal marker HF183 targeting bacteria closely related to *Bacteroides dorei* have been widely applied in numerous MST studies (Bernhard and Field, 2000; Field and Samadpour, 2007; Boehm et al., 2013, 2015; Harwood et al., 2014; Mayer et al., 2018), with their use now mainstream as an EPA-validated method.

However, MST at urban or suburban beaches is complex, owing to numerous potential human fecal sources including sanitary sewers and storm drains (U.S. Environmental Protection Agency, 2004; Griffith et al., 2013; Cao et al., 2017). Leaking sewers and sewer overflows can occur in coastal areas (Sikich et al., 2018). Urban stormwater runoff can transport FIB and pathogens directly into coastal waters (U.S. Environmental Protection Agency, 2004). Watershed creeks can transport fecal contamination to beach waters, including through coastal lagoons or transitional waters (Ervin et al., 2014; Riedel et al., 2015). Discerning the numerous proximate and distal sources potentially contributing to surf zone fecal contamination therefore requires systematic sampling within study designs, as per MST guidelines in the California Microbial Source Identification Manual (Griffith et al., 2013). Additionally, bather shedding can contribute FIB and pathogens such as *Staphylococcus aureus* directly to marine water (Elmir et al., 2007, 2009; Plano et al., 2011), and thus acts as a non-point source of fecal contamination in recreational coastal waters.

Still, questions remain at many beaches, as evidenced by the low but chronic presence, with few actionable explanations, of human fecal markers along the California coastline (Santoro and Boehm, 2007; McQuaig et al., 2012; Russell et al., 2013; Riedel et al., 2015; Cao et al., 2017; Jennings et al., 2018). FIB surf zone exceedances can be rare in California (Sikich et al., 2018), yet surf zone HF183 markers may persist at low levels, even after management practices are implemented to address diverse potential fecal sources such as sewage spills, and contaminated drains, creeks, and rivers (Ervin et al., 2014; Goodwin et al., 2016; San Diego Regional Water Quality Control Board, 2017; Los Angeles Regional Water Quality Control Board, State of California, 2018; Sikich et al., 2018). Resolving lingering human marker-related questions after FIB sources are identified is needed to fully understand potential human health risks, and adopt additional management actions accordingly.

Here, a popular California recreational beach located in Goleta, CA with approximately 1.5 million visitors annually was studied to determine the origins of historically elevated surf zone water FIB[1]. A 2-year comprehensive MST investigation involved hypothesizing fecal sources, then sampling and analyzing FIB, several host fecal markers, and human pathogens (collectively abbreviated as "FMPs," for FIB, fecal markers, and pathogens). All conceivable sources and processes were investigated including watersheds, sanitary sewers, groundwater, intertidal beach sands, nearshore sediments, a wastewater treatment plant (WWTP) offshore outfall, and natural regional background levels. Impacts of surf zone human recreation during holidays or high visitation weekends and at different times of the day were also evaluated for their relationship to surf zone microbial water quality. By examining all potential fecal contamination sources and transmission routes, the origins of FIB contamination to Goleta Beach (GB) were revealed, together with the probable sources of low human fecal markers. The results herein shed light on a longstanding question regarding chronic surf zone HF183, and contribute to broadly understanding apparent fecal contamination in recreational beaches when most obvious or actionable sources are ruled out.

MATERIALS AND METHODS

Site Description, Sampling Locations, and Hypotheses

GB, located in Goleta, CA (latitude 34.4172186, and longitude −119.8295828) in the public GB Park, is popular for swimming and fishing, and near the University of California, Santa Barbara

[1]http://brc.healthebay.org/

campus point, which is popular for surfing. The beach and its location were recently described (Li et al., 2020), owing to GB's use as a disposal site for contaminated sediments following a regional debris flow that occurred after the MST study herein. Some results from the study herein were published to establish the 2017 background surf zone water quality for comparison to the study of 2018 sediment disposal effects (Li et al., 2020). However, the major findings of this study, and most data, are heretofore unpublished.

The watersheds upstream of GB, including Atascadero Creek, San Jose Creek, and San Pedro Creek, merge in an estuarine slough that discharges into the GB surf zone between sampling sites G01 and G02 when the beach berm is breached (**Supplementary Figure 1**). Historical FIB exceedances in GB surf zone waters and the upstream watershed waters including the slough were available prior to this study (see text footnote 1). Geographic information system (GIS) layers were compiled to display creek locations, and relevant infrastructure data for beach sanitary sewers and storm drains, septic systems, and hardscape (**Supplementary Figure 2**). Field reconnaissance at the beaches and into upstream watersheds revealed visible potential fecal sources including groups of shorebirds and wrack on the beach, dogs and wildlife, homeless encampments in the watersheds, sanitary sewer manholes, storm drains, and restrooms. Based on background information and field reconnaissance, the following potential summertime sources and routes of fecal contamination to the surf zone were investigated: gulls, dogs, watershed waters and sediments, horses along Atascadero creek, runoff-generating rainfall events, intertidal beach sands, infrastructure including bathrooms, sewer lines and pump stations at the beach, nearshore and offshore marine water, treated WWTP effluent discharging through an ocean outfall (**Supplementary Figure 2**), and nearshore marine sediments. Bather shedding on holiday or high visitation weekends and at different times of the day was also studied, with the numbers of bathers in the surf zone and people on the sand recorded at the time of sampling. Regional background levels of FMPs were determined in the surf zone and creek locations at a reference beach, Arroyo Hondo, which is a natural preserve with little human activity as described elsewhere (Li et al., 2021). Samples of water from the surf zones, watersheds, groundwaters, nearshore, offshore, and WWTP ocean outfall, as well as intertidal beach sands, sediments from the watersheds and nearshore were collected during the AB411 (mostly dry) season (April–October) during 2016–2017 for the analyses of FMPs (**Supplementary Figures 1, 3**). Dry weather was defined as < 0.1" of rainfall in the preceding 72 h. Rain events that occurred during the AB411 season were sampled if predicted to be ≥ 0.2". The details of the samples and results are in **Supplementary Tables 1–7**.

Dye Studies

Dye studies were performed to assess sanitary infrastructure integrity including sanitary sewer lines, laterals, and pressurized force mains as potential sources of fecal contamination, and to determine if any leaking sewers were impacting groundwater or surface waters through high velocity groundwater or preferential flow pathways. A non-toxic fluorescent dye (2.5% Rhodamine WT dye, Cole-Parmer, Vernon Hills) was flushed down toilets in the five bathrooms and added directly to the three recreational vehicle (RV) sewer connections and two sewage lift stations located at GB Park. A total of 116 groundwater, 84 surf zone, and 63 slough water samples were collected from 6 groundwater, 4 surf zone, and 3 slough sampling locations after the addition of dye for dye testing (**Supplementary Figure 4**). One sample from each groundwater well was also collected and analyzed for FMPs. The details of studied infrastructure, temporary groundwater monitoring wells, dye addition, water sampling, and dye analysis are in the SI Methods.

Sample Collection, Physicochemical Characterization, and FIB Analyses

The water sampling procedures for the surf zone, watershed, groundwater, nearshore, offshore, and WWTP ocean outfall diffuser are described in the SI Methods. Dissolved oxygen, electrical conductivity, and temperature of water samples were recorded in the field using an HQ40d multiparameter meter (Hach, Loveland, OH). The sampling details for watershed sediments, intertidal sands, and nearshore sediments, and the analyses of particle grain sizes, moisture, and total organic content are described in the SI Methods. Total coliform (TC), *E. coli* (EC) and enterococci (ENT) in water and solid samples were quantified using the IDEXX Quanti-Tray/2000 method, following the manufacturer's protocols (IDEXX Laboratories, Westbrook, MA), with details in the SI Methods. FIB exceedances were based on California single sample surf zone criteria for TC (10,000 MPN/100 ml), EC (400 MPN/100 ml), and ENT (104 MPN/100 ml) (Sikich et al., 2018).

DNA Extraction, PCR, qPCR, and ddPCR

The approaches for water and solid sample processing and DNA extraction have been described elsewhere (Li et al., 2020), and are detailed in the SI Methods. Briefly, DNA was extracted from water samples using the DNeasy PowerWater Kit (Qiagen, Carol Stream, IL) or the RNeasy PowerWater Kit (Qiagen) following the manufacturer's protocols. The DNeasy PowerSoil Kit (Qiagen, Carol Stream, IL) was used to extract DNA from sand and sediment samples. DNA concentrations were quantified using the Quant-iT dsDNA Broad-Range Assay Kit (Invitrogen, Carlsbad, CA) on Cytation3 Cell Imaging Multi-Mode Reader (BioTek Instruments, Inc., Winooski, VT).

The presence of human, dog, and gull fecal materials was determined using qPCR, and the presence of horse feces using conventional PCR, with HF183 (Bernhard and Field, 2000; Green et al., 2014) and HumM2 (Shanks et al., 2009) as human-associated fecal markers, DogBact (Dick et al., 2005; Sinigalliano et al., 2010) as the dog marker, Gull2Taqman (Lu et al., 2008; Sinigalliano et al., 2010) as the gull fecal marker, and HoF597 (Dick et al., 2005) as the horse marker. These host-specific markers were selected based on their performances as thoroughly evaluated in a previous study (Boehm et al., 2013). A TaqMan version of the HF183 assay that incorporates an internal amplification control was used as the primary assay

to detect human fecal waste (Green et al., 2014). The HumM2 qPCR assay is considered more specific but less sensitive to human feces compared to the HF183 assay (Boehm et al., 2013), and was only performed for HF183 positive samples in this study, and detected in 19.1% of HF183 positive samples (**Supplementary Tables 1, 2, 5, 6**). The Entero1A marker was used to quantify ENT in units of cell equivalents (c.eq.) per 100 ml for comparison to culture-based measurements by assuming a ribosomal RNA copy number as 6 for *Enterococcus* (Oana et al., 2002; Haugland et al., 2012). The *ttr* gene was used to quantify *Salmonella* spp. bacteria (Malorny et al., 2004). Sample inhibition was assessed using an internal amplification control (IAC) incorporated in the HF183 TaqMan qPCR assay (Green et al., 2014). All qPCR assays were performed using the TaqMan Environmental Master Mix 2.0 (Life Technologies, Grand Island, NY) on a Bio-Rad CFX96 Real-Time PCR Detection System (Hercules, CA). Synthesized plasmid DNA containing qPCR targeted sequences were serially diluted to generate standard curves for all qPCR assays. All samples and standards were analyzed in triplicate with triplicate no-template controls included for each plate. Human adenovirus was quantified using ddPCR with a Bio-Rad QX200 Droplet Digital PCR System (Hercules, CA) as previously published (Steele et al., 2018). The conventional PCR, qPCR and ddPCR procedures are described in the SI Methods. Samples with two or more replicates amplifying within the range of the standard curve were considered to be within the range of quantification (ROQ) and were quantified. Samples with two or more replicates amplifying below the lowest standard were considered detected but not quantifiable (DNQ), and samples with one or zero replicates amplifying were considered not detected (ND), as described previously (Ervin et al., 2014).

Bather Shedding

For studying bather shedding as a potential source of FMPs, surf zone water samples were collected on high visitation weekends or holidays (Jul 3rd, Aug 12th, and Sep 4th, 2017) during the morning and afternoon. Surf zone waters were also sampled at site G03 at different times of the day (late afternoon, and again in the early morning and mid-afternoon of the following day) on five occasions for evaluating if water defecation was occurring by potential campers overnight (**Supplementary Table 6**). During bather shedding studies, the counts of people were grouped into people "in the water" including swimmers or anyone recreating in the surf zone (bather), and people "on the sand," meaning people recreating but not in the water (i.e., walking, sitting on the sand, and exercising).

Statistical Analyses

Statistical analyses included Wilcoxon tests (Mann-Whitney for two categories, or Kruskal-Wallis with Steel-Dwass pairwise comparisons for three or more categories) and Spearman's rho rank correlation performed using JMP10 (SAS, Cary, NC). Over or under range values were adjusted prior to statistical analysis to be above the highest, and below the lowest, quantified values, respectively. FIB values were treated as follows, $< 10 = 0$ (log scale), and $> 24{,}196 = 25{,}000$. For qPCR assay results, DNQ values were set to 1.8 (log scale) and ND values were set to 1.3 (log scale).

RESULTS

FMPs in Surf Zone Waters

Surf zone waters were sampled at 5 GB sites in 2016 and 2017 (**Supplementary Figure 1**). During dry weather ($n = 90$ samples), there were no exceedances of single sample AB411 criteria for TC or ENT, but 17.5 and 18% of the samples exceeded EC criteria in 2016 and 2017, respectively (**Supplementary Table 1**). Gull markers were detected in all surf zone water samples, with 97.5 and 96% of samples at quantifiable levels in 2016 and 2017, respectively; dog markers were also detected in 67.5 and 52% of water samples (in 2016, and 2017, respectively) at DNQ or quantifiable levels (**Figure 1**). The human marker HF183 was detected in 25 and 14% of surf zone water samples in 2016 and 2017, respectively, all at DNQ levels (**Figure 1**). *Salmonella* spp. bacteria were not found in any sample, and human adenovirus was detected in one sample across the 2 years (**Supplementary Table 1**). In another regional study, the potential for non-specific detection of HF183 was ruled out, since there were no detections when performing the assay using DNA extracted from gull feces (SI Results).

Although the estuarine slough was bermed with sand and thus closed to the surf zone in 2016, it breached and flowed into the surf zone in 2017; still, the breaching did not appear to affect surf zone water quality as FMPs were similar across 2016 and 2017 (**Supplementary Table 1** and **Figure 1**) except for higher concentrations of TC in 2017 (Mann-Whitney, $p < 0.001$, $n = 90$). There was also no significant difference in FIB or host fecal marker concentrations across the 5 surf zone sampling locations across both years (**Supplementary Table 1** and **Figure 1**; Kruskal-Wallis, all $p > 0.05$, $n = 90$), except for ENT by Entero1A ($p < 0.001$). When site to site variations were evaluated within each study year, FMPs did not vary among the 5 surf zone locations in 2016, and only TC and ENT by Entero1A showed significant variance among the 5 locations in 2017 (both $p < 0.03$, $n = 50$). Taken together, these results indicate that the slough breaching did not affect fecal markers in the surf zone. Across the 2016 and 2017 dry weather results, the surf zone concentrations of TC, EC, and ENT by Entero1A were closely correlated (Spearman's ρ ranging from 0.24 to 0.76, all $p < 0.03$, $n = 90$). Further, the FIB concentrations correlated with gull marker concentrations (ρ ranging from 0.28 to 0.46, all $p < 0.01$, $n = 90$), but not with dog or human marker HF183. Such correlation and the prevalence of gull markers indicated that gulls were the major sources of FIB in the surf zone during dry weather.

For the three rain events occurring during the study periods ($n = 9$ samples), exceedances of single sample AB411 criteria in 2016 were recorded once (17%) for TC and ENT in surf zone waters, and for 33% of EC measurements; there were no FIB criteria exceedances in 2017 (**Supplementary Table 1**). The correlations among gull markers, ENT, and ENT by Entero1A were stronger for wet weather samples (ρ ranging from 0.75 to

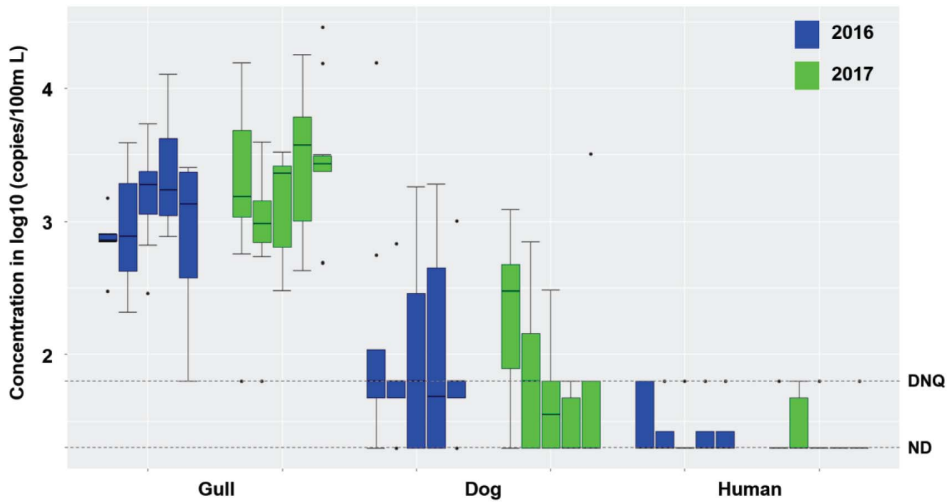


FIGURE 1 | Boxplots of gull, dog and HF183 human fecal marker concentrations in surf zone waters collected from 5 sites (G01-G05, from left to right in each set of boxes) of GB in 2016 (8 samples per site and 40 samples in total) and 2017 (10 samples per site and 50 samples in total) during dry weather (**Supplementary Table 1**). DNQ (1.8 on log10 scale) and ND (1.3 on log10 scale) levels are shown in the figure as dotted lines.

0.89, all $p < 0.05$, $n = 9$), indicating that rain events transported gull fecal deposits into surf zone waters.

FMPs in Watershed Waters and Sediments

Inland waters contained much higher concentrations of FIB under dry weather conditions compared to surf zone waters, with 70 and 75% of TC, 65 and 57.5% of EC, and 22.5 and 30% of ENT exceeding AB411 single sample criteria in 2016 and 2017, respectively (**Supplementary Table 2**, Mann-Whitney test, all $p < 0.001$, $n = 80$). Gull markers were present in 7.5 and 52.5% of inland water samples in 2016 (40 samples) and 2017 (40 samples), respectively, and dog markers were present in 2.5 and 27.5% water samples in 2016 and 2017, respectively (**Figure 2**). The concentrations of gull and dog markers were both significantly lower in inland waters than in surf zone waters (**Figure 2**; Mann-Whitney test, both $p < 0.0001$, $n = 170$). Human marker HF183 was present in 5 and 7.5% of inland water samples in 2016 and 2017, respectively, while *Salmonella* spp. bacteria were not present in any dry weather sample, and human adenovirus was present in only one sample (2.5%) in 2016 and one sample (2.5%) in 2017, respectively (**Supplementary Table 2**). Horse markers were not detected in water samples from the Atascadero Creek and lower slough sites G07 through G09 (**Supplementary Figure 1**).

During dry weather, inland waters had significantly higher concentrations of gull and dog markers in 2017 relative to 2016 (Mann-Whitney test, both $p < 0.05$, $n = 80$), while human marker HF183 and FIB did not vary significantly (**Supplementary Table 2**). FIB concentrations significantly differed across the sampling locations including the slough, and upstream into Atascadero Creek, San Jose Creek, and San Pedro Creek (Kruskal-Wallis, all $p < 0.002$, $n = 80$), with the highest average concentrations in San Jose Creek and San Pedro Creek waters, followed by the slough, and Atascadero Creek. Gull marker concentrations also significantly differed across the slough and the three creeks (**Figure 2**; Kruskal-Wallis, $p = 0.038$, $n = 80$), with the slough containing the highest average concentrations. Dog and human marker HF183 detections did not significantly differ across the slough and creeks (**Figure 2**). Although the slough contained significantly higher FIB concentrations than surf zone waters (**Supplementary Tables 1, 2**, Mann-Whitney test, all $p < 0.01$, $n = 80$), there were significantly fewer gull and dog markers in the slough than surf zone waters (**Supplementary Tables 1, 2**; Mann-Whitney test, both $p < 0.01$, $n = 80$), and the average human marker HF183 detections appeared lower than surf zone waters although not significantly (**Supplementary Tables 1, 2**, $p = 0.2$). These results indicated that the slough was not the source of host fecal markers in surf zone waters. Overall, significant correlations were found between the concentrations of TC and EC (Spearman's $\rho = 0.46$, $p < 0.00001$, $n = 80$), ENT and ENT by Entero1A ($\rho = 0.67$, $p < 0.00001$, $n = 80$), and gull and dog markers ($\rho = 0.42$, $p = 0.0003$, $n = 80$) in inland water samples.

Rain events resulted in increased FMPs in inland waters, with gull markers detected in 60 and 83.3% of samples, dog markers in 70 and 50% of samples, and human marker HF183 in 40 and 50% of samples in 2016 ($n = 10$ samples) and 2017 ($n = 6$ samples), respectively, mostly at quantifiable levels (**Figure 2** and **Supplementary Table 2**). *Salmonella* spp. were present in 40 and 33.3% of water samples in 2016 and 2017, respectively, and human adenovirus was found in 33.3% of water samples in 2017. The concentrations of host fecal markers and *Salmonella* spp. were higher during rain events compared to dry weather conditions (Mann-Whitney test, all $p < 0.02$, $n = 96$), and dog markers, human marker HF183, and *Salmonella* spp. were correlated with each other (Spearman's ρ ranging from 0.65 to 0.84, $p < 0.01$, $n = 16$). These results indicated that most of the assayed fecal markers, and *Salmonella* spp., were mobilized

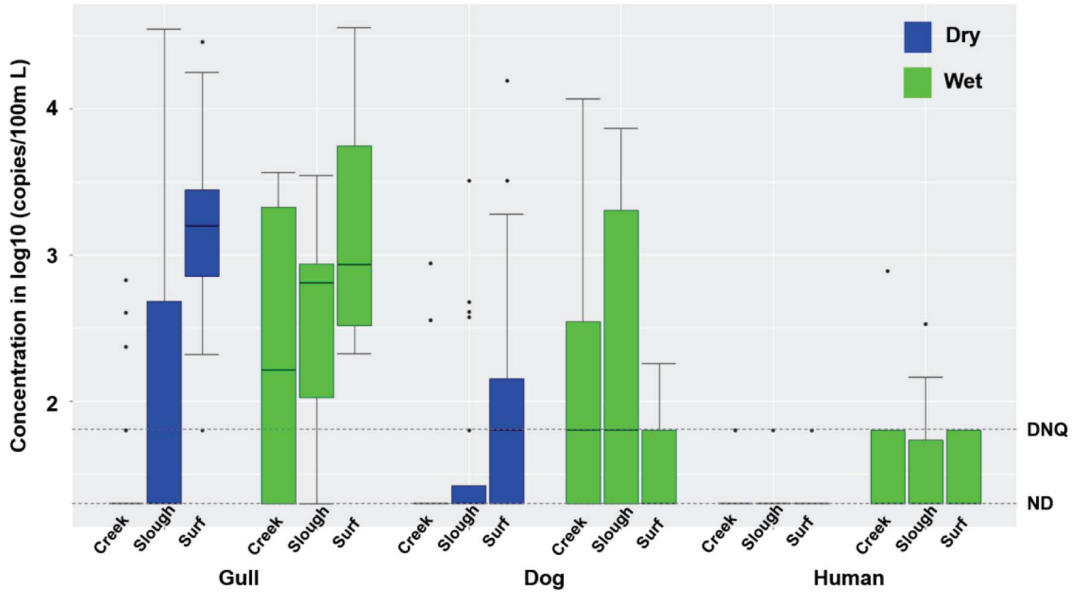


FIGURE 2 | Boxplots of gull, dog and HF183 human fecal marker concentrations in creek, slough, and surf zone (from left to right) water samples collected from watershed and surf zones of GB in 2016 and 2017 during dry and wet weathers (**Supplementary Tables 1, 2**). The number of creek, slough, and surf zone samples was 40, 40, and 90 during dry wet weather, respectively, and 9, 7, and 9 during wet weather, respectively. DNQ (1.8 on log10 scale) and ND (1.3 on log10 scale) levels were shown in the figure as dotted lines.

with fecal deposits into inland waters during runoff-generating rainfall events. The only exceptions were horse markers which were not detected. There was mostly no significant interannual difference for FMPs across 2016 and 2017, except that EC was higher in 2016.

In contrast to sampled waters, watershed sediments contained much lower levels of fecal markers, with no human HF183 or dog markers detected in any sediment sample, and gull markers detected at 2 locations in the slough (G06/G06A and G13 in **Supplementary Figure 1** and **Supplementary Table 3**). FIB were quantifiable in many samples, which is common in freshwater sediments (Pachepsky and Shelton, 2011).

Studies of Beach Sands and Infrastructure at GB

In intertidal beach sands (**Supplementary Figure 1**), FIB concentrations were very low or non-detectable, and human HF183 or dog markers were not detected (**Supplementary Table 3**). Gull markers were present in 6.7% (1 of 15) and 13.3% (2 of 15) of sand samples in 2016 and 2017, respectively, and human adenovirus was found in 1 sand sample in 2016. There were no above background levels of dye detected in groundwater, surf zone, and slough water samples collected after the addition of dye to infrastructure including the bathrooms, RV sewer connections, and sewage lift stations at GB (**Supplementary Figures 4–7**). Furthermore, no human marker HF183, dog markers, *Salmonella* spp., or human adenovirus were detected in any groundwater samples collected from groundwater wells; gull markers were present in 50% of groundwater samples (**Supplementary Table 4**).

Nearshore and Offshore Waters, Sediments, and Treated Wastewater Effluent

Synchronous sampling of surf zone and, from watercraft, nearshore and offshore waters in both 2016 and 2017 (**Supplementary Figure 3**) revealed decreasing concentrations of FIB from the surf zone to nearshore and further offshore (**Supplementary Table 5**, Kruskal-Wallis, all $p < 0.03$, $n = 96$). This decreasing spatial trend was also observed for gull markers in both 2016 and 2017 (both $p < 0.0001$), and dog markers in 2017 ($p < 0.03$), but not for human marker HF183 in either 2016 or 2017, possibly due to low detection frequencies (**Figure 3**).

When combining the 2016 and 2017 data, the FIB, gull and dog fecal marker concentrations decreased spatially from the surf zone to nearshore and further offshore (**Figure 3**; Kruskal-Wallis, all $p < 0.01$, $n = 96$). Human marker HF183 also decreased in the same spatial pattern, with detection ratios of 10, 8, and 6% in the surf zone, nearshore, and offshore, respectively (**Figure 3**). The FIB or fecal marker concentrations within the 5 surf zone sampling sites, 3 nearshore sites, or 4 offshore sites were mostly spatially invariant (Kruskal-Wallis, all $p > 0.05$), except for the gull marker within the 5 surf zone sites ($p = 0.047$, $n = 40$). Overall, FIB and ENT by Entero1A concentrations across all surf zone, nearshore, and offshore water samples correlated significantly with each other (Spearman's ρ ranging from 0.36 to 0.73, $p < 0.001$, $n = 96$), and further with gull markers (ρ ranging from 0.23 to 0.66, $p < 0.03$, $n = 96$). *Salmonella* spp. and human adenovirus were not detected in any surf zone, nearshore, or offshore sample of either 2016 or 2017. No significant interannual variation was observed for any host fecal marker across 2016 and

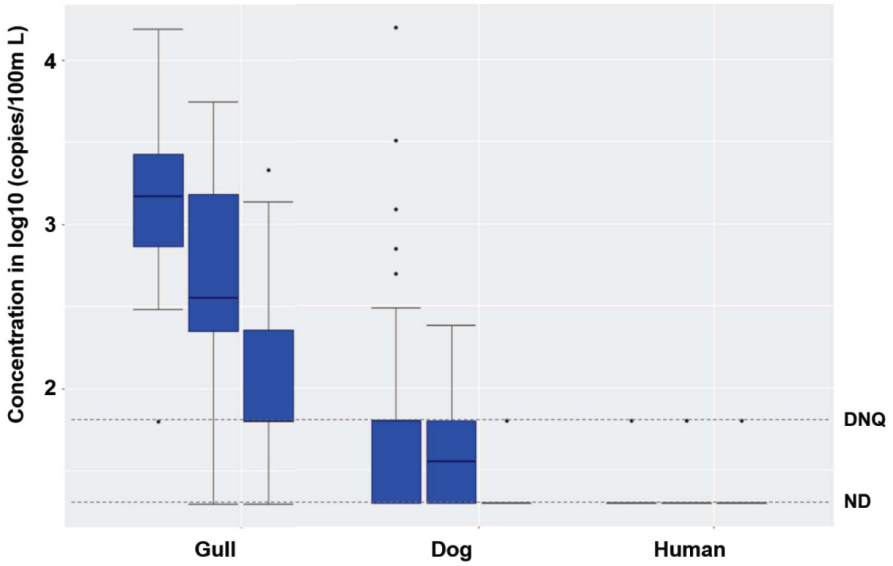


FIGURE 3 | Boxplots of gull, dog and HF183 human fecal marker concentrations in surf zone, nearshore, and offshore (from left to right in each box set, one set per marker) water samples in 2016 and 2017. The number of surf zone, nearshore, and offshore samples was 40, 24, and 32, respectively. DNQ (1.8 on log10 scale) and ND (1.3 on log10 scale) levels were shown in the figure as dotted lines.

2017 (Mann-Whitney test, all $p > 0.05$, $n = 96$). Nearshore marine sediments were mostly devoid of fecal markers, except that gull marker was detected in one sample; sediments contained very low to non-detectable FIB (**Supplementary Table 3**).

For treated effluent collected from a diffuser port on the WWTP outfall and marine water collected over the outfall diffuser at 1, 9, and 18 m depth from the surface (**Supplementary Figure 3**), the concentrations of EC, ENT, ENT by Entero1A, and dog markers were similar or lower when compared to nearshore and offshore waters, and significantly lower than surf zone waters (**Supplementary Table 5**, Mann-Whitney test, all $p < 0.01$). The concentrations of gull markers were also significantly lower in waters from and over the outfall diffuser than nearshore, offshore, and surf zone waters (all $p < 0.001$), and human marker HF183 levels were similar among these water samples (all $p > 0.05$) (**Figure 3** and **Supplementary Table 5**). Only one sample with human marker HF183 was detected from the diffuser port on the WWTP outfall, and this was at the DNQ level (**Supplementary Table 5**). In addition, there were only two detections of HF183 in the water over the outfall diffuser (one DNQ and 131 copies/100 ml, both at 18 m depth). These results indicated that the treated wastewater effluent discharging from the ocean outfall was unlikely to be the major source of chronic human fecal markers detected in surf zone waters. However, given complex nearshore ocean circulation patterns, the influence of the WWTP effluent outfall discharge on surf zone HF183 detections remained uncertain.

Results of Studies to Determine the Role of Bather Shedding

The possible effects of bather shedding on surf zone water quality were studied by counting bathers in the water at the time of sampling surf zone waters, including in the mornings and afternoons of holidays and high visitation weekends (**Supplementary Tables 6, 7**) when many people were recreating at the beach and in the water. There was also a sub-study intended to determine if water defecation might be happening overnight in the surf zone by beach campers; this was investigated by comparing HF183 human marker detections in the morning with those in the previous late afternoon (ca. 6–7 pm), and also with the following mid-afternoon (ca. 3–4 pm) (**Supplementary Tables 6, 7**). This timing was selected expecting that overnight water defecation would lead to higher morning human marker detections; it also allowed for inferring possible marker decay due to sunlight during the day. Besides the sub-studies of holidays and high visitation weekends, and of possible water defecation, the human beach visitor census at the time of surf zone water sampling was recorded during all other morning sampling campaigns for surf zone waters during dry weather in 2017 (**Supplementary Table 7**). The bather census was not recorded during 2016 surf zone water sampling.

In the studies of surf zone waters during holidays or high visitation weekends, human marker HF183 was detected in 1 of 15 morning samples, and 2 of 15 afternoon samples. Thus, as with FIB concentrations (**Supplementary Table 6**), the human marker HF183 concentrations did not vary significantly across the mornings and afternoons (**Supplementary Table 6**, Mann-Whitney test, all $p > 0.05$, $n = 30$). No human marker HF183 was detected in any water sample collected during the water defecation study, and FIB concentrations did not vary across the three different times of day (Kruskal-Wallis, $p > 0.05$, $n = 15$). However, HF183 markers were detected in 7 of 50 other morning surf zone samples during dry weather in 2017 (watershed to surf zone transect, and surf zone to nearshore to offshore transect;

Supplementary Table 7) when the human visitor census was recorded at the time of sampling.

Across all surf zone water data for which the human census was recorded at the time of sampling, HF183 concentrations and total counts of people—whether in the water (bather) or on the sand—were uncorrelated (**Supplementary Table 7**, Spearman's one tailed $\rho < 0.1$, both $p > 0.29$, $n = 93$), particularly in the morning ($\rho < 0$, both $p > 0.58$, $n = 70$). There was no relationship between human marker HF183 detections and the counts of people on the sand in the afternoon ($\rho = 0.19$, $p = 0.19$, $n = 23$). However, there was a significant correlation between the number of afternoon bathers vs. surf zone water sample HF183 concentrations ($\rho = 0.36$, $p = 0.04$, $n = 23$). Furthermore, the correlation between HF183 concentrations and bather counts appeared to be stronger when only afternoon samples from holidays and high visitation weekends were considered ($\rho = 0.46$, $p = 0.04$, $n = 15$), i.e., when there were more people bathing at GB (2.7 bathers in average per sampling location). This was in contrast to morning samples from holidays and high visitation weekends wherein, with only 0.5 bathers in average per location, there was no correlation between HF183 concentrations and bather counts observed ($\rho < 0$, $p = 0.71$, $n = 15$). These results suggest that people recreating in the water during the afternoon contributed to human fecal markers observed in the surf zone at GB, but leaves open that another source could be responsible for observations of HF183 in morning surf zone water samples.

DISCUSSION

Animal fecal materials, particularly from gulls and dogs, have been identified as major sources of fecal contamination in recreational beach surf zone waters in previous studies (Wright et al., 2009; Converse et al., 2012; Goodwin et al., 2016) as well as in this study, possibly due to many resident gulls and dog walking at GB. Corresponding management practices such as using falconry or dogs to control gulls can dramatically reduce the levels of FIB in surf zone waters (Converse et al., 2012; Goodwin et al., 2016). However, due to host specificity, pathogens from the animal feces may pose less serious risks to public health than human feces (Schoen and Ashbolt, 2010), although some risk does stem from gull fecal exposure (Brown et al., 2017a). Thus, such FIB control actions, even if successful, may not fully achieve the desired level of health risk reduction for swimmers.

Leaking sewers, storm drains, creeks, and rivers can deliver FMPs of human concern to surf zone waters. Also, although specific sources such as treated wastewater or reclaimed water may be confounded by dead microbial cells or free human marker DNA, sources such as human fecal materials or sewage from leaking sewer lines carry a risk to human health from associated pathogens. Thus, sources of even low levels of human fecal markers should be discerned for better evaluating potential human health risks and for guiding suitable management actions.

In other studies, beach berms breaching at coastal sloughs significantly controlled watershed influences on surf zone water quality (Riedel et al., 2015; Sikich et al., 2018). Here, the slough was bermed in 2016 and breached in 2017 (**Supplementary Figure 1**). However, no significant variations in human fecal marker HF183 and pathogen levels in surf zone waters under dry weather were observed across 2016 and 2017, or among the 5 sampling locations (**Figure 1**). This might owe to the lower concentrations of human fecal marker HF183 in the slough compared to the surf zone waters during dry weather. Although rain events can mobilize human and dog fecal materials to watershed waters, thereby increasing levels of human and dog fecal markers as well as pathogens (e.g., *Salmonella* spp.) (Steele et al., 2018), runoff-generating rain events were rare in this study area (twice in 2016 and once in 2017), and the slough was continuously bermed in 2016. In other studies, creek sediments were persistent sources of contamination including human fecal markers to creek waters (Frey et al., 2015; Kim and Wuertz, 2015), but watershed sediments, intertidal sands, and nearshore sediments in this study were all devoid of human and dog markers (**Supplementary Table 3**). This indicated that the fecal markers mobilized into creek waters during rain events were not accumulating in creek sediments; they were also not stored in beach sands and thereby becoming long-term slow-release sources into surf zone waters. Taken together, all evidence would suggest that human fecal markers detected in the GB surf zone during this study did not originate in the watershed.

Infrastructure such as sewer lines in this study were not leaking. The low concentrations of human fecal marker HF183 in treated effluent discharging from the WWTP offshore ocean outfall (67% ND, 33% DNQ) and measured in marine water above the outfall diffuser at 18 m depth from the surface [60% ND, 20% DNQ, 20% quantifiable levels (131 copies/100 ml)]—with likely further marker dilution, decay, and predation during ocean transport (Carneiro et al., 2018)—did not appear to fully explain GB surf zone fecal contamination. Further, in a parallel study, it was concluded that there were no regional background levels of HF183 human markers or pathogens in surf zone waters, based on the results of sampling the Arroyo Hondo reference beach located up-coast of the Santa Barbara region (SI Results).

The results of this study point to bather shedding as a source of HF183 human fecal markers in surf zone waters, particularly in the afternoon when more bathers were present. Bather shedding has been demonstrated to be a possible source of FIB such as enterococci and human pathogens such as methicillin resistant *Staphylococcus aureus* to surf zone waters (Elmir et al., 2007, 2009; Plano et al., 2011). Correlations between bather density and levels of human waterborne pathogens as well as enterococci in marine recreational beach waters have also been observed in summer weekends previously (Graczyk et al., 2010). However, direct correlations between bathers and human fecal markers have not been reported yet. Although the low levels of human marker HF183 in the surf zone waters (at the DNQ level) of this study were well below public health risk thresholds (Boehm et al., 2015; Brown et al., 2017b), there was a significant correlation between bather counts and human fecal marker HF183 for afternoon samples, especially during holidays and high visitation weekends. The human markers detected in the morning could have originated from bather shedding in the previous afternoon, but it is also possible that the WWTP outfall was influential on human marker detections due to diffuser discharge ocean circulation patterns overnight, which were unstudied. This is worthy of further consideration given that HF183 was detected in

11.4% of all morning samples (8 of 70, **Supplementary Table 7**), which was higher than for the afternoon samples (2 of 23, or 8.7%), and similar to that for holiday and high visitation weekend afternoon samples (2 of 15, or 13.3%). There were few bathers in the surf zone in the morning (**Supplementary Table 7**) which may point to sources of HF183 varying at different times.

In another study in the region (Li et al., 2021), bather counts significantly correlated with human fecal marker HF183 in surf zone waters, particularly in the afternoon, while effects of the WWTP outfalls were unknown and could not be ruled out. However, in a prior MST study of another high visitation regional beach (Ervin et al., 2014), low level HF183 markers were chronically detected in the surf zone and there was no nearby WWTP outfall. Although swimmers were not counted in that prior study, the absence of a WWTP outfall at this other high visitation beach with chronic low surf zone HF183 would reinforce—based on this study's results—that bather shedding might have been responsible. Bather shedding might further explain the low but chronic presence of human fecal markers at many beaches along the California coastline (Santoro and Boehm, 2007; McQuaig et al., 2012; Russell et al., 2013; Riedel et al., 2015; Cao et al., 2017; Jennings et al., 2018). Our results indicate that afternoon sampling should be included and compared with morning sampling, rather than assume that afternoon sunlight-mediated decay confounds human marker detections. Yet, additional research should be performed to further understand relationships between bather shedding and human fecal contamination in recreational surf zone waters, including at different times of the day, such that impacts of bather shedding on microbial water quality and potential associated public health implications can be understood. Regardless, the possibility that multiple, permanent sources—in this case a WWTP outfall, and bathers routinely using the beach—can explain low chronic HF183 surf zone detections suggests that there are likely HF183 thresholds that are the lowest achievable at popular bathing beaches.

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

DL and LV planned and performed research, analyzed data, and co-authored manuscript. BS and JM planned research and co-authored manuscript. JE and ND planned and performed research and co-authored manuscript. PH secured project funding; led and planned the research; co-analyzed the results; co-authored the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the State of California Clean Beach Initiative, using funds from Proposition 84, and from funds provided by Henry (Sam) Wheeler.

ACKNOWLEDGMENTS

We acknowledge the assistance of Tania Gomez Ayala, Marina Feraud, Mitchell Maier, Patrick Roehrdanz, Ying Wang, Monika Mortimer, Christoph Pierre, Christian Orsini, and Eric Hessell at UCSB as well as Geosyntec Consultants' researchers including Stacy Isaac, Adam Questad, and Maia Colyar. We also thank Hillary Hauser of Heal the Ocean, Cathleen Garnand of Santa Barbara County, Project Clean Water, Willie Brummett of Santa Barbara County, and Environmental Health Services. We acknowledge the generous consultation of several researchers including Yiping Cao (previously Southern California Coastal Water Research Project, now Orange County Sanitation District), Professor Libe Washburn (UCSB), and Carter Ohlmann (UCSB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.673190/full#supplementary-material

REFERENCES

Ahmed, W., Gyawali, P., Sidhu, J. P., and Toze, S. (2014). Relative inactivation of faecal indicator bacteria and sewage markers in freshwater and seawater microcosms. *Lett. Appl. Microbiol.* 59, 348–354. doi: 10.1111/lam.12285

Arnold, B. F., Schiff, K. C., Griffith, J. F., Gruber, J. S., Yau, V., Wright, C. C., et al. (2013). Swimmer illness associated with marine water exposure and water quality indicators: impact of widely used assumptions. *Epidemiology* 24, 845–853. doi: 10.1097/01.ede.0000434431.06765.4a

Bernhard, A. E., and Field, K. G. (2000). 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

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

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

Brown, K. I., Graham, K. E., and Boehm, A. B. (2017a). Risk-based threshold of gull-associated fecal marker concentrations for recreational water. *Environ. Sci. Technol. Lett.* 4, 44–48. doi: 10.1021/acs.estlett.6b00473

Brown, K. I., Graham, K. E., Soller, J. A., and Boehm, A. B. (2017b). Estimating the probability of illness due to swimming in recreational water with a mixture

of human- and gull-associated microbial source tracking markers. *Environ. Sci. Process Impacts* 19, 1528–1541. doi: 10.1039/c7em00316a

Cao, Y., Raith, M. R., Smith, P. D., Griffith, J. F., Weisberg, S. B., Schriewer, A., et al. (2017). Regional assessment of human fecal contamination in southern California coastal drainages. *Int. J. Environ. Res. Public Health* 14, E874.

Carneiro, M. T., Cortes, M. B. V., and Wasserman, J. C. (2018). Critical evaluation of the factors affecting *Escherichia coli* environmental decay for outfall plume models. *Rev. Ambient. Água.* 13, 1–18. doi: 10.4136/ambi-agua.2106

Colford, J. M., Wade, T. J., Schiff, K. C., Wright, C. C., Griffith, J. F., Sandhu, S. K., et al. (2007). Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. *Epidemiology* 18, 27–35. doi: 10.1097/01.ede.0000249425.32990.b9

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

Dick, L. K., Bernhard, A. E., Brodeur, T. J., Santo Domingo, J. W., Simpson, J. M., Walters, S. P., et al. (2005). Host distributions of uncultivated fecal bacteroidales bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71, 3184–3191. doi: 10.1128/aem.71.6.3184-3191.2005

Elmir, S. M., Shibata, T., Solo-Gabriele, H. M., Sinigalliano, C. D., Gidley, M. L., Miller, G., et al. (2009). Quantitative evaluation of enterococci and Bacteroidales released by adults and toddlers in marine water. *Water Res.* 43, 4610–4616. doi: 10.1016/j.watres.2009.07.006

Elmir, S. M., Wright, M. E., Abdelzaher, A., Solo-Gabriele, H. M., Fleming, L. E., Miller, G., et al. (2007). Quantitative evaluation of bacteria released by bathers in a marine water. *Water Res.* 41, 3–10. doi: 10.1016/j.watres.2006.10.005

Ervin, J. S., Van De Werfhorst, L. C., Murray, J. L., and Holden, P. A. (2014). Microbial source tracking in a coastal California watershed reveals canines as controllable sources of fecal contamination. *Environ. Sci. Technol.* 48, 9043–9052. doi: 10.1021/es502173s

Field, K. G., and Samadpour, M. (2007). Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res.* 41, 3517–3538. doi: 10.1016/j.watres.2007.06.056

Frey, S. K., Gottschall, N., Wilkes, G., Grégoire, D. S., Topp, E., Pintar, K. D., et al. (2015). Rainfall-induced runoff from exposed streambed sediments: an important source of water pollution. *J. Environ. Qual.* 44, 236–247. doi: 10.2134/jeq2014.03.0122

Goodwin, K. D., Gruber, S., Vondrak, M., and Crumpacker, A. (2016). Watershed assessment with beach microbial source tracking and outcomes of resulting gull management. *Environ. Sci. Technol.* 50, 9900–9906. doi: 10.1021/acs.est.6b02564

Graczyk, T. K., Sunderland, D., Awantang, G. N., Mashinski, Y., Lucy, F. E., Graczyk, Z., et al. (2010). Relationships among bather density, levels of human waterborne pathogens, and fecal coliform counts in marine recreational beach water. *Parasitol. Res.* 106, 1103–1108. doi: 10.1007/s00436-010-1769-2

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

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. Microb.* 80, 3086–3094. doi: 10.1128/aem.04137-13

Griffith, J. F., Layton, B. A., Boehm, A. B., Holden, P. A., Jay, J. A., Hagedorn, C., et al. (2013). *The California Microbial Source Identification Manual: A Tiered Approach to Identifying Fecal Pollution Sources to Beaches. Southern California Coastal Water Research Project Technical Report 804*. Costa Mesa, CA: Southern California Coastal Water Research Project.

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. A., Siefring, S., Lavender, J., and Varma, M. (2012). Influences of sample interference and interference controls on quantification of enterococci fecal indicator bacteria in surface water samples by the qPCR method. *Water Res.* 46, 5989–6001. doi: 10.1016/j.watres.2012.08.017

Jennings, W. C., Chern, E. C., O'Donohue, D., Kellogg, M. G., and Boehm, A. B. (2018). Frequent detection of a human fecal indicator in the urban ocean: environmental drivers and covariation with enterococci. *Environ. Sci. Process Impacts* 20, 480–492. doi: 10.1039/c7em00594f

Kim, M., and Wuertz, S. (2015). Survival and persistence of host-associated Bacteroidales cells and DNA in comparison with *Escherichia coli* and Enterococcus in freshwater sediments as quantified by PMA-qPCR and qPCR. *Water Res.* 87, 182–192. doi: 10.1016/j.watres.2015.09.014

Li, D., Van De Werfhorst, L. C., Dunne, T., Devarajan, N., Ayala, T. G., and Holden, P. A. (2020). Surf zone microbiological water quality following emergency beach nourishment using sediments from a catastrophic debris flow. *Water Res.* 176:115733. doi: 10.1016/j.watres.2020.115733

Li, D., Van De Werfhorst, L. C., Steets, B., Ervin, J., Murray, J. L. S., Blackwell, A., et al. (2021). Sources of low level human fecal markers in recreational waters of two Santa Barbara, CA beaches: roles of WWTP outfalls and swimmers. *Water Res.* (in press). doi: 10.1016/j.watres.2021.117378

Los Angeles Regional Water Quality Control Board, State of California (2018). Available online at: https://www.waterboards.ca.gov/losangeles/water_issues/programs/stormwater/municipal/watershed_management/ (accessed May 13, 2021).

Lu, J. R., 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

Malorny, B., Paccassoni, E., Fach, P., Bunge, C., Martin, A., and Helmuth, R. (2004). Diagnostic real-time PCR for detection of *Salmonella* in food. *Appl. Environ. Microbiol.* 70, 7046–7052. doi: 10.1128/aem.70.12.7046-7052.2004

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

Mayer, R. E., Reischer, G. H., Ixenmaier, S. K., Derx, J., Blaschke, A. P., Ebdon, J. E., et al. (2018). Global distribution of human-associated fecal genetic markers in reference samples from six continents. *Environ. Sci. Technol.* 52, 5076–5084. doi: 10.1021/acs.est.7b04438

McQuaig, S., Griffith, J., and Harwood, V. J. (2012). Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. *Appl. Environ. Microbiol.* 78, 6423–6432. doi: 10.1128/aem.00024-12

Oana, K., Okimura, Y., Kawakami, Y., Hayashida, N., Shimosaka, M., Okazaki, M., et al. (2002). Physical and genetic map of Enterococcus faecium ATCC19434 and demonstration of intra- and interspecific genomic diversity in enterococci. *FEMS Microbiol. Lett.* 207, 133–139. doi: 10.1111/j.1574-6968.2002.tb11041.x

Pachepsky, Y. A., and Shelton, D. R. (2011). *Escherichia coli* and fecal coliforms in freshwater and estuarine sediments. *Crit. Rev. Environ. Sci. Technol.* 41, 1067–1110. doi: 10.1080/10643380903392718

Plano, L. R., Garza, A. C., Shibata, T., Elmir, S. M., Kish, J., Sinigalliano, C. D., et al. (2011). Shedding of Staphylococcus aureus and methicillin-resistant Staphylococcus aureus from adult and pediatric bathers in marine waters. *BMC Microbiol.* 11:5. doi: 10.1186/1471-2180-11-5

Riedel, T. E., Thulsiraj, V., Zimmer-Faust, A. G., Dagit, R., Krug, J., Hanley, K. T., et al. (2015). Long-term monitoring of molecular markers can distinguish different seasonal patterns of fecal indicating bacteria sources. *Water Res.* 71, 227–243. doi: 10.1016/j.watres.2014.12.037

Russell, T. L., Sassoubre, L. M., Wang, D., Masuda, S., Chen, H., Soetjipto, C., et al. (2013). A coupled modeling and molecular biology approach to microbial source tracking at Cowell Beach, Santa Cruz, CA, United States. *Environ. Sci. Technol.* 47, 10231–10239.

San Diego Regional Water Quality Control Board (2017). *Cost-Benefit Analysis: San Diego Region Bacteria Total Maximum Daily Loads*. San Diego, CA: San Diego Regional Water Quality Control Board, 416.

Santoro, A. E., and Boehm, A. B. (2007). Frequent occurrence of the human-specific *Bacteroides* fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators. *Environ. Microbiol.* 9, 2038–2049. doi: 10.1111/j.1462-2920.2007.01319.x

Schoen, M. E., and Ashbolt, N. J. (2010). Assessing pathogen risk to swimmers at non-sewage impacted recreational beaches. *Environ. Sci. Technol.* 44, 2286–2291. doi: 10.1021/es903523q

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

Shuval, H. (2003). Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *J. Water Health* 1, 53–64. doi: 10.2166/wh.2003.0007

Sikich, S. A., Luce, S., Pease, K., King, M., and Vu, K. (2018). *Heal the Bay's 2011-2012 Annual Beach Report Card*. Available online at: https://healthebay.org/wp-content/uploads/2018/07/BRC_2017-2018_07-12-18.pdf (accessed May 13, 2021).

Sinclair, R. G., Jones, E. L., and Gerba, C. P. (2009). Viruses in recreational water-borne disease outbreaks: a review. *J. Appl. Microbiol.* 107, 1769–1780. doi: 10.1111/j.1365-2672.2009.04367.x

Sinigalliano, C. D., Fleisher, J. M., Gidley, M. L., Solo-Gabriele, H. M., Shibata, T., Plano, L. R. W., et al. (2010). Traditional and molecular analyses for fecal indicator bacteria in non-point source subtropical recreational marine waters. *Water Res.* 44, 3763–3772. doi: 10.1016/j.watres.2010.04.026

Steele, J. A., Blackwood, A. D., Griffith, J. F., Noble, R. T., and Schiff, K. C. (2018). Quantification of pathogens and markers of fecal contamination during storm events along popular surfing beaches in San Diego, California. *Water Res.* 136, 137–149. doi: 10.1016/j.watres.2018.01.056

U.S. Environmental Protection Agency (2004). *Report to Congress: impacts and control of CSOs and SSOs*. Washington, DC: U.S. Environmental Protection Agency, Office of Water.

Wright, M. E., Solo-Gabriele, H. M., Elmir, S., and Fleming, L. E. (2009). Microbial load from animal feces at a recreational beach. *Mar. Pollut. Bull.* 58, 1649–1656. doi: 10.1016/j.marpolbul.2009.07.003

Conflict of Interest: BS and JE were employed by the company Geosyntec Consultants.

The remaining 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.

Copyright © 2021 Li, Van De Werfhorst, Steets, Ervin, Murray, Devarajan and Holden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 June 2021
doi: 10.3389/fmicb.2021.661954

Identifying Sources of Faecal Contamination in a Small Urban Stream Catchment: A Multiparametric Approach

*Liam J. Reynolds[1], Niamh A. Martin[1], Laura Sala-Comorera[1], Kevin Callanan[2], Padraig Doyle[3], Clare O'Leary[2], Paul Buggy[4], Tristan M. Nolan[1], Gregory M. P. O'Hare[5], John J. O'Sullivan[6] and Wim G. Meijer[1]**

[1] UCD School of Biomolecular and Biomedical Science, UCD Earth Institute, and UCD Conway Institute, University College Dublin, Dublin, Ireland, [2] Central Laboratory, Dublin City Council, Dublin, Ireland, [3] Drainage Planning, Policy and Development Control, Dublin City Council, Dublin, Ireland, [4] Municipal Services, Dún Laoghaire-Rathdown County Council, Dublin, Ireland, [5] UCD School of Computer Science, UCD Earth Institute, University College Dublin, Dublin, Ireland, [6] UCD School of Civil Engineering, UCD Dooge Centre for Water Resources Research, UCD Earth Institute, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Tarja Pitkänen, Water Microbiology Laboratory, National Institute for Health and Welfare, Finland
Gaurav Saxena, Shoolini University of Biotechnology and Management Sciences, India
Brian Badgley, Virginia Tech, United States

***Correspondence:**
Wim G. Meijer
wim.meijer@ucd.ie

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 31 January 2021
Accepted: 24 May 2021
Published: 29 June 2021

Citation:
Reynolds LJ, Martin NA, Sala-Comorera L, Callanan K, Doyle P, O'Leary C, Buggy P, Nolan TM, O'Hare GMP, O'Sullivan JJ and Meijer WG (2021) Identifying Sources of Faecal Contamination in a Small Urban Stream Catchment: A Multiparametric Approach. Front. Microbiol. 12:661954. doi: 10.3389/fmicb.2021.661954

Small urban streams discharging in the proximity of bathing waters may significantly contribute to the deterioration of water quality, yet their impact may be overlooked. This study focuses on the Elm Park stream in the city of Dublin that is subject to faecal contamination by unidentified sources. The aim of the study was to identify a minimum number of "sentinel" sampling stations in an urban catchment that would provide the maximum amount of information regarding faecal pollution in the catchment. Thus, high-resolution sampling within the catchment was carried out over the course of 1 year at 11 stations. Faecal indicator bacteria were enumerated and microbial source tracking (MST) was employed to evaluate human pollution. In addition, ammonium, total oxidised nitrogen, and phosphorus levels were monitored to determine if these correlated with faecal indicator and the HF183 MST marker. In addition, the effect of severe weather events on water quality was assessed using automated sampling at one of the identified "sentinel" stations during baseflow and high flow conditions over a 24-h period. Our results show that this urban stream is at times highly contaminated by point source faecal pollution and that human faecal pollution is pervasive in the catchment. Correlations between ammonium concentrations and faecal indicator bacteria (FIB) as well as the human MST marker were observed during the study. Cluster analysis identified four "sentinel" stations that provide sufficient information on faecal pollution in the stream, thus reducing the geographical complexity of the catchment. Furthermore, ammonium levels strongly correlated with FIB and the human HF183 MST marker under high flow conditions at key "sentinel" stations. This work demonstrates the effectiveness of pairing MST, faecal indicators, and ammonium monitoring to identify "sentinel" stations that could be more rapidly assessed using real-time ammonium readouts to assess remediation efforts.

Keywords: faecal contamination, multiparametric, microbial source tracking, urban stream, ammonia, total oxidised nitrogen, phosphorous, sentinel sampling

INTRODUCTION

Faecal pollution in the environment, including rivers and streams, is a growing problem globally. This is in part due to an increase in urbanisation: 55% of the world's population currently live in urban areas and this is projected to increase to 68% by 2050 (UN, 2019). Numerous studies have demonstrated that rivers and streams flowing through urban areas are impacted to a greater extent by faecal pollution than their upstream catchments (Goto and Yan, 2011; Paule-Mercado et al., 2016). As faecal pollution may introduce various pathogens, the degradation of such streams and rivers and the bathing waters they discharge into represents an increased infection risk for individuals (Arnold et al., 2017; Kauppinen et al., 2017). For example, a survey of 654 surfers in San Diego revealed increased numbers of wound and gastroenteritis infections following seawater exposure, particularly after rainfall (Arnold et al., 2017). Furthermore, faecal contamination of rivers and streams may negatively impact water used for purposes other than recreational activities, for example irrigation and drinking water intake.

Understanding how faecal pollution enters urban rivers and streams is vital if remediation measures are to be effectively implemented by water management authorities (Walker et al., 2015; Tillett et al., 2018). However, monitoring and remediating faecal pollution is difficult as pollutants can enter via multiple point and diffuse sources (Muller et al., 2020). For example, stream banks, impervious surface cover, and streambed sediments may act as reservoirs of faecal bacteria (Brinkmeyer et al., 2015; Baral et al., 2018; Fluke et al., 2019). Combined sewage overflows (CSOs) and misconnected sanitation pipes also contribute to faecal contamination particularly in cities reliant on antiquated subterranean sewage infrastructure (Brinkmeyer et al., 2015; Balleste et al., 2020). As such, devising efficient and effective sampling and monitoring programmes is vital if authorities are to successfully improve water quality in rivers and streams. A limited number of studies have attempted catchment level surveys to identify high impacted sites in catchments (Sauer et al., 2011; Tillett et al., 2018). Tillett et al., for instance coupled the monitoring of *Escherichia coli*, ammonium and the BacHum (human marker) to identify multiple point sources of human faecal pollution in the Frankston and Mornington Peninsula, Australia (Tillett et al., 2018). However, such sampling campaigns are labour-, time-, and cost-intensive and so water management authorities typically monitor only the discharge points of rivers and streams which provides no data on how and where faecal pollution enters.

Faecal pollution in water bodies is commonly determined by quantifying levels of fecal indicator bacteria (FIB), namely *E. coli* and intestinal enterococci (Saxena et al., 2015). Indeed, the EU bathing water directive uses only these indicators to determine bathing water quality (EU, 2006; Fewtrell and Kay, 2015). As part of the water framework directive (WFD), which aims to improve the ecological and chemical quality of European waters, ammonium, oxidised nitrogen compounds, and phosphorus concentrations are monitored in rivers and streams (EC, 2000). Several studies have identified strong correlations between ammonium, its oxidised derivatives (nitrites and nitrates) and phosphates with FIB (Cabral and Marques, 2006; Mallin and McIver, 2012; Napier et al., 2017; Baral et al., 2018). Although nutrient markers may have non-faecal origins, they offer utility as cost-effective faecal indicators in well characterised catchments where their association with faecal pollution has been validated. Further, monitoring of nutrients can be accomplished by installing probes at relevant stations in a catchment which allows for more rapid identification of faecal pollution events than FIB enumeration. Monitoring only FIB and nutrients, however, provides no information on the biological source of pollution which may be human or zoonotic in origin. Microbial source tracking (MST) methods rely on faecal bacteria being more closely associated with one animal host than another to identify the biological sources of faecal pollution (Harwood et al., 2014). Thus, coupling MST with quantitative PCR (qPCR) can identify the most abundant of these bacteria to provide insight into the primary biological sources of faeces in an environment (Gourmelon et al., 2007; Unno et al., 2018).

The complexity involved in identifying the sources of faecal pollution highlights the need to design and implement effective and efficient monitoring programmes of rivers and streams. Thus, the aim of this study was to use hierarchical clustering to identify a minimum number of "sentinel" stations that provide maximum information about the pollution status of a complex urban stream. Furthermore, as monitoring FIB and MST markers are time consuming, correlation analyses were conducted to determine if specific nutrients could be used as indicators for human faecal pollution events. This study therefore describes methodology to identify a limited number of "sentinel" stations in a complex catchment and an approach to validate ammonium as a marker for faecal contamination within the catchment. This would allow the assessment of water quality in real-time.

MATERIALS AND METHODS

Study Area and Sampling

Dublin bay is a UNESCO Biosphere as it is home to rare wildlife species and is unique in that it encompasses an area of a capital city: Dublin, Ireland. There are also a number of bathing areas within Dublin bay which are popular with tourists and the city's 560,000 population. Several small streams with completely urban catchments flow through the city and discharge into the bay. The 3.8 km long Elm Park stream, for example, flows through an urban area with a population of approximately 40,000 people before discharging onto Sandymount Strand, a designated bathing water, and the recently declassified Merrion Strand.

Between October 2019 and September 2020, water samples (n = 209) were collected from 11 stations along the Elm Park stream catchment (**Figure 1**). Faecal pollution in this stream is regularly monitored at its discharge point only. Our sampling regime represented a more detailed assessment of the catchment which we refer to as high-resolution sampling throughout the text. Of these stations, M1–M3 (upstream), together with GR1–3 and EP2–3 (downstream) were located on the main trunk of the stream. Stations G1, RK1, and RV1 were tributaries entering the main trunk (**Supplementary Table 1**). Duplicate

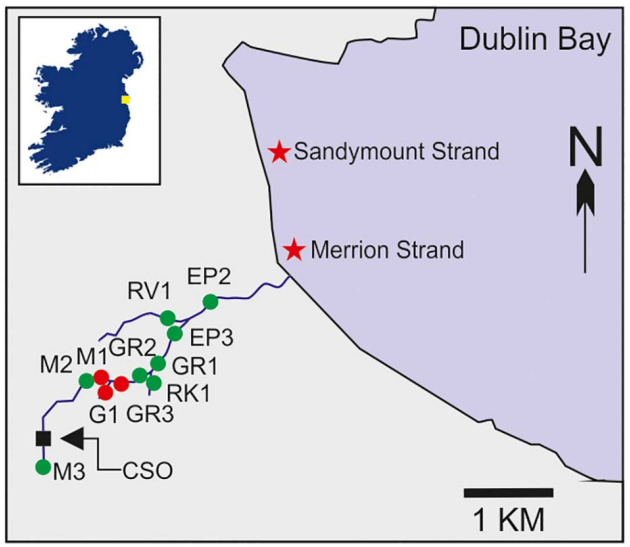


FIGURE 1 | Map of the Elm Park stream catchment indicating the 11 stations sampled as part of the high-resolution survey of the catchment (labelled green and red) as well as the Larchfield combined sewer overflow (CSO). Coordinates of all stations are listed in **Supplementary Table 1**. The inset box shows the location of Dublin Bay in Ireland indicated by a yellow square. Stations M1, G1, and GR3, which were sampled intensively during stormflow and baseflow, are highlighted in red. Two nearby bathing waters are indicated by stars.

grab samples were collected in sterile 1 L bottles, stored at 4°C, and processed within 6 h.

Automatic samplers with refrigeration units were launched at three upstream stations on the Elm Park Stream (M1, G1, and GR3) during stormflow conditions in February 2020 and baseflow conditions in June 2020 (**Figure 1**). Baseflow conditions do not significantly vary (P = 0.1–0.6) throughout the year. These stations are immediately downstream of a CSO and frequent sampling was conducted to determine how faecal indicators responded to stormflow (CSO active) and baseflow (CSO inactive). During stormflow 23 and 22 samples were collected at M1 and G1, respectively. Due to a machine error 9 samples were collected at GR3, but the peak of stormflow was still captured. During baseflow 24 samples were collected at M1 and G1 and 23 samples were collected at GR3. During these events 900 ml of water was collected hourly from each station in individual sterile bottles and kept at 4°C. Samples were received and processed within 18 h. Modelled rainfall data was obtained from the Met Éireann Re-Analysis database.

Enumeration of Faecal Indicator Bacteria

Faecal indicator bacteria were enumerated for samples M1–M3, G1, and GR3 by the Dublin City Council Central Laboratory. *E. coli* were enumerated for these samples using the IDEXX Colilert-18 assay according to ISO 9308-2:2012. Briefly, 10 and 100 ml of sample were mixed with dehydrated Colilert medium and incubated in Quanti-Trays at 36.0 ± 2.0°C. *E. coli* concentrations for these samples are determined as most probable number per 100 ml (MPN/100 ml). *E. coli* concentrations for samples GR1–2, EP2–3, RK1, and RV1 were determined using standard filtration methods in University College Dublin according to ISO 16649-1:2018. Samples were passed through 0.45 μm nitrocellulose filter membranes (Nalgene, Thermo Scientific) and then incubated on Tryptone Bile X-Glucuronide agar (TBX; Sigma-Aldrich) for 4 h at 37°C followed by incubation for 18 h at 44°C to enumerate *E. coli* as colony forming units per 100 ml (CFU/100 ml). Previously, split sampling of stations in the Elm Park stream catchment and subsequent analysis by both laboratories demonstrated these protocols to be comparable, therefore *E. coli* concentrations are referred to as CFU/100 ml throughout (**Supplementary Figure 1**). For all samples, intestinal enterococci were enumerated by incubating filter membranes on Slanetz and Bartley agar (Oxoid) at 37°C for 44 h. Filter membranes were then transferred to Bile Aesculin Agar (Sigma-Aldrich) and incubated at 44°C for 2 h to enumerate intestinal enterococci as CFU/100 ml according to ISO 7899-2:2000. For sites M1–M3, G1, and GR3 the upper limit of quantification was 20,000 MPN/100 ml.

Extraction of DNA and MST Quantification

Water samples (100 ml) were concentrated by filtration through 0.22 μm nitrocellulose filter membranes and stored at −20°C in cold lysis buffer (5 M guanidine isothiocyanate, 100 mM EDTA [pH 8.0], 0.5% [w/v] sodium lauroyl sarcosinate). A previously described modified DNeasy Blood and Tissue kit protocol (Qiagen) was used to extract DNA from filter membranes (Gourmelon et al., 2007).

Previous work has demonstrated that the HF183 marker is sensitive and specific to human faecal pollution (Boehm et al., 2013; Balleste et al., 2020). The HF183 MST marker was amplified by PCR and cloned into a pBLUE plasmid. This plasmid was linearized and used to create standard curves between 10^6 and 10^0 gene copies in each qPCR run to quantify target gene levels in each sample (**Supplementary Table 2**; Seurinck et al., 2005). Ten minute preincubations at 95°C and melting curve analyses were included in all qPCR cycles. The efficiency of each reaction was determined using the $E = 10^{(1/slope)} - 1$ equation (Rutledge and Côté, 2003). The limit of detection was determined as the lowest concentration of DNA detected in 95% or more of replicates and the limit of quantification was determined as the lowest concentration of DNA quantified within 0.5 SDs of the $\log_{10}$ concentration (**Supplementary Table 2**; Rutledge and Stewart, 2008). All standard curves had an R^2 value greater than 0.985.

Immediately following DNA extraction, duplicate qPCR reactions from undiluted and 10-fold diluted samples for each DNA extraction were conducted. All 20 μl reaction mixtures contained 10 μl of SYBR Green I Master (Roche), 500 nM each of the forward and reverse primer (**Supplementary Table 2**), 1 μl of sample, and were run on the Roche Lightcycler 96 platform (Roche). HF183 MST marker concentrations were expressed as gene copies per 100 ml (gc/100 ml).

Nitrogen Compound and Phosphorus Analysis

All nitrogen and phosphorus compounds were analysed using Gallery Plus instrumentation with an automated discrete colorimetric analyser. Phosphorus pollution was determined by measuring orthophosphate levels using the automated ascorbic acid reduction method with absorbance measured at 880 nm. Orthophosphate concentrations were derived as mg/L (as P) and the quantification limit for this assay is 0.01 mg/L (as P) (Baird and Bridgewater, 2017).

All nitrogen compound concentrations were derived as mg/L (as N). Ammonium concentration in all samples was determined using the automated phenate method with absorbance being monitored at 660 nm, the quantification limit for this assay is 0.01 mg/L (as N) (Baird and Bridgewater, 2017). Nitrite concentrations were determined using the automated sulphanilamide – *N*-1-Naphthylethylene-diamine dihydrochloride (NEDD) colorimetric method, with absorbance measured at 540 nm, the quantification limit for this assay is 0.005 mg/L (as N) (Baird and Bridgewater, 2017). Total oxidized nitrogen (TON) was monitored by first reducing the nitrate in the sample to nitrite. The produced nitrite and the nitrite present in the sample was then quantified using the automated sulphanilamide – NEDD colorimetric method and the quantification limit for this assay is 0.01 mg/L (as N). The nitrate concentrations of samples were derived by subtracting the nitrite concentration from the TON concentration (Baird and Bridgewater, 2017).

Multivariate and Statistical Analyses

Hierarchical clustering is a method that groups objects such that intra-cluster objects are more similar than between cluster objects. Principle component analysis is a process of data reduction that can identify the main variables contributing to variation of a dataset. By coupling hierarchical clustering and principal component analysis, "sentinel" stations and the primary variables contributing to them can be identified. Hierarchical clustering and principal component analysis were accomplished using scaled and log transformed data in RStudio 3.5.1. Hierarchical clustering of stations based on the geometric means of the assessed variables was conducted using Ward's hierarchical linkage of Euclidean distances (Suzuki and Shimodaira, 2006). Multiscale bootstrapping was employed to determine "true" clusters (>0.90 significance cut-off). Principal component analysis was conducted with Horn's Parallel Analysis to determine the minimum number of significant principal components. Varimax rotation was employed on the resulting principal components to identify the primary variables contributing to each principal component.

Spearman correlation analyses of the levels of FIB, MST markers, and nutrient concentrations were performed using Prism GraphPad software (version 9.1.0.221). A significance cut-off of $p \leq 0.05$ was used for all analyses. When culture, molecular, and chemical assays had an upper limit of quantification this upper limit was used in statistical analyses when reached. When culture, molecular, and chemical assays where below the quantification limit, a value of half the quantification limit was used in further analyses.

RESULTS AND DISCUSSION

High-Resolution Sampling Identifies Hotspots of Human Faecal Pollution and Informs the Identification of "Sentinel" Stations

Faecal pollution in urban streams and rivers can have a detrimental impact on the bathing waters into which they discharge (Molina et al., 2014). For remediation measures to be successful in protecting bathing waters understanding faecal pollution and devising efficient monitoring programmes are paramount (Walker et al., 2015). Thus, high-resolution sampling of the Elm Park stream was conducted to better understand faecal pollution in the catchment and to inform the identification of "sentinel" stations. All sampling stations in the Elm Park stream were highly polluted as FIB concentrations varied up to four orders of magnitude (5.90×10^1–2.42×10^5 and $<$10–2.65×10^4 CFU/100 ml for *E. coli* and intestinal enterococci, respectively) (**Figures 2A,B**). On occasion, these levels surpassed those observed in treated wastewater impacted rivers and streams that have been studied (Paule-Mercado et al., 2016; Balleste et al., 2020). High-resolution sampling of the Elm Park catchment revealed station M2, which is located directly downstream of a CSO, to have the highest concentrations of *E. coli* (geometric mean of 5.16×10^3 CFU/100 ml) and intestinal enterococci (geometric mean of 1.54×10^3 CFU/100 ml). Although the concentrations of FIB generally decreased along the course after this point of the stream, the geometric mean of concentrations of *E. coli* and intestinal enterococci never fell below 1.4×10^3 and 3×10^2 CFU/100 ml, respectively, in the main trunk of the stream. These data suggest that there are no significant point sources of pollution between station M2 and GR3. However, as the levels of FIB remained high between these stations, they are likely impacted by upstream flow and surface run-off from the embankment (Muller et al., 2020). Interestingly, *E. coli* concentrations increased again at station EP3 (geometric mean of 3,054 CFU/100 ml). As there are no documented CSOs between GR1 and EP3, this increase indicates the presence of an unknown point source of faecal pollution. The tributary stations, G1 and RV1, typically had FIB concentrations lower than, or at comparable levels to downstream stations on the main trunk. The RK1 tributary station on the other hand had a higher mean concentration of both faecal indicators than GR1 which is immediately downstream and may be contributing to faecal pollution in the main trunk.

In addition to identifying the geographical sources of faecal pollution, it is important to determine the primary biological sources of faecal pollution, as human pollution poses a greater risk to human health than animal faecal pollution (Soller et al., 2010). The human faecal marker, HF183, could be quantified in all but two samples (RV1 samples collected in August 2020),

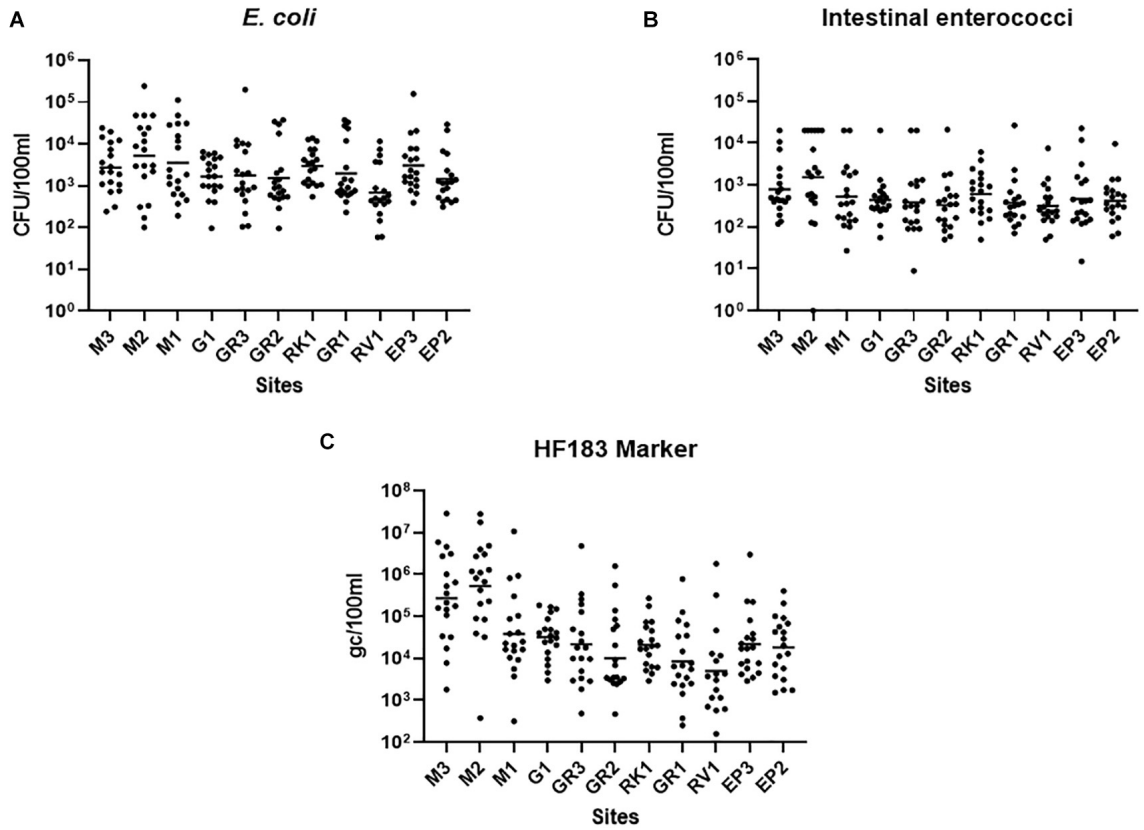


FIGURE 2 | Faecal Indicator bacteria concentrations in the Elm Park Catchment. Shown are the levels of panels **(A)** *E. coli*, **(B)** intestinal enterococci, and **(C)** the HF183 (human) microbial source tracking marker at sampling station. The dot plots show the concentration of each sample and the geometric means are indicated as a horizontal line.

demonstrating that human faecal pollution is pervasive in the Elm Park catchment (**Figure 2C**). HF183 concentrations along the Elm Park stream followed a similar pattern to that of the FIB, although, the variation observed was greater, up to 10^6-fold at some stations. Station M3 and M2 exhibited the highest HF183 concentrations (geometric means of 5.28×10^5 and 2.74×10^5 gc/100 ml, respectively) providing further evidence of upstream pollution in the Elm Park catchment and that M2 is directly impacted by the CSO discharging human faecal waste. Station M3 is not impacted by any documented point sources so human faecal pollution is likely entering the stream from diffuse leakages and misconnections. On occasion, these two upstream stations had HF183 concentrations approaching that documented for raw sewage (10^8 gc/100 ml) (Hughes et al., 2017; Balleste et al., 2020). Although HF183 concentrations decreased along the main trunk of the Elm Park stream by greater than 10-fold at M1 (3.81×10^4 gc/100 ml) and greater than 60-fold at GR1 (8.00×10^3 gc/100 ml), these levels are still considered high and within the range of other human waste impacted streams (Olds et al., 2018).

In this study, high-resolution sampling was coupled with MST analysis in the Elm Park catchment to identify point and diffuse sources of human faecal pollution at stations M3, M2, and EP3. Such an approach may prove useful to identify locations of faecal pollution in other urban stream or larger river catchments (Tillett et al., 2018).

High-Resolution Sampling Reveals Catchment Level Nitrogen and Phosphorus Contamination

Ammonium is a parameter that is monitored as part of the WFD and has also been used to identify incidences of faecal pollution (EC, 2000; Cabral and Marques, 2006; Mallin and McIver, 2012). Nutrient markers such as ammonium, when used alongside FIB and MST, can assist in identifying the biological source of faecal pollution and may offer an alternative and more rapid method of monitoring faecal pollution (Lim et al., 2017). Ammonium concentrations varied up to four orders of magnitude at some stations in the Elm Park catchment (**Figure 3A**). The highest concentrations of ammonium were observed at stations M3 and M2 [0.01–13.91 mg/L (as N); geometric mean of 0.3298 mg/L (as N)] which are comparable to treated wastewater impacted rivers (Jin et al., 2017; Preisner, 2020). This agrees with the high concentrations of FIB and HF183 observed at these upstream stations suggesting that ammonium levels may be indicative of human faecal pollution in this catchment. Moving downstream, concentrations of ammonium generally decreased in the Elm

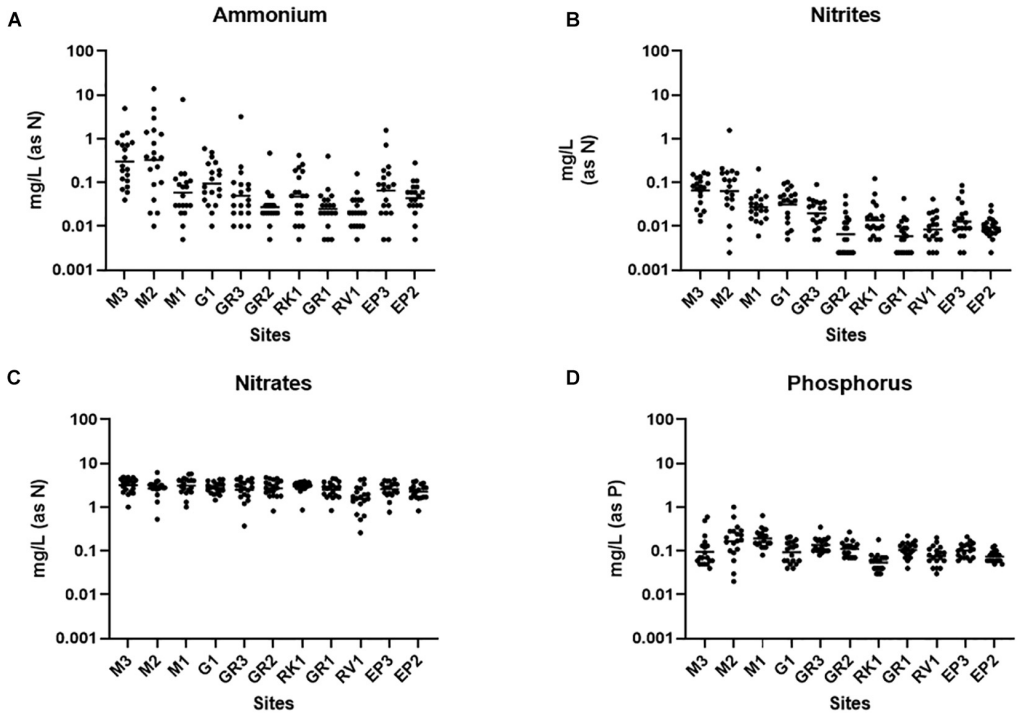


FIGURE 3 | Nutrient concentrations in the Elm Park Catchment. Shown are the levels of panels **(A)** ammonium, **(B)** nitrite, **(C)** nitrate, and **(D)** phosphorus for each sampling station. The dot plots show the concentration of each sample.

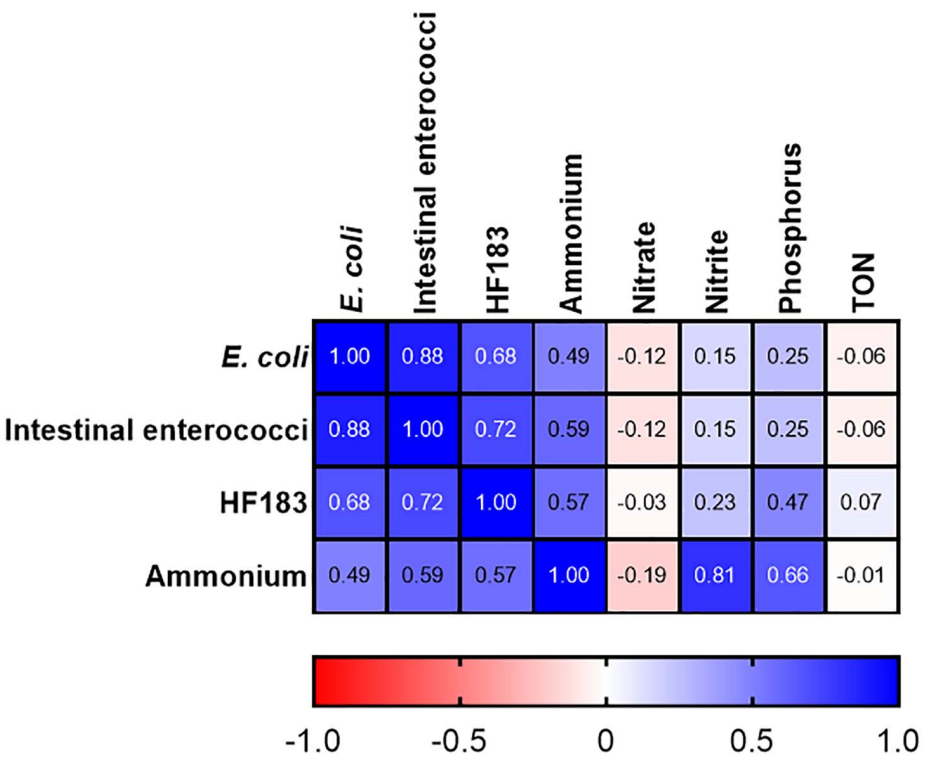


FIGURE 4 | Heatmap showing correlations between FIB, the HF183 marker and nutrients at station M2. TON refers to total oxidised nitrogen. Dark blue and dark red indicate strong positive and negative correlations, respectively. R^2 coefficients > 0.49 were considered statistically significant ($p < 0.05$).

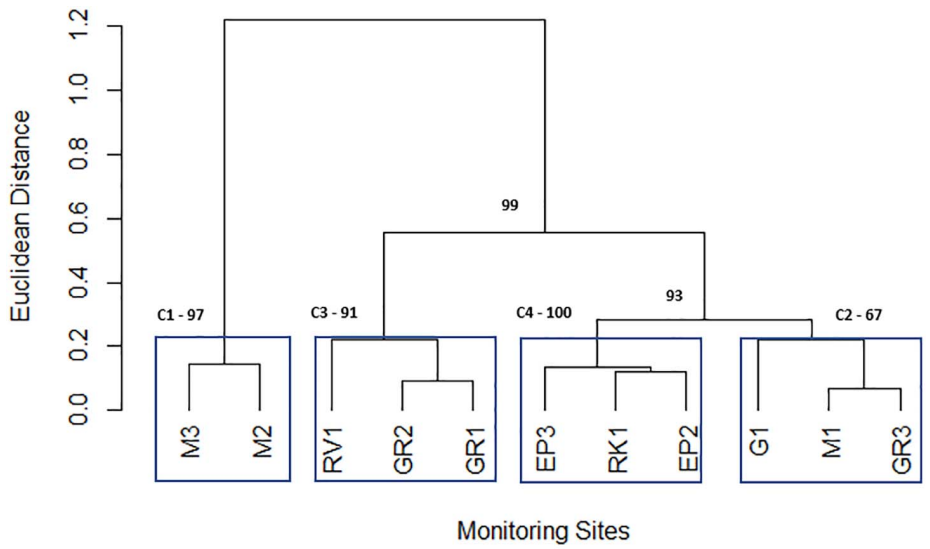


FIGURE 5 | Dendrogram depicting the grouping of Elm Park catchment sampling stations obtained from the geometric means of variables using the Ward hierarchical clustering of Euclidean distances. Clusters are labelled C1, C2, C3, and C4 and surrounded by blue boxes. The numbers highlighted in bold indicate the *p*-values (approximately unbiased *p*-value; cut-off > 90%) used to identify "true" clusters.

Park catchment such that all downstream stations had geometric means ≤ 0.1 mg/L (as N). Nitrite in surface waters can result from faecal pollution and from the oxidation of ammonium by ammonium oxidising bacteria and archaea (He et al., 2018). Nitrite levels in the Elm Park catchment followed the same trend as ammonium with concentrations ranging over three orders of magnitude [>0.001 to >1 mg/L (as N)] (**Figure 3B**). Thus, increased nitrite concentrations may be indicative of faecal pollution in the Elm Park catchment. In contrast, little variation was observed in nitrate levels within the catchment [geometric mean of 1.539–3.133 mg/L (as N)] which typically made up greater than 96% of measured TON (**Figure 3C**). Nitrates are present in human excreta and are the terminal oxidation product of ammonium. They may also enter surface water from decaying plant matter and atmospheric deposition (Vrzel et al., 2016). The temporal and geographical stability of nitrate concentrations in the Elm Park catchment indicates that it has little use as a faecal indicator. Phosphorus is also present in human excreta, but it can also enter surface water by mineral erosion and organic decomposition (van der Kooij et al., 2020). Phosphorus concentrations exhibited little variation and so may not prove useful as a faecal indicator in the Elm Park catchment [geometric means of 0.05–0.19 mg/L (as P)] (**Figure 3D**).

Faecal Indicator Bacteria, HF183, and Ammonium Strongly Correlate at Highly Polluted Stations

If alternative nutrient parameters are to be successfully used as rapid alternatives to monitor faecal pollution it is important to determine their relationship with classical FIB and MST markers. We therefore carried out correlation analyses. At station M3, which is highly impacted by human faecal pollution, no correlations between FIB, MST markers, and nutrients were observed. Faecal pollution at this station originates from diffuse human sources which may be sporadic in nature. Thus, variations in pollutant discharge coupled with different die-off rates between FIB and MST harbouring *Bacteroides* spp. will compound correlations between these indicators and nutrient parameters. Similarly, no correlations were observed at tributary stations G1 and RK1.

At station M2 statistically significant correlations between FIB, the human MST marker, and ammonium were observed (R^2 = 0.491–0.879; *p*-value <0.05) (**Figure 4**). Previous studies have demonstrated a strong correlation between ammonium, *E. coli*, intestinal enterococci, and HF183 in pooled WWTP influent/effluent (Mayer et al., 2016). As M2 is directly impacted on by a CSO, the strong correlations documented between ammonium concentrations and FIB as well as HF183 demonstrates the usefulness of this nutrient as a proxy for recent faecal pollution that can be measured in a rapid and cost-effective manner. In well characterised catchments, where ammonium has been shown to correlate with faecal indicators, this would allow for real-time monitoring of remediation efforts that would complement regular monitoring of FIB and MST markers. Downstream of station M2, on the main trunk of the stream, *E. coli* and intestinal enterococci concentrations correlated (R^2 = 0.507–0.800; *p*-value <0.05), however, correlations with HF183 and ammonium were more sporadic (**Supplementary Table 3**). Indeed, no correlations with these indicators are observed at the most downstream EP3 and EP2 stations. As these stations are furthest geographically from the M2 CSO this may be a result of dilution of the CSOs effluent.

At stations GR2 and GR3 FIB levels correlated with ammonium but only with HF183 concentrations at M1 (R^2 = 0.537–0.703; *p*-value <0.05). At the tributary stations,

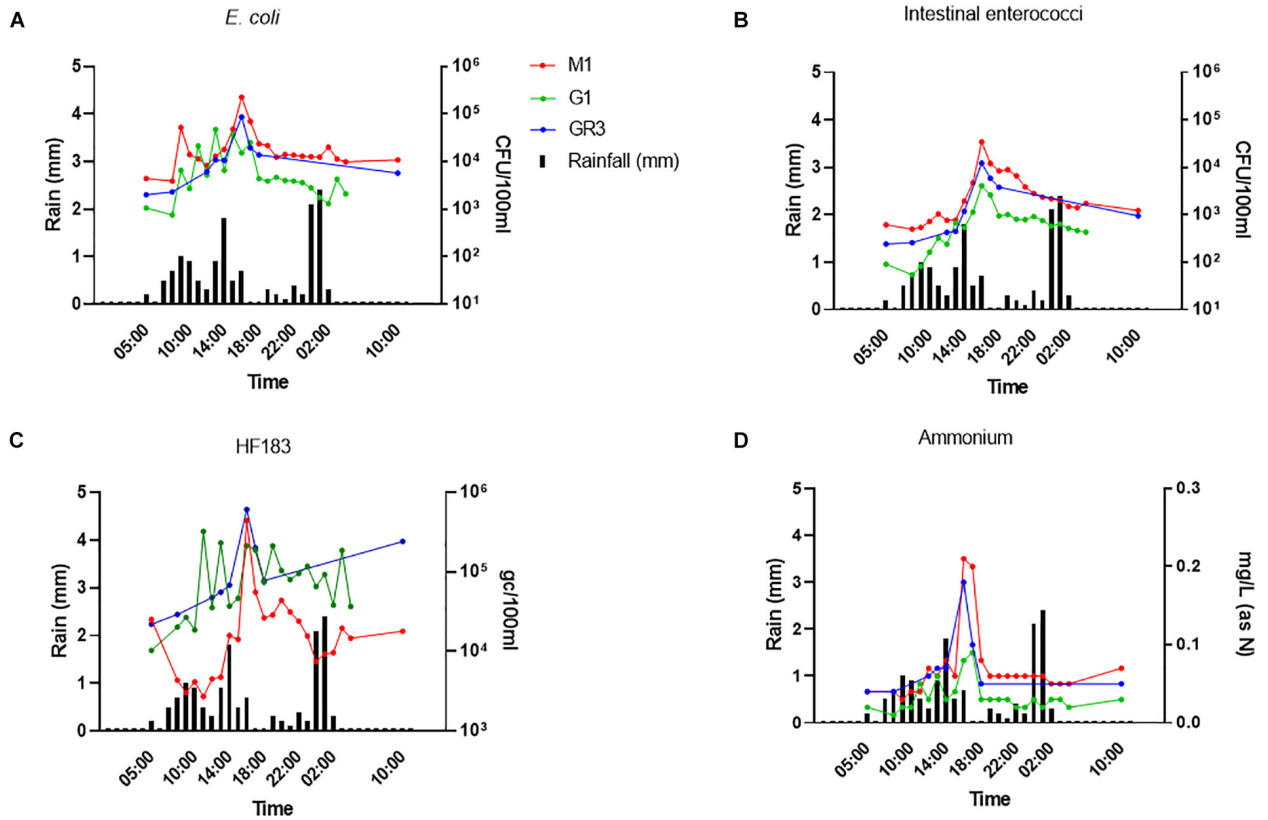


FIGURE 6 | The response of *E. coli* **(A)**, intestinal enterococci **(B)**, HF183 **(C)**, and **(D)** ammonium concentrations to rainfall at stations GR3, G1, and M1. Modelled rainfall is depicted as mm of rain per hour.

only correlations between intestinal enterococci and *E. coli* concentrations were observed at RV1 ($R^2 = 0.606$; p-value <0.05). No statistically significant correlations were observed between nitrate, TON and Phosphorus, and the faecal indicators at any stations. As these nutrients exhibited little variation this is expected and further supports these parameters being unsuitable as faecal indicators in the Elm Park catchment. These results indicate the usefulness of monitoring ammonium to determine faecal pollution in the catchment, particularly, following recent pollution events.

Cluster Analysis of High-Resolution Sampling Data Identifies "Sentinel" Stations

Cluster analysis of the geometric mean of each variable for each station resulted in the 11 stations being grouped into 4 statistically significant clusters. These four clusters represent the geographical grouping of the sampling stations of the Elm Park catchment. Cluster 1 (M3 and M2) represents the most upstream stations in the catchment, both cluster 2 (M1, G1, and GR3) and cluster 3 (GR2, GR1, and RV1) are in the middle of the catchment and cluster 4 (EP3, EP2, and RK1) typifies the culverted and downstream stations of the stream (**Figure 5**). Sampling of representative "sentinel" stations from each cluster may allow more efficient and rapid monitoring of faecal pollution in the Elm Park stream going forward as just four stations would be sampled rather than all 11. Nnane et al., employed a similar approach on the River Ouse in the United Kingdom whereby they reduced the 14 stations to 5 or 6 "sentinel" stations (Nnane, 2011; Nnane et al., 2011). Furthermore, as the tributaries (G1, RK1, and RV1) didn't group into distinct clusters this further supports the observations that faecal pollution is occurring predominantly in the main trunk of the Elm Park stream.

Principal component analysis was performed on individual clusters to identify the main influential variables of each (**Figure 5**). Three principal components were retained for each cluster according to Horn's Parallel analysis. Interestingly, for all four clusters, PC2 represented between 19 and 26% of the total variation and was driven by moderate loadings (>0.65) of nitrates and phosphorus (**Supplementary Table 4**). These variables may represent embankment erosion indicating that all clusters are impacted by diffuse runoff to comparable extents (Mellander et al., 2018). PC1 of cluster 1 (which represented 37% of variance) was primarily accounted for by positive moderate ammonium, nitrite, and TON loadings (>0.56). The positive loading of ammonium in this principal component indicates that it represents domestic sewage pollution in cluster 1 (Cabral and Marques, 2006; Baral et al., 2018). Indeed, M2 is downstream of a CSO and M3 is likely impacted by

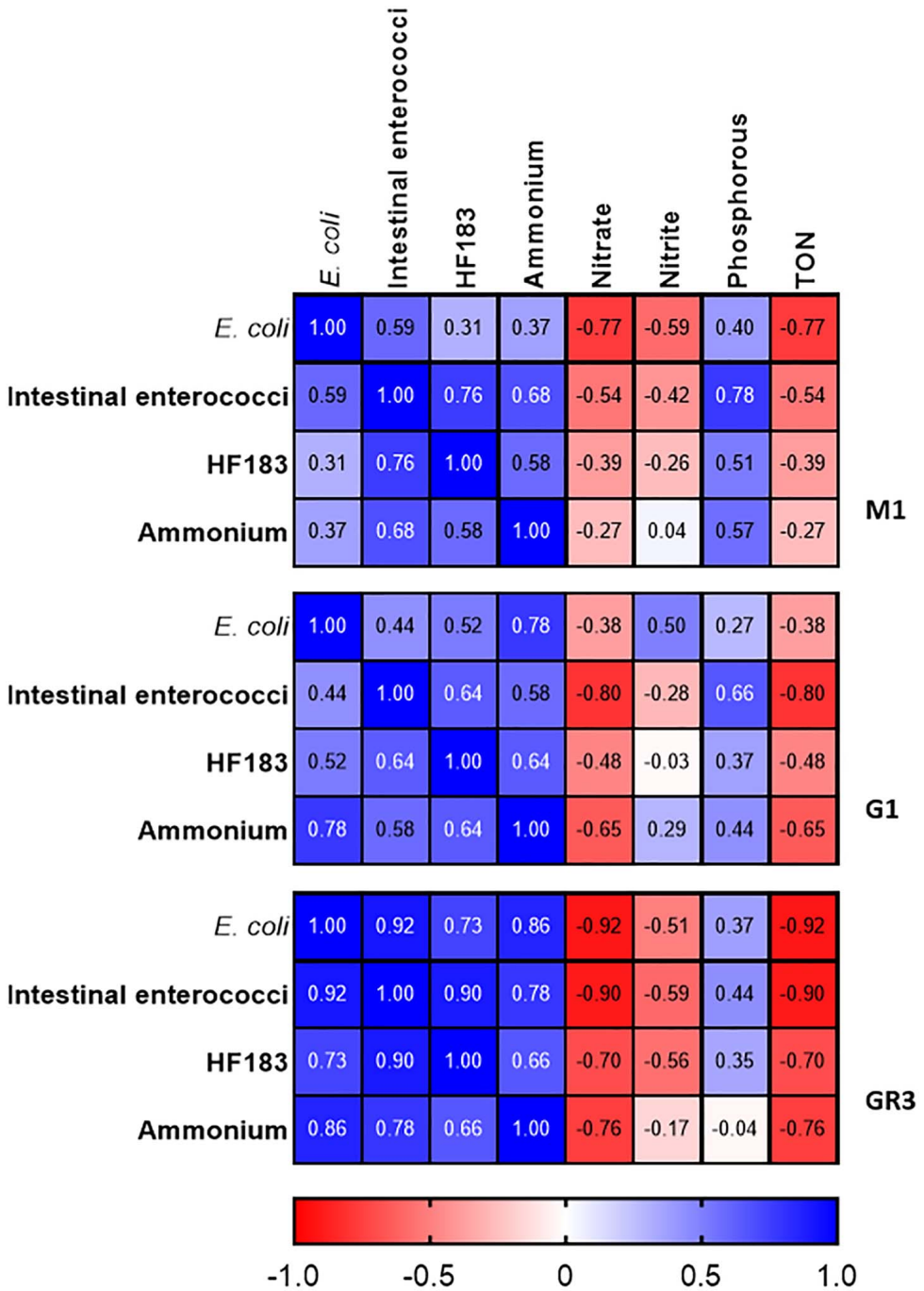


FIGURE 7 | Heatmap showing correlations between FIB, the HF183 marker and nutrients at station M1, G1, and GR3 during stormflow. Dark blue and dark red indicate strong positive and negative correlations, respectively. R^2 coefficients > 0.49 were considered statistically significant ($p < 0.05$).

diffuse domestic faecal pollution. PC1 represents 37 and 60% of the variation of cluster 2 and 3, respectively. In both cases moderate negative (−0.638 to −0.537) loadings of *E. coli* and intestinal enterococci are the primary factors. This indicates that stream volume, that dilutes these FIB, is a key factor for these clusters, particularly cluster 3. PC1 of cluster 4 (explaining 32% of variation) has moderate negative loadings (−0.600 to −0.583) of ammonium and nitrite. This cluster includes the stations EP3 and EP2 that are furthest from the CSO (which is key contributor of ammonium to the catchment) suggesting FIB influx at these stations is not concomitant with ammonium pollution.

This approach of using hierarchical clustering of high-resolution sampling data to identify "sentinel" stations can

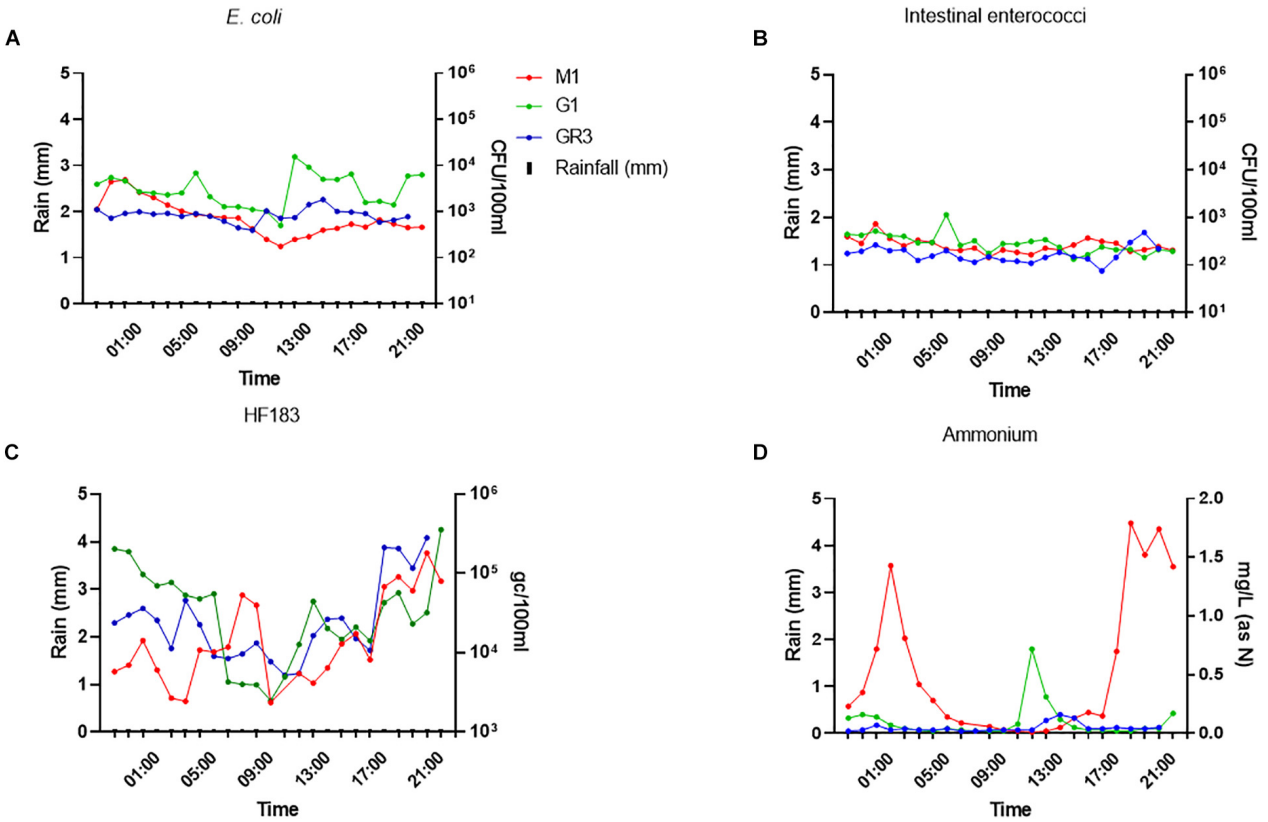


FIGURE 8 | Variation of panel **(A)** *E. coli* **(B)** intestinal enterococci, **(C)** the HF183 marker, and **(D)** ammonium concentrations during baseflow at stations GR3, G1, and M1. Modelled rainfall is depicted as mm of rain per hour.

provide water management authorities with a novel and more efficient faecal pollution monitoring system in other catchments.

Stormflow Influences Faecal Indicator Pollution and Correlates With Ammonium Concentrations in the Upstream Elm Park Catchment

To further assess the utility of ammonium and other nutrient parameters as rapid alternatives to FIB and MST marker monitoring, intensive sampling under stormflow (February 2020) and baseflow (June 2020) conditions at the three cluster 2 stations (M1, G1, and GR3) was undertaken. Over the stormflow period, three peaks in rainfall were modelled, whereas no rainfall was modelled during baseflow. During stormflow conditions, M1 and the downstream GR3 station responded similarly to increased rainfall. Peak concentrations of FIB in response to rainfall were observed at hour 10 of the storm at both stations, for *E. coli* (4.70×10^4–2.25×10^5 CFU/100 ml), intestinal enterococci (4.10×10^3–3.47×10^4 CFU/100 ml), HF183 (3.24×10^5–6.10×10^5 gc/100 ml), and ammonium [0.09–0.21 mg/L (as N)] (**Figures 6A–D**). Faecal indicator concentrations increased between 10- and 100-fold whereas ammonium concentrations increased over threefold. At station G1 intestinal enterococci and ammonium concentrations followed a similar response to rainfall as observed at M1 and GR3. Although *E. coli* and HF183 concentrations increased in response to rainfall, their levels were more variable at this station. Such intra-storm variability has been described previously for faecal indicators and antibiotic resistance genes in urban streams (Liao et al., 2015; Garner et al., 2017). Statistically significant correlations (R^2 = 0.437–0.917; $p < 0.05$) between FIB, HF183, and ammonium concentrations were observed at all stations during stormflow (**Figure 7** and **Supplementary Table 5**). Conversely, nitrate, nitrite, and TON followed the opposite trend to ammonium at all three stations in response to rainfall as concentrations decreased until hour 10 of the storm before increasing again. In fact, statistically significant negative correlations between nitrate and TON and FIB and the HF183 marker were observed ($R^2 = -0.917$ to -0.538; $p < 0.05$). Together, these data suggest that increased stream flow during stormflow conditions rapidly introduces ammonium rich human faeces into the Elm Park catchment which concomitantly dilutes oxidised nitrogen compounds. Thus, monitoring ammonium in real-time can be used to identify faecal contamination following storm events which can subsequently be confirmed by standard FIB enumeration.

Escherichia coli and intestinal enterococci concentrations at all three stations were more stable during baseflow than stormflow, ranging between 10^2–10^4 and 10^2–10^3 CFU/100 ml, respectively (**Figures 8A,B**). In contrast, HF183 concentrations at all stations

were more variable and had a diurnal pattern as they decreased until 9–10 a.m. before increasing again until the evening. A similar pattern of diurnality was observed for ammonium concentrations but only at station M1 (**Figures 8C,D**).

Surprisingly, ammonium concentrations exceeded 1.4 mg/L (as N) twice during baseflow sampling at M1, over sevenfold greater than the peak observed during stormflow conditions. In fact, nitrite and phosphorus concentrations were up to 10-fold higher at all stations during baseflow, which likely results from dilution of these nutrients during rainfall. At station M1, *E. coli* and intestinal enterococci correlated (R^2 = 0.578–0.704; $p < 0.0038$) with nitrate and nitrite concentrations during baseflow, which contrasts with stormflow observations (**Supplementary Table 5**). This station is closest to the M2 CSO and so these correlations indicate that the CSO may be slowly discharging sewage into the stream during baseflow. Under such conditions the sewage associated ammonium will be oxidised to nitrate and nitrite in the environment.

Together the monitoring of cluster 2 sampling stations during stormflow and baseflow revealed that all stations exhibit similar patterns of variability in faecal indicator levels, particularly in response to rainfall. This provides further evidence that sampling a single "sentinel" station from a cluster provides maximum information for that cluster. Furthermore, the strong correlations observed between FIB, MST, and ammonium levels highlight the effectiveness of monitoring ammonium as a rapid alternative to FIB and MST.

CONCLUSION

To conclude, we have demonstrated the usefulness of conducting high-resolution sampling to obtain a detailed picture of faecal pollution in a catchment. Further to this, we have demonstrated how these data can be used to reduce the geographical complexity of a catchment by employing hierarchical clustering to identify "sentinel" sampling stations. This "sentinel" sampling approach coupled with real-time monitoring of nutrients that correlate with faecal pollution in the catchment, such as ammonium, provides an alternative, and efficient approach to water management authorities to assess remediation efforts.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

LR wrote the draft of the manuscript. LR, NM, LS-C, TN, KC, PD, JO'S, CO'L, and PB carried out the experiments. LR, WM, PD, PB, CO'L, KC, JO'S, and LS-C conceived and planned the experiments. WM received funding for the project and supervised the project. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research (Acclimatize) was part funded by the European Regional Development Fund through the Ireland Wales Cooperation Programme.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.661954/full#supplementary-material

REFERENCES

Arnold, B. F., Schiff, K. C., Ercumen, A., Benjamin-Chung, J., Steele, J. A., Griffith, J. F., et al. (2017). Acute illness among surfers after exposure to seawater in dry- and wet-weather conditions. *Am. J. Epidemiol.* 186, 866–875. doi: 10.1093/aje/kwx019

Baird, R., and Bridgewater, L. (2017). *Standard Methods for the Examination of Water and Wastewater*, 23rd Edn. Washington, DC: American Public Health Association.

Balleste, 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

Baral, D., Speicher, A., Dvorak, B., Admiraal, D., and Li, X. (2018). Quantifying the relative contributions of environmental sources to the microbial community in an urban stream under dry and wet weather conditions. *Appl. Environ. Microbiol.* 84:e00896-18. doi: 10.1128/AEM.00896-18

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828.

Brinkmeyer, R., Amon, R. M., Schwarz, J. R., Saxton, T., Roberts, D., Harrison, S., et al. (2015). Distribution and persistence of *Escherichia coli* and enterococci in stream bed and bank sediments from two urban streams in houston, tx. *Sci. Total Environ.* 502, 650–658. doi: 10.1016/j.scitotenv.2014.09.071

Cabral, J. P., and Marques, C. (2006). Faecal coliform bacteria in febros river (northwest portugal): temporal variation, correlation with water parameters, and species identification. *Environ. Monit. Assess.* 118, 21–36.

EC (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. Available online at: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32000L0060 (accessed January 20, 2021).

EU (2006). *Directive 2006/7/EC of the European Parliament and the Council. Official Journal of the European Union, 4.3.2006, L64/37.* Available online at: http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_064/l_06420060304en00370051.pdf (accessed March 20, 2007).

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

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

Garner, E., Benitez, R., von Wagoner, E., Sawyer, R., Schaberg, E., Hession, W. C., et al. (2017). Stormwater loadings of antibiotic resistance genes in an urban stream. *Water Res.* 123, 144–152. doi: 10.1016/j.watres.2017.06.046

Goto, D. K., and Yan, T. (2011). Effects of land uses on fecal indicator bacteria in the water and soil of a tropical watershed. *Microbes Environ.* 26, 254–260. doi: 10.1264/jsme2.ME11115

Gourmelon, M., Caprais, M. P., Segura, R., Le Mennec, C., Lozach, S., Piriou, J. Y., et al. (2007). Evaluation of two library-independent microbial source tracking methods to identify sources of fecal contamination in french estuaries. *Appl. Environ. Microbiol.* 73, 4857–4866. doi: 10.1128/AEM.03003-06

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

He, H., Zhen, Y., Mi, T., Fu, L., and Yu, Z. (2018). Ammonia-oxidizing archaea and bacteria differentially contribute to ammonia oxidation in sediments from adjacent waters of Rushan bay, China. *Front. Microbiol.* 9:116. doi: 10.3389/fmicb.2018.00116

Hughes, B., Beale, D. J., Dennis, P. G., Cook, S., and Ahmed, W. (2017). Cross-comparison of human wastewater-associated molecular markers in relation to fecal indicator bacteria and enteric viruses in recreational beach waters. *Appl. Environ. Microbiol.* 83:e0028-17. doi: 10.1128/AEM.00028-17

Jin, Z., Zhang, X., Li, J., Yang, F., Kong, D., Wei, R., et al. (2017). Impact of wastewater treatment plant effluent on an urban river. *J. Freshw. Ecol.* 32, 697–710. doi: 10.1080/02705060.2017.1394917

Kauppinen, A., Al-Hello, H., Zacheus, O., Kilponen, J., Maunula, L., Huusko, S., et al. (2017). Increase in outbreaks of gastroenteritis linked to bathing water in finland in summer 2014. *Euro. Surveill.* 22:30470. doi: 10.2807/1560-7917.ES.2017.22.8.30470

Liao, H., Krometis, L. H., Cully Hession, W., Benitez, R., Sawyer, R., Schaberg, E., et al. (2015). Storm loads of culturable and molecular fecal indicators in an inland urban stream. *Sci. Total Environ.* 530-531, 347–356. doi: 10.1016/j.scitotenv.2015.05.098

Lim, F., Ong, S., and Hu, J. (2017). Recent advances in the use of chemical markers for tracing wastewater contamination in aquatic environment: a review. *Water* 9:143. doi: 10.3390/w9020143

Mallin, M. A., and McIver, M. R. (2012). Pollutant impacts to cape Hatteras National Seashore from urban runoff and septic leachate. *Mar. Pollut. Bull.* 64, 1356–1366. doi: 10.1016/j.marpolbul.2012.04.025

Mayer, R. E., Bofill-Mas, S., Egle, L., Reischer, G. H., Schade, M., Fernandez-Cassi, X., et al. (2016). Occurrence of human-associated bacteroidetes genetic source tracking markers in raw and treated wastewater of municipal and domestic origin and comparison to standard and alternative indicators of faecal pollution. *Water Res.* 90, 265–276. doi: 10.1016/j.watres.2015.12.031

Mellander, P. E., Jordan, P., Bechmann, M., Fovet, O., Shore, M. M., McDonald, N. T., et al. (2018). Integrated climate-chemical indicators of diffuse pollution from land to water. *Sci. Rep.* 8:944. doi: 10.1038/s41598-018-19143-1

Molina, M., Hunter, S., Cyterski, M., Peed, L. A., Kelty, C. A., Sivaganesan, M., et al. (2014). Factors affecting the presence of human-associated and fecal indicator real-time quantitative pcr genetic markers in urban-impacted recreational beaches. *Water Res.* 64, 196–208. doi: 10.1016/j.watres.2014.06.036

Muller, A., Osterlund, H., Marsalek, J., and Viklander, M. (2020). The pollution conveyed by urban runoff: a review of sources. *Sci. Total Environ.* 709:136125. doi: 10.1016/j.scitotenv.2019.136125

Napier, M. D., Haugland, R., Poole, C., Dufour, A. P., Stewart, J. R., Weber, D. J., et al. (2017). Exposure to human-associated fecal indicators and self-reported illness among swimmers at recreational beaches: a cohort study. *Environ Health.* 16:103.

Nnane, D. E. (2011). Sustainable microbial water quality monitoring programme design using phage-lysis and multivariate techniques. *Sci. Total Environ.* 409, 5188–5195. doi: 10.1016/j.scitotenv.2011.08.057

Nnane, D. E., Ebdon, J. E., and Taylor, H. D. (2011). Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. *Water Res.* 45, 2235–2246. doi: 10.1016/j.watres.2011.01.018

Olds, H. T., Corsi, S. R., Dila, D. K., Halmo, K. M., Bootsma, M. J., and McLellan, S. L. (2018). High levels of sewage contamination released from urban areas after storm events: a quantitative survey with sewage specific bacterial indicators. *PLoS Med.* 15:e1002614. doi: 10.1371/journal.pmed.1002614

Paule-Mercado, M. A., Ventura, J. S., Memon, S. A., Jahng, D., Kang, J. H., and Lee, C. H. (2016). Monitoring and predicting the fecal indicator bacteria concentrations from agricultural, mixed land use and urban stormwater runoff. *Sci. Total Environ.* 550, 1171–1181. doi: 10.1016/j.scitotenv.2016.01.026

Preisner, M. (2020). Surface water pollution by untreated municipal wastewater discharge due to a sewer failure. *Environ. Process.* 7, 767–780. doi: 10.1007/s40710-020-00452-5

Rutledge, R. G., and Côté, C. (2003). Mathematics of quantitative kinetic pcr and the application of standard curves. *Nucleic Acids Res.* 31:e93. doi: 10.1093/nar/gng093

Rutledge, R. G., and Stewart, D. (2008). Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time pcr. *BMC Mol. Biol.* 9:96. doi: 10.1186/1471-2199-9-96

Sauer, E. P., VandeWalle, J. L., Bootsma, M. J., and McLellan, S. L. (2011). Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. *Water Res.* 45, 4081–4091. doi: 10.1016/j.watres.2011.04.049

Saxena, G., Bharagava, R. N., Kaithwas, G., and Raj, A. (2015). Microbial indicators, pathogens and methods for their monitoring in water environment. *J. Water Health* 13, 319–339. doi: 10.2166/wh.2014.275

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 *Bacteroides* 16s rRNA genetic marker with real-time pcr for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* 7, 249–259. doi: 10.1111/j.1462-2920.2004.00702.x

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

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542.

Tillett, B. J., Sharley, D., Almeida, M., Valenzuela, I., Hoffmann, A. A., and Pettigrove, V. (2018). A short work-flow to effectively source faecal pollution in recreational waters - a case study. *Sci. Total Environ.* 644, 1503–1510. doi: 10.1016/j.scitotenv.2018.07.005

UN (2019). *World Urbanization Prospects: The 2018 Revision. Department of Economic and Social Affairs, Population Division (ST/ESA/SER.A/420).* New York: United Nations.

Unno, T., Staley, C., Brown, C. M., Han, D., Sadowsky, M. J., and Hur, H.-G. (2018). Fecal pollution: new trends and challenges in microbial source tracking using next-generation sequencing. *Environ. Microbiol.* 20, 3132–3140.

van der Kooij, S., van Vliet, B. J. M., Stomph, T. J., Sutton, N. B., Anten, N. P. R., and Hoffland, E. (2020). Phosphorus recovered from human excreta: a socio-ecological-technical approach to phosphorus recycling. *Resour. Conserv. Recycl.* 157:104744. doi: 10.1016/j.resconrec.2020.104744

Vrzel, J., Vukovic-Gacic, B., Kolarevic, S., Gacic, Z., Kracun-Kolarevic, M., Kostic, J., et al. (2016). Determination of the sources of nitrate and the microbiological sources of pollution in the Sava river basin. *Sci. Total Environ.* 573, 1460–1471. doi: 10.1016/j.scitotenv.2016.07.213

Walker, J. W., van Duivenboden, R., and Neale, M. W. (2015). A tiered approach for the identification of faecal pollution sources on an Auckland urban beach. *N. Z. J. Mar. Freshw. Res.* 49, 333–345. doi: 10.1080/00288330.2015.1014376

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.

Copyright © 2021 Reynolds, Martin, Sala-Comorera, Callanan, Doyle, O'Leary, Buggy, Nolan, O'Hare, O'Sullivan and Meijer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 July 2021
doi: 10.3389/fmicb.2021.668778

Genetic Microbial Source Tracking Support QMRA Modeling for a Riverine Wetland Drinking Water Resource

Julia Derx[1†], Katalin Demeter[2†], Rita Linke[2†], Sílvia Cervero-Aragó[3†], Gerhard Lindner[1†], Gabrielle Stalder[4], Jack Schijven[5,6], Regina Sommer[3†], Julia Walochnik[7], Alexander K. T. Kirschner[3,8†], Jürgen Komma[1], Alfred P. Blaschke[1†] and Andreas H. Farnleitner[2,8†]*

[1] Institute of Hydraulic Engineering and Water Resources Management, Vienna, Austria, [2] Research Group Environmental Microbiology and Molecular Diagnostics E166/5/3, Institute of Chemical, Environmental and Bioscience Engineering, Vienna, Austria, [3] Institute for Hygiene and Applied Immunology, Medical University of Vienna, Vienna, Austria, [4] Institute of Wildlife Ecology, University of Veterinary Medicine, Vienna, Austria, [5] Department of Statistics, Informatics and Modelling, National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands, [6] Faculty of Geosciences, Department of Earth Sciences, Utrecht University, Utrecht, Netherlands, [7] Institute of Specific Prophylaxis and Tropical Medicine, Medical University of Vienna, Vienna, Austria, [8] Division Water Quality and Health, Department of Pharmacology, Physiology, and Microbiology, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Matthew Verbyla, San Diego State University, United States
Victor Gannon, Public Health Agency of Canada (PHAC), Canada

***Correspondence:**
Julia Derx
derx@hydro.tuwien.ac.at

† Interuniversity Cooperation Centre Water and Health (www.waterandhealth.at)

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *17 February 2021*
Accepted: *18 June 2021*
Published: *14 July 2021*

Citation:
Derx J, Demeter K, Linke R, Cervero-Aragó S, Lindner G, Stalder G, Schijven J, Sommer R, Walochnik J, Kirschner AKT, Komma J, Blaschke AP and Farnleitner AH (2021) Genetic Microbial Source Tracking Support QMRA Modeling for a Riverine Wetland Drinking Water Resource. Front. Microbiol. 12:668778. doi: 10.3389/fmicb.2021.668778

Riverine wetlands are important natural habitats and contain valuable drinking water resources. The transport of human- and animal-associated fecal pathogens into the surface water bodies poses potential risks to water safety. The aim of this study was to develop a new integrative modeling approach supported by microbial source tracking (MST) markers for quantifying the transport pathways of two important reference pathogens, *Cryptosporidium* and *Giardia,* from external (allochthonous) and internal (autochthonous) fecal sources in riverine wetlands considering safe drinking water production. The probabilistic-deterministic model QMRAcatch (v 1.1 python backwater) was modified and extended to account for short-time variations in flow and microbial transport at hourly time steps. As input to the model, we determined the discharge rates, volumes and inundated areas of the backwater channel based on 2-D hydrodynamic flow simulations. To test if we considered all relevant fecal pollution sources and transport pathways, we validated QMRAcatch using measured concentrations of human, ruminant, pig and bird associated MST markers as well as *E. coli* in a Danube wetland area from 2010 to 2015. For the model validation, we obtained MST marker decay rates in water from the literature, adjusted them within confidence limits, and simulated the MST marker concentrations in the backwater channel, resulting in mean absolute errors of < 0.7 $\log_{10}$ particles/L (Kruskal–Wallis $p > 0.05$). In the scenarios, we investigated (i) the impact of river discharges into the backwater channel (allochthonous sources), (ii) the resuspension of pathogens from animal fecal deposits in inundated areas, and (iii) the pathogen release from animal fecal deposits after rainfall (autochthonous sources). Autochthonous and allochthonous human and animal sources

resulted in mean loads and concentrations of *Cryptosporidium* and *Giardia* (oo)cysts in the backwater channel of 3–13 × 10^9 particles/hour and 0.4–1.2 particles/L during floods and rainfall events, and in required pathogen treatment reductions to achieve safe drinking water of 5.0–6.2 $\log_{10}$. The integrative modeling approach supports the sustainable and proactive drinking water safety management of alluvial backwater areas.

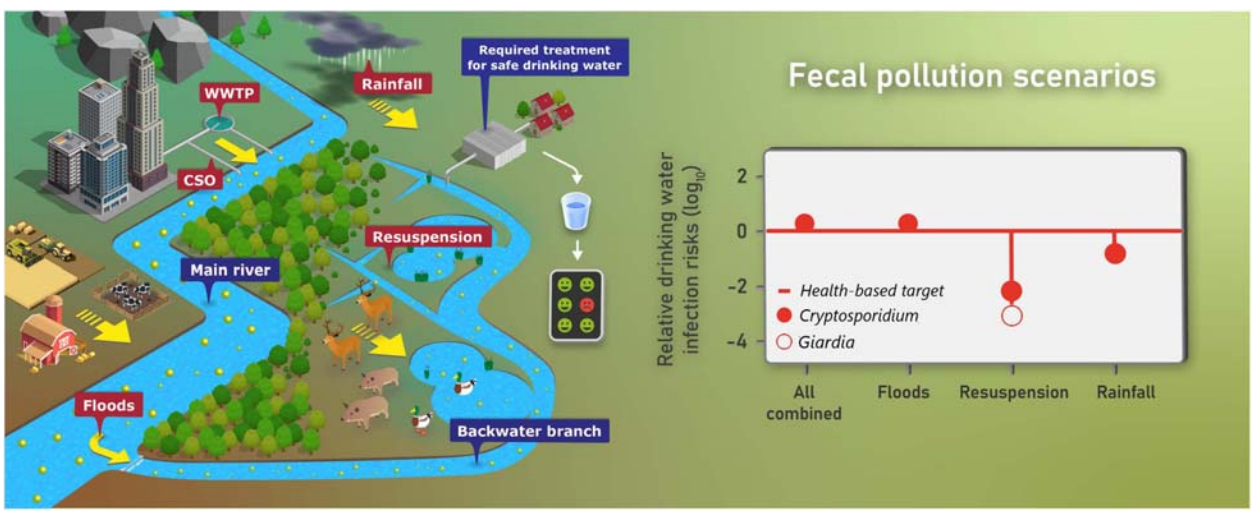
Keywords: genetic microbial source tracking markers, microbial fate and transport model, hydrodynamic model, *Cryptosporidium*, *Giardia*, QMRA, microbial decay in environment

INTRODUCTION

Alluvial backwater areas along large rivers are important as natural habitats, for flood protection, and contain valuable resources for drinking water production. The impacts of urban wastewater discharges, and livestock and wildlife fecal deposits pose substantial hazards to these ecosystems concerning water supply (Hogan et al., 2012; Frick et al., 2020). River water impacted by human wastewater and/or diffusive animal sources may transport fecal pathogens into the backwater area from outside (allochthonous sources). Inside the backwater area, pathogens may be released from animal fecal deposits following rainfall events, or may be resuspended in inundated areas during floods (autochthonous sources). Several authors have studied the impact of allochthonous and anthroponotic pathogens, such as human-specific viruses, on the microbiological quality of wetland water resources considering safe drinking water production (Sanders et al., 2005; Daniels et al., 2014; Liu et al., 2015; Derx et al., 2016). In this respect, however, there is little known about the relative contribution of animal versus human and autochthonous versus allochthonous sources of fecal pollution.

The microbial transport and removal mechanisms in wetlands have been primarily studied for fecal indicator organisms (FIO) (Sanders et al., 2005; Liu et al., 2015). Solely relying on FIO, however, does not allow differentiating the impact of different sources, as FIO occur in all human and animal sources. Microbial source tracking markers (MST) provide therefore immensely valuable information to identify fecal pollution sources for wetlands. This was demonstrated, e.g., by Frick et al. (2020), who linked FIO data with host-associated MST data and river connectivity in an alluvial Danube wetland. If combined with pathogen data, host-associated microbial source tracking markers of sufficient specificity and sensitivity can greatly support the source-targeted calibration of microbial fate and transport models and support a health risk assessment, as shown by Derx et al. (2016) and Demeter et al. (2021). So far, the spectrum of available animal and human MST marker assays has not yet been exploited in this context. In combination with FIO and zoonotic pathogen data, host-associated MST markers could significantly support microbial fate and transport modeling and microbial infection risk assessments in wetlands.

Important transport processes in wetlands are the advection, release and resuspension, decay and settling of microbial particles or chemical substances (Pavlik et al., 1999; Grant et al., 2001; Hogan et al., 2012; Peterson and Hanna, 2016). Further influencing factors for the retention mechanisms are water temperature, turbidity, salinity, and vegetation cover (Daniels et al., 2014). The flow patterns driving these transport processes in riverine or coastal wetlands vary in space and time, and are commonly simulated based on multi-dimensional, hydrodynamic flow and transport models (Sanders et al., 2005;



GRAPHICAL ABSTRACT

Liu et al., 2015). Due to the uncertainty of the source and transport variables, several studies conducted in wetlands used probabilistic-deterministic approaches to model microbial fate and transport. Daniels et al. (2014) developed a deterministic pathogen transport model for wetlands within a Bayesian statistical framework. Schijven et al. (2015) developed the probabilistic-deterministic microbial fate, transport and infection risk model QMRAcatch. This model was later applied by Derx et al. (2016) for evaluating the impact of human fecal pollution in a large riverine wetland. Daniels et al. (2014), Schijven et al. (2015), and Derx et al. (2016), however, did not account for the complex spatiotemporal variations of the flow and transport processes in wetlands.

The primary aim of this study was to develop a new integrative modeling approach supported by MST markers for quantifying the impact of human and animal, as well as, autochthonous and allochthonous fecal sources on the microbiological quality of an alluvial backwater channel considering safe drinking water production. The model should be able to account for the transient spatiotemporal variations of flow patterns and for the uncertainty of the source and transport variables. The secondary aim was to test the developed model at a Danube backwater area supported by MST markers and *E. coli* for quantifying the concentrations and loads of *Cryptosporidium* and *Giardia* (oo)cysts considering safe drinking water. These are important reference protozoa occurring ubiquitously in human and animal fecal sources (Stalder et al., 2011). *Cryptosporidium* and *Giardia* (oo)cysts are highly persistent in the environment and infectious at low dose (de Regnier et al., 1989; Boyer et al., 2009), making them suitable to address our research question. To meet our aims, we modified and extended QMRAcatch (v 1.1 python backwater). We defined event-driven scenarios of microbial transport into the backwater channel, (i) via the river entering the backwater during floods (allochthonous sources), (ii) via the resuspension from fecal deposits during flooding and inundation of the backwater area, and, (iii) via the release and runoff from fecal deposits during rainfall within the backwater area (autochthonous sources). These were assumed to be the most important transport pathways, and human wastewater, ruminants, wild boar (pigs) and birds were considered as the most important fecal pollution sources, based on the findings of Frick et al. (2020) in our study area. To validate this assumption, we used a comprehensive 6-year monthly monitoring dataset of human and animal MST markers and *E. coli*.

MATERIALS AND METHODS

Study Area

The investigated alluvial model backwater area is situated at the Danube at the downstream end of the city of Vienna, in Austria (**Figure 1**, see Frick et al., 2020 for a detailed description). The catchment upstream of Vienna is home to approximately 11 million inhabitants (Schreiber et al., 2005). Considering that 99% of the human population in the study area is connected to a WWTP (European Commission, 2018), urban wastewater is the main source of human pollution. Due to the regulation of the Danube, the backwater area has been almost completely disconnected from the main stream. The backwater presently consists of a channel network, whose main lateral branch has a surface connection with the Danube River at its lowermost end through a levee opening ('entry point during floods', **Figure 1**). Floods that enter the floodplain via the backflow connection move upstream along the main lateral branch, creating a distinct gradient in hydrologic connectivity in the various waterbodies with distance to the inflow (Reckendorfer et al., 2013). The backwater channel connects with the Danube River if the discharge exceeds 2200 m^3/s (mean discharge: 1900 m^3/s). Danube water enters the backwater channel and its lateral branches during the rising limb of flood events, causing inundation of parts of the surface area (**Figure 1**). When the flood peak is reached, the flow direction reverses, and the water flows back toward the Danube River.

The backwater area is 14 km^2 in size, and represents an important water resource for drinking water supply for Vienna (Hein et al., 2006). There are five drinking water wells (riverbank filtrate) situated in the area. It is also part of a national park that plays a strategic role as a wilderness area and for recreation (Arnberger et al., 2009). There is no livestock in the considered backwater area, but there is a considerable population of wild animals, such as ruminants, wild boars and birds (Frühauf and Sabathy, 2006a,b; Parz-Gollner, 2006; Arnberger et al., 2009). Population sizes estimated for 2010 resulted in 180 red deer (*Cervus elaphus*), 44 roe deer (*Capreolus capreolus*), 17 fallow deer (*Dama dama*), 20 European mouflon (*Ovis orientalis musimon*) and 150 wild boars (*Sus scrofa*) (Government of the City of Vienna, Alexander Faltejsek, personal communication). Hunting is allowed in the area. The bird abundance is nearly 2500 at maximum in total (Frühauf and Sabathy, 2006a,b; Parz-Gollner, 2006; Schulze and Schütz, 2013). In addition, 600,000 visitors visit the national park area every year (Hinterberger et al., 2000). Bathing is prohibited at the considered backwater area, and visitors may only use selected paths for walking and cycling. Therefore, the possibility for human fecal input from visitors is considered to be low (Frick et al., 2018).

Modeling Approach

For this paper, we used a modified and extended version of the probabilistic-deterministic microbial fate and transport and infection risk model QMRAcatch (Schijven et al., 2015; Demeter et al., 2021; **Figure 2**). We took a two-step modeling approach: (1) To test the assumption of the most relevant fecal sources and transport processes, we performed a source-targeted model validation using measured concentrations of host-associated MST markers and *E. coli* at the study site (Section "Validation of the Microbial Fate and Transport Model"). (2) To evaluate the importance of animal versus human, or autochthonous versus allochthonous sources in terms of potential health relevance, we then simulated various pollution scenarios (Section "Scenario Load and Infection Risk Assessment"). In these scenarios, we simulated the concentrations and loads of *Cryptosporidium* and *Giardia* in the backwater channel, and the required pathogen treatment reduction and daily drinking

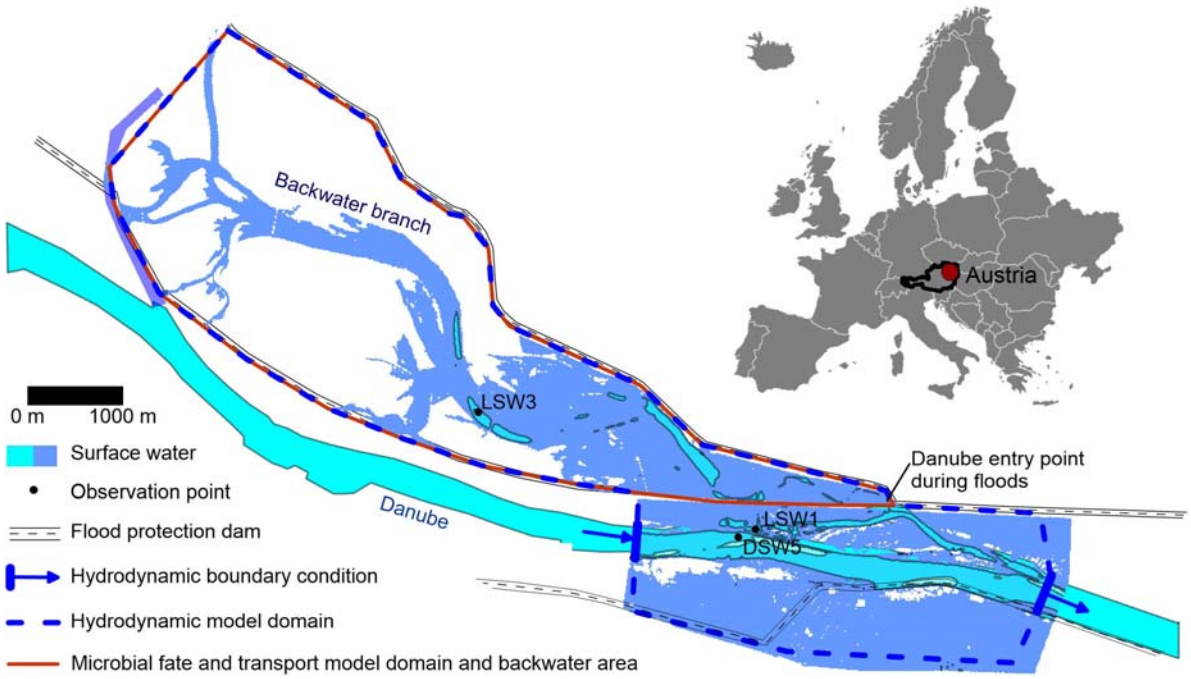


FIGURE 1 | Study area with observation points DSW5 (Danube), LSW1 (affected by Danube discharge and inundation) and LSW 3 (affected by rainfall-release from fecal deposits) (Frick et al., 2020), QMRAcatch and hydrodynamic model domain; the surface water area simulated during the peak of the flood event in January 2011 is indicated in blue.

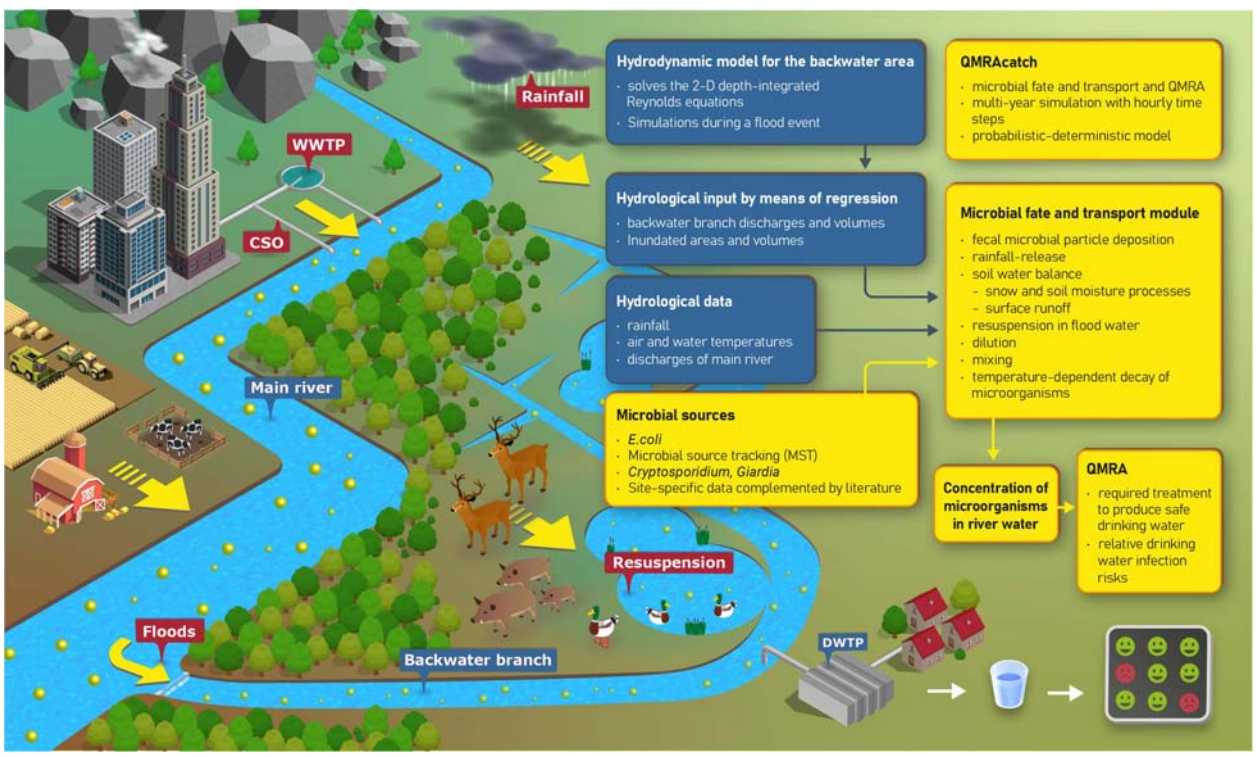


FIGURE 2 | Microbial sources and transport processes in the alluvial backwater area (graphics). Components and data flows of the integrative modeling approach (flow chart).

water infection risks relative to a health-based benchmark. The model domain of QMRAcatch encompasses the total backwater area, delimited by the Danube along the southern boundary and a flood protection dam on its northern end (**Figure 1**). Hydrological monitoring data together with 2-D hydrodynamic flow simulations during a flood event (Section "Hydrological and Hydrodynamic Flow Situation") as well as measured microbial source data complemented with literature data (Section "Microbiological Source Characterization") served as input to the model.

QMRAcatch

Model Overview

The probabilistic-deterministic microbial fate and transport and infection risk model QMRAcatch (Schijven et al., 2015) was extended for this study and coded as open source (v1.1 Python backwater). QMRAcatch was used to simulate the microbial concentrations of the backwater channel (Sections "Microbial Fate and Transport Module" to "Microbiological Source Characterization") and the daily drinking water infection risks after further treatment of the source water (Section "Quantitative Microbial Risk Assessment Module"). The model comprises the functionality of the version QMRAcatch 1.0 Python (Demeter et al., 2021) with the following extensions:

- Simulation time steps of 1 h in contrast to 1 day in the previous version.
- The concentration of MST markers, fecal indicators and pathogens in the main river (Danube) water is described by a statistical distribution according to the observed data set.
- Microbial particle release from animal fecal deposits is described as a function of precipitation and elapsed time since the start of precipitation according to Bradford and Schijven (2002).
- Hourly discharges and volumes of the backwater channel, flooded areas and floodplain volumes are additional input variables.
- Model equations were included to calculate surface runoff based on transient evaporation and soil moisture processes using air temperature and rainfall as input variables.
- Microbial decay in animal feces is described based on a uniformly distributed first order rate coefficient μ_f. In the earlier version, the decay rates in water and feces were not differentiated.
- The prevalence of the reference pathogens *Cryptosporidium* and *Giardia* in animal waste is described by a mixture of beta distributions from reported studies following the methodology of Dorner et al. (2004).
- Data on times of consumption and consumed volumes of unboiled drinking water per person per day by the Dutch National Food Consumption Survey 2007–2010 (DNFCS) (Van Rossum et al., 2011) are used to calculate the daily cumulative dose and the daily drinking water infection risks.

Microbial Fate and Transport Module

Microbial contamination of the backwater channel occurs from the following three reservoirs: Danube, non-flooded and flooded area. Microbial particles are transported via inflows of Danube water during floods, by resuspension from animal deposits in inundated areas, and during rainfall causing release and runoff from animal fecal deposits in non-flooded areas.

Transport via the Danube entering the backwater during floods

Microorganisms carried by Danube water are subjected to mixing with the backwaters and temperature-dependent decay. Assuming steady-state conditions and complete mixing on each hour, the analytical solution for the microbial concentration in the backwater channel, $C_{r\rightarrow bw}$ [particles/L] at time step t is (Schijven et al., 2015):

$$C_{r\rightarrow bw}(t) = \frac{Q_{bw}}{Q_{bw}+\mu_w(T_{bw})V_{bw}}C_r + \left(C_{r\rightarrow bw}(t-1) - \frac{Q_{bw}}{Q_{bw}+\mu_w(T_{bw})V_{bw}}C_r\right)\exp\left(-\frac{Q_{bw}+\mu_w(T_{bw})V_{bw}}{V_{bw}}t\right) \quad (1)$$

where C_r [particles/L] is the microbial particle concentration of the Danube river, Q_{bw} and V_{bw} are the discharge and volume of the backwater channel (Section "Hydrological and Hydrodynamic Flow Situation," **Supplementary Table 2**). The degree of reduction of microorganisms during the transport depends on the travel time or flow rate. Decay during transport is described as a first order reaction, where the decay rate in water (μ_w [1/d]) is a function of the water temperature (T_{bw} [°C]):

$$\mu_w(t) = \frac{ln10}{10^{a_0+a_1T_{bw}}} \quad (2)$$

where a_0 [$\log_{10}$ day] and a_1 [$\log_{10}$ day/°C] are microorganism-specific decay rate parameters (Bertrand et al., 2012). We conducted an extensive literature review on the microorganism-specific decay rates in water and adjusted the values within prediction intervals (**Supplementary Table 4**, **Figure 4**). The rates implicitly included additional removal or regrowth processes, or effects of other environmental factors, such as temperature, pH, TOC (van Elsas et al., 2011; Ahmed et al., 2019).

Microbial particle deposition

Microbial particle loadings from animals were determined using the method described by **Dorner et al. (2004)** and **Sterk et al. (2016)**. For pathogens, the fraction of animals infected by *Cryptosporidium* or *Giardia* was derived by random sampling from several (equally weighted) beta-distributions, describing the probability that an animal is positive (prev):

$$prev \sim \beta(\alpha, \beta) \quad (3)$$

Parameters a and b for each of the b-distributions are based upon prevalence studies (see **Supplementary Table 1**). In the selection process, studies were prioritized based on their recentness, number of samples and location. Studies conducted in temperate, high-income regions were given priority.

TABLE 1 | Microorganism-specific model input parameters in QMRAcatch.

Parameter	Dimension	Details	Distribution type	Microorganism	Value	References
Concentrations of feces C_f (mean, 95^{th} percentile)	N/g	Deer sources	Gamma	*E. coli*	(3, 16) × 10^7	Farnleitner et al., 2010
				Human MST	(3, 11) × 10^3	Farnleitner et al., 2014
				Ruminant MST	(2, 6) × 10^9	Farnleitner et al., 2014
				Pig MST	(3, 20) × 10^3	Farnleitner et al., 2014
				Duck MST	(7, 52) × 10^3	Farnleitner et al., 2014
				C. bovis + *C. ryanae*	103, 225	Garcia-Presedo et al., 2013
				Giardia duodenalis A-II	89, 320	Garcia-Presedo et al., 2013
		Pig sources	Gamma	*E. coli*	(0.04, 2.5) × 10^8	Frick et al., 2018
				Human MST	(7, 33) × 10^3	Farnleitner et al., 2014
				Ruminant MST	0, 0	Farnleitner et al., 2014
				Pig MST	(2, 10) × 10^{10}	Farnleitner et al., 2014
				Duck MST	0, 0	Farnleitner et al., 2014
				Cryptosporidium	70, 133	Castro-Hermida et al., 2011
				Giardia	8, 10	Castro-Hermida et al., 2011
		Bird sources	Gamma	*E. coli*	(0.001, 3.2) × 10^8	Frick et al., 2018
				Human MST	(2, 6) × 10^3	Farnleitner et al., 2014
				Ruminant MST	(1, 9) × 10^3	Farnleitner et al., 2014
				Pig MST	(8, 30) × 10^4	Farnleitner et al., 2014
				Duck MST	(1.3, 4.5) × 10^7	Farnleitner et al., 2014
				Cryptosporidium parvum	288, 686	for Canada migratory geese (Graczyk et al., 1998)
				Giardia	405, 786	for Canada migratory geese (Graczyk et al., 1998)
a_0, a_1	log_{10} day, log_{10} day/ °C	Parameters to describe first order decay in water as function of water temperature	Constant	Human MST marker	0.78, −0.03	This study (best fit)
				Ruminant MST marker	0.79, −0.037	This study (best fit)
				Pig MST marker	1.7, −0.039	This study (best fit)
				Bird MST marker	1.85, −0.039	This study (best fit)
				E. coli	1.04, −0.017	Franz et al., 2014
				Cryptosporidium	3.3, −0.076	Ives et al., 2007
				Giardia	2.16, −0.07	de Regnier et al., 1989
μ_f minimum, maximum	1/d	first order decay in feces in all animal sources based on reported values in bovine feces	Uniform	Ruminant MST marker	−0.13, −0.11	Oladeinde et al., 2014
				E. coli	−0.3, +0.1	Oladeinde et al., 2014
				Cryptosporidium	−0.05, −0.03	Olson et al., 1999
				Giardia	−0.38, −0.11	Olson et al., 1999
a, β	1/mm, -	release parameters		All microorganisms	0.1, 2.0	Bradford and Schijven, 2002; Guber et al., 2015
Dose-response model parameters						
α, β	-	hypergeometric		*Cryptosporidium*	0.3, 1.1	Schijven et al., 2011, 2015
r	-	exponential		*Giardia*	0.02	Regli et al., 1991

Decay rate coefficients in fresh water and in feces (Equations 2 and 9), release parameters (Equation 4), and dose-response parameters α and β (Equation 13).

For each infected animal, the microbial particle numbers shed per hour were determined by multiplying the mass of feces per dropping m_f (kg, normally distributed), the number of droppings per hour per animal (Poisson distributed) and the microbial concentration (**Tables 1**, **2**). The microbial concentration in feces is described by a gamma distribution based on mean and 95th percentile values reported in the literature (**Table 1**). Fecal deposition only takes place in the non-flooded area of the floodplain.

Transport in non-flooded area

Rainfall-induced microbial particle release was modeled iteratively using the function for the release of *Cryptosporidium*

TABLE 2 | Input parameter settings for the animal sources in QMRAcatch.

Parameter	Unit	Value	References	Value	References	Value	References
		Deer		*Wild boar*		*Birds*	
Population size	N	240	Vierheilig et al., 2013; Böhm, 2016	200	Vierheilig et al., 2013, MA 49, personal communication	2,500	Frühauf and Sabathy, 2006a,b; Parz-Gollner, 2006; Schulze and Schütz, 2013
Weight (Mean, 95%)	g	15,30	von Oheimb et al., 2005	10, 20	Schmidt et al., 2004	0.5, 1	Hahn et al., 2007
Defecation rate	1/h/animal	0.63	von Oheimb et al., 2005	0.2	Schmidt et al., 2004	2.1	Hahn et al., 2007

and *Giardia* from dairy cattle manure of Bradford and Schijven (2002). The release rate is given by:

$$\omega(t) = aP(t)[1 + a\beta P\,(t)\,t_{rain}\,(t)]^{-(1+\frac{1}{\beta})} \quad (4)$$

according to Guber et al. (2015), where P (mm/h) is the amount of rainfall at time step t, t_{rain} is the time passed since the start of the rainfall event, a (1/mm) controls the initial release rate, and β (–) determines the shape of the release curve (**Table 1**). Release from the deposits occurs in the non-flooded area (A_{dep} [m^2]). First, the available number of microbial particles in animal fecal deposits N_{deptot} is determined from the newly deposited numbers (N_{dep}) plus the residual deposits of the previous time step (N_{depres}, Equation 9):

$$N_{deptot}\,(t) = N_{dep}\,(t) + N_{depres}(t-1)\triangle A_{dep}(t) \quad (5)$$

where

$$\triangle A_{dep}(t) \;=\; Min[1, A_{dep}(t)/A_{dep}(t-1)]. \quad (6)$$

If the size of the non-flooded area decreases, $\triangle A_{dep}$ is smaller than one, otherwise it is one. The number of microbial particles that are released in the non-flooded area are then calculated from the total number of microbial particles in animal fecal deposits (N_{deptot}):

$$N_{rel,nonflooded}(t) = \begin{cases} P_Q = 0 \rightarrow 0 \\ P_Q > 0 \;\rightarrow \omega N_{deptot}\,(t) \cdot \exp[\mu_w\,(t)] \end{cases} \quad (7)$$

where μ_w is determined according to Equation 2. We calculated the surface water runoff volume P_Q [m^3/h] as function of precipitation (P, including rain water and snow melt), evaporation and soil moisture processes according to Blöschl et al. (2008), Equations 1 – 4. To calculate the changes in soil moisture we used Equations 5, 7 and 8 given by Blöschl et al. (2008), with parameter settings according to Demeter et al. (2021) for the study site (**Supplementary Table 3**).

Transport in flooded area

Animal fecal deposits are completely resuspended in floodwater. The number of resuspended microbial particles, $N_{rel,flooded}$, is:

$$N_{rel,flooded}(t) = \begin{cases} \triangle A_{dep} = 1 \rightarrow 0 \\ \triangle A_{dep} \;< 1 \;\rightarrow N_{depres} \cdot (1 - \triangle A_{dep}) \cdot \exp[\mu_w\,(t)] \end{cases} \quad (8)$$

where μ_w and $\triangle A_{dep}$ are determined according to Equations (2) and (6). The number of released microbial particles is then subtracted from the total deposited numbers and reduced by first-order decay in feces, μ_f [1/h]:

$$N_{depres}\,(t) = [N_{deptot}\,(t) - \omega(t)N_{deptot}\,(t)] \cdot \exp[\mu_f\,(t)] \quad (9)$$

For μ_f, we took ranges of reported values for bovine feces as the boundaries of a uniform distribution according to Wu et al. (2020). This distribution was assumed to evenly represent the varying μ_f values with environmental conditions and time (**Table 1**). The number of microbial particles running off to the backwater channel in non-flooded and flooded areas are then added:

$$N_{vro}(t) = N_{rel,nonflooded}(t) + N_{rel,flooded}(t) \quad (10)$$

The numbers of microbial particles from each animal group are summed. The microbial particle concentrations in the backwater channel are then calculated:

$$C_{bw}\,(t) = \begin{cases} V_{fl}(t) + P_Q(t) > 0 \rightarrow \frac{C_{r\rightarrow bw}(t)\cdot V_{fl}(t) + N_{vro}(t)}{V_{fl}(t) + P_Q(t)} \\ V_{fl}(t) + P_Q(t) = 0 \;\rightarrow C_{r\rightarrow bw}\,(t) \end{cases} \quad (11)$$

where V_{fl} [m^3] is determined by means of regression (Section "Hydrological and Hydrodynamic Flow Situation," **Supplementary Table 2**).

Hydrological and Hydrodynamic Flow Situation

We selected the years 2010–2015 as study period. The lowest Danube discharges in this period occurred in 2011 (Q_{95}: 2400 m^3/s), and the highest discharges in 2013 (Q_{95}: 4100 m^3/s, **Figure 3**). Hourly discharge data of the Danube were available at gauge Wildungsmauer, which is located 12 km downstream of the study site. The annual precipitation was 452 mm in 2011 and 659 mm in 2013. Hourly precipitation data (mm/h) and air temperature was available at station Groß Enzersdorf, located at 6 km distance from the Danube along the upstream model boundary, **Figure 1**), the latter ranging from −20°C to 38°C (mean: 11.2°C, standard deviation: 8.9°C). The water temperature of the Danube (gauge Greifenstein, 41 km upstream of the study area) ranged from 0 to 23°C (mean: 11.2°C, standard deviation: 5.9°C).

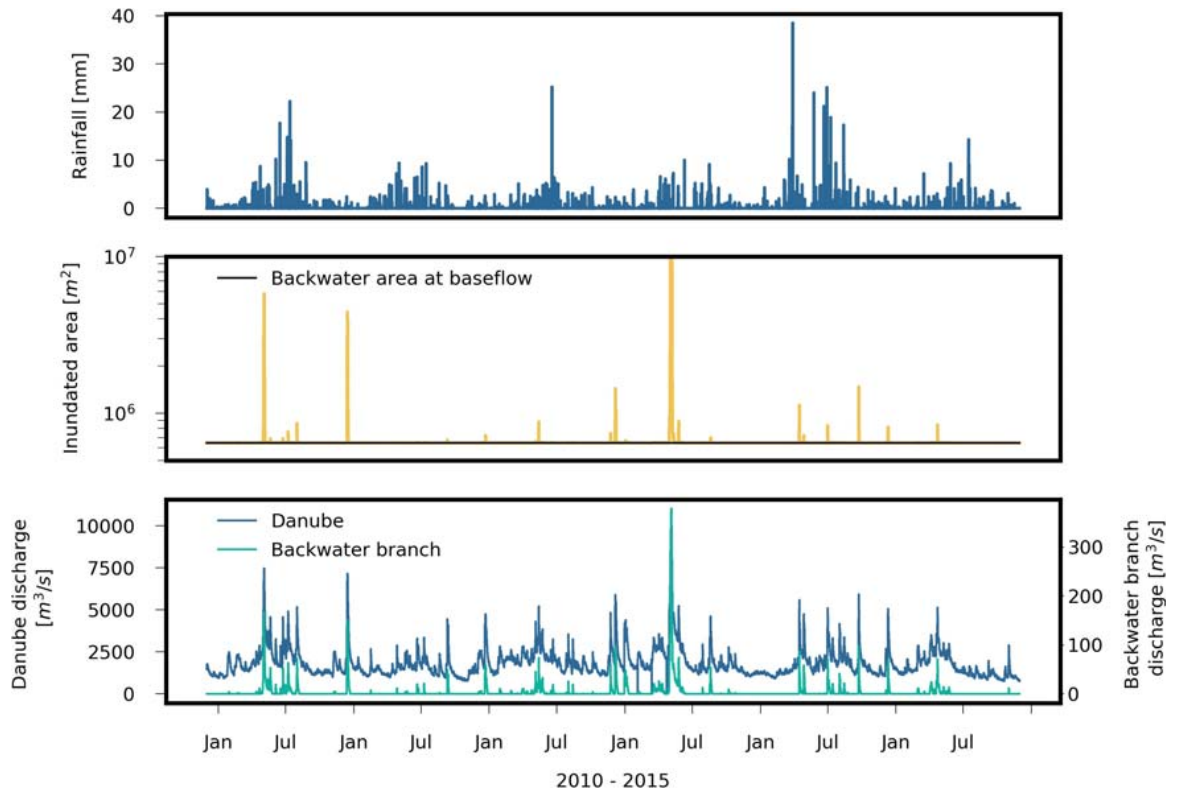


FIGURE 3 | Observed hourly rainfall and Danube discharge, and, simulated hourly discharge of the backwater channel and inundated area by means of regression (**Supplementary Table 2**) during the investigation period.

A 2-D hydrodynamic surface water model (CCHE2D Version 2.0, National Center for Computational Hydroscience and Engineering, University of Mississippi) was used to simulate the flow velocities and water levels of the backwater channel on an hourly basis during a flood event in January 2011 with an approximate 10-year return period. The model is described in detail by Gabriel et al. (2014) and Frick et al. (2020). In short, the model solves the two-dimensional formulation of the shallow water equations and uses depth integrated Reynolds equations. For temporal discretization, the implicit first order Euler's method was implemented and was able to simulate subcritical and supercritical flow conditions. The model domain covered an area of approximately 22 km^2 (**Figure 1**). The spatial distribution of the Manning roughness values were based on a detailed land use and vegetation classification. During model calibration the Manning roughness values were fine-adjusted to floods of the River Danube during August 2008 and June 2009, ranging from 0.024 to 0.125 $s \cdot m^{-0.3}$ within the model domain. The Nash and Sutcliffe coefficient of runoff model efficiency (Nash and Sutcliffe, 1970) at several water level gauges along the backwater branch from the inlet point of the Danube to LSW 3 (**Figure 1**) ranged from 0.92 to 0.98 for both calibration periods, and from 0.93 to 0.96 for the validation period during the flood event in January 2011.

The transient water quantities were determined by means of polynomial regression based on the hydrodynamic flow simulations during the rising limb of the flood event (**Supplementary Table 2**). The following variables were calculated and served as input to the microbial fate and transport model: hourly discharges and volumes of the backwater channel, and hourly volumes and areas of inundation (see **Supplementary Section** "Detailed Model Information" for more details). The shortest and longest period when the Danube discharged into the backwater channel were 16 days in 2011, and 91 days in 2013, respectively.

Microbiological Source Characterization

Microbial analyses of surface water

Surface water samples were collected monthly from 2010 to 2015. Samples were collected from one point at the Danube (DSW5), and from points LSW 1 and 3 along the backwater channel (**Figure 1**). The sampling location LSW1 is situated in a lateral branch outside the flood-protected area delineated by the dam and therefore represents waterbodies with high connectivity to the Danube River. The location LSW3 represents waterbodies along the main backwater channel, with an average water depth of ca. 170 cm. The MST markers were quantified in 500–600 mL water samples using quantitative PCR. The human marker HF183/BacR287 (Green et al., 2014), the ruminant marker BacR (Reischer et al., 2006), the pig marker Pig2Bac (Mieszkin et al., 2009), and the bird marker DuckBac (Kobayashi et al., 2013) were selected and applied as described

previously (Kirschner et al., 2017). As a robust approximation for the SLOD (sample limit of detection), which can only be determined by elaborate spiking processes to determine sample processing efficiencies on a sample-to-sample basis (filtration- and extraction efficiencies with representative MST mock communities), we applied the established threshold of detection (TOD) concept for MST field applications (Reischer et al., 2007, 2008). The filtration volume (200 - 300 mL), the use of 2.5 ml of diluted DNA extract in qPCR and the minimal amount of detectable targets per PCR reaction defines the detection threshold (Reischer et al., 2006, 2007). The quantitative microbial source tracking results were then expressed as marker equivalents per L (ME/L) to account for potential extraction losses (Reischer et al., 2007, 2008). The TOD covers sampling and sample processing information and also the efficiency of qPCR analysis. The mean TODs during the calibration and validation periods were 564 and 382 ME/L for the human, 419 and 327 ME/L for the ruminant, 490 and 337 ME/L for the pig, and 419 and 327 ME/L for the bird MST marker. The samples were analyzed for *E. coli* according to ISO 16649-1 (ISO, 2001) with a limit of detection (LOD) of 1 CFU/100 mL. Additional surface water samples were collected monthly at the Danube 23 km upstream of DSW 5 from June 2018 to August 2020 and analyzed for *Giardia* and *Cryptosporidium* (oo)cysts. *Giardia* spp. and *Cryptosporidium* spp. (oo)cysts were isolated from 10-L water samples, using an adaptation of the flat membrane method described in ISO (2006). Parasites were recovered from the filters and further analyzed as described in Demeter et al. (2021) using 50 mL of 1M glycine pH 5.5 solution and centrifuged at 1,550 × g for 15 min. Pellets were resuspended in 2 mL of ultrapure water. One mL of the suspension was used for the immunomagnetic separation of the parasites using the Dynabeads GC Combo kit (Thermo Fisher, United Kingdom). Concentrates were stained with the EasyStain kit (BTF Pty. Ltd., Biomerieux, Australia) and quantified as described by Stevenson et al. (2015). The LOD of *Giardia* and *Cryptosporidium* in surface waters was 0.4 (oo)cysts/L.

Microbial concentrations in animal feces

Samples of fecal matter of ruminants (hunted herbivores), wild boar, avian fecal matter from great cormorant (*Phalacrocorax carbo*), wild duck (*Anas platyrhynchos*) and other *Anatidae*, common tern (*Sterna hirundo*), and *Charadriiformes* were previously collected and analyzed for MST markers and *E. coli* in the study area (Vierheilig et al., 2013; Farnleitner et al., 2014; Frick et al., 2018). Marker concentrations associated with human (n = 19), ruminant (n = 20), porcine (n = 18), and bird fecal pollution (n = 11) from these samples were determined via qPCR (Farnleitner et al., 2014; **Table 1**). For *Cryptosporidium* and *Giardia*, reported values were used (**Table 1**).

Data analysis

The fecal indicator and pathogen concentrations in the Danube, C_r, used in Equation 1, were described by selected statistical distributions (**Table 3**). The parameters of the distributions were obtained from fits to the observed dataset at point DSW 5 (**Figure 1**). We performed Kruskal–Wallis tests for the selection of the distribution types ($p > 0.05$, **Table 3**). During all model simulations, random values were drawn from the distributions for each time step and Monte Carlo run. A substantial fraction of the measured microbial concentrations were left-censored values (i.e., 20, 60, 76, 30% for the human, ruminant, pig, and bird MST markers, and, 32 and 39 % for *Giardia* and *Cryptosporidium*), meaning that the concentration was known only to be lower than the LOD (*Giardia* and *Cryptosporidium*) or the TOD (MST markers). Non-detects (ND) were replaced by half of the TOD in case of the MST markers and by half of the LOD in case of *Cryptosporidium* and *Giardia*. This continues to be the most common procedure within the disciplines of environmental sciences to deal with non-detects (Helsel, 2006). Concerning the MST data, we used these in the calibration process. Since we treated both simulated and observed values the same way (i.e., observed non-detects and simulated values < TOD were both replaced by TOD/2), the chosen performance metrics are not affected, leading to the best calibration possible. Concerning the *Giardia* and *Cryptosporidium* data, we compared different levels for the substitution, i.e., substitution by zero, LOD/2 and LOD. It was shown that the choice of the level did not affect the results of the Kruskal–Wallis tests ($p > 0.05$), so the method was justifiable in our case. For modeling the concentrations of *Giardia* and *Cryptosporidium* in the backwater branch and the QMRA, we merely used the data in the Danube as boundary condition. We did not use the data in the backwater river for this purpose, where the concentrations would be lower than the LOD in most cases. For the data analysis, Python 3.7 and Scipy package 1.3.1 were used.

Quantitative Microbial Risk Assessment Module

Exposure to the pathogens is given as the dose D [L/d], the number of ingested pathogens per person per day. For calculating D, the Monte Carlo samples of pathogen concentrations in the backwater branch (C_{bw} [particles/L]), recovery (R, [-]), pathogen treatment reduction (log reduction value, LRV), and consumption data (V, [L]) are multiplied according to Equation 12. Data on times of consumption and consumed volumes of unboiled drinking water per person (V_i in equation 12) during a day were available from the Dutch National Food Consumption Survey 2007–2010 (DNFCS) (Van Rossum et al., 2011). The cumulative dose per person per day is:

$$D = \sum_{i=1}^{d} \left(C_{bw,i} \times \frac{1}{R} \times 10^{LRV} \times V_i \right) \tag{12}$$

where i denotes the hourly time step. Recovery rates were determined in the laboratory, resulting in mean values of 0.65 for *Giardia* (standard deviation: 0.28, n = 17) and 0.53 for *Cryptosporidium* (standard deviation: 0.27, n = 8). Beta distributions were fitted to the recovery data (α: 0.87, β: 0.53 for *Cryptosporidium* and α: 1.03, β: 0.44 for *Giardia*). Daily probabilities of infection for *Cryptosporidium* can be estimated

TABLE 3 | Observed values and descriptive statistics for the microbial concentrations of the Danube from 2010 to 2015 (C_r in Equation 1) and the Kruskal–Wallis test results showing that the simulated and observed values were not significantly different ($p \geq 0.05$).

Parameter	Observed dataset detected/n	Observed median (minimum, maximum) [particles/L]	Distribution	Descriptive statistical parameter values gamma distribution: shape/location/scale; normal distribution: mean/standard deviation	Kruskal–Wallis *p*	Comment
Human MST marker	50/62	5.39×10^3 (60, 1.41×10^6)	Gamma	0.23 / 58 / 80,000	0.05	After model optimization based on OF (Equation 14)
Ruminant MST marker	21/60	250 (60, 4.54×10^4)	Gamma	0.23 / 58.5 / 10,766	0.14	
Pig MST marker	14/60	220 (60, 1.58×10^5)	Gamma	0.25 / 58.5 / 3,700	0.11	
Bird MST marker	16/23	1.21×10^4 (60, 2.19×10^5)	Gamma	0.45 / 58 / 20,000	0.05	
E. coli	64/64	570 (20, 4.9×10^4)	Gamma	0.19 / 100 / 27,000	0.33	
Giardia	21/31	0.8 (0.2, 4.4)	Normal	1.39 / 1.17	0.45	stats fit function in Python 3.7
Cryptosporidium	19/31	0.8 (0.2, 6.44)	Normal	1.29 / 1.42	0.40	

Non-detects were replaced by half of the TOD (MST markers) or half of the LOD (E. coli, Giardia, and Cryptosporidium).

using a hypergeometric dose–response relation (Teunis and Havelaar, 2000):

$$P_{inf} = 1 - {}_1F_1(\alpha, \alpha + \beta, D) \quad (13)$$

where α and β are infectivity parameters that are pathogen-specific and ${}_1F_1$ is the confluent hypergeometric function. The dose-response model parameters for *Cryptosporidium* and *Giardia* were taken from the literature (**Table 1**). As daily health based target (hbt), $1 \cdot 10^{-6}$ infections/person/d was adopted in this study (Signor and Ashbolt, 2009). LRV was estimated iteratively until the criterion $P_{inf} \leq$ hbt according to Equation (13) was fulfilled for both the mean and 95^{th} percentile values of P_{inf}.

Validation of the Microbial Fate and Transport Model

To prove that the model captured the most relevant fecal sources and transport processes, we evaluated the model performance based on monthly measured concentrations of the human and animal MST markers and *E. coli* during 2010–2015. We selected the observation dates, when Danube discharged into the wetland area, or when rainfall occurred. The mean absolute error (MAE) was used as a performance metric (Willmott and Matsuura, 2005; Demeter et al., 2021). Log_{10} transformed concentrations were used in the MAE computations because microorganisms typically follow a lognormal distribution and the use of logarithms minimizes the influence of outliers present in the data (Hong et al., 2018; Demeter et al., 2021). The Kruskal–Wallis test was used for the distribution comparisons of the simulated and observed datasets in the backwater channel and the p-value of the Kruskal–Wallis statistic was a metric of model performance. In order to ensure an optimum model performance, the optimization parameters were adjusted to minimize the objective function (OF) (Demeter et al., 2021).

$$OF = MAE + (1 - p) \quad (14)$$

Non-detects (ND) in the observed dataset and simulated values below that level were set to the half of the threshold of detection (for the MST markers) in the calculations. The optimization parameters were the distribution parameters describing the microbial concentrations in the Danube and the microorganism-specific decay rate parameters (C_r in Equation 1, a_0 and a_1 in Equation 2). The former were adjusted while ensuring that the Krukal-Wallis p-value was greater than 0.05 (**Table 3**). Decay rates of the MST markers were collected from a literature survey and summarized in **Supplementary Table 4**. An ordinary least square method was used to fit the time-to-first-log (TFL) as a function of water temperature (dashed lines in **Figure 4**), using the Python 3.7 package statsmodels (0.10.1). During the adjustment of intercept a_0 (used in Equation 2, solid lines in **Figure 4**), it was ensured that the decay as function of temperature obtained lay within the prediction interval of the ordinary least square regressions (**Figure 4**, shaded area). The Bradford-Schijven release parameter a and β were kept the same for all animal sources, as their effects on the simulated concentrations was small in comparison with the optimization parameters (**Table 1**). We performed a stepwise, source-targeted model optimization of QMRAcatch:

- The model was validated, considering individual, presumably important fecal sources using measured concentrations of the respective MST markers. We simulated concentrations of the human MST marker in the backwater channel and compared them with the measured dataset on days when the Danube discharged into the backwater channel, i.e., during floods. We selected the data during days when the Danube discharged into the backwater channel (Section "Study Area," **Figure 1**). The distribution parameters describing the human-associated MST marker concentrations in the Danube and the decay rate coefficient a_0 were adjusted to minimize OF (Equation 14). The same procedure was applied consecutively for the ruminant, pig, and bird associated MST markers, except

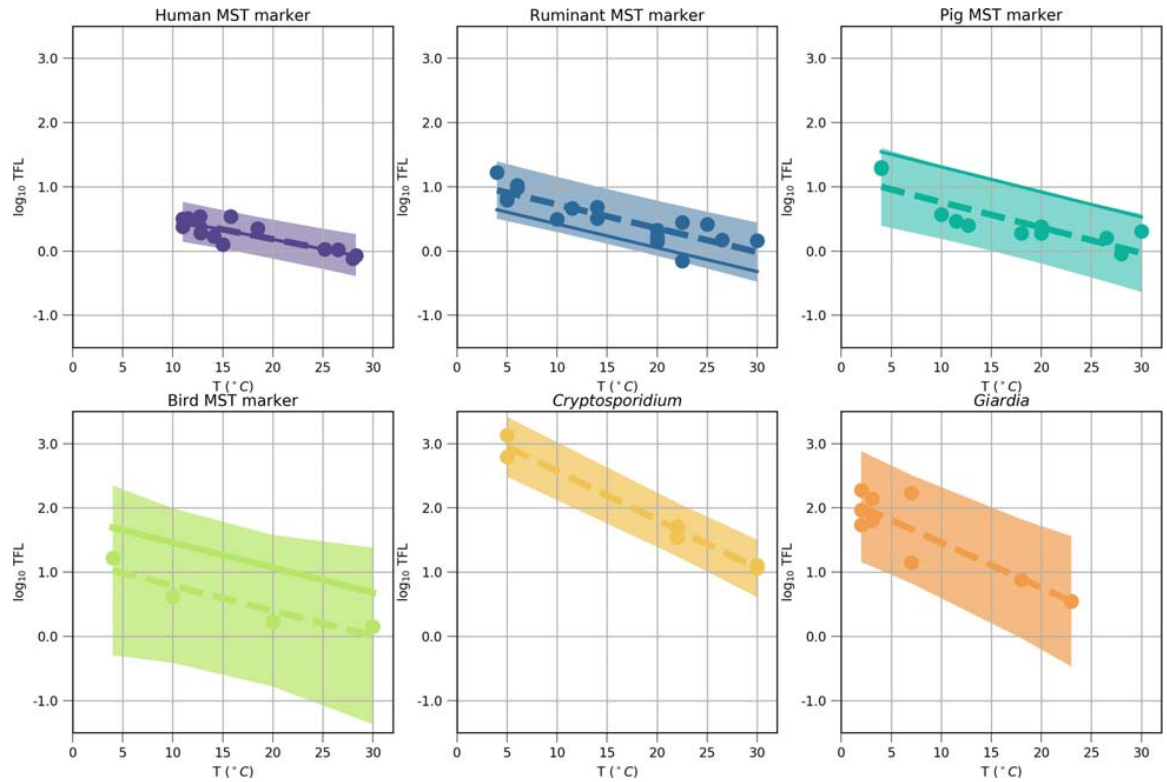


FIGURE 4 | Inactivation of the MST markers, *Cryptosporidium*, and *Giardia* after model optimization (solid lines) and as reported in experimental studies (dots, **Supplementary Table 4**), plotted as time to first $\log_{10}$ reduction (TFL, days) values $\log_{10}$-transformed ($\log_{10}$ (TFL) in function of the temperature. Ordinary-least-square regressions (dashed lines) were fitted to the literature values, shown with their 95% prediction intervals (shaded). The intercept and the slope of the solid lines were the result of the model optimization. These values were used as model input parameters a_0 and a_1 in Equation 2 (**Table 1**).

that animal fecal deposits were additionally considered as microbiological sources (**Table 1**). As for the human MST marker, we used the data collected during floods. In addition, we used data collected on days, when the backwater area was partially inundated or when it was raining (Section "Study Area," **Figure 1**). The observation sites were selected based on the findings of Frick et al. (2020) who conducted a comprehensive analysis of the spatial distribution of human and animal fecal pollution in the study area. To validate the model during floods, we selected the site, which was influenced by floods (LSW 1). To validate the model during days of rainfall or inundation of the area, we used the site, which was impacted by wildlife (LSW 3, **Figure 1**).

- In the second step, the model was validated using measured concentrations of *E. coli*. The same procedure was applied as for the MST markers, except that we considered all fecal sources combined and the decay rate coefficients were not adjusted but taken from the literature (**Table 1**).

Scenario Load and Infection Risk Assessment

We defined the following event-driven scenarios for quantifying the effects of fecal sources on the microbiological quality of the backwater channel considering safe drinking water:

- As allochthonous source, we considered *Cryptosporidium* and *Giardia* transport via Danube discharges into the backwater channel (scenario FLOODS).
- As autochthonous sources, we considered the resuspension of *Cryptosporidium* and *Giardia* from fecal deposits in inundated areas (scenario RESUSP), and the rainfall-release and runoff of *Cryptosporidium* and *Giardia* from fecal deposits (scenario RAIN).
- All of the above scenarios were considered simultaneously (scenario all combined).

Using QMRAcatch, we simulated the concentrations and loads of *Cryptosporidium* and *Giardia* in the backwater channel, and the drinking water infection risks relative to a health based target. We considered different hydrological conditions and fecal sources in the scenarios, as indicated in **Table 4**. To simulate no connection of Danube and backwater (**Table 4**), the Danube discharge was set to the mean flow rate, consequently there was no discharge into the backwater channel. All other parameter settings were taken from **Tables 1–3** and **Supplementary Material**.

TABLE 4 | Fecal sources and hydrological conditions in the event-driven scenarios investigating (i) the wastewater-impacted river water entering the backwater during floods (FLOODS), (ii) the resuspension of pathogens from fecal deposits in inundated areas (RESUSP), and (iii) the pathogen release and runoff from fecal deposits (RAIN), and (iv) all combined.

	Wastewater impacted river water (FLOODS)	Resuspension of pathogens from fecal deposits in inundated areas (RESUSP)	Pathogen release and runoff from fecal deposits (RAIN)	All combined
Hydrological conditions				
Rain	−	−	+	+
Connection between Danube and backwater	+	+	−	+
Fecal pollution sources				
River	+	−	−	+
Ruminants	−	+	+	+
Wild boar	−	+	+	+
Birds	−	+	+	+

+ taken into account, – not applied.

RESULTS

In order to test if we considered the most relevant fecal sources and transport pathways, we tested the applicability of the human and selected animal MST markers (Section "Applicability of the MST Marker Specificity for Modeling"), and validated the model based on measured concentrations of these MST markers and *E. coli* in the backwater channel (Section "Performance of the Microbial Fate and Transport Model"). We then simulated the *Giardia* and *Cryptosporidium* concentrations, loads and drinking water infection risks relative to a health-based benchmark for the defined fecal pollution scenarios and reference pathogens (Section "Contribution of Fecal Sources to the Reference Pathogen Impact on the Backwater Resource Considering Safe Drinking Water Production).

Applicability of the MST Marker Specificity for Modeling

The selected human, ruminant, pig and bird MST marker assays (Section "Microbiological Source Characterization") are primarily associated with their respective target sources. However, low numbers may also occur in the non-target pollution sources. According to the analysis of fecal samples, the reported MST marker concentrations in the non-target pollution sources were more than five orders of magnitude lower than those in the target pollution sources (**Table 1**). Nevertheless, the impact on false-positive MST marker detection rates in the backwater area may become significant, when a large non-target animal population is the source. To evaluate the applicability of the MST markers, we investigated the impact on the simulated concentrations from non-target animal sources in the floodplain river with QMRAcatch. The concentrations of each MST marker in the backwater channel were described by a gamma distribution based on the reported mean and 95^{th} percentile fecal source concentrations according to **Table 1**, and considering (i) both target (i.e., correct positive detections) and non-target host groups (i.e., false positive detections) and (ii) only the target group. The parameter settings were used according to **Tables 1–3** in the simulations.

For each measured MST marker, the simulated mean and 95^{th} percentile concentrations in the backwater branch during the simulation period were compared for the two cases. For all MST markers, the simulated concentrations considering both target and non-target groups differed by 0 - 5 % from the simulated concentrations considering only the target group. This means that at least 95 % of the simulated concentrations in the floodplain river were associated with the target pollution sources in the catchment. The animal and human associated qPCR assays and the measured concentrations at our study site were thus considered to be useful for a source–targeted evaluation of the microbial fate and transport model.

TABLE 5 | Model performance based on the observed microbial concentrations in 2010–2015 during days when the Danube discharged into the backwater branch and when rainfall occurred.

Parameter	Observed dataset	Observed median (minimum, maximum)	OF (Equation 14)	Mean absolute error	Kruskal–Wallis test
	Detected/n	[particles/L]		[$\log_{10}$ particles/L]	p
Human MST marker	9/10	4.1×10^3 (640, 9.3×10^4)	1.34	0.67	0.33
Ruminant MST marker	10/16	708 (56, 1.64×10^4)	0.70	0.61	0.91
Pig MST marker	6/15	231 (65, 1.29×10^4)	1.17	0.50	0.33
Bird MST marker	14/16	7.04×10^3 (89, 8.37×10^4)	1.53	0.69	0.16
E. coli	16/16	220 (10, 7.0×10^3)	1.19	0.62	0.44

Objective function values (OF), mean absolute error and Kruskal–Wallis p-values after comparison of the simulated and observed microbial concentrations in the backwater branch. Non-detects were replaced by half of the TOD (MST markers) or LOD (E. coli). Simulated values below the half of the LOD or the lowest TOD were set to the half of the LOD or lowest TOD.

Performance of the Microbial Fate and Transport Model

We validated the model based on measured concentrations of human-, ruminant-, pig-, bird-associated MST marker and *E. coli* in the backwater branch. The model validation resulted in mean absolute errors ranging from 0.5 to 0.7 $\log_{10}$ for the MST markers and *E. coli* (**Table 5**). The objective function values ranged from 0.7 to 1.5 (OF in Equation 14). The cumulative distribution plot of the simulated and observed concentrations confirmed the general good agreement (**Figure 5**). The majority (70–90 %) of the simulated concentrations of all fecal indicators resulted in errors ranging from −1.0 to 1.0 $\log_{10}$ particles/L after model optimization (**Figure 6**).

Contribution of Fecal Sources to the Reference Pathogen Impact on the Backwater Resource Considering Safe Drinking Water Production

For the scenarios, we evaluated the simulated concentrations of *Cryptosporidium* and *Giardia* during time steps when

- The Danube discharged into the backwater branch (for the FLOODS scenario),
- Part of the backwater area was inundated (for the RESUSP scenario),
- Rainfall occurred (for the RAIN scenario), and
- All of the above combined.

A more detailed definition of how we defined these events is given in the **Supplementary Information Section** "Detailed Definition of the Events for the Scenarios." The FLOODS and RAIN scenarios and all scenarios combined resulted in similar ranges of concentrations of *Cryptosporidium* and *Giardia* in the backwater branch (0.4–1.2 particles/L for the mean and 1.8–5.1 particles/L for the 95^{th} percentiles, **Figure 7**). The concentrations were more than one $\log_{10}$ lower for the RESUSP scenario. The concentrations were the same for *Giardia* and *Cryptosporidium* for the FLOODS scenario, while they were 70–90 % smaller for *Giardia* than for *Cryptosporidium* for the RAIN and RESUSP scenario due the higher inactivation.

We further conducted a source apportionment by calculating the mean pathogen loads of *Cryptosporidium* and *Giardia* in

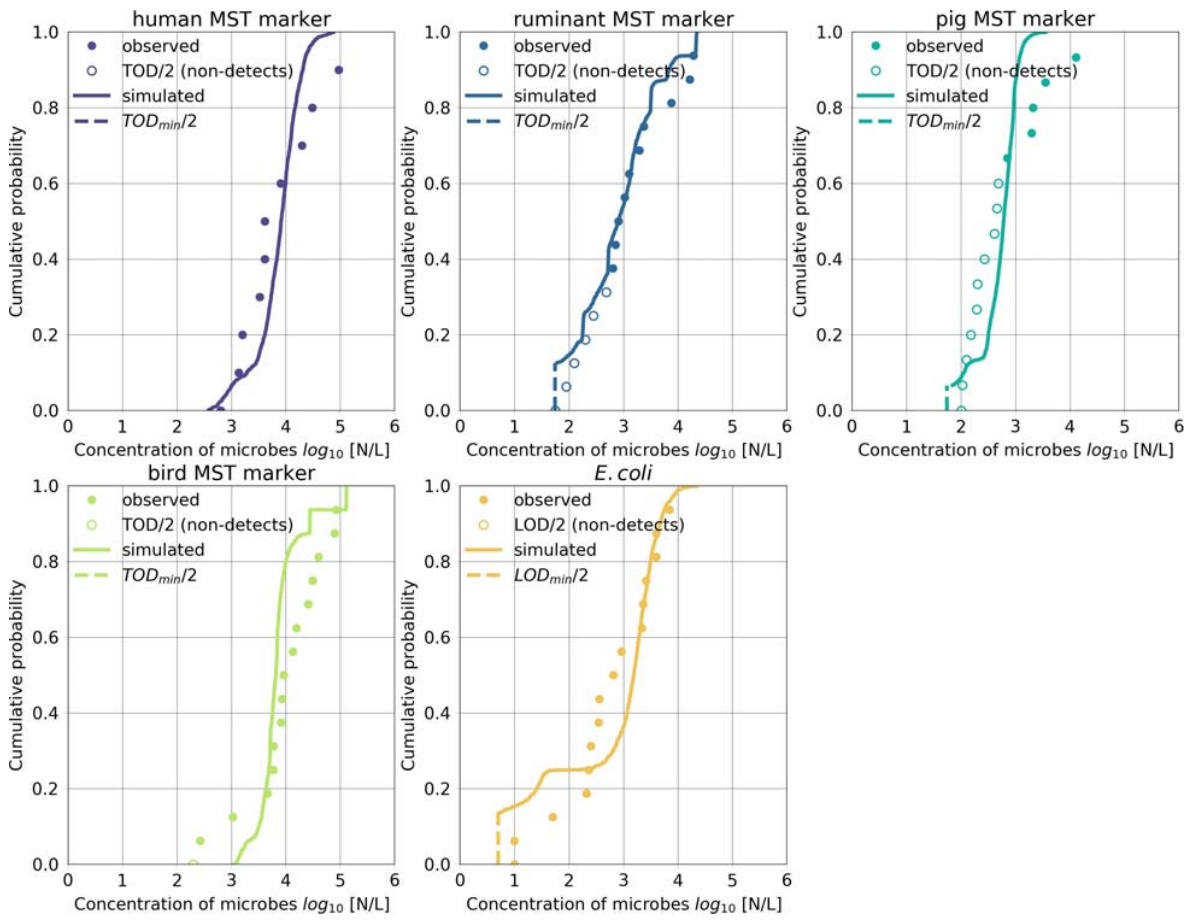


FIGURE 5 | Microbial fate and transport model performance. Simulated and observed concentrations of the host-associated MST markers and of *E. coli* during days when Danube water discharged into the backwater branch or on rainy days. Non-detects were replaced by half of the TOD (MST markers) or LOD (*E. coli*). Simulated values below the LOD or the lowest TOD (TOD_{min}) were set to half of the LOD or the TOD_{min}.

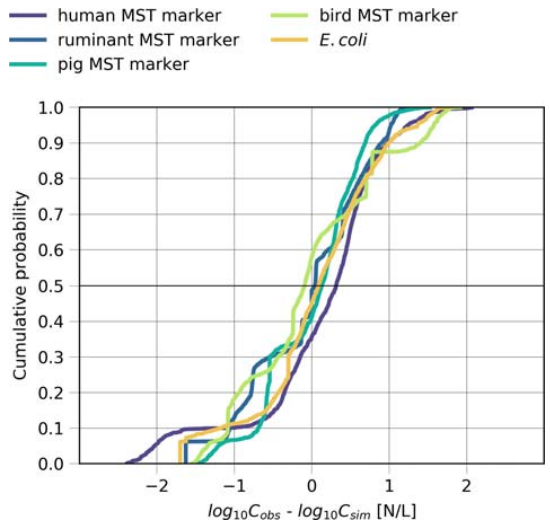


FIGURE 6 | Model performance for the source-targeted microbial fate and transport. Cumulative distribution plot of the difference between the simulated and measured concentrations (log_{10}-transformed) of the human, ruminant, pig and bird-associated MST markers and *E. coli* during days when Danube water discharged into the backwater branch or on rainy days.

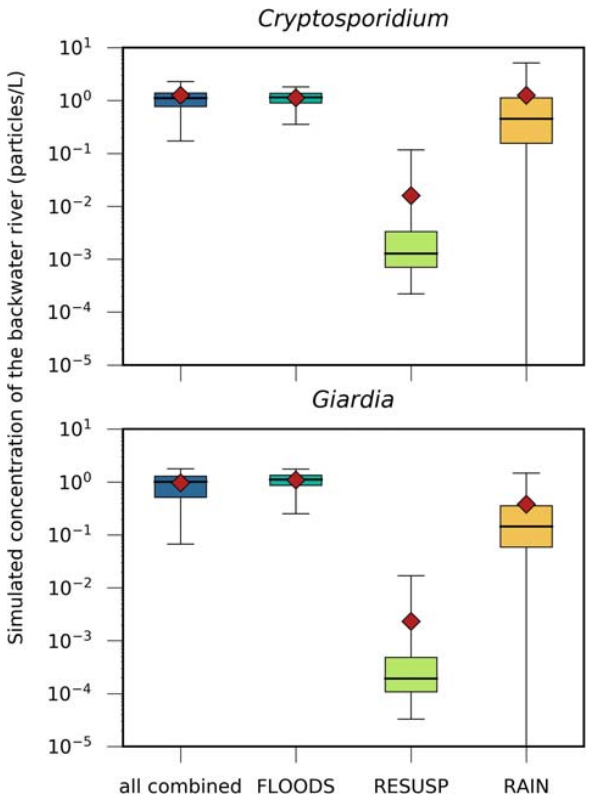


FIGURE 7 | Simulated concentrations of *Cryptosporidium* and *Giardia*, transported via (i) the discharge of Danube water (FLOODS), (ii) the resuspension of pathogens from fecal deposits in inundated areas (RESUSP), (iii) the rainfall-release and runoff from fecal deposits (RAIN), and all scenarios combined. Black horizontal lines, red diamonds, and whiskers mark the median, mean, and 95th percentile values, respectively.

the backwater branch. For that, we multiplied the simulated hourly concentrations by the hourly discharges and evaluated the mean loads for the selected time steps (Section "Scenario Load and Infection Risk Assessment"). The simulated mean loads of *Cryptosporidium* and *Giardia* were again in a similar range for the FLOODS and RAIN scenarios and all scenarios combined (3–13 $\times$ 10^9 particles/h), and were at least one log_{10} lower for the RESUSP scenario. The FLOODS scenario occurred during 20 % of the 6-year time period (**Figure 8**). The RAIN scenario, which occurred only during 7 % of the time, resulted in higher standard deviations and peaks of loads than the FLOODS scenario in case of *Cryptosporidium*. The RESUSP scenario resulted in the smallest source attribution in comparison, occurring during 8 % of the time.

The drinking water infection risks relative to a health based target of $\leq 1 \cdot 10^{-6}$ infections/person/d were estimated assuming a value of 6.2 and 6.0 as treatment reduction of *Cryptosporidium* and *Giardia* from backwater river water (LRV, Equation 13). The mean drinking water infection risks for *Cryptosporidium* and *Giardia* resulted in values one log_{10} below to close to the health based target for the FLOODS and RAIN scenarios, and all scenarios combined. For the RESUSP scenarios, the mean drinking water infection risks were $>$ 3 log_{10} below the health based target (**Figure 9**).

DISCUSSION

Strengths and Limitations of the Integrative Approach

In this study, we presented a new integrative modeling approach for evaluating the impact of fecal sources and transport pathways on the microbiological quality of a riverine wetland considering safe drinking water. By integrating measured concentrations of human and the most relevant animal MST markers, the approach allowed for the first time quantifying the relative drinking water infection risks from external (allochthonous, i.e., river water inflows) and internal fecal sources (autochthonous, i.e., wild boar, ruminants, birds in the backwater study area). This would not have been possible based on FIO data alone, which are sum indicators in contrast to MST markers (Zhang et al., 2019). The approach also allowed assessing if a given MST marker is appropriate in the study area, given its fecal specificity and fecal sensitivity. These performance characteristics of MST markers can be highly regional- and site-specific (Reischer et al., 2013). Furthermore the required MST performance criteria depend on the relative abundance of the specific fecal sources to be detected amongst the sum of total fecal pollution occurring at the investigation site, e.g., % fraction of human fecal pollution in relation to the sum of human and animal fecal pollution (Reischer et al., 2011). For example, if the animal numbers were different to our study site, the resulting non-target concentrations of the MST markers could render the selected MST marker assays inapplicable. In this case, other

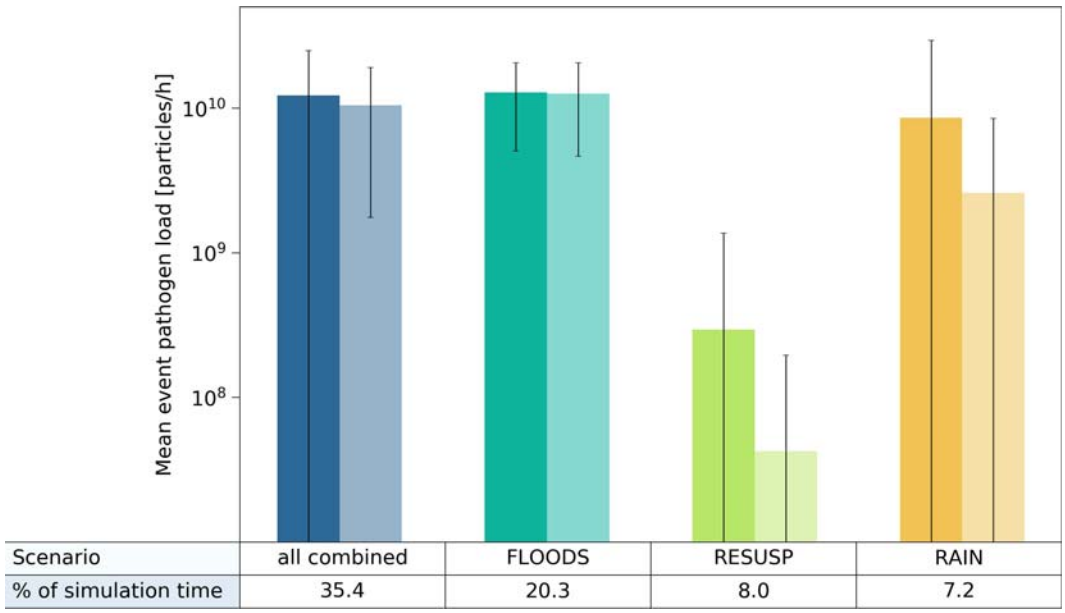


FIGURE 8 | Mean load attribution of *Cryptosporidium* (left columns) and *Giardia* (right columns) via (i) the discharge of Danube water (FLOODS), (ii) the resuspension from fecal deposits in inundated areas (RESUSP), (iii) the rainfall-release and runoff from fecal deposits (RAIN), and all scenarios combined from 2010 to 2015. Whiskers indicate the standard deviations.

markers with appropriate performance characteristics for the specific situation and question have to be chosen. A trade-off between fecal source sensitivity and specificity for MST qPCR assays often exists (Layton et al., 2013; Raith et al., 2013; Mayer et al., 2016). To illustrate this relationship, for recent fecal pollution detection bacterial MST qPCR assays often show high sensitivity but limited specificity as in contrast to many viral qPCR MST assay often showing high source specificity but limited sensitivity (Mayer et al., 2016). As an exemplary test case for the selected human MST marker assay at our study site, we assumed a theoretical increase of the non-target population number of animals by 10 fold, i.e., of ruminants, boar and birds, while leaving the human fecal sources unchanged (Section "Applicability of the MST Marker Specificity for Modeling"). We then simulated the resulting concentration signal of the human MST marker in the backwater river (i.e., the sum of the total qPCR signal from the correct positive (humans) and false negative (animals) DNA targets). The scenarios showed that the selected human MST markers would still be applicable at our study site and for the calibration of the model, even if the animal population number increased drastically by 10-fold (error $< 5\%$, results not shown).

The model was validated in two steps over a 6-year time period, considering (i) individual, presumably important fecal sources using measured concentrations of the respective MST markers, and (ii) all fecal sources combined using measured concentrations of *E. coli*. This was to ascertain that we accounted for the most relevant fecal sources and transport pathways. Interestingly, the measured and simulated MST marker and *E. coli* concentrations were similar, even though other sources of FIO may have potentially contributed (**Figure 5**). Frick et al. (2018) identified poikilothermic animals (earthworms, gastropods, frogs, and fish) as further potential autochthonous reservoirs of bacterial fecal indicators in our study area. To validate the transient concentration changes during floods and rainfall, it would be advantageous to collect microbial data at high temporal resolution. Current advances in online monitoring techniques may provide this opportunity in the near future (Stadler et al., 2019).

In this study, we modified and extended the microbial fate, transport and infection risk model QMRAcatch (v 1.1 python backwater) to simulate rainfall - runoff and mixing with released microbial particles from animal fecal deposits as functions of transient soil moisture processes according to Bradford and Schijven (2002) and Blöschl et al. (2008). The QMRA framework was fine-tuned for making use of the simulated exposure concentrations of *Cryptosporidium* and *Giardia* at hourly time steps based on human drinking water consumption data (Van Rossum et al., 2011). The discharge rates, volumes and surface water areas simulated by a validated hydrodynamic model allowed accounting for the spatiotemporal changes of these hydrological variables by means of polynomial regression. This was an essential input information for accurately predicting the microbial fate and transport in the alluvial wetland, as pointed out by Sanders et al. (2005) and Liu et al. (2015). Integrating a probabilistic Monte Carlo framework into the model analysis allowed accounting for the uncertainty of the source and transport variables and conducting a microbial infection risk assessment (Liao et al., 2016). One limitation of our model was that it did not account for the microbial particle interaction with the riverbed sediments. While Sanders et al. (2005) showed that the sediment erodibility parameters,

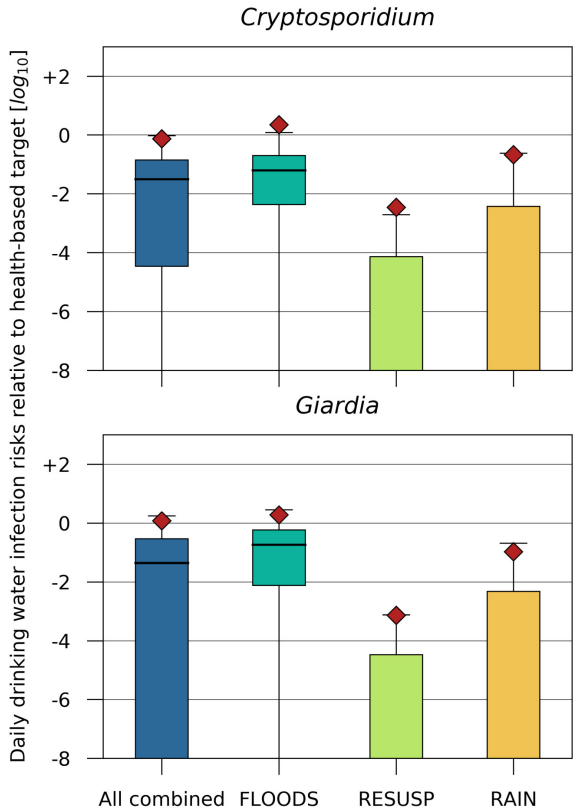


FIGURE 9 | *Cryptosporidium* and *Giardia* daily drinking water infection risks relative to a health-based benchmark of –6 $\log_{10}$/person/d (Signor and Ashbolt, 2009) for the scenarios (i) discharge of Danube water (FLOODS), (ii) resuspension from fecal deposits in inundated areas (RESUSP), (iii) rainfall-release and runoff from fecal deposits (RAIN), and all scenarios combined during 2010–2015. The treatment reduction of *Cryptosporidium* and *Giardia* from backwater river water (LRV, Equation 13) was 6.2 and 6.0 $\log_{10}$.

and sediment concentrations were important for FIO transport in a coastal wetland, sediment erosion was presumably of minor importance at our study site. Our model predicted 70–80 % of the observed concentrations within acceptable error limits (± 1 $\log_{10}$ particles/L), and the simulated and observed cumulative concentrations were not significantly different (Kruskal–Wallis $p > 0.05$). The settling of microbial particles and sediment transport simulations may be included for future applications. To estimate pathogen source loads from animal fecal deposits, we assumed that they were evenly distributed in the backwater area. This simplifying assumption was justifiable in our 14 km^2 sized model area. In larger wetlands, the spatial distribution of animals may need to be accounted for (Kay et al., 2007).

Impact of Event-Driven Fecal Pollution Sources and Pathways

The new integrative modeling approach allowed determining the transfer rates of pathogens from diverse fecal sources into wetlands during storm events and floods. Such weather extremes are of increasing concern due to climate change in many parts of the world. Several studies identified links of severe rainfall and flood events to elevated concentrations of pathogens such as *Giardia* and *Cryptosporidium* in rivers (Atherholt et al., 1998), the associated drinking water infection risks (Tolouei et al., 2019), or, to the number of outbreaks and sporadic cases of waterborne illness (Galway et al., 2015; Chhetri et al., 2019). For quantifying the impact of such events on the microbiological water quality of wetlands, modeling frameworks were developed either for pathogen transport via the rainfall-induced release and runoff (Guber et al., 2013), or via floods and resuspension (Sanders et al., 2005; Daniels et al., 2014; Liu et al., 2015). Our integrated modeling approach was developed to quantitatively compare these transfer pathways in a probabilistic framework. Our study showed that rainfall-induced pathogen release from animal fecal deposits, and floods can result in similar ranges of concentrations and loads of *Cryptosporidium* and *Giardia* and required reductions to achieve safe drinking water. This implies for water safety planning, that the autochthonous, homeothermic animal sources, such as ruminants, wild boar and birds, can be similarly important fecal pollution sources as the allochthonous human wastewater. This also implies that additional treatment may be required for drinking water production in wetlands inhabiting abundant wildlife, even in the absence of human wastewater discharges from upstream. According to our estimates, a 5–6 $\log_{10}$ reduction of *Cryptosporidium* and *Giardia* is required to achieve safe drinking water during floods and rainfall events. Demeter et al. (2021) considered only human wastewater sources to calculate the required reductions of *Cryptosporidium* to achieve safe drinking water at the Danube study site. Our estimation during floods is 0.5 $\log_{10}$ higher due to the additional contribution of diffuse animal sources in the Danube catchment.

For the estimation of infection risks, we used a mixture of beta distributions for the prevalence of the reference pathogens *Cryptosporidium* and *Giardia* in animal waste. We conducted a comprehensive literature survey, and selected values from the most recent, data-intensive studies conducted in temperate, high-income regions as our study area. The human infection risks from the animal fecal sources, however, may still be an overestimate, as we assumed the same dose-response models as for the human wastewater sources. To date, there are no reports about dose-response studies including different genotypes of *Cryptosporidium* or *Giardia* in the scientific literature. However, as long as this information is missing, it seems acceptable for risk assessment to choose this conservative risk assessment approach. Besides the reference pathogens *Cryptosporidium* and *Giardia*, other zoonotic pathogens such as EHEC and *Salmonella* spp. could be included in future analysis. These bacteria are also important reference pathogens occurring both in human and animal sources (Stalder et al., 2011), and their effects will depend on region and microorganism-specific source concentrations, prevalence and decay.

The modeling approach is transferrable to other riverine wetlands worldwide, even though the results of our study are

site-specific. To support water safety planning, it is important to integrate site-specific data into the modeling analysis and to validate the different transfer pathways of pathogens. In contrast to our local-scale approach, larger scale modeling studies previously identified hot spots of fecal pollution or evaluated the impact of system changes on the microbiological water quality (Medema and Schijven, 2001; Vermeulen et al., 2015; Sterk et al., 2016). These studies commonly made generalizing assumptions about the pathogen source and transport parameters as well as the hydrological and environmental boundary conditions and were not validated on real-world data.

CONCLUSION

- This study presents a new integrative modeling approach for determining the transfer rates of pathogens from diverse fecal sources into alluvial wetlands during storm events and floods considering safe drinking water supply.
- The modified and extended QMRAcatch (v1.1 Python backwater) combines microbial source tracking (MST) with 2-D hydrodynamic flow, rainfall-runoff, microbial fate and transport, and QMRA.
- The modeling approach allowed assessing the applicability of the chosen MST markers for the targeted fecal pollution source in relation to the total sum of all fecal pollution sources, considering fecal sensitivity and fecal specificity. They were found fully applicable for the modeling requirements and the research question in this study. The model captured the most relevant fecal sources and transport pathways, as proven by the model validation based on MST markers and *E. coli*.
- Allochthonous and autochthonous fecal sources during floods and rainfall events contributed similar ranges of concentrations and loads of *Cryptosporidium* and *Giardia* in the backwater branch, and drinking water infection risks relative to a health-based target.

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

JD extended the computer code of QMRAcatch, did the computational and model analyses, and wrote the manuscript. KD contributed to the design of the study. KD and RL organized the microbial source tracking database. RL, SC-A, and KD did the molecular and microbiological analyses, supported by RS and JW. GL took the water samples in the field. GS conducted the literature survey on pathogen prevalence and source concentrations. JS gave support with QMRA modeling and provided data on the cumulative doses to calculate daily drinking water infection risks. JK did the hydrodynamic model simulations and provided support regarding the hydrological model. SC-A, RS, JW, AK, AB, and AF contributed to conception and in the acquisition of funding support. SC-A and KD wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the Vienna Science and Technology Fund (WWTF) (Grant Number ESR17-070), the Austrian Academy of Science (Grant Number JF_2019_15, project name 'Swim City'), and by the Niederösterreichische Forschungs- und Bildungsgesellschaft (NFB) (Grant Number LSC 19-016 "Future Danube"). Additional support came from a research cooperation between Vienna Water and the ICC Water & Health in the frame of the Vienna Water resource systems project (ViWa 2020+).

ACKNOWLEDGMENTS

We thank the viadonau for providing data about the discharge and water temperature of the Danube, and the Central Institute for Meteorology and Geodynamics of Austria (ZAMG) for providing the precipitation and air temperature data. We also thank Hannes Gabriel for support with the hydrodynamic flow simulations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.668778/full#supplementary-material

REFERENCES

Ahmed, W., Zhang, Q., Kozak, S., Beale, D., Gyawali, P., Sadowsky, M. J., et al. (2019). Comparative decay of sewage-associated marker genes in beach water and sediment in a subtropical region. *Water Res.* 149, 511–521. doi: 10.1016/j.watres.2018.10.088

Arnberger, A., Frey-Roos, F., Eder, R., Muralt, G., Nopp-Mayr, U., Tomek, H., et al. (2009). *Ökologische und soziale Tragfähigkeiten als Managementherausforderungen für suburbane Biosphärenparke am Beispiel Untere Lobau (Ecological and Social Carrying Capacities as Management Challenges for Peri-Urban Biosphere Reserves). Final report*. Vienna: University of Natural Resources and Life Sciences.

Atherholt, T. B., LeChevallier, M. W., Norton, W. D., and Rosen, J. S. (1998). Effect of rainfall on Giardia and Crypto. *J. Am. Water Works Assoc.* 90, 66–80. doi: 10.1002/j.1551-8833.1998.tb08499.x

Bertrand, I., Schijven, J. F., Sanchez, G., Wyn-Jones, P., Ottoson, J., Morin, T., et al. (2012). The impact of temperature on the inactivation of enteric viruses in food and water: a review. *J. Appl. Microbiol.* 112, 1059–1074. doi: 10.1111/j.1365-2672.2012.05267.x

Blöschl, G., Reszler, C., and Komma, J. (2008). A spatially distributed flash flood forecasting model. *Environ. Mod. Softw.* 23, 464–478. doi: 10.1016/j.envsoft.2007.06.010

Böhm, J. (2016). *Grundlagen für die Optimierung des Wildmanagements im Nationalpark Donau-AuenEin Vergleich verschiedener Schutzgebiete unter besonderer Berücksichtigung des Managements von Wildruhegebieten*. San Francisco, CA: Wildtierökologie und Wildtiermanagement, University of Natural Applied Sciences (BOKU).

Boyer, D. G., Kuczynska, E., and Fayer, R. (2009). Transport, fate, and infectivity of *Cryptosporidium parvum* oocysts released from manure and leached through macroporous soil. *Environ. Geol.* 58, 1011–1019. doi: 10.1007/s00254-008-1580-x

Bradford, S. A., and Schijven, J. (2002). Release of *Cryptosporidium* and *Giardia* from dairy calf manure: impact of solution salinity. *Environ. Sci. Technol.* 36, 3916–3923. doi: 10.1021/es025573l

Castro-Hermida, J. A., Garcia-Presedo, I., Gonzalez-Warleta, M., and Mezo, M. (2011). Prevalence of *Cryptosporidium* and *Giardia* in roe deer (*Capreolus capreolus*) and wild boars (*Sus scrofa*) in Galicia (NW, Spain). *Veterinary Parasitol.* 179, 216–219. doi: 10.1016/j.vetpar.2011.02.023

Chhetri, B. K., Galanis, E., Sobie, S., Brubacher, J., Balshaw, R., Otterstatter, M., et al. (2019). Projected local rain events due to climate change and the impacts on waterborne diseases in Vancouver, British Columbia, Canada. *Environ. Health* 18:116. doi: 10.1186/s12940-019-0550-y

Daniels, M. E., Hogan, J., Smith, W. A., Oates, S. C., Miller, M. A., Hardin, D., et al. (2014). Estimating environmental conditions affecting protozoal pathogen removal in surface water wetland systems using a multi-scale, model-based approach. *Sci. Total Environ.* 493, 1036–1046. doi: 10.1016/j.scitotenv.2014.06.053

de Regnier, D. P., Cole, L., Schupp, D. G., and Erlandsen, S. L. (1989). Viability of *Giardia* cysts suspended in lake, river, and tap water. *Appl. Environ. Microbiol.* 55, 1223–1229. doi: 10.1128/aem.55.5.1223-1229.1989

Demeter, K., Derx, J., Komma, J., Parajka, J., Schijven, J., Sommer, R., et al. (2021). Modelling the interplay of future changes and wastewater management measures on the microbiological river water quality considering safe drinking water production. *Sci. Total Environ.* 768:144278. doi: 10.1016/j.scitotenv.2020.144278

Derx, J., Schijven, J., Sommer, R., Zoufal-Hruza, C. M., van Driezum, I. H., Reischer, G., et al. (2016). QMRAcatch: human-associated fecal pollution and infection risk modeling for a river/floodplain environment. *J. Environm. Q.* 45, 1205–1214. doi: 10.2134/jeq2015.11.0560

Dorner, S. M., Huck, P. M., and Slawson, R. M. (2004). Estimating potential environmental loadings of *Cryptosporidium* spp. and *Campylobacter* spp. from livestock in the Grand River Watershed, Ontario, Canada. *Environ. Sci. Technol.* 38, 3370–3380. doi: 10.1021/es035208

European Commission (2018). *Urban Waste Water Website: Austria*. Available online at: http://uwwtd.oieau.fr/Austria/ (accessed May 25, 2018).

Farnleitner, A., Derx, J., Frick, C., Reiner, P., Savio, D., Zoufal-Hruza, C., et al. (2014). *Water Connection (New) Danube. Lower Lobau (Nationalpark Donauauen), Water Quality Report for Microbiology/Water Hygiene*. Vienna: Municipal Department MA45.

Farnleitner, A. H., Ryzinska-Paier, G., Reischer, G. H., Burtscher, M. M., Knetsch, S., Kirschner, A. K. T., et al. (2010). *Escherichia coli* and enterococci are sensitive and reliable indicators for human, livestock and wildlife faecal pollution in alpine mountainous water resources. *J. Appl. Microbiol.* 109, 1599–1608. doi: 10.1111/j.1365-2672.2010.04788.x

Franz, E., Schijven, J., Husman, A. M. D., 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

Frick, C., Vierheilig, J., Linke, R., Savio, D., Zornig, H., Antensteiner, R., et al. (2018). Poikilothermic animals as a previously unrecognized source of fecal indicator bacteria in a backwater ecosystem of a large river. *Appl. Environ. Microbiol.* 84:aem.00715-00718. doi: 10.1128/aem.00715-18

Frick, C., Vierheilig, J., Nadiotis-Tsaka, T., Ixenmaier, S., Linke, R., Reischer, G. H., et al. (2020). Elucidating fecal pollution patterns in alluvial water resources by linking standard fecal indicator bacteria to river connectivity and genetic microbial source tracking. *Water Res.* 184:116132. doi: 10.1016/j.watres.2020.116132

Frühauf, J., and Sabathy, E. (2006a). Untersuchung an Schilf- und Wasservögeln in der unteren Lobau, Teil I: Bestände und Habitat. *Wissenschaftliche Reihe* 23, 1–68. doi: 10.1515/9783111624150-002

Frühauf, J., and Sabathy, E. (2006b). Untersuchung an Schilf- und Wasservögeln in der unteren Lobau, Teil II: Arten. *Wissenschaftliche Reihe* 24, 1–80. doi: 10.1515/9783486784503-004

Gabriel, H., Blaschke, A. P., Taschke, R., and Mayr, E. (2014). *Water Connection (New) DanubeLower Lobau (Nationalpark Donauauen), Water Quantity Report for Surface Water*. Vienna: Municipial Department MA45 Vienna Waters.

Galway, L. P., Allen, D. M., Parkes, M. W., Li, L., and Takaro, T. K. (2015). Hydroclimatic variables and acute gastro-intestinal illness in British Columbia, Canada: a time series analysis. *Water Resour. Res.* 51, 885–895. doi: 10.1002/2014WR015519

Garcia-Presedo, I., Pedraza-Diaz, S., Gonzalez-Warleta, M., Mezo, M., Gomez-Bautista, M., Ortega-Mora, L. M., et al. (2013). The first report of *Cryptosporidium bovis*, *C. ryanae* and *Giardia duodenalis* sub-assemblage A-II in roe deer (*Capreolus capreolus*) in Spain. *Veter. Parasitol.* 197, 658–664. doi: 10.1016/j.vetpar.2013.07.002

Graczyk, T. K., Fayer, R., Trout, J. M., Lewis, E. J., Farley, C. A., Sulaiman, I., et al. (1998). *Giardia* sp. cysts and infectious *Cryptosporidium parvum* oocysts in the feces of migratory Canada geese (*Branta canadensis*). *Appl. Environ. Microbiol.* 64, 2736–2738. doi: 10.1128/aem.64.7.2736-2738.1998

Grant, S. B., Sanders, B. F., Boehm, A. B., Redman, J. A., Kim, J. H., Mrše, R. D., et al. (2001). Generation of enterococci bacteria in a coastal saltwater marsh and its impact on surf zone water quality. *Environ. Sci. Technol.* 35, 2407–2416. doi: 10.1021/es0018163

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

Guber, A. K., Fry, J., Ives, R. L., and Rose, J. B. (2015). *Escherichia coli* survival in, and release from, white-tailed deer feces. *Appl. Environ. Microbiol.* 81, 1168–1176. doi: 10.1128/Aem.03295-14

Guber, A. K., Pachepsky, Y. A., Dao, T. H., Shelton, D. R., and Sadeghi, A. M. (2013). Evaluating manure release parameters for nonpoint contaminant transport model KINEROS2/STWIR. *Ecol. Mod.* 263, 126–138. doi: 10.1016/j.ecolmodel.2013.05.008

Hahn, S., Bauer, S., and Klaassen, M. (2007). Estimating the contribution of carnivorous waterbirds to nutrient loading in freshwater habitats. *Fresh. Biol.* 52, 2421–2433. doi: 10.1111/j.1365-2427.2007.01838.x

Hein, T., Blaschke, A. P., Haidvogl, G., Hohensinner, S., Kucera-Hirzinger, V., Preiner, S., et al. (2006). Optimised management strategies for the Biosphere reserve Lobau, Austria, based on a multi criteria decision support system. *Ecohydrol. Hydrobiol.* 6, 25–36. doi: 10.1016/S1642-3593(06)70123-9

Helsel, D. R. (2006). Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65, 2434–2439. doi: 10.1016/j.chemosphere.2006.04.051

Hinterberger, B., Arnberger, A., Brandenburg, C., and Cermak, P. (2000). *Besucherstromanalyse für den Wiener Bereich des Nationalpark Donau-Auen - Lobau: GIS-Implementierung und erste Ergebnisse*. Wien: Magistrat der Stadt Wien, Magistratsabteilung, 49.

Hogan, J. N., Daniels, M. E., Watson, F. G., Conrad, P. A., Oates, S. C., Miller, M. A., et al. (2012). Longitudinal poisson regression to evaluate the epidemiology of *Cryptosporidium, Giardia*, and fecal indicator bacteria in coastal California wetlands. *Appl. Environ. Microbiol.* 78, 3606–3613. doi: 10.1128/AEM.00578-12

Hong, E. M., Park, Y., Muirhead, R., Jeong, J., and Pachepsky, Y. A. (2018). Development and evaluation of the bacterial fate and transport module for the Agricultural Policy/Environmental eXtender (APEX) model. *Sci. Total Environ.* 615, 47–58. doi: 10.1016/j.scitotenv.2017.09.231

ISO (2006). ISO 1553:Water Quality Isolation and Identification of *Cryptosporidium* oocysts and *Giardia* cysts from water. Technical Committee: ISO/TC 147/SC 4 Microbiological methods, ICS : 07.100.20 Microbiology of water.

ISO (2001). ISO 16649-1: Microbiology of food and animal feeding stuffs – Horizontal method for the enumeration of beta-glucuronidase-positive Escherichia coli – Part 1: Colony-count technique at 44 degrees C using

membranes and 5-bromo-4-chloro-3-indolyl beta-D-glucuronide. Technical Committee : ISO/TC 34/SC 9 Microbiology, ICS : 07.100.30 Food microbiology.

Ives, R. L., Kamarainen, A. M., John, D. E., and Rose, J. B. (2007). Use of cell culture to assess *Cryptosporidium parvum* survival rates in natural groundwaters and surface waters. *Appl. Environ. Microbiol.* 73, 5968–5970. doi: 10.1128/Aem.00347-07

Kay, D., Aitken, M., Crowther, J., Dickson, I., Edwards, A. C., Francis, C., et al. (2007). Reducing fluxes of faecal indicator compliance parameters to bathing waters from diffuse agricultural sources: the Brighouse Bay study, Scotland. *Environ. Pollut.* 147, 138–149. doi: 10.1016/j.envpol.2006.08.019

Kirschner, A. K. T., Reischer, G. H., Jakwerth, S., Savio, D., Ixenmaier, S., Toth, E., et al. (2017). Multiparametric monitoring of microbial faecal pollution reveals the dominance of human contamination along the whole Danube river. *Water Res.* 124, 543–555. doi: 10.1016/j.watres.2017.07.052

Kobayashi, A., Sano, D., Hatori, J., Ishii, S., and Okabe, S. (2013). Chicken and duck-associated *Bacteroides*-Prevotella genetic markers for detecting fecal contamination in environmental water. *Appl. Microbiol. Biotechnol.* 97, 7427–7437. doi: 10.1007/s00253-012-4469-2

Layton, B. A., Cao, Y., Ebentier, D. L., Hanley, K., Ballesté, E., Brandão, J., et al. (2013). Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res.* 47, 6897–6908. doi: 10.1016/j.watres.2013.05.060

Liao, H., Krometis, L.-A. H., and Kline, K. (2016). Coupling a continuous watershed-scale microbial fate and transport model with a stochastic dose-response model to estimate risk of illness in an urban watershed. *Sci. Total Environ.* 55, 668–675. doi: 10.1016/j.scitotenv.2016.02.044

Liu, W.-C., Chan, W.-T., and Young, C.-C. (2015). Modeling fecal coliform contamination in a tidal Danshuei river estuarine system. *Sci. Total Environ.* 502, 632–640. doi: 10.1016/j.scitotenv.2014.09.065

Mayer, R. E., Sofill-Mas, S., Egle, L., Reischer, G. H., Schade, M., Fernandez-Cassi, X., et al. (2016). Occurrence of human-associated *Bacteroidetes genetic* source tracking markers in raw and treated wastewater of municipal and domestic origin and comparison to standard and alternative indicators of faecal pollution. *Water Res.* 90, 265–276. doi: 10.1016/j.watres.2015.12.031

Medema, G. J., and Schijven, J. F. (2001). Modelling the sewage discharge and dispersion of *Cryptosporidium* and *Giardia* in surface water. *Water Res.* 35, 4307–4316. doi: 10.1016/S0043-1354(01)00161-0

Mieszkin, S., Furet, J. P., Corthier, G., and Gourmelon, M. (2009). Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045–3054. doi: 10.1128/AEM.02343-08

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part Ia discussion of principles. *J. Hydrol.* 10, 282–290. doi: 10.1016/0022-1694(70)90255-6

Oladeinde, A., Bohrmann, T., Wong, K., Purucker, S. T., Bradshaw, K., Brown, R., et al. (2014). Decay of fecal indicator bacterial populations and bovine-associated source-tracking markers in freshly deposited cow pats. *Appl. Environ. Microbiol.* 80, 110–118. doi: 10.1128/aem.02203-13

Olson, M. E., Goh, J., Phillips, M., Guselle, N., and McAllister, T. A. (1999). *Giardia* cyst and *Cryptosporidium* oocyst survival in water, soil, and cattle feces. *J. Environ. Q.* 28, 1991–1996. doi: 10.2134/jeq1999.00472425002800060040x

Parz-Gollner, R. (2006). Zur Situation der Kormoranschlafplätze im Nationalpark Donau-Auen (NÖ) - Auswirkungen der Uferrückbauten im Bereich des Schlafplatzes Turnhaufen. Universität für Bodenkultur, Department für Integrative Biologie und Biodiversitätsforschung , Vienna, Austria.

Pavlik, M., Wildeman, T., Kohn, K., Emerick, J., and Robinson, R. (1999). Fate and transport of metals in a natural wetland receiving mine drainage. *J. Am. Soc. Mining Reclamat.* 1999, 563–578. doi: 10.21000/JASMR99010563

Peterson, E. W., and Hanna, L. A. (2016). Estrogen reduction in a coupled wetland and ground water flow-through system. *Environ. Earth Sci.* 75:384. doi: 10.1007/s12665-016-5259-4

Raith, M. R., Kelty, C. A., Griffith, J. F., Schriewer, A., Wuertz, S., Mieszkin, S., et al. (2013). Comparison of PCR and quantitative real-time PCR methods for the characterization of ruminant and cattle fecal pollution sources. *Water Res.* 47, 6921–6928. doi: 10.1016/j.watres.2013.03.061

Reckendorfer, W., Funk, A., Gschöpf, C., Hein, T., and Schiemer, F. (2013). Aquatic ecosystem functions of an isolated floodplain and their implications for flood retention and management. *J. Appl. Ecol.* 50, 119–128. doi: 10.1111/1365-2664.12029

Regli, S., Rose, J. B., Haas, C. N., and Gerba, C. P. (1991). Modeling the risk from giardia and viruses in drinking-water. *J. Am. Water Works Assoc.* 83, 76–84. doi: 10.1002/j.1551-8833.1991.tb07252.x

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Astrom, J., et al. (2013). Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* 47, 8548–8556. doi: 10.1021/es304367t

Reischer, G. H., Haider, J. M., Sommer, R., Stadler, H., Keiblinger, K. M., Hornek, R., et al. (2008). Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. *Environ. Microbiol.* 10, 2598–2608. doi: 10.1111/j.1462-2920.2008.01682.x

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

Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., and Farnleitner, A. H. (2006). Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72, 5610–5614. doi: 10.1128/Aem.00364-06

Reischer, G. H., Kollanur, D., Vierheilig, J., Wehrspaun, C., Mach, R. L., Sommer, R., et al. (2011). Hypothesis-driven approach for the identification of fecal pollution sources in water resources. *Environ. Sci. Technol.* 45, 4038–4045. doi: 10.1021/es103659s

Sanders, B. F., Arega, F., and Sutula, M. (2005). Modeling the dry-weather tidal cycling of fecal indicator bacteria in surface waters of an intertidal wetland. *Water Res.* 39, 3394–3408. doi: 10.1016/j.watres.2005.06.004

Schijven, J., Derx, J., Husman, A. M. D., Blaschke, A. P., and Farnleitner, A. H. (2015). QMRAcatch: microbial quality simulation of water resources including infection risk assessment. *J. Environ. Q.* 44, 1491–1502. doi: 10.2134/jeq2015.01.0048

Schijven, J. F., Teunis, P. P. M., Rutjes, S. A., Bouwknegt, M., and Husman, A. M. D. (2011). QMRAspot: a tool for quantitative microbial risk assessment from surface water to potable water. *Water Res.* 45, 5564–5576. doi: 10.1016/j.watres.2011.08.024

Schmidt, M., Sommer, K., Kriebitzsch, W. U., Ellenberg, H., and von Oheimb, G. (2004). Dispersal of vascular plants by game in northern Germany. part I: roe deer (*Capreolus capreolus*) and wild boar (*Sus scrofa*). *Eur. J. Forest Res.* 123, 167–176. doi: 10.1007/s10342-004-0029-3

Schreiber, H., Behrendt, H., Constantinescu, L. T., Cvitanic, I., Drumea, D., Jabucar, D., et al. (2005). Nutrient emissions from diffuse and point sources into the River Danube and its main tributaries for the period of 1998–2000 results and problems. *Water Sci. Technol.* 51, 283–290. doi: 10.2166/wst.2005.0602

Schulze, C. H., and Schütz, C. (2013). Gewässervernetzung (Neue) Donau Untere Lobau (Nationalpark Donau Auen). Wissenschaftliche Beweissicherung Lausgrund: Erhebung der Brutvögel mit Gewässerbindung 2012 (Unpublished Project Report). Wasser Cluster Lunz.

Signor, R. S., and Ashbolt, N. J. (2009). Comparing probabilistic microbial risk assessments for drinking water against daily rather than annualised infection probability targets. *J. Water Health* 7, 535–543. doi: 10.2166/wh.2009.101

Stadler, P., Loken, L. C., Crawford, J. T., Schramm, P. J., Sorsa, K., Kuhn, C., et al. (2019). Spatial patterns of enzymatic activity in large water bodies: Ship-borne measurements of beta-D-glucuronidase activity as a rapid indicator of microbial water quality. *Sci. Total Environ.* 651, 1742–1752. doi: 10.1016/j.scitotenv.2018.10.084

Stalder, G. L., Farnleitner, A., Sommer, R., Beiglbock, C., and Walzer, C. (2011). Hazard- and risk based concepts for the assessment of microbiological water quality-part 2. *Wiener Tierarztliche Monatsschrift* 98, 54–65.

Sterk, A., Schijven, J., Husman, A. M. D., and de Nijs, T. (2016). Effect of climate change on runoff of *Campylobacter* and *Cryptosporidium* from land to surface water. *Water Res.* 95, 90–102. doi: 10.1016/j.watres.2016.03.005

Stevenson, M. E., Blaschke, A. P., Toze, S., Sidhu, J. P., Ahmed, W., van Driezum, I. H., et al. (2015). Biotin- and glycoprotein-coated microspheres as surrogates for studying filtration removal of *Cryptosporidium parvum* in a granular

limestone aquifer medium. *Appl. Environ. Microbiol.* 81, 4277–4283. doi: 10.1128/AEM.00885-15

Teunis, P. F., and Havelaar, A. H. (2000). The beta poisson dose-response model is not a single-hit model. *Risk Anal.* 20, 513–520. doi: 10.1111/0272-4332.204048

Tolouei, S., Dewey, R., Snodgrass, W. J., Edge, T. A., Andrews, R. C., Taghipour, M., et al. (2019). Assessing microbial risk through event-based pathogen loading and hydrodynamic modelling. *Sci. Total Environ.* 693:133567. doi: 10.1016/j.scitotenv.2019.07.373

van Elsas, J. D., Semenov, A. V., Costa, R., and Trevors, J. T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5, 173–183. doi: 10.1038/ismej.2010.80

Van Rossum, C. T. M., Fransen, H. P., Verkaik-Kloosterman, J., Buurma-Rethans, E. J. M., and Ocké, M. C. (2011). Durch National Food Consumption Survey 2007-2010: Diet of Children and Adults Aged 7 to 69 Years. *National Institute of Public Health and the Environment (RIVM) Report 350050006/2011*.

Vermeulen, L. C., de Kraker, J., Hofstra, N., Kroeze, C., and Medema, G. (2015). Modelling the impact of sanitation, population growth and urbanization on human emissions of *Cryptosporidium* to surface waters-a case study for Bangladesh and India. *Environ. Res. Lett.* 10:094017. doi: 10.1088/1748-9326/10/9/094017

Vierheilig, J., Frick, C., Mayer, R. E., Kirschner, A. K. T., Reischer, G. H., Derx, J., et al. (2013). *Clostridium perfringens* is not suitable for the indication of fecal pollution from ruminant wildlife but is associated with excreta from nonherbivorous animals and human sewage. *Appl. Environ. Microbiol.* 79, 5089–5092. doi: 10.1128/Aem.01396-13

von Oheimb, G., Schmidt, M., Kriebitzsch, W.-U., and Ellenberg, H. (2005). Dispersal of vascular plants by game in northern Germany. part II: red deer (*Cervus elaphus*). *Eur. J. Forest Res.* 124, 55–65. doi: 10.1007/s10342-005-0053-y

Willmott, C. J., and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* 30, 79–82. doi: 10.3354/cr030079

Wu, B., Wang, C., Zhang, C., Sadowsky, M. J., Dzakpasu, M., and Wang, X. C. (2020). Source-associated gastroenteritis risk from swimming exposure to aging fecal pathogens. *Environ. Sci. Technol.* 54, 921–929. doi: 10.1021/acs.est.9b01188

Zhang, Q., Gallard, J., Wu, B., Harwood, V. J., Sadowsky, M. J., Hamilton, K. A., et al. (2019). Synergy between quantitative microbial source tracking (qMST) and quantitative microbial risk assessment (QMRA): a review and prospectus. *Environ. Internat.* 130:104703. doi: 10.1016/j.envint.2019.03.051

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.

Copyright © 2021 Derx, Demeter, Linke, Cervero-Aragó, Lindner, Stalder, Schijven, Sommer, Walochnik, Kirschner, Komma, Blaschke and Farnleitner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 July 2021
doi: 10.3389/fmicb.2021.697553

Comparative Analysis of Fecal Microbiomes From Wild Waterbirds to Poultry, Cattle, Pigs, and Wastewater Treatment Plants for a Microbial Source Tracking Approach

Amine M. Boukerb[1†], Cyril Noël[2], Emmanuelle Quenot[1], Bernard Cadiou[3], Julien Chevé[4], Laure Quintric[2], Alexandre Cormier[2], Luc Dantan[1] and Michèle Gourmelon[1*†]*

[1] IFREMER, RBE-SGMM-LSEM, Laboratoire Santé Environnement Microbiologie, Plouzané, France, [2] IFREMER – PDG-IRSI-SEBIMER, Plouzané, France, [3] Bretagne Vivante – SEPNB, Brest, France, [4] IFREMER, ODE-UL-LERBN, Laboratoire Environnement Ressource Bretagne Nord, Dinard, France

OPEN ACCESS

Edited by:
Saurabh Dhiman, South Dakota School of Mines and Technology, United States

Reviewed by:
Inderjeet Tyagi, Zoological Survey of India, India
Maite Muniesa, University of Barcelona, Spain

***Correspondence:**
Amine M. Boukerb
amine.boukerb@univ-rouen.fr
Michèle Gourmelon
michele.gourmelon@ifremer.fr

†Present address:
Amine M. Boukerb, Laboratoire de Microbiologie Signaux et Micro-Environnement, LMSM EA 4312, Université de Rouen-Normandie, Normandie Université, Evreux, France
Michèle Gourmelon, IFREMER, ODE-DYNECO-Pelagos, Laboratoire d'Ecologie Pélagique, Plouzané, France

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 19 April 2021
Accepted: 14 June 2021
Published: 14 July 2021

Citation:
Boukerb AM, Noël C, Quenot E, Cadiou B, Chevé J, Quintric L, Cormier A, Dantan L and Gourmelon M (2021) Comparative Analysis of Fecal Microbiomes From Wild Waterbirds to Poultry, Cattle, Pigs, and Wastewater Treatment Plants for a Microbial Source Tracking Approach. Front. Microbiol. 12:697553. doi: 10.3389/fmicb.2021.697553

Fecal pollution in coastal areas is of a high concern since it affects bathing and shellfish harvesting activities. Wild waterbirds are non-negligible in the overall signal of the detectable pollution. Yet, studies on wild waterbirds' gut microbiota focus on migratory trajectories and feeding impact on their shape, rare studies address their comparison to other sources and develop quantitative PCR (qPCR)-based Microbial Source Tracking (MST) markers to detect such pollution. Thus, by using 16S rRNA amplicon high-throughput sequencing, the aims of this study were (i) to explore and compare fecal bacterial communities from wild waterbirds (i.e., six families and 15 species, n = 275 samples) to that of poultry, cattle, pigs, and influent/effluent of wastewater treatment plants (n = 150 samples) and (ii) to develop new MST markers for waterbirds. Significant differences were observed between wild waterbirds and the four other groups. We identified 7,349 Amplicon Sequence Variants (ASVs) from the hypervariable V3–V4 region. Firmicutes and Proteobacteria and, in a lesser extent, Actinobacteria and Bacteroidetes were ubiquitous while Fusobacteria and Epsilonbacteraeota were mainly present in wild waterbirds. The clustering of samples in non-metric multidimensional scaling (NMDS) ordination indicated a by-group clustering shape, with a high diversity within wild waterbirds. In addition, the structure of the bacterial communities was distinct according to bird and/or animal species and families (Adonis R^2 = 0.13, p = 10^{-4}, Adonis R^2 = 0.11, p = 10^{-4}, respectively). The Analysis of Composition of Microbiomes (ANCOM) showed that the wild waterbird group differed from the others by the significant presence of sequences from Fusobacteriaceae (W = 566) and Enterococcaceae (W = 565) families, corresponding to the *Cetobacterium* (W = 1427) and *Catellicoccus* (W = 1427) genera, respectively. Altogether, our results suggest that some waterbird members present distinct fecal microbiomes allowing the design of qPCR MST markers. For instance, a swan- and an oystercatcher-associated markers (named Swan_2 and Oyscab, respectively) have been developed. Moreover, bacterial

genera harboring potential human pathogens associated to bird droppings were detected in our dataset, including enteric pathogens, i.e., *Arcobacter*, *Clostridium*, *Helicobacter*, and *Campylobacter*, and environmental pathogens, i.e., *Burkholderia* and *Pseudomonas*. Future studies involving other wildlife hosts may improve gut microbiome studies and MST marker development, helping mitigation of yet unknown fecal pollution sources.

Keywords: microbiome, wild waterbird, fecal pollution, microbial source tracking, enteric pathogens, environmental pathogens, NGS, qPCR

INTRODUCTION

Fecal pollution originating from urban areas, agriculture, and wildlife can significantly impair aquatic systems including coastal areas. This pollution can come from multiple sources, i.e., failing wastewater infrastructure and wastes from upstream livestock animals and wildlife, which are mainly routed by runoff following heavy rain events. Several outbreaks of food poisoning through the consumption of contaminated shellfish or surface water have been reported due to such pollution episodes (Potasman et al., 2002; Yoder et al., 2008; European Food Safety Authority (EFSA), and European Centre for Disease Prevention and Control (ECDC), 2019). In addition, this fecal pollution can also cause economic losses from bathing restrictions and closure of shellfish-harvesting areas.

Human microbial pathogens in water include bacterial pathogens represented by two categories, enteric (mainly attributed to fecal pollution) and environmental (autochthonous within aquatic and terrestrial habitats). Testing for major waterborne pathogens would give straightforward interpretation of human health risk (Harwood et al., 2014), but this is impossible since it is time and cost consuming. To avoid this, culture-based methods of fecal indicator bacteria (FIB; *Escherichia coli* and enterococci) are used to assess fecal pollution levels by regulatory agencies to limit exposition to impairment water bodies. Thus, these FIB have been detected in freshwaters, coastal waters, and shellfish as well as a selection of pathogenic bacteria such as *Salmonella* spp. and *Campylobacter* spp. by culture-based methods, or human viruses such as noroviruses by molecular methods (Yoder et al., 2008; Iwamoto et al., 2010; Westrell et al., 2010; Rince et al., 2018). These microorganisms mainly originated from different fecal sources (Pommepuy et al., 2006), while it remains difficult to determine their precise origin, which limits direct and cost-effective remediation efforts. Thus, molecular methods including qPCR Microbial Source Tracking (MST)-based methods are continually developed to identify fecal pollution sources by targeting specific markers.

The recent developments in MST methods helps in discriminating between human and non-human sources of fecal contamination in environmental waters and distinguishing contamination from different animal species and is particularly useful for non-point or multiple sources. This approach consists of investigating the presence of a source-specific target such as microorganisms (e.g., bacteria and viruses) or chemical compounds (e.g., stanols) associated with specific hosts (Boehm et al., 2013; Harwood et al., 2014; Jarde et al., 2018). An MST marker is considered as efficient if it detects its target when the intended host feces are present in the sample (sensitivity) or it does not detect its target in samples that do not contain the host feces (specificity). An MST marker is usually considered sensitive and specific if both these metrics exceed 80% (Boehm et al., 2013). Library-independent MST based on molecular methods targeting bacteria, such as quantification of host-associated Bacteroidales 16S rRNA gene marker by real-time PCR, was found efficient. A high number of qPCR MST markers were validated and already applied on environmental samples (Seurinck et al., 2005; Reischer et al., 2006; Mieszkin et al., 2009; Jarde et al., 2018). These markers target mainly human, pig, cattle, and pet sources while few markers targeting avian sources have yet been described (Reischer et al., 2006; Mieszkin et al., 2009; Ryu et al., 2012; Boehm et al., 2013; Ohad et al., 2016). For these latest sources, we can enumerate two general avian markers (GFD and AV4143), two gull markers targeting the bacterial species *Catellicoccus marimammalium* (Gull2 and Gull4), and poultry (AV43) and poultry litter (LA35) markers (Lu et al., 2008; Green et al., 2012; Ryu et al., 2012; Weidhaas and Lipscomb, 2013; Ohad et al., 2016).

With the rapid and cost-effective development of the Next-Generation Sequencing (NGS) technologies, our understanding of the enteric microbiomes of human, animals, and birds is significantly increasing (Ohad et al., 2016; Unno et al., 2018). Birds are the most diverse group of amniotic vertebrates with more than 10,000 and 20,000 described species and subspecies, respectively[1] (checked on February 24, 2021). Since most avian microbiome studies have focused on economically important species, such as chicken and turkey, or the impact of diet and behavior on their diversity, little is known on wild waterbirds (Grond et al., 2018; Youngblut et al., 2019). Expanding fecal microbiome data from the range of avian species, including wildlife, will improve our understanding of such microbiome diversity, for example, in relation to migratory or diet habits (Grond et al., 2018). Few studies described bacterial community structure in wild birds, e.g., a selection of four wild waterbird species in Israel (Laviad-Shitrit et al., 2019), Artic-breeding shorebirds (Artic and sub-Artic of North America) (Grond et al., 2019), Canada geese (Lu et al., 2009), shorebirds (red knots) (Grond et al., 2014; Ryu et al., 2014), gulls in United States (Lu et al., 2008; Koskey et al., 2014), or wild geese in China (Wang et al., 2018). Several studies revealed the

[1] avibase.bsc-eoc.org

dominance of Proteobacteria and Bacteroidetes within the avian gastrointestinal (GI) tract, with Firmicutes typically present in any avian fecal sample with different proportions (Grond et al., 2018; Sun et al., 2018). However, samples from captive poultry harbored higher proportions of Firmicutes than those from wild birds (Waite and Taylor, 2015). In Europe, and especially in France, studies on the characterization of such microbiomes are scarce. On the French coastal areas, a high variety of waterbird species including resident and migratory birds is present, with less than 500,000 birds during the breeding season and more than 1 million birds in winter (Issa and Muller, 2015).

From a microbial source tracking view, Ohad et al. (2016) were the latest to develop the AV4143 and AV43 markers targeting avian and chicken sources, respectively, using NGS data. Hence, such data can be used in studying bacterial community transfer from sources of pollution to sinks using the SourceTracker Bayesian approach (Knights et al., 2011). This method has been successfully used in different countries and conditions, helping in the identification of putative fecal sources of contamination (Henry et al., 2016; Brown et al., 2017).

Wild waterbirds are known as reservoirs of enteric bacterial pathogens such as *Campylobacter* spp. and *Salmonella* spp. (Waldenstrom et al., 2007; Chung et al., 2018; Smith et al., 2020). Studying wild waterbird fecal microbiomes could be useful to overview potential bacterial pathogens carried by those hosts. Thus, bacterial genera including potential human pathogens such as *Clostridium sensu stricto 1*, *Fusobacterium*, *Campylobacter*, and *Helicobacter* were identified within bacterial communities from waterbird species in Israel (i.e., great cormorants, little egrets, black-crowned night herons, and black-headed gulls; Laviad-Shitrit et al., 2019).

In the present study, we aimed to explore fecal microbiomes and develop new MST toolbox from a selection of wild waterbirds in France. The objectives of this study were (1) to characterize the fecal bacterial communities of several wild waterbirds within different families such as Anatidae, Laridae, Haematopodidae, Scolopacidae, Phalacrocoracidae, and Hydrobatidae; (2) to compare one fecal microbiome to another from livestock animals (cattle and pigs), poultry, and wastewater samples from wastewater treatment plants (WWTPs); (3) to develop novel bacterial MST qPCR markers targeting specific waterbird species; and (4) to evaluate the presence and distribution of bacterial genera (including enteric and environmental pathogens) within our metabarcoding dataset.

MATERIALS AND METHODS

Fecal Sample Collection and Location

Fresh fecal samples from wild waterbirds, poultry, cattle, pigs, and WWTPs (n = 425) were collected from July 2014 to November 2017 from different regions in France, mainly from Brittany: 80.4% of wild waterbirds, 40.3% of poultry, 100% of pigs, 100% of cattle, and 100% of WW. The samples from the other regions were collected from Nouvelle-Aquitaine (18.2% of wild waterbird and 29.8% of poultry samples), Normandy (25.4% of poultry samples), and Occitanie (0.4% of wild waterbird and 4.5% of poultry samples). Wild waterbird samples (n = 275) were collected from July 2016 to October 2017, and represented by six families and 15 species, with two genera for which the species was not always specified (i.e., *Larus* spp. and *Chroicocephalus* spp.; **Supplementary Tables 1, 2**). More precisely, the families of wild waterbirds are as follows: (i) Laridae [86; 83 Larinae (subfamily), seagulls and gulls: 45 *Larus* spp., 15 *Larus argentatus*, 7 *Larus marinus*, 3 *Chroicocephalus* spp., 13 *Chroicocephalus ridibundus*, and 3 Sternidae (subfamily): terns (*Thalasseus sandvicensis*)], (ii) Phalacrocoracidae [24 cormorants; 21 great cormorants (*Phalacrocorax carbo*) and 3 European shags (*Phalacrocorax aristotelis*)], (iii) Scolopacidae [5 dunlins (*Calidris alpina*), 11 red knots (*Calidris canutus*), and 11 curlews (*Numenius arquata*)] and Haematopodidae [23 oystercatchers (*Haematopus ostralegus*)] (these two families are part of wader-type birds), (iv) Hydrobatidae (two storm-petrels; *Hydrobates pelagicus*), and (v) Anatidae [113; 20 mallards (*Anas platyrhynchos*), 56 Brent goose (*Branta bernicla*), 12 wild swans (*Cygnus olor*), and 25 common shelducks (*Tadorna tadorna*)]. More details such as the geographical coordinates of the wild bird sample collection sites and their diets can be obtained from **Supplementary Tables 1, 2** and the metadata **Supplementary Table 3**, respectively.

In addition, this collection included fecal samples from poultry (droppings or litters; n = 67) and non-avian sources: livestock animals (33 cattle and 37 pigs; both feces and manure) and 13 (9 input and 4 output) WWTP water samples (**Supplementary Table 1**).

For wild waterbird samples, we distantly observed them until they defecated and flew away, and the fecal samples were immediately collected. For some species, samples were also collected from chicks in their nest. It should be noted that the overall number of samples for the different avian species is variable due to the difficulty in collecting feces from some wild waterbird species that are randomly distributed within the sampling dates. Poultry litters and cattle and pig manures were sampled in farms or in a research institute (ANSES, Ploufragan, France). Wastewater samples were collected at the input and output of three WWTPs. All samples were collected aseptically; then, they were transported to the laboratory on dry ice within 24 h of collection. The samples collected outside the Brittany region were sent to the laboratory on ice within 2 days. All samples were aseptically homogenized. For solid samples, aliquots of about 0.25 g wet weight (droppings, litters, slurries, or manures) were stored at −80°C prior to total DNA extraction. For liquid samples, 10 ml (6 out of the 24 pig slurry samples), 20 or 50 ml (influents of WWTP), and 50 or 100 ml (effluents of WWTP) were filtered onto 0.45-μm nitrocellulose filters. For the 18 additional pig slurry samples, 45 ml was centrifuged at 8,000 rpm for 15 min and supernatants were discarded. Filters and pellets were also stored at −80°C prior to total DNA extraction.

Total DNA Extraction

Microbial genomic DNA was extracted directly from about 0.25 g of fecal material for solid samples and from filters or pellets for liquid samples using the FastDNA^{TM} Spin Kit for Soil (MP Biomedicals, Illkirsh, France) according to the manufacturer's

recommendations. DNA was eluted in a final volume of 100 μl of sterile DNA/RNA-free water. The quality and quantity of DNA were determined using spectrophotometry (NanoDrop) and a Qubit fluorometric system (Thermo Fisher Scientific). All extracts were stored at −80°C prior to the 16S rRNA amplicon library preparation and sequencing.

16S rRNA Library Generation and MiSeq Sequencing

Amplification of the V3–V4 hypervariable region of the 16S rRNA loci was performed using the primer set PCR1F_460 (5′-ACGGRAGGCAGCAG-3′) and PCR1R_460 (5′-TACCAGGGTATCTAATCCT-3′) (Andersson et al., 2008; Liu et al., 2008). The 50-μl final-volume PCR1 reactions contained 5 × PCR buffer (Phusion), 10 mM of dNTP, 0.5 μM of each primer, 5 U of Taq Phusion, and 6 μl of genomic DNA. PCR conditions were as follows: one predenaturation step at 95°C for 5 min, 30 cycles of denaturation at 95°C for 20 s, annealing at 65°C for 30 s, and extension at 72°C for 30 s, one post-elongation step at 72°C for 5 min, and then 4°C forever. PCR product quality and integrity were determined using 1% agarose gel electrophoresis. PCR product purification and secondary PCR amplification for the addition of the Illumina compatible sequencing adapters and unique per-sample indexes were conducted at GenoToul facility (Toulouse, France). Barcoded amplicons were quantified, quality-checked, normalized, pooled, and sequenced within two sequencing runs (May and November 2017) using the 2 × 250 paired-end method on an Illumina MiSeq instrument with a MiSeq Reagent Kit V3 chemistry (Illumina), according to the manufacturer's recommendations.

Bioinformatics

Bacterial Community Analysis

Raw data were analyzed using the SAMBA v2.0.0 workflow[2], a Standardized and Automatized MetaBarcoding Analysis workflow using DADA2 (Callahan et al., 2016) and QIIME2 (Bolyen et al., 2019) with default parameters unless otherwise indicated. This workflow developed by the SeBiMER (Ifremer's Bioinformatics Core Facility) is an open-source modular workflow to process eDNA metabarcoding data. SAMBA was developed using the NextFlow workflow manager (Di Tommaso et al., 2017) and built around three main parts: data integrity checking, bioinformatics processes, and statistical analyses. Firstly, a SAMBA checking process allows one to verify the raw data integrity. Afterward, sequencing primers were trimmed from reads, and reads where primers were not found have been removed. Then, DADA2 was used to filter bad quality reads, correct sequencing errors, overlap paired reads, infer Amplicon Sequence Variants (ASVs), and remove chimeras. Due to the known diversity overestimation generated by DADA2, an additional step of ASV clustering [Operational Taxonomic Unit (OTU) calling] has been performed using dbOTU3 algorithm (Olesen et al., 2017). Taxonomy classification was achieved using the SILVA database 132 (Quast et al., 2013; Glockner et al., 2017). Finally, SAMBA performs extensive analyses of the alpha- and beta-diversities using homemade R scripts (R Core Team, 2020). ASV abundances for each sample were generated at the phylum, family, and genera taxonomic levels. Bacterial community indices describing the alpha diversity included Chao1 and Shannon indices. Beta diversity analyses were achieved by ordination method using non-metric multidimensional scaling (NMDS) with Bray–Curtis and weighted UniFrac distance matrices (Lozupone and Knight, 2005). Significant differences in variance for each index depending on the group/species were tested by Kruskal–Wallis rank sum test and Dunn's *post hoc* test. Differences in microbial mean taxa abundance according to group/species were detected using ANCOM (Analysis of Composition of Microbiomes), with W value corresponding to the number of times an ASV abundance is significantly different for a group of samples (Mandal et al., 2015). Unique and overlapping ASVs to sample groups were plotted with UpsetR v.1.4.0 (Conway et al., 2017) by highlighting the associated taxonomy in each set.

Identification of Potential New MST Markers Based on NGS Data

In order to develop new MST qPCR markers, we investigated unique and host-associated ASVs in a group of samples belonging to a bird species, a bird family, or to a main group of samples (i.e., waterbirds, poultry, cattle, pigs, or wastewaters) using ANCOM. A validation of the retrieved specific/host-associated sequences from non-targeted sources was achieved by building a phylogenetic tree using sequences from already validated markers retrieved from GenBank (i.e., Pig2Bac, Rum2Bac, and HF183).

The selected sequences were compared to the NCBI nucleotide database[3] to retrieve the 20 best hits with the highest host diversity. These sequences were aligned with ClustalX (v2.1) to determine the variable regions (specific to the target sequence) and constant regions (common to non-target sequences). When regions specific to the targeted host were identified, primers and probes were drawn manually or using Primer3 (v4.1.0[4]), OligoCalc (v3.27) (Kibbe, 2007), and Multiple Primer Analyzer (Thermo Fisher Scientific). Then, the sensitivity and specificity of the designed primers and probes were assayed on target and non-target fecal samples using qPCR assays.

qPCR Assays

We assayed three potential targets: one ASV belonging to the genus *Romboutsia* and found associated with swans, one belonging to the genus *Bacteroides* associated with oystercatchers, and one belonging to the species *Paeniclostridium sordellii* and associated with cormorants. Two of these proposed candidates did not allow primers drawing due to the absence of regions specific to the target hosts (Anatidae and cormorants). For the ASV sequence belonging to the genus *Romboutsia* (swans), three pairs of primers were designed. For the ASV sequence belonging to the genus *Bacteroides* (oystercatchers), a pair of primers and a probe have been designed.

[2]https://github.com/ifremer-bioinformatics/samba

[3]https://blast.ncbi.nlm.nih.gov/Blast.cgi

[4]https://primer3.ut.ee/

Details on the marker genes, primer/probe sequences, and qPCR reaction conditions are listed in **Table 1**. For the Swan_2 marker, quantitative PCR assays were performed using TaqMan Mix (Invitrogen) with the following conditions: 1 step at 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 60 s. Reactions were carried out in a final volume of 25 µl with a final concentration of 300 nM of each primer and 200 nM of probe (Eurogentec, France) and 2 µl of DNA template. DNA samples were tested at 1/10 and 1/100 dilutions, and the appropriate dilution (weaker dilution without inhibition) was retained. Negative controls (no template DNA) were performed in triplicates for each run. A targeted synthetic oligonucleotide, gBlocks Gene fragment (IDT, Integrated DNA Technology), containing a 400-bp partial sequence of 16S rRNA gene of *Romboutsia* was used as standard at 10-fold dilutions ranging from 10^5 to 10 copies/qPCR. Correlation coefficients (r^2) for all the standard curves were > 0.99 and PCR efficiency ranged between 93.7% and 101.6%. Inhibition tests were performed by running serially diluted DNA templates (10- and 100-fold dilution).

Identification of Pathogen Groups and a Selection of MST Markers Based on NGS Data

The presence of a selection of 37 bacterial genera known to include enteric or environmental pathogens plus three genera known to harbor validated MST markers were investigated in the whole sequencing dataset. The list for pathogens was validated according to the Canadian ePATHogen risk group database[5] (accessed on February 17, 2021, **Supplementary Table 4**).

Then, we investigated co-occurrences of pathogens and a selection of three genera known to harbor MST markers. Relationships among the considered variables were tested using Spearman's coefficient in R version 4.0.3 (R Core Team, 2020) with statistical significance set at $p < 0.05$, 0.01, and 0.001.

Data Availability

GenBank accession numbers (BioProject: PRJNA722421) of the 16 ASVs selected as potential MST markers are listed on **Supplementary Table 5**.

The 16S rRNA dataset generated for this study can be found in the Sequence Read Archive from NCBI (BioProject: PRJNA722421).

Raw data are available on Dataref at Ifremer[6].

RESULTS

The identification of bacterial DNA sequences associated with an avian host and used for the development of MST qPCR markers included the acquisition of an original dataset from high-throughput sequencing of 16S rRNA amplicon from avian droppings, livestock animal feces and manure, and WWTP

[5]https://health.canada.ca/en/epathogen

[6]https://doi.org/10.12770/ddce4bf7-2b71-4929-aa20-50a8fd0e706a

TABLE 1 | qPCR target genes, primer/probe sequences, and reaction conditions for MST candidates Swan_1, Swan_2, and Swan_3 targeting swans, and Oyscab targeting oystercatchers.

Marker	Primers/probe	Sequences (5'-3')	Product size (bp)	Reaction conditions	Sensitivity (*n* of samples)/specificity (*n* of samples)
Swan_2	Swan_2F	GTAATACGTAGGGGGCAAG	135	1 cycle of 10 min at 95°C and 40 cycles of 15 s at 95°C and 1 min at 60°C	Sensitivity of 75% (*n* = 16 swan feces) and a specificity of 90.2% (*n* = 116 samples)
	Swan_2R	TCTCCTGTACTCAAGTTTAAC			
	Swan_2P	(FAM)-TACGCATTTCACCGCTACAC-(TAMRA)			
Swan_1	Swan_1F	GCGGTTTAACAAGTCAGGAG	73	na	Sensitivity 50% (*n* = 6 target samples); specificity 40% (*n* = 5 non-target samples)
	Swan_1R	TACTCAAGTTTAACAGTTTCAAAA			
Swan_3	Swan_3F	GGCGGTTTAACAAGTCAGGA	118	na	Sensitivity 83% (*n* = 6 target samples); specificity 0% (*n* = 5 non-target samples)
	Swan_3R	TACGCATTTCACCGCTACAC			
	Swan_3P	(FAM)-ATAGTAAGCTTTTGAAACTGTTAA-(TAMRA)			
Oyscab	OYSCAB_F	AAACTCTACGTGTAGGGTCT	206	na	Sensitivity 71% (*n* = 24 target samples); specificity 91% (*n* = 22 non-target samples
	OYSCAB_R	TCAACCGCACTCAAGTACG			

na, not applicable; F, forward; R, reverse; P, probe.

samples. A particular focus was made on wild waterbirds for which less published data are available. More precisely, we collected 275 wild waterbird droppings from 15 different species and five bird families from coastal areas in France, 67 poultry dropping and litter samples, 70 livestock feces and manure (cattle and pigs), and 13 influents/effluents of wastewater samples from WWTPs. The number of samples per species varied from 1 to 56 (mean = 10; **Supplementary Table 1**).

Raw Data Primary Analysis

Illumina sequencing of the V3–V4 hypervariable region of the 16S rRNA loci resulted in a total of 17,083,374 reads. After quality checking (deleting low-quality sequences and primers, assembling, and removing chimeras) and ASV clustering, 6,698,670 high-quality reads (31.06%) with an average of 15,761 ± 8,115 reads per sample were retained for downstream analyses. The rarefaction curves of observed ASVs showed that sequencing depth was sufficient to encompass bacterial species richness (**Supplementary Figure 1**). A total of 7,349 ASVs across the 425 fecal samples have been obtained, where 7,327 ASVs were assigned at the phylum level, 6,999 at the family level, and 6,255 at the genus level.

The main part of the obtained reads originated from wild waterbird fecal samples (n = 275) collected from coastal areas [57.1% of the total reads; Anatidae (n = 113), 24.4%; Laridae (n = 86), 19.7%; Haematopodidae and Scolopacidae (wader birds; n = 50), 7.8%; Phalacrocoracidae (cormorants; n = 24), 4.5%; and Hydrobatidae (storm petrels; n = 2), 0.6%]. The remaining reads were distributed among poultry [n = 67; 15.6% of the total reads; domestic Anatidae (ducks, geese, and swans), 5.7%; chickens, 5.5%; turkeys, 2.2%; and guinea fowls, 2.1%], livestock animals (n = 70; 22.2% of the total reads; cattle, 8.5%; and pigs, 13.7%), and influents/effluents of wastewater samples (n = 13; 5.1% of the total reads).

Out of the 7,349 ASVs, 69.9% (5,138 ASVs) were specific to one of the five group samples (that is, 32.6% from wild waterbirds, 6.4% from poultry, 10.1% from cattle samples, 13.2% from pig samples, and 7.6% from wastewater samples) (**Figure 1**). A total of 3,892 ASVs were obtained in wild waterbirds whereas a lowest number of ASVs was obtained in both breeding animals (i.e., 2,162 ASVs and 1,784 ASVs in pig and cattle samples, respectively), in poultry (1,621 ASVs), and in wastewater (1,293 ASVs) samples. **Figure 1** shows ASVs shared between two, three, four, or all the five groups. Unique points indicate the signature of a specific ASV for the corresponding group.

Overall Taxonomic Composition

Two indices were calculated in order to investigate the alpha diversity within our dataset: Chao1 richness estimator (qualitative species richness) and Shannon index (non-parametric quantitative species richness/evenness) (**Figure 2**). According to the two indices, a low alpha diversity has been observed in both wild waterbird and poultry samples compared to cattle, pigs, and wastewater samples (Kruskal–Wallis rank sum test with Bonferroni's *post hoc* test; $p < 0.001$; **Figure 2**): e.g., Chao1 from cattle fecal samples ranged from 268 to 437 ASVs (median value of 358.4), and Shannon ranged from 4.9 to 5.5 (median value of 5.3), while in wild waterbird samples, Chao1 ranged from 5 to 414 ASVs (median value of 52) and Shannon ranged from 0.7 to 5.3 (median value of 2.3).

Overall, bacterial communities from all fecal samples analyzed in this study are composed of 31 different phyla (**Supplementary Table 6**), with 2 phyla with less than 10 reads, 11 phyla with 10–100 reads, 4 phyla with 100–1,000 reads, 5 phyla with 1,000–10,000 reads, and 9 phyla with >10,000 reads (i.e., Tenericutes, Cloacimonetes, Spirochaetes, Epsilonbacteraeota, Actinobacteria, Bacteroidetes, Fusobacteria, Proteobacteria, and Firmicutes). The six major phyla were represented by Firmicutes (51.8% of the total reads), Proteobacteria (16.6%), Fusobacteria (11.5%), Bacteroidetes (9.1%), Actinobacteria (8.2%), and Epsilonbacteraeota (1.31%).

At the bacterial phyla level, Firmicutes and Proteobacteria and, in a lesser extent, Actinobacteria and Bacteroidetes were present in all the five groups of samples, while Fusobacteria and Epsilonbacteraeota were mainly present in wild waterbirds (98.0% of the total reads of this first phylum; 59.5% in wild Anatidae and 24.3% in Laridae and 78.2% of the total reads of this second phylum; 63.6% in wild Anatidae; **Supplementary Figure 2**). Nevertheless, Proteobacteria was less represented in wastewater samples, while Bacteroidetes was ubiquitous within non-avian sources: breeding animals [pigs (34.0% of the total reads of this phylum) and cattle (25.1%)] and wild Anatidae (17.0%). Interestingly, Fusobacteria was mostly represented in wild waterbirds.

All other phyla were represented by less than 1% of the total reads. Among these rare phyla, we noticed that the phylum Acidobacteria (3,865 reads) was mainly observed in wild Anatidae (64.1% of the reads of this phylum) and in Laridae (17.5%), and the phylum Cyanobacteria (5,205 reads) was mainly present in Brent geese feces (Anatidae; 83.6%) and, to a lesser extent, in effluents of WWTPs (9.6%). The two phyla Cloacimonetes (16,970 reads) and Fibrobacteres (6,494 reads) were found mainly in pigs (99.7 and 84.8%, respectively), Spirochaetes (22,339 reads) was mainly distributed between pig and cattle samples (77.4 and 20.5%, respectively), and Deferribacteres (n = 7,255) was mainly found in wader birds (67%).

At the genera level, *Catellicoccus* (Firmicutes; 19.2%) and *Cetobacterium* (Fusobacteria; 17.5%) were the most prevalent genera within wild waterbirds (**Figure 3**). In poultry, the two dominant genera were *Lactobacillus* (Firmicutes; 26.4%) and *Romboutsia* (Firmicutes; 12.7%; **Supplementary Figure 4**). In cattle, *Ruminococcaceae* UCG-005 and *Ruminococcaceae* UCG-010 (Firmicutes) were the dominant genera in 61.5 and 38.5% of cattle feces, respectively, whereas *Acinetobacter* was the dominant genus in 71.4% of cattle manures (**Supplementary Figure 4B**). The main genera for pig feces were *Lactobacillus* (12.6%) and *Clostridium sensu stricto* 1 (Firmicutes; 11.1%), and for pig slurry and solid manure, *Clostridium sensu stricto 1* (25.7 and 7.7%, respectively). Finally, the main genera for influents were C39 (Proteobacteria; 6.4%) and *Acinetobacter* (Proteobacteria; 6.2%), and for effluents, *Arcobacter* (Proteobacteria; 8.7%) and *Mycobacterium* (Actinobacteria; 6.6%; **Supplementary Figure 4**).

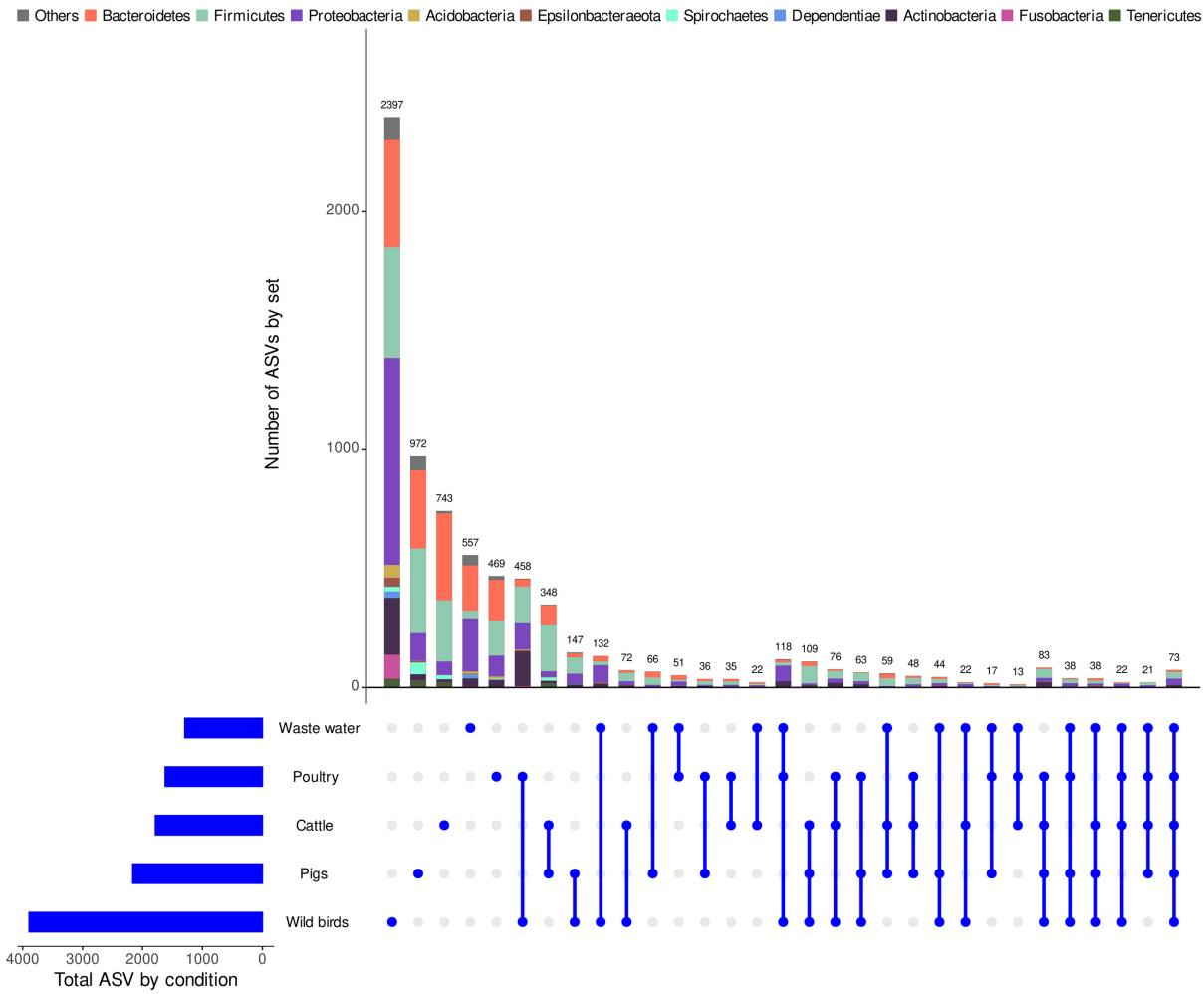


FIGURE 1 | UpSetR visualization of interactions between the obtained ASVs within the whole dataset. The grid along the bottom is used to identify interaction sets (analogous to a Venn diagram). Heavily colored and connected blue dots in the grid indicates that the key group shown on the left has contributed to the interaction set shown on the top. The number of ASVs per group and the size and taxonomy (at the phylum level) of each interaction set are represented by horizontal bars on the left and vertical bars on the figure above, respectively.

Comparative Analyses

Fecal Microbiome Composition of the Wild Waterbirds

Alpha diversity

Significant differences were observed where the highest bacterial community richness (Chao1) among the wild waterbird fecal samples was observed in gulls (*Chroicocephalus* spp.; Laridae; $n = 3$; median value of 199 ASVs), storm petrels ($n = 2$; mean value of 179 ASVs), and Brent geese (Anatidae; $n = 56$; median value of 149) (**Figure 2A**). The Chao1 index in Brent geese was significantly higher than the ones observed from wader birds such as curlews ($n = 11$) and oystercatchers ($n = 23$) (Scolopacidae and Haematopodidae, respectively), and from common shelducks (Anatidae) (e.g., median value of 24 ASVs in samples from curlews), which both present the lowest bacterial community richness (Kruskal–Wallis rank sum test with Bonferroni's *post hoc* test; $p < 0.001$). Furthermore, the largest range in values was observed for Brent geese (23–414 ASVs), followed by seagulls (*Larus* spp.; 7–318 ASVs).

The bacterial diversity at the ASV level with Shannon index (**Figure 2B**), which considers taxon diversity and abundance, also showed significant higher values for samples from Brent geese (median of 4.0) than for curlews (1.8) and common shelducks (1.9; Kruskal–Wallis rank sum test with Bonferroni's *post hoc* test; $p < 0.01$ and $p < 0.001$, respectively). A high range in Shannon index values (from 0.7, which corresponds to a gull fecal sample almost exclusively composed of the genus *Catellicoccus*, FO424, to 5.0) was observed in seagulls and black-headed gulls.

Taxon abundance

The six dominant phyla in wild waterbird fecal samples ($n = 275$) were Firmicutes (45.3% of the total reads of these birds), Proteobacteria (20.1%), Fusobacteria (19.7%), Actinobacteria (7.3%), Bacteroidetes (4.8%), and Epsilonbacteraeota (2.2%; **Figure 4** and **Supplementary Figure 2**). Indeed, Firmicutes

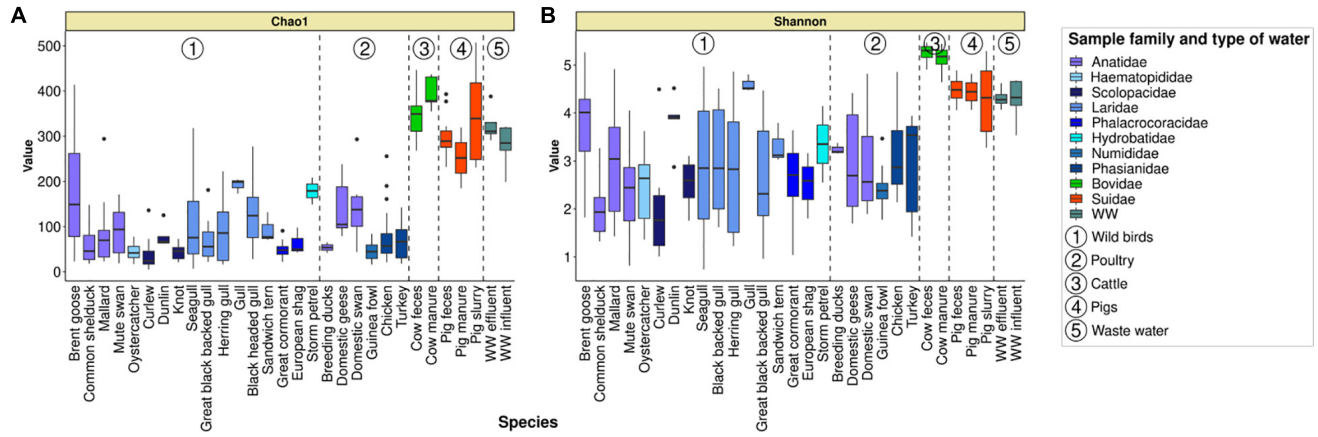


FIGURE 2 | Boxplots illustrating the alpha diversities with variations intra- and inter-groups for Chao1 richness estimator **(A)** and Shannon diversity index **(B)** computed from the ASV contingency table. WW, wastewater.

was the most abundant phylum in Anatidae (32.0%), Laridae (47.5%), shorebirds (i.e., Haematopodidae and Scolopacidae; 67.7%), and cormorants (Phalacrocoracidae; 69.3%). The second dominant phylum was Fusobacteria in Anatidae and shorebirds (28.0 and 13.8%, respectively), while Proteobacteria predominated in Laridae and cormorants (22.2 and 15.5%, respectively). The third phylum was Fusobacteria in Laridae and cormorants (14.1 and 11.8%, respectively). The phylum Actinobacteria was present in the two storm petrel samples (Hydrobatidae) and Laridae (39.4 and 13.5%, respectively).

Within wild Anatidae, differences were observed according to the bird species with the main presence of Proteobacteria in the Brent geese (53.6% of the samples), while Fusobacteria predominated in common shelducks (88%), mallards (45%), and mute swans (41.7%). Bacteroidetes was the dominant phylum in 30% of mallard samples and Firmicutes in 32.1% of Brent geese, 20% of mallard, and 41.7% of mute swan samples (**Supplementary Figure 2A**).

Within Laridae, the dominant phylum was Firmicutes in 68.9% of seagulls (i.e., *Larus* spp.), 71.4% of great black-backed gulls, and 86.7% of herring gulls (**Supplementary Figure 2B**). Actinobacteria predominated in all of the gull's samples (i.e., *Chroicocephalus* spp.) and 30.8% of black-headed gulls (*Chroicocephalus ridibundus*). Finally, Proteobacteria phylum was predominant in 11.1% of seagulls, 38.5% of black-headed gulls, and 66.7% of sandwich terns.

For wader birds, the dominant phylum was Firmicutes for 78.3% of oystercatchers, all curlews, 20% of dunlins samples, and 90.9% of knots (**Supplementary Figure 2C**). Bacteroidetes was dominant in 13% of oystercatcher samples while Actinobacteria was dominant in 60% of dunlins.

Concerning cormorants, the dominant phylum was Firmicutes for 90.5% of the analyzed samples and for 33.3% of European shag fecal samples (**Supplementary Figure 2D**). The two storm petrel samples were slightly different with the predominance of Firmicutes or Actinobacteria.

At the genus level, *Catellicoccus* (Enterococcaceae; 19.2%) and *Cetobacterium* (Fusobacteriaceae; 17.5%) were the most prevalent genera within wild waterbirds (**Figure 3**), followed by *Clostridium sensu stricto 1* (Clostridiaceae 1; 4.2%), *Lactobacillus* (Lactobacillaceae; 3.9%), *Bacteroides* (Bacteroidaceae; 2.7%), *Fusobacterium* (Fusobacteriaceae; 2.2%), *Romboutsia* (Peptostreptococcaceae; 1.6%), *Mycobacterium* (Mycobacteriaceae; 1.5%), *Psychrobacter* (Moraxellaceae; 1.4%), and uncultured *Ruminococcus* spp., GCA-900066225 (Ruminococcaceae; 1.3%).

In wild Anatidae (**Figure 3A**), *Cetobacterium* was the dominant genus in 8.9% of Brent geese, 92% of common shelducks, 45% of mallards, and 41.7% of mute swans. *Catellicoccus* was dominant in 7.1% of Brent geese, 8% of common shelducks, and 5% of mallard fecal samples. *Bacteroides* was the dominant genus in 40% of mallards and 3.6% of Brent geese, and *Romboutsia* was dominant in 8.3% of mute swans and 5.3% of Brent geese. There is a variable distribution from one sample to another for Brent geese with, e.g., *Lactobacillus* as the dominant genus in 19.6%, *Cetobacterium* in 8.9%, *Catellicoccus* in 7.1%, and *Campylobacter* in 3.6% of the fecal samples, while a quite homogeneous distribution was obtained in the common shelducks with *Cetobacterium* as the most dominant genus followed by *Catellicoccus*.

In Laridae, *Catellicoccus* was the dominant genus in 40% of seagulls (i.e., *Larus* spp.), 28.6% of great black-backed gulls, 40% of herring gulls, and 25% of black-headed gulls (**Figure 3B**). *Cetobacterium* was the dominant genus in 11.1% of seagulls (i.e., *Larus* spp.), 14.3% of great black-backed gull, and 38.5% of black-headed gulls, while *Clostridium sensu stricto 1* predominated in 11.1% of seagulls and 13.3% of herring gulls. *Mycobacterium* was dominant in 23.1% of black-headed gulls.

Like Laridae, the wader birds group presented the highest values for *Catellicoccus* in 34.8% of oystercatchers, 72.7% of curlews, 20% of dunlins, and 81.8% of knots (**Figure 3C**). *Cetobacterium* was the second dominant genus in only 0.9% of

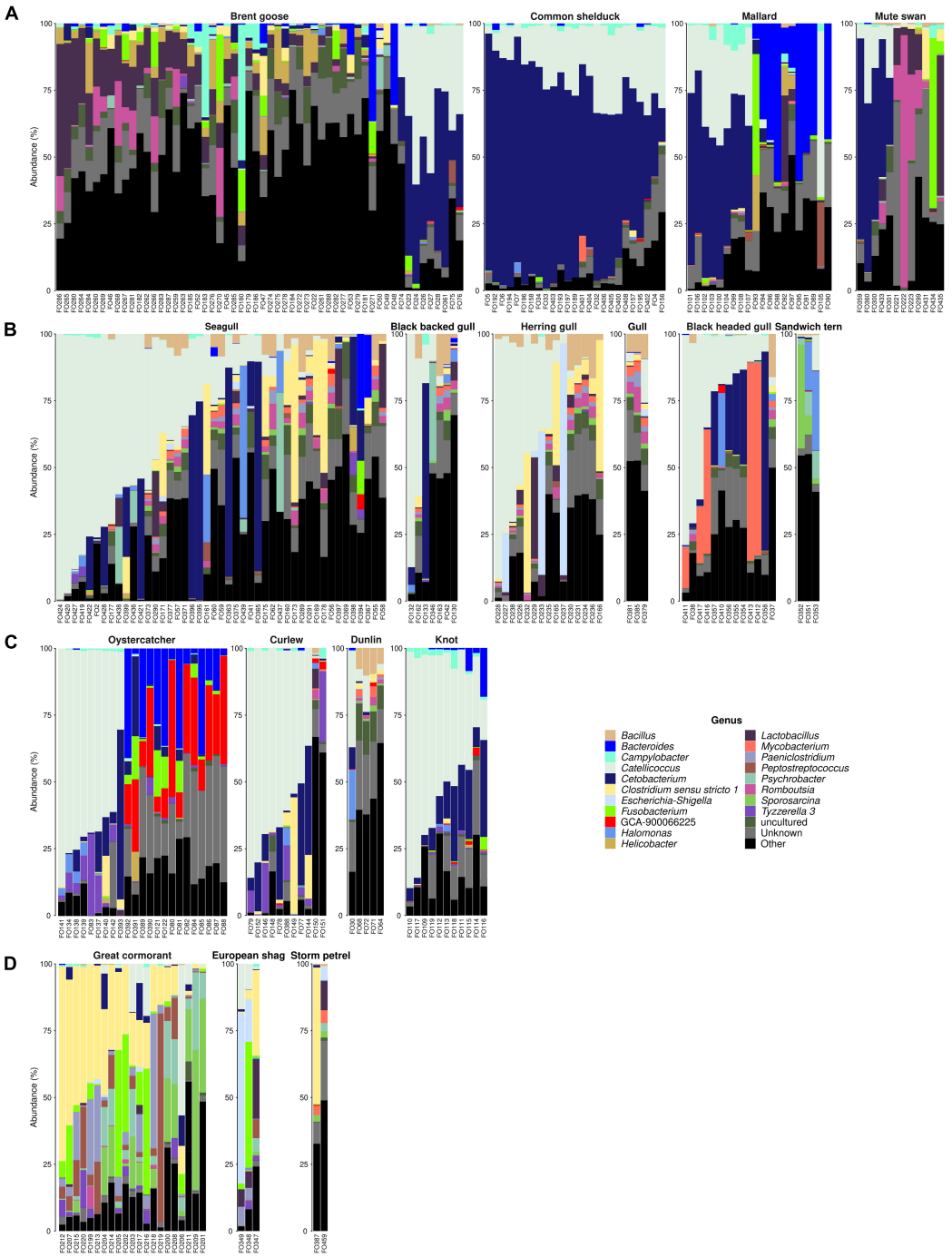


FIGURE 3 | Distribution of predominant bacterial genera in wild waterbird samples according to relative abundance obtained by the gene encoding 16S rRNA. Bacterial community compositions were grouped by wild waterbird families: **(A)** wild Anatidae, **(B)** Laridae, **(C)** Haematopodidae and Scolopacidae (wader birds), and **(D)** Phalacrocoracidae and Hydrobatidae (cormorants and storm petrels, respectively). Stacked bar plots represent the sequence abundances of the 17 most abundant genus-level taxa identified in the fecal samples. Percent sequence abundances given as the number of reads matching a given bacterial family per total reads for that sample.

oystercatchers, 9.1% of curlew, and 20% of knots; *Bacteroides* and GCA-900066225 (uncultured *Ruminococcus* spp. from the Ruminococcaceae family) were dominant in 26.1 and 34.8% of oystercatchers, respectively.

In cormorants, *Clostridium sensu stricto 1* was the dominant genus in 42.8% of great cormorants and 33.3% of European shags, followed by *Sporosarcina* (Planococcaceae), *Fusobacterium*, *Catellicoccus*, and *Paeniclostridium* (Clostridiaceae) that were

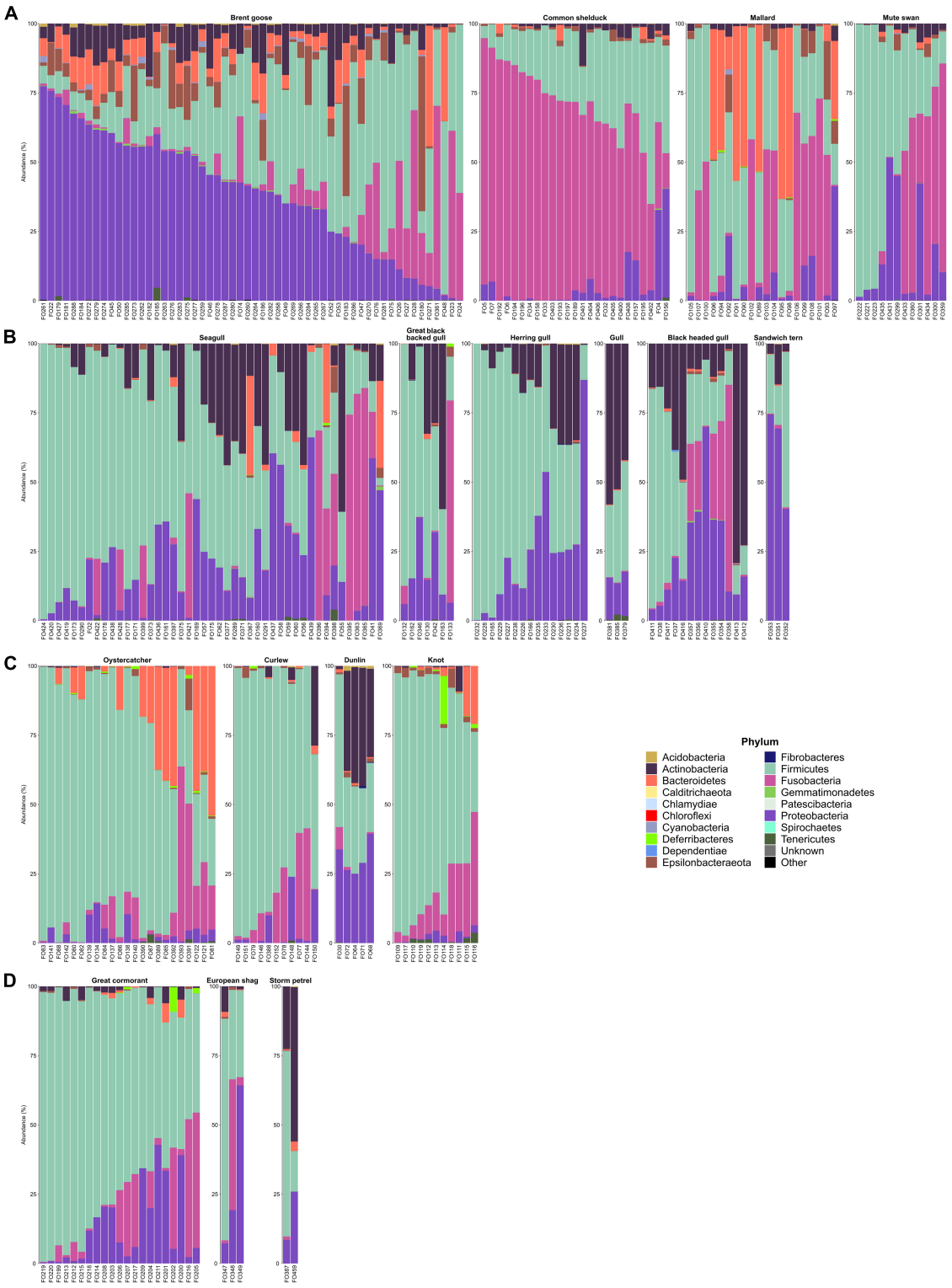


FIGURE 4 | Distribution of the 18 most predominant bacterial phyla in wild waterbird samples according to relative abundances obtained by the gene encoding 16S rRNA. Bacterial community compositions were grouped by wild waterbird families: **(A)** wild Anatidae, **(B)** Laridae, **(C)** Haematopodidae and Scolopacidae (wader birds), and **(D)** Phalacrocoracidae and Hydrobatidae (cormorants and storm petrels, respectively). Stacked bar plots represent the sequence abundances of the 18 most abundant phylum-level taxa identified in the fecal samples. Percent sequence abundances given as the number of reads matching a given bacterial family per total reads for that sample.

dominant in 19, 14.3, 4.8, and 4.8% of the fecal samples of great cormorants, respectively (**Figure 3D**).

Fecal Microbiome Composition in Poultry

Alpha diversity

The highest bacterial community Chao1 richness among the analyzed fecal samples from poultry was observed within domestic swans (median value of 137.5 ASVs) and domestic geese (105 ASVs), while the lowest bacterial community richness was observed in guinea fowls (45 ASVs) (**Figure 2A**).

The Shannon index showed the highest values for samples from turkeys (median value of 3.5) and the lowest values for guinea fowls (median value of 2.4) (**Figure 2B**).

Taxon abundance

Poultry samples included two types of samples: dropping samples of hens, turkeys, guinea fowls, and domestic Anatidae, and litter samples (from hens and turkeys). The three dominant phyla in all the samples were Firmicutes (67.9%), Actinobacteria (16.0%), and Proteobacteria (12.8%; **Supplementary Figures 2, 3**). Among domestic Anatidae, Firmicutes was the dominant phylum for all the three breeding duck samples, for 66.7% of domestic swans, and 66.6% of domestic geese, while Actinobacteria was the dominant phylum for 36.4% of geese and 16.7% of swans. In addition, Firmicutes was the dominant phylum in all guinea fowls, in 73.1% of chicken samples and all the turkeys. Actinobacteria was dominant in 22.2% of the chicken litter samples and Proteobacteria in 29.4% of the chicken dropping samples. Furthermore, Actinobacteria was in higher proportion in poultry litter samples than in droppings (mean percentage value of 4.4% in chicken droppings vs. 30.9% in chicken litter).

At the genus level, the two dominant genera were *Lactobacillus* (26.4%) and *Romboutsia* (12.7%; **Supplementary Figure 4**). In domestic Anatidae, *Romboutsia* was the dominant genus in 36.4% of geese and 50% of swans, *Turicibacter* (Erisipelotrichaceae) in 27.3% of geese and 33.3% of swans, and *Jeotgalibaca* (Carnobacteriaceae) in 66.7% of breeding ducks. *Lactobacillus* was the dominant genus in all of the 11 guinea fowls, in all of the three turkey feces, in 35.3% of the chicken dropping, and 85.7% of the turkey litter and 11.1% of the chicken litter samples.

A differential bacterial composition between dropping and litter samples was observed. In poultry droppings, the dominant genera were *Lactobacillus* (27.4 and 91.7% in chicken and turkey, respectively), followed by *Pseudomonas* (Pseudomonadaceae), *Acinetobacter* (Moraxellaceae), and *Romboutsia* (at 13.2, 12.6, and 9.4% in chicken, respectively). In chicken and turkey litters, they were represented by *Lactobacillus* (12.8 and 21.2%), *Staphylococcus* (18.2 and 10.0%), *Weissella* (Leuconostocaceae; 14.1 and 6.3%), *Corynebacterium 1* (10.7 and 7.4%), *Brachybacterium* (Dermabacteriaceae; 5.1 and 4.9%), and *Brevibacterium* (Brevibacteriaceae; 7.3 and 2.5%).

Fecal Microbiome Composition Within Livestock Samples (Cattle and Pigs)

Alpha diversity

As indicated above, the highest alpha diversity according to the two tested indices was observed in cattle fecal samples (see the overall bacterial composition section). These results were observed for both cow feces and cow manure samples (**Figure 2**).

In pig fecal samples, Chao1 richness showed significant high values than that of cattle samples (Kruskal–Wallis rank sum test with Bonferroni's *post hoc* test; $p < 0.001$), which was the opposite for Shannon's index. The greatest bacterial community richness was observed in pig slurry samples (median value of 339 ASVs vs. 289 ASVs in pig feces).

Taxon abundance

In breeding animal samples [cattle ($n = 33$) and pigs ($n = 37$)], Firmicutes (62.0 and 65.4%, respectively) and Bacteroidetes (26.9 and 22.5%) were the dominant phyla (**Supplementary Figures 2, 3B,C**). Among the cattle samples, Firmicutes was the dominant phylum in all the feces and 71.4% of manure samples, while Proteobacteria was dominant within the two other manure samples. Among pig samples, Firmicutes was the dominant phylum for all the feces, all the slurry, and in one pig manure sample, while Bacteroidetes was dominant in the other pig manure samples.

At the genus level, *Ruminococcaceae* UCG-005 and *Ruminococcaceae* UCG-010 were the dominant genera in 61.5% and 38.5% of cattle feces, respectively, whereas *Acinetobacter* was the dominant genus in 71.4% of cattle manures (**Supplementary Figure 4B**). In pigs, *Clostridium sensu stricto 1* was the main genus in 18.2% of feces, one (50%) manure, and 87.5% of the slurry samples, while *Lactobacillus* was the main dominant genus in 54.5% of feces. DMER64 genus from the Rikenellaceae family was dominant in 12.5% of the slurry samples. *Prevotella 9* and *Escherichia–Shigella* were dominant in two (18.2%) and one (0.9%) feces, respectively, whereas *Acinetobacter* was dominant in the second (50%) pig manure sample (**Supplementary Figure 4C**).

Fecal Microbiome Composition of the Influents and Effluents of WWTPs

Alpha diversity

In wastewater samples, Chao1 richness and Shannon index presented significantly higher values than in wild waterbird samples (Kruskal–Wallis rank sum test with Bonferroni's *post hoc* test; $p < 0.001$; **Figure 2**). In fact, median values of 296 ASVs for Chao1 estimator and 4.3 for Shannon index were obtained.

Taxon abundance

Proteobacteria (38.8%), Firmicutes (22.1%), Actinobacteria (19.2%), and Bacteroidetes (12.4%) were the dominant phyla in wastewater samples ($n = 13$; **Supplementary Figures 2, 3D**). Proteobacteria was the dominant phylum in all the four effluents and in 55.6% of influents, whereas Firmicutes was dominant in 33.3% of influents and Actinobacteria in one (11.1%) effluent sample.

At the genus level, *Acinetobacter* was the dominant genus in one effluent and three influent samples. *Hypnocyclicus* (Fusobacteriaceae), *Methylotenera* (Methylophilaceae), and *Arcobacter* (Campylobacteriaceae) were the dominant genera in one effluent sample, while C39 (Rhodocyclaceae), *Trichococcus* (Carnobacteriaceae), and *Mycobacterium* (Mycobacteriaceae)

were dominant in three, two, and one influent sample, respectively (**Supplementary Figure 4D**).

Beta Diversity

Within the Overall Dataset

Non-metric multidimensional scaling analysis of the whole dataset using Bray–Curtis distance metric showed that individuals of the same group (i.e., wild waterbirds, poultry, cattle, pigs, or wastewaters) clustered together, with however a greater diversity within wild waterbird samples (stress value 0.177; Adonis $R^2 = 0.14$; $p = 10^{-4}$; **Figure 5A**). The structure of the bacterial communities was also distinct both according to bird and/or animal species and according to bird and/or animal families (Adonis $R^2 = 0.13$; $p = 10^{-4}$; Adonis $R^2 = 0.11$; $p = 10^{-4}$, respectively). This suggests that each group harbored a distinct bacterial community profile completed by the shared part with the other groups (**Figure 5A**).

The ANCOM analysis applied to these five groups showed that the wild waterbird group differs mainly from the other groups by the significant presence of sequences from Fusobacteriaceae ($W = 566$) and Enteroccocaceae ($W = 565$) families, more precisely from the *Cetobacterium* ($W = 1427$) and *Catellicoccus* ($W = 1427$) genera, respectively. We noted the low abundance of the Rikenellaceae ($W = 564$) and Christensenellaceae ($W = 563$) families that are mainly detected in pigs and cattle. Poultry can be outlined from the other groups by sequences from Lactobacillaceae ($W = 566$) (mainly from *Lactobacillus*), Dermatobacteriaceae ($W = 564$) (mainly from *Bradybacterium*), and Staphylococcaceae ($W = 564$) (mainly from *Staphylococcus*). Cattle were found characterized by the very low abundance of sequences from Fusobacteriaceae ($W = 564$), Enteroccocaceae ($W = 563$), and Lactobacillaceae ($W = 561$), while Rikenellaceae ($W = 565$) and Christensenellaceae ($W = 563$) were highly present in pigs, with low abundance of sequences from Fusobacteriaceae ($W = 566$). Wastewater samples can be distinguished from the others by the abundance of sequences from Rhodocyclaceae ($W = 566$) and Arcobactericeae ($W = 562$).

Within the Wild Waterbird's Microbiota

Within this group, the samples clustered according to their families (**Figure 5B** and **Supplementary Figure 5**). In the same way, the structure of bacterial communities was distinct both according to bird and/or animal species and according to the bird and/or animal families using either Bray–Curtis (stress value 0.179; Adonis $R^2 = 0.15$; $p = 10^{-4}$; Adonis $R^2 = 0.14$; $p = 10^{-4}$, respectively) or weighted-Unifrac distance metrics (stress value 0.126; Adonis $R^2 = 0.10$; $p = 0.0035$; Adonis $R^2 = 0.11$; $p = 0.0005$, respectively). For Bray–Curtis distance metric, a high diversity was obtained inside each bird family with, however, a weaker diversity for samples from *Phalacrocoracidae*. For the weighted-Unifrac distance metric, most of the samples clustered together, with a greater diversity for Phalacrocoracidae and Anatidae. Indeed, few samples from geese feces differ from overall other avian samples (**Figure 5B** and **Supplementary Figure 5**).

Identification of Potential New MST Markers From the NGS Data

Following the analysis of the bacterial communities within our dataset, an investigation of the unique and/or host-associated ASVs in a group of samples belonging to the same group or to a bird species was carried out to identify host-associated sequences and thus potentially develop new MST qPCR-based markers.

Among these host-associated ASVs, 17 candidates were retained with four that were further investigated in this study: one ASV belonging to the genus *Romboutsia* and found strongly associated with swans, one belonging to the genus *Bacteroides* and associated with oystercatchers (**Supplementary Table 5**), one belonging to the species *P. sordellii* and associated with cormorants, and one belonging to the family Peptostreptococcaceae and associated with Anatidae. Two of these proposed candidates did not allow primers drawing due to the absence of regions specific to the target hosts (Anatidae and cormorants). For the ASV sequence belonging to the genus *Romboutsia* (swans), three pairs of primers with the corresponding probes have been designed (**Table 1**). In addition, a pair of primers has been designed for the ASV sequence belonging to the genus *Bacteroides* (oystercatchers).

The preliminary sensibility and specificity tests carried out for these four pairs of primers on target and non-target fecal samples led us to retain only two pairs, the one targeting swans and named Swan_2, and the one targeting oystercatchers and named Oyscab (**Table 1**). These markers presented a sensitivity of 75 and 71% and a specificity of 90.2 and 91%, respectively.

Presence of Potential Pathogenic Bacterial Groups Within the Whole Dataset

We further investigated the distribution of a selection of 37 bacterial genera harboring enteric or environmental human pathogens (**Figure 6**, **Supplementary Figures 6, 7,** and **Supplementary Table 4**). In wild waterbirds ($n = 275$), the members of the genera *Clostridium sensu stricto 1*, *Campylobacter*, *Fusobacterium*, and *Helicobacter* were dominant and present in 67.3, 55.6, 48.7, and 33.8% of the dropping samples, respectively (**Supplementary Figure 6B**). In poultry, five pathogenic genera were found to be dominant: *Streptococcus*, *Enterococcus*, *Staphylococcus*, *Psychrobacter*, and *Corynebacterium* (79.1, 68.7, 53.7, 50.1, and 49.2%, respectively).

In cattle samples, members of the genus *Bacteroides* were ubiquitous, *Clostridium sensu stricto 1*, *Prevotella*, and *Treponema* were highly prevalent (97%), and *Escherichia–Shigella* in a lesser extent (42.4%). In pigs, *Clostridium sensu stricto 1* and *Treponema* were ubiquitous; *Streptococcus* was present in 81.1% and *Bacteroides* and *Corynebacterium* were present in 75.7% of samples (**Supplementary Figure 6B**). In wastewater samples, *Bacteroides*, *Arcobacter*, *Flavobacterium*, and *Aeromonas* were ubiquitous, and *Clostridium sensu stricto 1*, *Escherichia-Shigella*, and *Mycobacterium* were present in 92.3% of the 13 samples (**Supplementary Figure 6B**).

Interestingly, the dominant bacterial genera that include human pathogens were different in wild and domestic

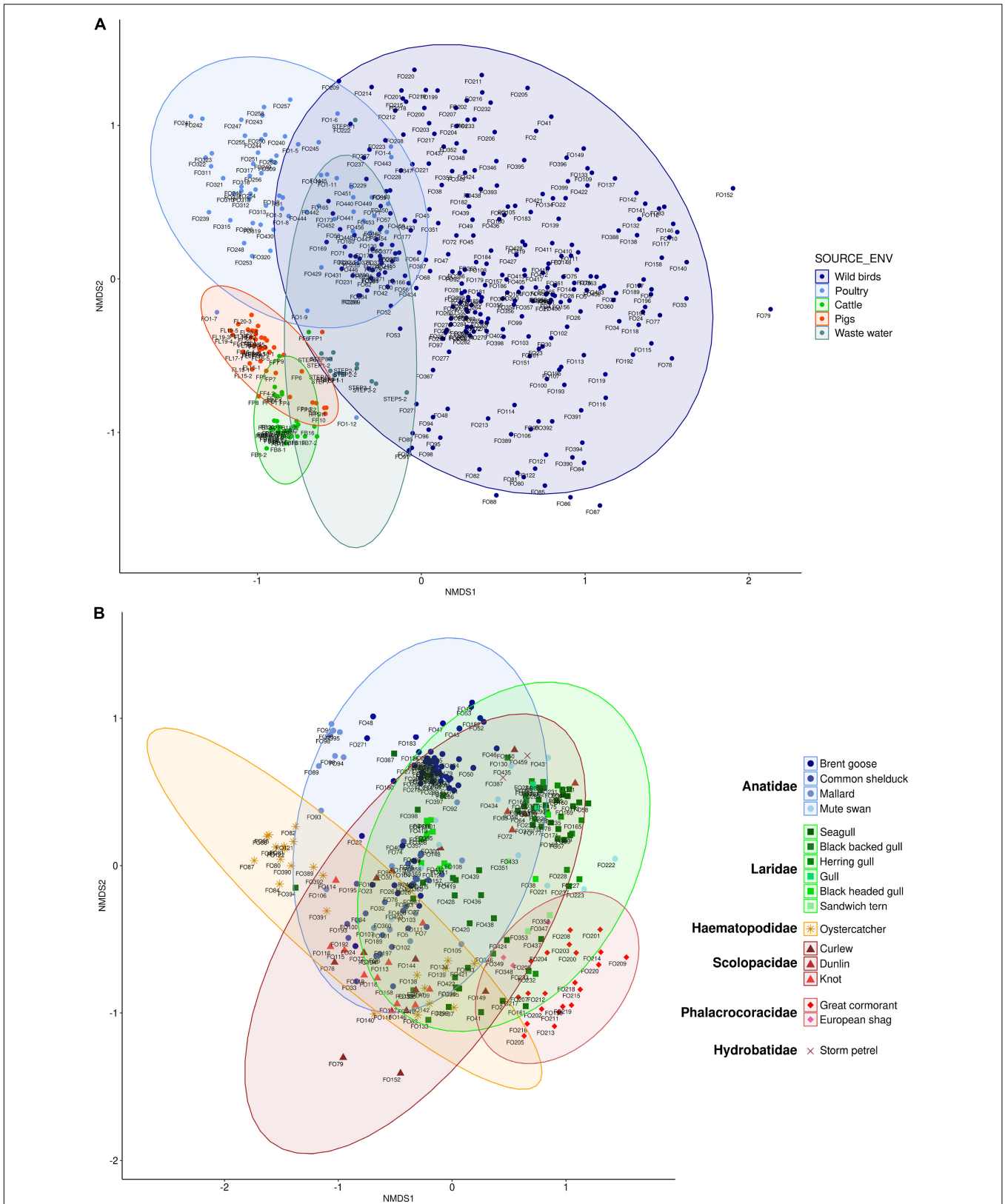


FIGURE 5 | Non-metric multidimensional scaling (NMDS) plots based on Bray–Curtis (**A**, stress = 0.177; **B**, stress = 0.179) distance metrics in relation to the whole dataset **(A)** or the wild waterbird groups **(B)**. Colors represent host classes. Ellipses represent 95% confidence intervals of centroids of each point.

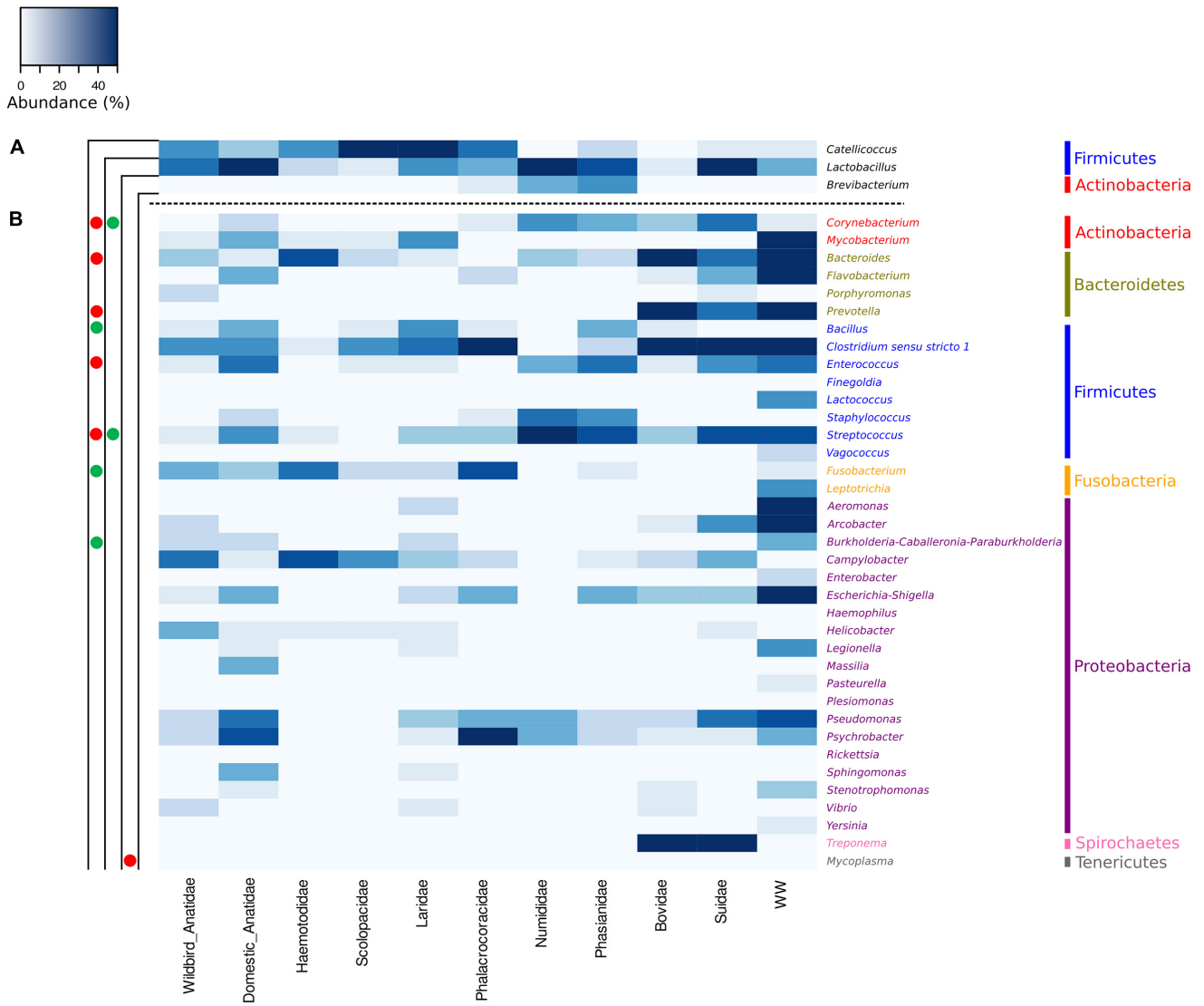


FIGURE 6 | Heatmaps based on the number of reads of a selection of **(A)** 3 genera harboring known MST markers and **(B)** 37 bacterial genera harboring potential pathogens, and derived from the whole dataset at the bird or livestock animal family levels including wastewaters (WW).

Anatidae. In fact, *Campylobacter*, *Clostridium sensu stricto 1*, *Fusobacterium*, and *Helicobacter* were present in 70.8, 64.6, 54.9, and 54.9% in wild Anatidae, respectively, whereas *Psychrobacter*, *Pseudomonas*, and *Enterococcus* were present in 85, 75, and 70% in domestic Anatidae, respectively. In Laridae, the four dominant genera were *Clostridium sensu stricto 1*, *Bacillus*, *Mycobacterium*, and *Pseudomonas* (75.6, 58.1, 55.8, and 39.5%, respectively). In Phalacrocoracidae, *Clostridium sensu stricto 1* (100%), *Psychrobacter* (91.7%), *Fusobacterium* (83.3%), and *Escherichia–Shigella* and *Pseudomonas* (50%) were dominant. For wader birds and within the Haematopodidae and Scolopacidae families, *Campylobacter* was the dominant bacterial genus that includes human enteric pathogens (78.3 and 63%, respectively), whereas *Bacteroides* (78.3%) and *Fusobacterium* (73.9%) were the most prevalent genera in Haematopodidae and *Clostridium sensu stricto 1* (63%) in Scolopacidae (**Figure 6B**).

The relationships between the number of reads of ASVs harboring pathogens and ASVs harboring MST markers [i.e., *Catellicoccus* (gull marker), *Lactobacillus* (general avian marker or pig marker, *Lactobacillus amylovorus*), and *Brevibacterium* (poultry litter marker)] were examined by Spearman's correlation tests.

At the detailed species level, *Brevibacterium* ($p < 0.05$) and *Lactobacillus* ($p < 0.05$) were positively correlated with *Corynebacterium*, while a strong negative correlation for this latest genus was observed with *Catellicoccus* ($p < 0.001$). Only *Catellicoccus* presented negative correlation with three of the four tested genera within Bacteroidetes ($p < 0.001$) (**Supplementary Figure 7**).

While no correlation was observed for *Lactobacillus* and the tested bacterial genera harboring pathogens, *Brevibacterium* had negative correlations with *Vagococcus* ($p < 0.05$), *Leptotrichia* ($p < 0.05$), and *Pasteurella* ($p < 0.05$) at the five group levels (**Supplementary Figure 6**). *Catellicoccus* presented negative correlation with *Prevotella* ($p < 0.05$), but positive correlations with *Haemophilus* ($p < 0.05$) and *Mycoplasma* ($p < 0.05$), and strong correlation with *Fusobacterium* ($p < 0.001$).

At the bird or livestock animal family levels, *Lactobacillus* harboring general avian and pig MST markers presented positive correlations with *Corynebacterium* ($p < 0.05$) and *Streptococcus* ($p < 0.05$), while the gull-associated genus *Catellicoccus* had negative correlations with those two genera. In addition, *Catellicoccus* had negative correlations with *Bacteroides* ($p < 0.05$), *Prevotella* ($p < 0.05$), and *Enterococcus* ($p < 0.05$) and positive correlations with *Bacillus* ($p < 0.01$), *Fusobacterium* ($p < 0.01$), and the *Burkholderia–Caballeronia–Paraburkholderia* group ($p < 0.05$) (**Figure 6**).

DISCUSSION

Although the 16S rRNA gene amplicon method is a gold standard being used in most microbiome-based studies, only recent works focused on its use in MST (Microbial Source Tracking) development, while its accuracy is questionable in detecting pathogens in routine tests. In this study, we generated an original dataset of highly diverse and predominant wild waterbirds (275/342; 80% of bird fecal samples) with inter-individual replicates for all wild waterbird species investigated (2–56), with different diets (i.e., herbivore, omnivore, and carnivore; **Supplementary Table 1**) and from different geographical locations in France (at least six different departments). This dataset also considers the other potential sources of fecal bacteria to temperate coastal areas (oceanic climate; Brittany, France) from individual feces of poultry, pigs, and cattle to composite samples under the influence of farm environment (i.e., poultry litter and pig and cattle manure samples) and influents/effluents from WWTPs. In addition, for the Anatidae family, we collected samples from both wild and captive birds.

Wild waterbirds are fascinating animals that present special behavior, dietary patterns, and flight capacities that could influence the diversity of their gut microbiota. This diversity was suggested to be higher than that of mammals, since birds are more diverse in terms of species, with a greater dependence on microbes for digestion and their adaptation to various terrestrial and aquatic environments and to long-distance flights (Hird et al., 2015; Grond et al., 2018). They are an integral part of the aquatic ecosystems, and their fecal pollution should not be neglected since they harbor and spread relevant pathogens and antibiotic resistance determinants (Ewbank et al., 2021). One of the first 16S rRNA gene amplicon studies that attempt to investigate the GI microbiome of wild birds was done by Wienemann et al. (2011), which analyzed seasonal changes in the gut microbial community of wild and captive capercaillie (*Tetrao urogallus*) (Wienemann et al., 2011). A few studies on wild birds followed this publication, where mainly Proteobacteria, Firmicutes, and Actinobacteria were found as the most abundant bacterial phyla (Kreisinger et al., 2015; Laviad-Shitrit et al., 2019; Fu et al., 2020). From an MST view, few studies developed markers for wild waterbird species including ducks (Devane et al., 2007), geese (Hamilton et al., 2006), gulls (Lu et al., 2008), and cranes (Ryu et al., 2012).

Our dataset (>35 different bird or animal species) shows few taxonomic groups to be present in all samples (i.e., Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria), in agreement with the dataset on vertebrate microbiome obtained by Youngblut et al. (2019). In addition, this dataset makes available data on avian species that had not been considered in their study (i.e., common shelducks, storm petrels, and European shags). However, the main feature of this dataset is the presence of taxonomic groups that are different between birds, other animals, and wastewaters, and also between wild and captive birds. This differential detection is very useful for developing MST markers targeting specific DNA sequences of a group, family, or species for the purpose of identifying sources of fecal contamination in the environment. It is also noteworthy that in most cases we found high intra-species diversity, suggesting the need to sample enough individuals within the same species. This is particularly relevant for Brent goose and seagulls, whereas individual differences were much lower for common shelducks. Interestingly, our data suggest a lower diversity of the fecal microbiota of wild waterbird compared to that of farm animals and humans (reflected by WW analysis). Dietary and environmental conditions are different between captive and wild animals and might impact this microbiome diversity (Hird, 2017). In addition, species physiology may play a major role with birds presenting a smaller (in terms of surface) and less evolved GI tractus compared to that of farm animals and humans (Reese and Dunn, 2018). Furthermore, diversity was generally higher in herbivores [such as ruminants (i.e., cattle in our study)] than omnivores or carnivores (Reese and Dunn, 2018). Additional data from other sites and individuals may improve our understanding of major drivers of fecal microbiome diversity.

For example, the ANCOM identified a specific ASV belonging to the Fusobacteria phylum and more precisely to the genus *Cetobacterium*, to be associated with wild birds, which may be particularly relevant to the development of a new general avian marker (**Supplementary Table 5**). Consistent with this observation, Youngblut et al. (2019) also found that OTUs of the genus *Cetobacterium* were associated with animals other than mammals, such as birds. A perspective to this study will be the identification of a qPCR marker from this ASV to complement the general avian MST toolbox.

Indeed, general avian markers have already been developed, and if they are specific (>94%), they often lack significant sensitivity [30–58% for the GFD marker targeting *Helicobacter* in United States and Australia (Green et al., 2012; Ahmed et al., 2016) and 46.5% for the AV4143 marker targeting the *Lactobacillus* genus in France ($n = 144$; wild bird and poultry samples; Brittany; data not shown)]. However, this dataset confirms that the targeted sequences are present in several avian species. *Helicobacter* spp., the genus targeted by the GFD marker, was mostly found in wild birds (> 20% of the samples

from Anatidae, Laridae, and Haematopodidae). The *Lactobacillus* genus targeted by the AV4143 marker was detected in both wild (mainly Brent goose, mallards, and mute swans) and domestic (geese and swans) Anatidae, in poultry (mainly chickens, turkeys, and guinea fowls) and in seagulls, curlews, and cormorants, confirming its status as a general avian source marker.

Fecal Microbiomes of Wild Waterbirds and MST Markers

In this study, we described the bacterial communities' structure of a large number of fecal samples ($n = 425$), most of them ($n = 275$) representing the main wild waterbird families (i.e., Anatidae, Laridae, Scolopacidae, Haematopodidae, Phalacrocoracidae, and Hydrobatidae) that are present on coastal areas in Brittany (France) and likely to be a source of pathogenic microbes to bathing waters or shellfish harvesting areas. We compared them to fecal microbiomes from a selection of poultry, livestock animals (cattle and pigs), and influents/effluents of WWTPs ($n = 150$) with the secondary aim of selecting potential MST qPCR markers.

Our bioinformatic analysis used the Nextflow-based SAMBA pipeline[7], which is based on the approach of ASVs rather than OTUs. The ASV concept is based on sequence variant that provides finer resolution consistent with biological significance (Callahan et al., 2016). In fact, to avoid the ecological limitations of the 97% threshold-based OTU method, we applied ASVs as recommended by Callahan et al. (2016), which is a threshold-free metric of classification. In addition, this workflow performs a distribution-based OTU calling after DADA2, which reduces the bias in identifying false-positive ASVs and thus enables the real diversity of the samples to be described as accurately as possible.

Here, sequencing data were analyzed at the bacterial phylum (**Figure 4** and **Supplementary Figures 2, 3**), family (**Supplementary Figures 8, 9**), and genus (**Figure 3** and **Supplementary Figure 4**) classification levels. From our results, we stated that fecal bacterial communities' structure within the 15 wild waterbird species was significantly divergent compared to the other investigated fecal sources, suggesting that DNA sequences specific to wild birds could be identified and allowed the development of markers associated with certain wild avian hosts. Firmicutes, Proteobacteria, and Fusobacteria were the dominant phyla in wild waterbird samples, which corroborates findings from previous studies (Waite and Taylor, 2015; Ohad et al., 2016; Wang et al., 2018; Liao et al., 2019).

The Proteobacteria phylum was ubiquitous, with a predominance in wild waterbirds compared to chicken droppings and cattle and pig feces. Fusobacteria was found predominant in the common shelduck group (*Tadorna tadorna*, Anatidae), which differs from the other bird species. However, this result was different from common shelduck fecal samples analyzed in China where Proteobacteria was found as the dominant phylum (90.6%) (Cao et al., 2020). The Actinobacteria phylum is widely distributed in the environment, including soils, fresh and marine waters, and GI tracts of animals (Janssen, 2006; Barka et al., 2016). In our study, it was the fifth dominant phylum in the whole dataset, while scarce data are available on its distribution within wild birds. One of the aims of this study was to explore fecal bacterial communities of poorly investigated bird species such as storm petrels or European shags. Interestingly, the Hydrobatidae family represented by storm petrels in our dataset was found dominated by Actinobacteria (40.8%), followed by Firmicutes (38.7%) and Proteobacteria (17.3%).

Within the Firmicutes phylum, *C. marimammalium* is the bacterial species targeted by several gull-associated qPCR markers (i.e., Gull2 and Gull4). In our study, it was detected mainly in seagulls (great black-backed, herring, and black-headed gulls, and *Larus* sp. and *Chroicocephalus* spp.) but also in shorebirds (oystercatchers, curlews, and knots) and, in a lesser extent, in species belonging to wild Anatidae (Brent goose, common shelducks, mallards, and mute swans), indicating that this bacterial species may have a larger broad-spectrum presence in GI of wild waterbirds. These results agreed with data from Laviad-Shitrit et al. (2019), where black-headed gulls harbored high abundance of the genus *Catellicoccus* (58.8%), while Gull2 and Gull4 markers were found with high incidence in gulls and shorebirds in the United States (Ryu et al., 2012, 2014). In the same way, *Helicobacter* spp., the genus targeted by the general avian marker GFD (Green et al., 2012), was found mostly in the wild birds of our dataset (>20% of the samples from Anatidae, Laridae, and Haematopodidae).

Rare studies developed avian-associated MST markers from the phylum Bacteroidetes (Kobayashi et al., 2013), even if members of Bacteroidetes have been found in relatively high abundance in fecal microbiomes of few waterbirds like wild geese (Wang et al., 2018). In our study, a low relative abundance (<5%) of this phylum was observed in birds except oystercatchers (15.6%) and mallards (*Anas*, 20.1%), which corroborates data from Grond et al. (2018), which attributed it to dietary differences.

In our study, the ANCOM analysis allowed us to identify several potential new MST markers from wild waterbirds (**Supplementary Table 5**) including Swan_2 and Oyscab targeting *Romboutsia* (Firmicutes phylum) and *Bacteroides* genera, respectively, which were the most promising ones. While the *Romboutsia* genus was mainly found in wild waterbirds, it was predominant in swans, which was confirmed by the ANCOM analysis. The qPCR assays targeting Swan_2 gave a sensitivity of 75% and a specificity of 90.2%. A broader validation with extensive sample size and geographic sampling locations will improve the obtained data.

Fecal Microbiomes of Poultry, Livestock Animals, and WW, and MST Markers

Poultry fecal microbiomes were dominated by Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria. This is in line with previously published metataxonomic data for the chicken and turkey GI microbiomes (Qu et al., 2008; Yeoman et al., 2012; Oakley et al., 2013; Wei et al., 2013; Choi et al., 2014; Mohd Shaufi et al., 2015; Borda-Molina et al., 2016). However, a higher average relative abundance of Firmicutes was observed in poultry than in wild waterbirds as previously described by

[7]https://github.com/ifremer-bioinformatics/samba

Grond et al. (2018). At our best knowledge, no study addressing Firmicutes function in wild birds was available, while in domestic chicken, several studies found a positive relationship between Firmicutes abundance and mass gain and immune function, suggesting similar roles of Firmicutes between mammals and birds (Liao et al., 2015; Zhang et al., 2015).

Only one bacterial poultry marker has been developed to our knowledge: the AV43 marker developed by Ohad et al. (2016) from high-throughput sequencing data that was found to be specific in their study (91% sensitivity). However, the taxonomic affiliation only to phylum Firmicutes did not allow searching for this marker or the targeted DNA sequence in our dataset.

Within Actinobacteria, *Brevibacterium*, the genus targeted for the poultry litter qPCR marker LA35 (Weidhaas et al., 2010), was ubiquitous in poultry litter samples (in chicken, from 1 to 18% of reads per sample, and in turkey, from 1 to 4%), whereas it was only present in 10 out of the 17 chicken droppings (from 0.1 to 2.2%), and absent in the turkey droppings, confirming the specificity of the *Brevibacterium* genus to poultry litter samples.

Most MST effort was made on human and livestock animals where bacteria within the phylum Bacteroidetes, and in particular the bacterial order *Bacteroidales*, were the major focus (Bernhard and Field, 2000; Mieszkin et al., 2009; Boehm et al., 2013; Reischer et al., 2013). In our study, this phylum was found in all cattle samples, in more than 70% of the pig samples, and wastewater samples.

In our study, the ANCOM analysis confirmed the status of the well-known human (HF183), pig (Pig2Bac), and cattle (Rum2Bac) MST markers. Furthermore, this ANCOM analysis showed that ASVs belonging to the Ruminococcaceae family (Firmicutes) could be an interesting target for new MST markers for cattle, in agreement with Youngblut et al. (2019) who find genera of this family exclusively in mammals.

Genera With Bacterial Pathogens and MST Markers

Although wild bird droppings appear to harbor fewer pathogenic bacteria and in a weaker frequency than the wastes from poultry, livestock animals, and the influents/effluents of WWTPs, their incidence should not be neglected (Benskin et al., 2009). Studying wild waterbird fecal microbiomes could be useful to have an overview of potential bacterial pathogens harbored by these hosts. This method makes it possible to screen a larger number of pathogens than by using culture-based methods targeting independently pathogens. We found that the composition and the diversity of potential pathogen genera vary significantly between hosts, as observed by Fu et al. (2020). Thus, we observed frequent presence of certain genera that include pathogenic species such as *Clostridium sensu stricto 1*, *Campylobacter*, *Fusobacterium*, and *Helicobacter* in the bird droppings. Such genera were also identified within bacterial communities from several waterbird species including great cormorants, little egrets, black-crowned night herons, and black-headed gulls (Laviad-Shitrit et al., 2019).

As reported in several studies using either 16S rRNA gene sequencing (high-throughput or cloning; Grond et al., 2014; Ryu et al., 2014; Laviad-Shitrit et al., 2019) or by culture-based methods (Waldenstrom et al., 2007; More et al., 2017; Vogt et al., 2018), *Campylobacte*r spp. were found to be frequent in bird fecal microbiomes analyzed in this study. Thus, using a culture-based method, we isolated several *Campylobacte*r spp. strains in half of the wild waterbird droppings samples including Brent goose, common shelducks, mute swans, oystercatchers, curlews, knots, great black-backed, and black-headed gulls (data not shown). Even if the high-throughput sequencing method allows the detection of members of this genus, it rarely allows the taxonomic assignation at the species level (i.e., only *Campylobacter canadensis* identified here). Furthermore, the presence of ASVs that belong to members of genera including potentially human pathogens does not necessarily mean that these ASVs belong to pathogenic species, as suggested by Laviad-Shitrit et al. (2019). Thus, the culture-based isolation of the *Campylobacter* spp. is needed to define the species pathogenic level and evaluate their pathogenicity by *in silico* (screening antibiotic resistance and virulence genes) or *in vitro* (insect or animal models) assays. *Clostridium sensu stricto 1*, another genus harboring human pathogens, was mainly found in wild Anatidae, Phalocrocoracidae, and Scolopacidae (>50% of the samples). This is consistent with previous studies where this genus was commonly found in migratory wild waterbirds (hooded crane and greater white-fronted goose) in China (Xiang et al., 2019) and Israel (Laviad-Shitrit et al., 2019). Members of the genus *Fusobacterium* were present in more than half of the droppings of Brent geese, mallards, oystercatchers, black-headed gulls, and great cormorants, while its relative abundance was very low in black-headed gull samples (<1%). These results are consistent with the results obtained in the study of Laviad-Shitrit et al. (2019) in which *Fusobacterium* was found ubiquitous in the studied bird species with high prevalence (19.8–32.4%) except for the black-headed gulls (0.01%) (Grond et al., 2019). In contrast, other genera were minor, including *Mycoplasma* that was detected in breeding ducks and mute swans. While few species are known as pathogens in birds, *M. gallisepticum* was detected in numerous wild bird species (Sawicka et al., 2020).

The detection and enumeration of all pathogenic microorganisms potentially present is technically impossible. Thus, for routine water quality monitoring, FIB are usually enumerated to evaluate the level of microbial contamination from a fecal source (Ashbolt et al., 2001). However, it is well known that they have not the same distribution as pathogens (Stewart et al., 2008), so it is necessary to develop new MST markers of yet undefined sources. Here, we identified key correlations, i.e., strong correlation between *Catellicoccus*, which harbors bird markers, and *Fusobacterium*, which is widely detected in birds and known to harbor pathogens. While most studies on the relationships between MST markers and pathogens were achieved using qPCR assays for environmental sinks (Vadde et al., 2019), their status for intra-fecal sources using microbiome data in our study is less clear and needs to be further investigated.

In summary, several hypotheses could be stated to explain microbiome diversity and pathogen distribution within the hosts,

such as the bird species diversity, and their variable life-history, such as migratory behavior, diet, and physiology, all of which may impact gut microbiota (Grond et al., 2018). While our dataset included 15 different wild waterbird species, the main limitation was the difficulty in collecting similar numbers of samples per bird species. Simultaneously, culture-based methods were applied for enriching *Campylobacter* members from the collected fecal samples, which helps in evaluating the prevalence of pathogenic species from this genus and their pathogenicity. Future work will include metabarcoding data from other sites, including waters from bathing and shellfish areas that may help in evaluating the coalescence of bacterial communities using the SourceTracker Bayesian approach and mitigation fecal pollution from yet unknown sources.

CONCLUSION

This study provides a comprehensive snapshot of the gut bacterial diversity in a selection of wild waterbirds compared to poultry, cattle, pigs, and wastewater samples, using the NGS-based 16S rRNA gene amplicon method. Although fecal markers have been identified and developed for source tracking applications in many countries, the microorganism communities in GI tracts can be dissimilar for each host species in different geographical areas, which may be due to climate, food, behavior, antibiotics, and other region-specific factors. This potential variation can lead to variability in the performance of fecal markers among regions. High abundances of phyla including Firmicutes, Proteobacteria, and Fusobacterium (especially in wild waterbirds, for this latest phylum), which encompass several genera including potential pathogens, were detected. Moreover, alpha diversity indices and NMDS plots indicated more similarity in the gut microbiota within waterbirds, while they were highly different from the other sources. Comparison to the non-bird sources led us to select several unique and host-associated ASVs related to one or multiple groups of wild waterbirds to be used as MST candidates including the Swan_2 marker. Further analyses are required to improve our findings, while other strategies (i.e., SourceTracker) can be used to help in developing prediction models.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MG and AB coordinated and designed the experiments. AB, BC, JC, and MG participated in the collection of fecal samples. CN, LQ, and AC developed the SAMBA workflow. AB, MG, LQ, and CN performed the bioinformatic data analysis. AB, EQ, LD, and MG performed the experiments and contributed to the analysis of the datasets. AB and MG prepared the manuscript with contributions from all co-authors. All the authors read and approved the final version of the manuscript.

FUNDING

This study was supported by Agence de l'Eau Adour–Garonne, Syndicat mixte du Bassin d'Arcachon (SIBA) and Communauté d'Agglomération Pays Basque (BacTrac project, 2016 –2020).

ACKNOWLEDGMENTS

We thank Adeline Thevand (Siba, Arcachon), Leila Lachkar (Laboratoire des Pyrénées et des Landes, Lagor), Jenna Coton and Alassane Keita (ANSES, Ploufragan), Anthony Sturbois (VivArmor Nature, Réserve Naturelle Baie de Saint Brieuc), Joelle Serghine and Véronique Loiseau (Ifremer, Plouzané), and Stéphane Robert (Ifremer, La Tremblade) as well as farmers and WWTP operators for their contribution to the sample collection. We also thank Patrick Durand and Laura Leroi (Ifremer, Plouzané) for their participation in the data management of DATARMOR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.697553/full#supplementary-material

REFERENCES

Ahmed, W., Harwood, V. J., Nguyen, K., Young, S., Hamilton, K., and Toze, S. (2016). Utility of *Helicobacter* spp. associated GFD markers for detecting avian fecal pollution in natural waters of two continents. *Water Res.* 88, 613–622. doi: 10.1016/j.watres.2015.10.050

Andersson, A. F., Lindberg, M., Jakobsson, H., Backhed, F., Nyren, P., and Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS One* 3:e2836. doi: 10.1371/journal.pone.0002836

Ashbolt, N. J., Grabow, W. O., and Snozzi, M. (2001). "Indicators of microbial water quality," in *Water Quality: Guidelines, Standards and Health*, eds L. Fewtrell and J. Bartram (Geneva: World Health Organization (WHO)), 290–315.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., et al. (2016). Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* 80, 1–43. doi: 10.1128/mmbr.00019-15

Benskin, C. M., Wilson, K., Jones, K., and Hartley, I. R. (2009). Bacterial pathogens in wild birds: a review of the frequency and effects of infection. *Biol. Rev. Camb. Philos. Soc.* 84, 349–373. doi: 10.1111/j.1469-185x.2008.00076.x

Bernhard, A. E., and Field, K. G. (2000). 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

Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., et al. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. *Water Res.* 47, 6812–6828. doi: 10.1016/j.watres.2012.12.046

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.

Borda-Molina, D., Vital, M., Sommerfeld, V., Rodehutscord, M., and Camarinha-Silva, A. (2016). Insights into Broilers' Gut microbiota fed with phosphorus, calcium, and phytase supplemented diets. *Front. Microbiol.* 7:2033. doi: 10.3389/fmicb.2016.02033

Brown, C. M., Staley, C., Wang, P., Dalzell, B., Chun, C. L., and Sadowsky, M. J. (2017). A high-throughput DNA-Sequencing approach for determining sources of fecal bacteria in a lake superior estuary. *Environ. Sci. Technol.* 51, 8263–8271. doi: 10.1021/acs.est.7b01353

Callahan, B. J., Mcmurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cao, J., Hu, Y., Liu, F., Wang, Y., Bi, Y., Lv, N., et al. (2020). Metagenomic analysis reveals the microbiome and resistome in migratory birds. *Microbiome* 8:26.

Choi, J. H., Kim, G. B., and Cha, C. J. (2014). Spatial heterogeneity and stability of bacterial community in the gastrointestinal tracts of broiler chickens. *Poult. Sci.* 93, 1942–1950. doi: 10.3382/ps.2014-03974

Chung, D. M., Ferree, E., Simon, D. M., and Yeh, P. J. (2018). Patterns of bird-bacteria associations. *Ecohealth* 15, 627–641. doi: 10.1007/s10393-018-1342-5

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Devane, M. L., Robson, B., Nourozi, F., Scholes, P., and Gilpin, B. J. (2007). A PCR marker for detection in surface waters of faecal pollution derived from ducks. *Water Res.* 41, 3553–3560. doi: 10.1016/j.watres.2007.06.043

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35, 316–319. doi: 10.1038/nbt.3820

European Food Safety Authority (EFSA), and European Centre for Disease Prevention and Control (ECDC) (2019). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2018. *EFSA J.* 17:5926.

Ewbank, A. C., Esperon, F., Sacristan, C., Sacristan, I., Krul, R., Cavalcante De Macedo, E., et al. (2021). Seabirds as anthropization indicators in two different tropical biotopes: a One Health approach to the issue of antimicrobial resistance genes pollution in oceanic islands. *Sci. Total Environ.* 754:142141. doi: 10.1016/j.scitotenv.2020.142141

Fu, R., Xiang, X., Dong, Y., Cheng, L., and Zhou, L. (2020). comparing the intestinal bacterial communies of sympatric wintering hooded crane (*Grus monacha*) and domestic goose (*Anser anser* domesticus). *Avian Res.* 11:13.

Glockner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., et al. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi: 10.1016/j.jbiotec.2017.06.1198

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

Grond, K., Ryu, H., Baker, A. J., Santo Domingo, J. W., and Buehler, D. M. (2014). Gastro-intestinal microbiota of two migratory shorebird species during spring migration staging in Delaware Bay, USA. *J. Ornithol.* 4, 969–977. doi: 10.1007/s10336-014-1083-3

Grond, K., Sandercock, B. K., Jumpponen, A., and Zeglin, L. H. (2018). The avian gut microbiota: community, physiology and function in wild birds. *J. Avian Biol.* 49:e01788. doi: 10.1111/jav.01788

Grond, K., Santo Domingo, J. W., Lanctot, R. B., Jumpponen, A., Bentzen, R. L., Boldenow, M. L., et al. (2019). Composition and drivers of gut microbial communities in arctic-breeding shorebirds. *Front. Microbiol.* 10:2258. doi: 10.3389/fmicb.2019.02258

Hamilton, M. J., Yan, T., and Sadowsky, M. J. (2006). Development of goose- and duck-specific DNA markers to determine sources of *Escherichia coli* in waterways. *Appl. Environ. Microbiol.* 72, 4012–4019. doi: 10.1128/aem.02764-05

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

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., et al. (2016). Into the deep: evaluation of sourcetracker for assessment of faecal contamination of coastal waters. *Water Res.* 93, 242–253. doi: 10.1016/j.watres.2016.02.029

Hird, S. M. (2017). Evolutionary biology needs wild microbiomes. *Front. Microbiol.* 8:725. doi: 10.3389/fmicb.2017.00725

Hird, S. M., Sanchez, C., Carstens, B. C., and Brumfield, R. T. (2015). Comparative gut microbiota of 59 neotropical bird species. *Front. Microbiol.* 6:1403. doi: 10.3389/fmicb.2015.01403

Issa, N., and Muller, Y. (2015). *Atlas des Oiseaux de France Métropolitaine : Nidification et Présence Hivernale*. Paris: Delachaux et Niestlé.

Iwamoto, M., Ayers, T., Mahon, B. E., and Swerdlow, D. L. (2010). Epidemiology of seafood-associated infections in the United States. *Clin. Microbiol. Rev.* 23, 399–411. doi: 10.1128/cmr.00059-09

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi: 10.1128/aem.72.3.1719-1728.2006

Jarde, E., Jeanneau, L., Harrault, L., Quenot, E., Solecki, O., Petitjean, P., et al. (2018). Application of a microbial source tracking based on bacterial and chemical markers in headwater and coastal catchments. *Sci. Total Environ.* 610-611, 55–63. doi: 10.1016/j.scitotenv.2017.07.235

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* 35, W43–W46.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., et al. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763. doi: 10.1038/nmeth.1650

Kobayashi, A., Sano, D., Hatori, J., Ishii, S., and Okabe, S. (2013). Chicken- and duck-associated *Bacteroides*-Prevotella genetic markers for detecting fecal contamination in environmental water. *Appl. Microbiol. Biotechnol.* 97, 7427–7437. doi: 10.1007/s00253-012-4469-2

Koskey, A. M., Fisher, J. C., Traudt, M. F., Newton, R. J., and Mclellan, S. L. (2014). Analysis of the gull fecal microbial community reveals the dominance of *Catellicoccus* marimammalium in relation to culturable *Enterococci*. *Appl. Environ. Microbiol.* 80, 757–765. doi: 10.1128/aem.02414-13

Kreisinger, J., Cizkova, D., Kropackova, L., and Albrecht, T. (2015). Cloacal microbiome structure in a long-distance migratory bird assessed using deep 16sRNA pyrosequencing. *PLoS One* 10:e0137401. doi: 10.1371/journal.pone.0137401

Laviad-Shitrit, S., Izhaki, I., Lalzar, M., and Halpern, M. (2019). Comparative analysis of intestine microbiota of four wild waterbird species. *Front. Microbiol.* 10:1911. doi: 10.3389/fmicb.2019.01911

Liao, F., Gu, W., Li, D., Liang, J., Fu, X., Xu, W., et al. (2019). Characteristics of microbial communities and intestinal pathogenic bacteria for migrated *Larus ridibundus* in southwest China. *Microbiologyopen* 8:e00693. doi: 10.1002/mbo3.693

Liao, X. D., Ma, G., Cai, J., Fu, Y., Yan, X. Y., Wei, X. B., et al. (2015). Effects of Clostridium butyricum on growth performance, antioxidation, and immune function of broilers. *Poult. Sci.* 94, 662–667. doi: 10.3382/ps/pev038

Liu, Z., Desantis, T. Z., Andersen, G. L., and Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* 36:e120. doi: 10.1093/nar/gkn491

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

Lu, J., Santo Domingo, J. W., Hill, S., and Edge, T. A. (2009). Microbial diversity and host-specific sequences of Canada goose feces. *Appl. Environ. Microbiol.* 75, 5919–5926. doi: 10.1128/aem.00462-09

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

Mandal, S., Van Treuren, W., White, R. A., Eggesbo, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* 26:27663.

Mieszkin, S., Furet, J. P., Corthier, G., and Gourmelon, M. (2009). Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045–3054. doi: 10.1128/aem.02343-08

Mohd Shaufi, M. A., Sieo, C. C., Chong, C. W., Gan, H. M., and Ho, Y. W. (2015). Deciphering chicken gut microbial dynamics based on high-throughput 16S rRNA metagenomics analyses. *Gut Pathog.* 7:4. doi: 10.1186/s13099-015-0051-7

More, E., Ayats, T., Ryan, P. G., Naicker, P. R., Keddy, K. H., Gaglio, D., et al. (2017). Seabirds (Laridae) as a source of *Campylobacter* spp., *Salmonella* spp. and antimicrobial resistance in South Africa. *Environ. Microbiol.* 19, 4164–4176. doi: 10.1111/1462-2920.13874

Oakley, B. B., Morales, C. A., Line, J., Berrang, M. E., Meinersmann, R. J., Tillman, G. E., et al. (2013). The poultry-associated microbiome: network analysis and farm-to-fork characterizations. *PLoS One* 8:e57190. doi: 10.1371/journal.pone.0057190

Ohad, S., Ben-Dor, S., Prilusky, J., Kravitz, V., Dassa, B., Chalifa-Caspi, V., et al. (2016). The development of a novel qPCR assay-set for identifying fecal contamination originating from domestic fowls and waterfowl in Israel. *Front. Microbiol.* 7:145. doi: 10.3389/fmicb.2016.00145

Olesen, S. W., Duvallet, C., and Alm, E. J. (2017). dbOTU3: a new implementation of distribution-based OTU calling. *PLoS One* 12:e0176335. doi: 10.1371/journal.pone.0176335

Pommepuy, M., Hervio-Heath, D., Caprais, M. P., Gourmelon, M., and Le Guyader, F. (2006). "Fecal contamination in coastal areas: an engineering approach," in *Oceans and Health: Pathogens in the Marine Environment*, eds S. Belkin and R. R. Colwell (Cambridge, MA: Academic Press), 331–359. doi: 10.1007/0-387-23709-7_14

Potasman, I., Paz, A., and Odeh, M. (2002). Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective. *Clin. Infect. Dis.* 35, 921–928. doi: 10.1086/342330

Qu, A., Brulc, J. M., Wilson, M. K., Law, B. F., Theoret, J. R., Joens, L. A., et al. (2008). Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS One* 3:e2945. doi: 10.1371/journal.pone.0002945

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reese, A. T., and Dunn, R. R. (2018). Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. *mBio* 9:e01294-18.

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Astrom, J., et al. (2013). Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* 47, 8548–8556. doi: 10.1021/es304367t

Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., and Farnleitner, A. H. (2006). Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72, 5610–5614. doi: 10.1128/aem.00364-06

Rince, A., Baliere, C., Hervio-Heath, D., Cozien, J., Lozach, S., Parnaudeau, S., et al. (2018). Occurrence of bacterial pathogens and human noroviruses in shellfish-harvesting areas and their catchments in France. *Front. Microbiol.* 9:2443. doi: 10.3389/fmicb.2018.02443

Ryu, H., Grond, K., Verheijen, B., Elk, M., Buehler, D. M., and Santo Domingo, J. W. (2014). Intestinal microbiota and species diversity of Campylobacter and *Helicobacter* spp. in migrating shorebirds in Delaware Bay. *Appl. Environ. Microbiol.* 80, 1838–1847. doi: 10.1128/aem.03793-13

Ryu, H., Lu, J., Vogel, J., Elk, M., Chavez-Ramirez, F., Ashbolt, N., et al. (2012). Development and evaluation of a quantitative PCR assay targeting sandhill crane (*Grus canadensis*) fecal pollution. *Appl. Environ. Microbiol.* 78, 4338–4345. doi: 10.1128/aem.07923-11

Sawicka, A., Durkalec, M., Tomczyk, G., and Kursa, O. (2020). Occurrence of Mycoplasma gallisepticum in wild birds: a systematic review and meta-analysis. *PLoS One* 15:e0231545. doi: 10.1371/journal.pone.0231545

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* 7, 249–259. doi: 10.1111/j.1462-2920.2004.00702.x

Smith, O. M., Snyder, W. E., and Owen, J. P. (2020). Are we overestimating risk of enteric pathogen spillover from wild birds to humans? *Biol. Rev.* 3, 652–679. doi: 10.1111/brv.12581

Stewart, J. R., Gast, R. J., Fujioka, R. S., Solo-Gabriele, H. M., Meschke, J. S., Amaral-Zettler, L. A., et al. (2008). The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. *Environ. Health* 7(Suppl. 2):S3.

Sun, C. H., Liu, H. Y., Zhang, Y., and Lu, C. H. (2018). Comparative analysis of the gut microbiota of hornbill and toucan in captivity. *Microbiologyopen* 8:e786.

Unno, T., Staley, C., Brown, C. M., Han, D., Sadowsky, M. J., and Hur, H. G. (2018). Fecal pollution: new trends and challenges in microbial source tracking using next-generation sequencing. *Environ. Microbiol.* 20, 3132–3140. doi: 10.1111/1462-2920.14281

Vadde, K. K., Mccarthy, A. J., Rong, R., and Sekar, R. (2019). Quantification of microbial source tracking and pathogenic bacterial markers in water and sediments of tiaoxi river (Taihu Watershed). *Front. Microbiol.* 10:699. doi: 10.3389/fmicb.2019.00699

Vogt, N. A., Pearl, D. L., Taboada, E. N., Mutschall, S. K., Janecko, N., Reid-Smith, R., et al. (2018). Epidemiology of *Campylobacter*, *Salmonella* and antimicrobial resistant *Escherichia coli* in free-living Canada geese (*Branta canadensis*) from three sources in southern Ontario. *Zoonoses Public Health* 65, 873–886.

Waite, D. W., and Taylor, M. W. (2015). Exploring the avian gut microbiota: current trends and future directions. *Front. Microbiol.* 6:673. doi: 10.3389/fmicb.2015.00673

Waldenstrom, J., On, S. L., Ottvall, R., Hasselquist, D., and Olsen, B. (2007). Species diversity of campylobacteria in a wild bird community in Sweden. *J. Appl. Microbiol.* 102, 424–432.

Wang, W., Liu, Y., Yang, Y., Wang, A., Sharshov, K., Li, Y., et al. (2018). Comparative analyses of the gut microbiota among three different wild geese species in the genus *Anser*. *J. Basic Microbiol.* 58, 543–553. doi: 10.1002/jobm.201800060

Wei, S., Morrison, M., and Yu, Z. (2013). Bacterial census of poultry intestinal microbiome. *Poult. Sci.* 92, 671–683. doi: 10.3382/ps.2012-02822

Weidhaas, J. L., Macbeth, T. W., Olsen, R. L., Sadowsky, M. J., Norat, D., and Harwood, V. J. (2010). Identification of a *Brevibacterium* marker gene specific to poultry litter and development of a quantitative PCR assay. *J. Appl. Microbiol.* 109, 334–347. doi: 10.1111/j.1365-2672.2010.04666.x

Weidhaas, J., and Lipscomb, E. (2013). A new method for tracking poultry litter in the Potomac Basin headwaters of West Virginia. *J. Appl. Microbiol.* 115, 445–454. doi: 10.1111/jam.12231

Westrell, T., Dusch, V., Ethelberg, S., Harris, J., Hjertqvist, M., Jourdan-Da Silva, N., et al. (2010). Norovirus outbreaks linked to oyster consumption in the United Kingdom, Norway, France, Sweden and Denmark, 2010. *Euro Surveill.* 15, 19524.

Wienemann, T., Schmitt-Wagner, D., Meuser, K., Segelbacher, G., Schink, B., Brune, A., et al. (2011). The bacterial microbiota in the ceca of Capercaillie (*Tetrao urogallus*) differs between wild and captive birds. *Syst. Appl. Microbiol.* 34, 542–551. doi: 10.1016/j.syapm.2011.06.003

Xiang, X., Zhang, F., Fu, R., Yan, S., and Zhou, L. (2019). Significant differences in bacterial and potentially pathogenic communities between sympatric hooded crane and greater white-fronted goose. *Front. Microbiol.* 10:163. doi: 10.3389/fmicb.2019.00163

Yeoman, C. J., Chia, N., Jeraldo, P., Sipos, M., Goldenfeld, N. D., and White, B. A. (2012). The microbiome of the chicken gastrointestinal tract. *Anim. Health Res. Rev.* 13, 89–99. doi: 10.1017/s1466252312000138

Yoder, J. S., Hlavsa, M. C., Craun, G. F., Hill, V., Roberts, V., Yu, P. A., et al. (2008). Surveillance for waterborne disease and outbreaks associated with recreational

water use and other aquatic facility-associated health events–United States, 2005-2006. *MMWR Surveill. Summ.* 57, 1–29.

Youngblut, N. D., Reischer, G. H., Walters, W., Schuster, N., Walzer, C., Stalder, G., et al. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat. Commun.* 10:2200.

Zhang, L., Li, J., Yun, T. T., Qi, W. T., Liang, X. X., Wang, Y. W., et al. (2015). Effects of pre-encapsulated and pro-encapsulated *Enterococcus faecalis* on growth performance, blood characteristics, and cecal microflora in broiler chickens. *Poult. Sci.* 94, 2821–2830. doi: 10.3382/ps/pev262

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.

Copyright © 2021 Boukerb, Noël, Quenot, Cadiou, Chevé, Quintric, Cormier, Dantan and Gourmelon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 July 2021
doi: 10.3389/fmicb.2021.665664

Influence of Filter Pore Size on Composition and Relative Abundance of Bacterial Communities and Select Host-Specific MST Markers in Coastal Waters of Southern Lake Michigan

Muruleedhara N. Byappanahalli[1], Meredith B. Nevers[1], Dawn Shively[1,2], Cindy H. Nakatsu[3], Julie L. Kinzelman[4†] and Mantha S. Phanikumar[2]*

[1] U.S. Geological Survey, Great Lakes Science Center, Chesterton, IN, United States, [2] Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI, United States, [3] Department of Agronomy, Purdue University, West Lafayette, IN, United States, [4] Public Health Department Laboratory, Racine, WI, United States

Water clarity is often the primary guiding factor in determining whether a prefiltration step is needed to increase volumes processed for a range of microbial endpoints. In this study, we evaluate the effect of filter pore size on the bacterial communities detected by 16S rRNA gene sequencing and incidence of two host-specific microbial source tracking (MST) markers in a range of coastal waters from southern Lake Michigan, using two independent data sets collected in 2015 (bacterial communities) and 2016–2017 (MST markers). Water samples were collected from river, shoreline, and offshore areas. For bacterial communities, each sample was filtered through a 5.0-μm filter, followed by filtration through a 0.22-μm filter, resulting in 70 and 143 filter pairs for bacterial communities and MST markers, respectively. Following DNA extraction, the bacterial communities were compared using 16S rRNA gene amplicons of the V3–V4 region sequenced on a MiSeq Illumina platform. Presence of human (*Bacteroides* HF183) and gull (Gull2, *Catellicoccus marimammalium*) host-specific MST markers were detected by qPCR. Actinobacteriota, Bacteroidota, and Proteobacteria, collectively represented 96.9% and 93.9% of the relative proportion of all phyla in the 0.22- and 5.0-μm pore size filters, respectively. There were more families detected in the 5.0-μm pore size filter (368) than the 0.22-μm (228). There were significant differences in the number of taxa between the two filter sizes at all levels of taxonomic classification according to linear discriminant analysis (LDA) effect size (LEfSe) with as many as 986 taxa from both filter sizes at LDA effect sizes greater than 2.0. Overall, the Gull2 marker was found in higher abundance on the 5.0-μm filter than 0.22 μm with the reverse pattern for the HF183 marker. This discrepancy could lead to problems with identifying microbial sources of contamination. Collectively, these results highlight the importance of analyzing pre- and

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Jinjun Kan, Stroud Water Research Center, United States
Lisa Paruch, Norwegian Institute of Bioeconomy Research (NIBIO), Norway
Warish Ahmed, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

***Correspondence:**
Muruleedhara N. Byappanahalli
byappan@usgs.gov
orcid.org/0000-0001-5376-597X

† Present address:
Julie L. Kinzelman, College of Natural and Health Sciences, University of Wisconsin–Parkside, Kenosha, WI, United States

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *08 February 2021*
Accepted: *24 May 2021*
Published: *15 July 2021*

Citation:
Byappanahalli MN, Nevers MB, Shively D, Nakatsu CH, Kinzelman JL and Phanikumar MS (2021) Influence of Filter Pore Size on Composition and Relative Abundance of Bacterial Communities and Select Host-Specific MST Markers in Coastal Waters of Southern Lake Michigan. Front. Microbiol. 12:665664. doi: 10.3389/fmicb.2021.665664

final filters for a wide range of microbial endpoints, including host-specific MST markers routinely used in water quality monitoring programs. Analysis of both filters may increase costs but provides more complete genomic data *via* increased sample volume for characterizing microbial communities in coastal waters.

Keywords: bacterial community composition, 16S rRNA sequencing, filter pore size, filter processing, Great Lakes, host-specific MST markers, sample volume

INTRODUCTION

Significant, recent advancements in the field of molecular biology have contributed to increased application of molecular (i.e., genetic) tools in water quality monitoring programs. Notable examples include the use of polymerase chain reaction (PCR) and quantitative PCR (qPCR) to monitor fecal indicator bacteria and/or human pathogens and to identify contamination sources (Bej et al., 1990; Mahbubani et al., 1991; Scott et al., 2002; Brinkman et al., 2003). More recent advances include high-throughput DNA sequencing (e.g., 16S rRNA amplicon sequencing or similar targets) to better identify and characterize contamination sources and potential interactions among waterways using microbial communities as an index of water quality (McLellan et al., 2009; Newton et al., 2013; Staley and Sadowsky, 2016). These molecular methods and other emerging technologies (e.g., environmental DNA; high-throughput sequencing) are increasingly incorporated into monitoring programs—as a stand-alone method or in conjunction with traditional monitoring methods—with endpoints ranging from measuring shoreline water quality (Wade et al., 2006) to assessing population of fish and other aquatic biota (Goldberg et al., 2016; Lacoursière-Roussel et al., 2016). For water, membrane filtration remains the most common method to concentrate cells and/or DNA from organisms whether the targets are select bacteria, host-specific MST markers, microbial communities, or other biota of interest (e.g., fish and macroinvertebrates) (Venter et al., 2004; Harwood et al., 2014; Goldberg et al., 2016). Selection of filter pore size is largely dependent on water clarity, sample type, and volume as well as characterizing biotic communities by size, composition, and their association with particulate matter (Fuhrman et al., 1988; Venter et al., 2004; Walsh et al., 2009; Francy et al., 2013; Padilla et al., 2015).

Increasingly, a prefiltration step is included as part of sample processing protocols to overcome clogging issues when samples are generally turbid and to ensure that sufficient cells (biota)/DNA are collected from samples by increasing the sample volume for a range of applications (e.g., to differentiate particle-bound and free or unbound cells) (Venter et al., 2004; Staley et al., 2013; Padilla et al., 2015). However, there is little consistency in selecting pore-size filters across water types; often, prevailing water conditions (e.g., dissolved organic matter, turbidity) and measured microbial endpoints (e.g., viruses) (Mull and Hill, 2012; Francy et al., 2013; Xu and Guo, 2017) dictate filter selection. Further, processing of prefilters varies widely across studies: from size-specific, independent extractions to co-extractions (Staley et al., 2013; Padilla et al., 2015; Jain and Krishnan, 2017), which makes study comparisons difficult. Thus, the goal of the current study was to evaluate the influence of filter pore size on composition and relative abundance of bacterial communities and select host-specific microbial source tracking (MST) markers detected in a range of coastal waters along southern Lake Michigan (Wisconsin, Illinois, and Indiana). For this purpose, we used two independent data sets collected over two different time periods: 2015 for 16S rRNA gene sequencing and 2016–2017 for host-specific MST markers to identify human (HF183) and gull (Gull2) fecal contamination.

MATERIALS AND METHODS

Study Area and Sampling Design

The study area is located along the southwestern shoreline of Lake Michigan within the following cities: Racine, Wisconsin; Chicago, Illinois, and East Chicago and Whiting, Indiana. The sampling sites in Racine included the Root River mouth (RRM), an upstream location on the Root River (REC), an offshore location at the marina harbor wall (NBW), and an engineered wetland outlet at North Beach (NB; IEB overflow) as well as a nearshore location at NB, a recreational beach. In Chicago, the 63rd Street Beach was the sole sampling location. In East Chicago, samples were collected from the Grand Calumet River at Columbus Drive (GCR), the mouth of the GCR (GCM), three shoreline locations at Jeorse Park (JP, a recreational beach), and two offshore locations: north (GCN) and east (GCE) of a constructed peninsula; in Whiting, samples were collected from the Whihala Beach (WHW). Additional details of sample collection are available elsewhere (Nakatsu et al., 2019; Kinzelman et al., 2020; Nevers et al., 2020).

Water samples were collected over three summers, 2015–2017. There were 70 water samples collected in 2015 (0.22- and 5.0-μm pore size filters) for 16S rRNA gene sequence analysis and 143 water samples collected in 2016–2017 for MST marker analysis; all water samples were filtered through 0.22- and 5.0-μm pore size filters (see **Table 1**). Bacterial community data gathered from the 0.22-μm pore size filters have been published elsewhere in two separate studies: 63 filters from Illinois and Indiana locations (Nakatsu et al., 2019) and 54 filters from Wisconsin locations only (Kinzelman et al., 2020). Similarly, the MST data from the 0.22-μm pore size filters (143), except for samples from NB2, NB4, and RRM, are published in a separate study (Nevers et al., 2020). The data related to the 5.0-μm pore size filters—70 (i.e., from 16S rRNA sequencing) and 143 (MST)—are independent and unpublished work. Results from the 0.22-μm pore size filters are used for comparison purposes only and acknowledged appropriately here and elsewhere in this manuscript.

TABLE 1 | Study locations, sampling years, and the measured microbial endpoints (16S rRNA gene sequencing and two host-specific MST markers) in coastal waters along southern Lake Michigan.

		Analysis		Year		
Site	**Coordinates (Lat/Long)**	**16S rRNA gene sequencing**	**Host-specific MST markers**	**2015**	**2016**	**2017**
JP	41.650478/−87.433551	X	X	X	X	X
GCE	41.667120/−87.404940	X		X		
GCN	41.692170/−87.415550	X		X		
GCM	41.687190/−87.439740	X		X		
GCR	41.639400/−87.471276	X	X	X	X	X
WHW	41.685118/−87.492283	X	X	X		
63rd	41.782209/−87.572926	X	X	X	X	X
IEB outlet	42.746607/−87.781631	X		X		
NB	42.741666/−87.779915	X	X	X	X	X
NBW	42.737361/−87.774867	X		X		
RRM	42.733424/−87.771695	X	X	X	X	X
REC	42.724360/−87.795930	X	X	X	X	X

Water samples were collected (as denoted by symbol X) from Racine, Wisconsin [lake nearshore, NB; lake offshore, NBW, river (Root River), REC; river mouth (Root River Mouth), RPM; and wetland outlet, IEB outlet], Illinois (63rd Street Beach, 63rd); and East Chicago, Indiana (lake nearshore, JP and WHW; lake offshore, GCE and GCN; Grand Calumet River, GCR; and river mouth, GCM) as shown in ***Figure 1****.*

Sample Collection

In the summer of 2015, water samples (1000 ml) were collected in triplicate during three independent events on the following dates (month/day): Illinois and Indiana: 08/12, 09/01, and 09/21; Wisconsin: 08/03, 08/27, and 09/02. The 11 locations represented five sources: (i) nearshore lake, (ii) offshore lake, (iii) river, (iv) river mouth, and (v) wetland outlet (**Figure 1** and **Table 1**). In 2016, individual water samples (1,000 ml) for filter comparison were collected during three events; and in 2017, individual water samples were collected 1 day per week for 11 weeks (06/08–08/10 from Wisconsin, Illinois, and Indiana locations; see **Table 1**).

Water from GCR was collected by tossing a sterile collection bucket from the bridge crossing at Columbus Drive, offshore surface water samples (GCM, GCN, and GCE) were collected from a boat by dipping sterile collection bottles (1,000 ml capacity) below the surface, and nearshore samples (JP, WHW, 63rd) were similarly collected by dipping collection bottles (1,000 ml capacity) below the surface in 45-cm-deep water. Nearshore water samples at NB were collected in sterile Whirl-Pak™ bags (Nasco, Fort Atkinson, WI, United States) at a depth of 0.65–0.70 m and approximately 0.3 m below the water surface. Samples from the REC and IEB outlet sites were collected using sterile Whirl-Pak bags and a sampling pole, typically at 10 cm below the surface; samples from RPM and an offshore location (NBW) were collected in sterile Nalgene bottles (1,000 ml capacity) using a sampling line (Kinzelman et al., 2020). All samples were stored on ice directly after collection for return to the laboratory and processed within 6 h of collection.

DNA Extraction

All water samples (1,000 ml) were first filtered through a 5.0-μm nitrocellulose filter (47 mm, EMD Millipore, Billerica, MA, United States); filtrates were then filtered through 0.22-μm nitrocellulose filter (47 mm, EMD Millipore, Billerica, MA, United States), and both filters were placed in separate DNeasy PowerWater Kit bead tubes (Qiagen, Inc., formerly MOBIO PowerWater) and held at −80°C until processing. DNA was extracted from each filter using a DNeasy PowerWater Kit according to manufacturer's instructions but with one exception: the final DNA elution was performed twice using 50 μl of DNA elution buffer each time for a final extraction volume of 100 μl. DNA concentrations were measured by fluorometric quantitation using a Qubit® instrument and a High Sensitivity dsDNA HS Assay kit (Qubit, Thermo Fisher Scientific, Waltham, MA, United States). Nucleic acid quality (i.e., 260/280 ratio) was measured with a Nanophotometer® Pearl (Implen Inc., Westlake Village, CA, United States).

qPCR

DNA extracts from water samples collected during 2016–2017 were analyzed for human (HF183) (Green et al., 2014) and gull (Gull-2) (Lu et al., 2008) markers. The primers and probes used, reaction conditions, and assay controls are explained in greater detail elsewhere (Nevers et al., 2018). qPCR assays for the human and gull MST markers were performed using the Bio-Rad CFX™ Connect Real-time PCR Detection System (Bio-Rad, Hercules, CA, United States) in clear 96-well PCR plates containing 25 μl reaction mixtures. All qPCR assays included appropriate positive and no template controls as well as other quality assurance parameters, such as inhibition controls and extraction, filter, and field blanks to confirm that the assays were working within the conditions defined. For instance, each qPCR assay was tested for inhibition by analyzing diluted and undiluted DNA extracts (randomly chosen, 20%); dilutions consisted of 5× to achieve a concentration range of 2–5 ng total DNA.

DNA extracts were stored at −20°C (short-term, typically <1 week) or −80°C (long-term, up to several months) in case reanalysis was required. For each qPCR assay, amplification efficiency and R^2 were calculated; standard curves for all runs had an $R^2 \geq 0.99$ with amplification efficiencies ranging

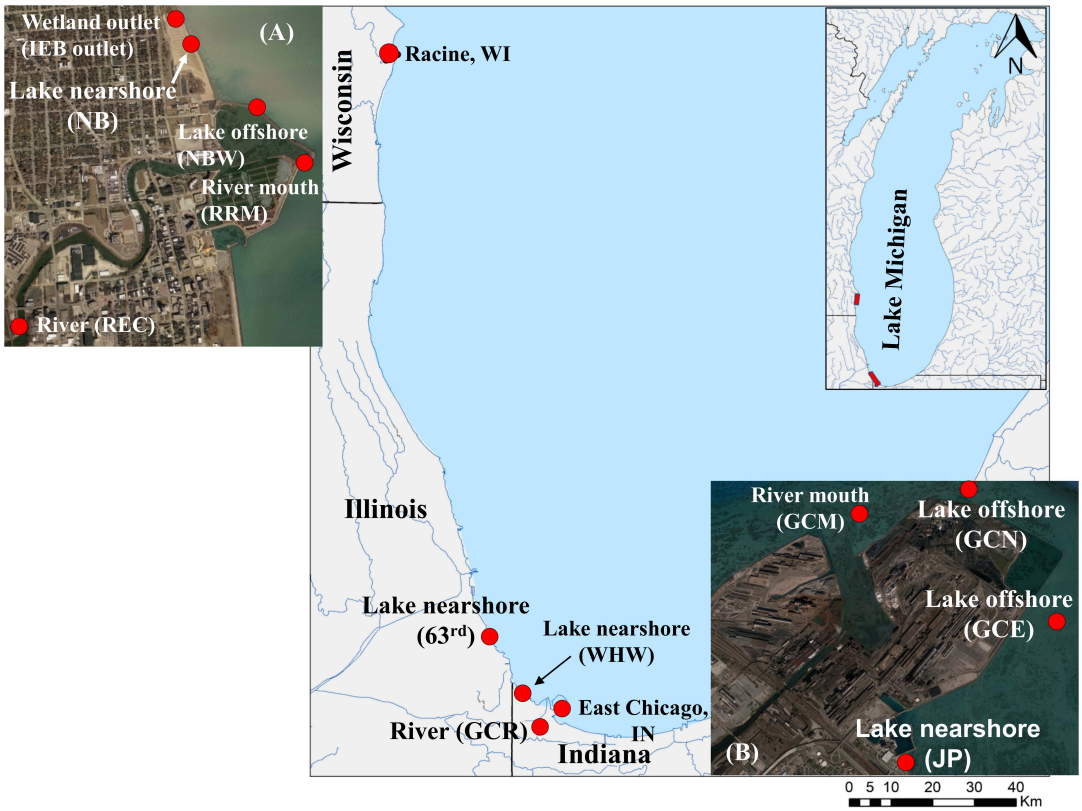


FIGURE 1 | Map of southern Lake Michigan depicting the sampling locations. Water samples were collected from river, shoreline, and offshore sites from Racine, Wisconsin; Chicago, Illinois; and East Chicago, Indiana, locations bordering southern Lake Michigan. The graphical illustration is adapted from Nevers et al. (2020). The sampling sites for river, river mouth, lake near shore, and offshore as well as wetland are shown in the enlarged insets: **(A)** Racine and **(B)** Jeorse Park.

between 92–98% and 92–94% for HF183 and Gull2 assays, respectively. Samples were considered inhibited if Cq values (diluted and undiluted) were dissimilar relative to dilution. Cq is the quantification cycle or the cycle number in which the DNA fluorescence increases above any background fluorescence. The lower limit of quantification (LLOQ) was established by averaging the Cq values obtained from the highest dilutions from standard curves across all qPCR runs in which at least two of the three technical replicates were detected. This resulted in an LLOQ of Cq = 37 for Gull2 and HF183 assays (Nevers et al., 2018), corresponding to an average of 33CN/rx (Gull2) and 18 CN/rx (HF183) (Nevers et al., 2020), which were subsequently used to determine if a sample Cq value falls within the range of quantification (ROQ) and detected but not quantifiable (DNQ). Samples were considered positive if the Cq value was within the ROQ and DNQ; samples were considered negative non-detect (ND) if there was no exponential curve crossing the threshold value before cycle 40.

Illumina 16S rRNA Gene Sequencing

The 16S rRNA gene sequencing of DNA extracts from water samples is explained in greater detail elsewhere (Nakatsu et al., 2019; Kinzelman et al., 2020). Briefly, PCR amplification of the 16S rRNA gene was performed using V3–V4 region primers (343-forward TAC GGR AGG CAG CAG and 804-reverse CTA CCR GGG TAT CTA ATC C) and ∼10 ng of template DNA. Index tags were added using the step out protocol following the manufacturer's instructions (Illumina, San Diego, CA, United States) (Gloor et al., 2010). All PCR included no template controls. PCR amplicons were quantified using a Nanodrop 3000 fluorospectrometer after staining with the QuantiFluor dsDNA System (Promega^{TM} Corporation, Madison, WI, United States). Equimolar amounts of amplicons from each sample were combined and sent to the Purdue Genomics core facilities, West Lafayette, IN, United States for 2 × 250 paired end sequencing using a MiSeq Illumina system.

16S rRNA Gene Sequence Analysis

One set of river samples (designated by W06 and W123, 0.2- and 5.0-μm pore size filters, respectively) was excluded from sequence analysis because of low reads in W06. Sequences from 140 samples were analyzed using the QIIME 2 pipeline version 2021.2 (Bolyen et al., 2019). Demultiplex sequences were trimmed for quality, denoised, paired ends merged, and sorted into representative amplicon sequence variants (ASVs) using DADA2 (Callahan et al., 2016). Taxonomic assignment to ASVs were made using a classifier trained using the SILVA data set (version 138, 99% OTUs from the V3/V4 16S rRNA

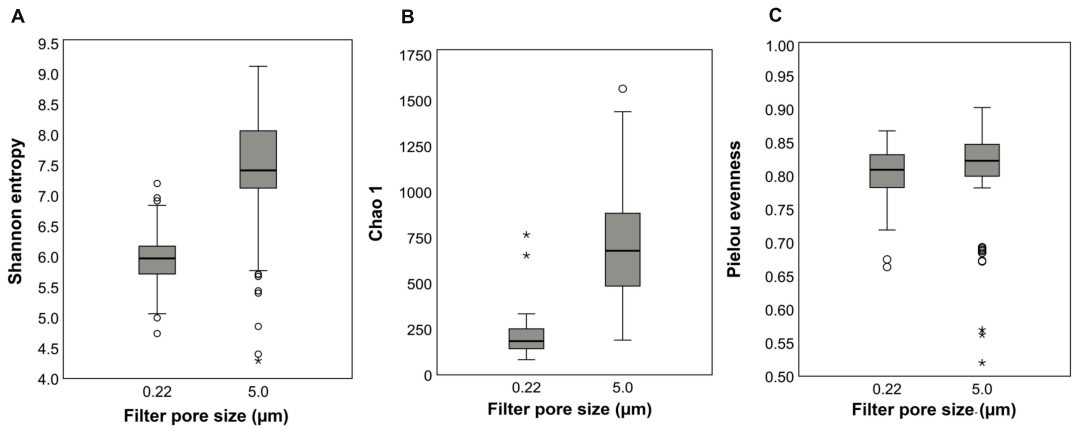


FIGURE 2 | Alpha diversity indices: Boxplots of means and standard deviations (SD) of within-sample diversity from all sampling locations for 0.22- and 5.0-µm pore size filters; **(A)** Shannon (richness and evenness), **(B)** Chao1 richness estimate, and **(C)** Pielou's community evenness.

gene region) (Yilmaz et al., 2014). For phylogenetic analyses, trees were generated using the multi-alignment program MAFFT (Katoh et al., 2002) and FastTree (Price et al., 2010). Rare ASVs (<0.1% of average reads), unclassified ASVs, and chloroplast reads were filtered out of the data set. Sufficient read coverage of samples was determined by rarefaction analysis as well as estimation of Good's coverage. All 140 samples were included in diversity analyses and were rarified to the same sampling depth of 6400 reads. Alpha diversity was analyzed using Shannon (richness and evenness), and also, Faith's phylogenetic diversity (PD) was used for phylogenetic richness and Chao1 for species richness estimates and Pielou's evenness. Beta diversity analysis included non-phylogenetic metrics Bray–Curtis and Jaccard and phylogenetic metrics weighted and unweighted Unifrac (Lozupone et al., 2011).

Statistical Analyses

QIIME 2 software was used to determine statistically significant differences between 0.22- and 5.0-µm pore size filters for all alpha and beta diversity metrics tested. All statistical tests were nonparametric for non-normally distributed data. Pairwise Kruskal–Wallis tests were performed for alpha diversity metrics, ANOSIM for non-phylogeny and PERMANOVA (999 permutations) (Anderson, 2001) for phylogeny-associated beta diversity metrics. Also, PERMDISP (Anderson, 2006) was used to determine if dispersion contributed to a difference between the two filter groups. Differential abundances in taxonomic groups between filter groups were determined using analysis of composition of microbiomes (ANCOM) (Mandal et al., 2015) and linear discriminant analysis effect size (LEfSe) (Segata et al., 2011).

RESULTS AND DISCUSSION

Sequence Information

Illumina sequencing produced a total of 7,399,117, non-chimeric, paired-end reads after quality filtering, denoising, and merging with an average read of 52,850.8 ± 47,221.9 (SD), ranging from 6614 to 311,659 reads from 140 samples (70 samples of each of 0.22- and 5.0-µm pore size filters and a control), representing five major sources: nearshore lake, offshore lake, river, river mouth, and wetland outlet (**Figure 1** and **Table 1**). After filtering out read contaminants, alpha and beta diversity comparisons were made using data rarefied to 6400 reads with an average Good's coverage of 99.9% (range 99–100%).

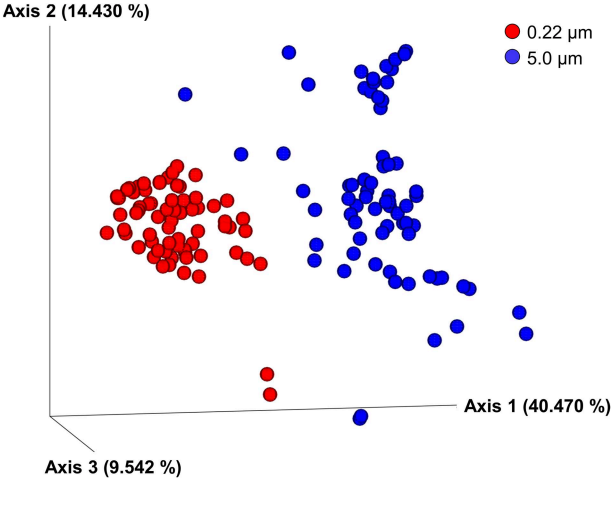


FIGURE 3 | Principal coordinate analysis (PCoA) of weighted Unifrac metrics of samples labeled according to filter pore size (i.e., 0.22- and 5.0-µm filters). Differences among bacterial communities were significant according to PERMANOVA and ANOSIM tests (q = 0.001).

Alpha Diversity

Rarefaction curves indicated saturation was reached for Shannon, Pielou's evenness, and Chao1 indices at read numbers used for statistical comparisons. Shannon index, a measure of richness and evenness, was statistically higher in the 5.0-µm (filter) fraction than the 0.22-µm fraction (H = 58.42, $q \leq 0.0001$;

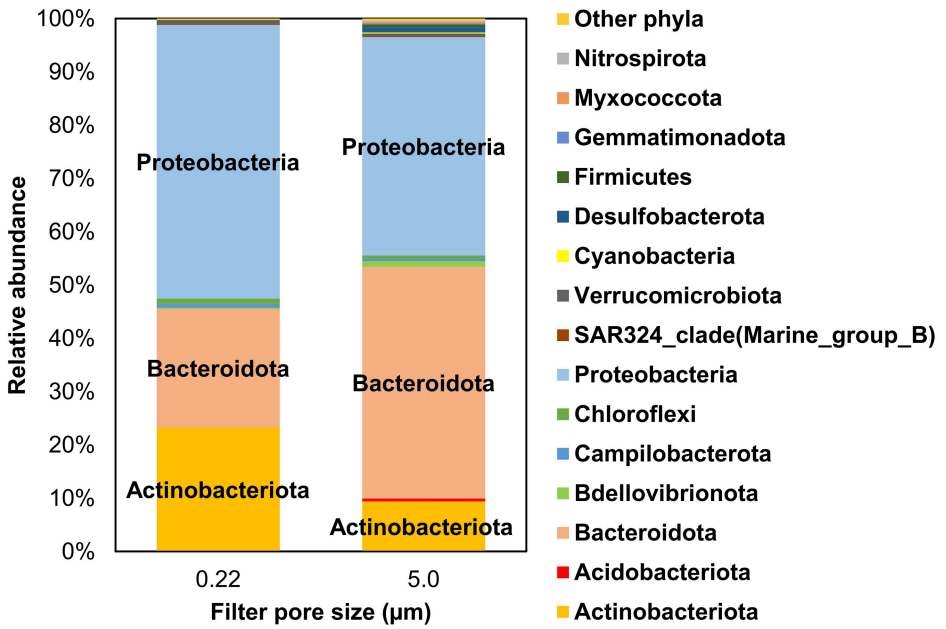


FIGURE 4 | Mean relative abundance of bacterial phyla identified in water samples from all sampling locations by filter pore size (i.e., 0.22- and 5.0-µm filters). Only phyla >0.1% by relative abundance are included in this illustration.

Figure 2A). Both richness and evenness differences contributed as indicated by differences in Pielou's community evenness index (H = 5.74, q = 0.0165; **Figure 2C**) and Chao1 richness estimate (H = 93.24, $q \leq 0.0001$; **Figure 2B**). These results show that, when multiple filters are used for processing water samples for microbial community analysis (e.g., pre- and final filters in the current study), relying on results from only one filter size (pre- or final filter) underestimates bacterial communities both qualitatively and quantitatively (Mestre et al., 2017). One of the main reasons for adding a prefiltration step in the current study was to increase the sample volume to 1,000 ml for the final filtration through the 0.22-µm pore size filter for 16S rRNA sequencing as reported elsewhere (Nakatsu et al., 2019; Kinzelman et al., 2020). Analysis of the 5.0-µm filters (current study) provided an opportunity to compare results between the two filter fractions and to characterize the observed differences at different taxonomic levels. Further, portioning the bacterial communities into free-living or attached to particulate matter was not an objective of this study; however, the higher community richness in the 5.0-µm filter fraction might be attributed to bacteria most likely associated with plankton and other particulate matter (e.g., soil, sediments) collected on the larger pore size filter (Padilla et al., 2015; Mestre et al., 2017; Borrego et al., 2020).

Beta Diversity

Clear differences between communities of the 5.0- and 0.22-µm fractions were seen in PCoA plots using all the beta diversity indices tested (results are shown only for weighted Unifrac; **Figure 3**). The main difference seen was in PCoA axis 1, accounting for 40.47% of the variation for weighted Unifrac; collectively, the three axes accounted for more than 64% of the variation. Both PERMANOVA and ANOSIM indicated these differences were significant for all metrics (q = 0.001). Despite the significant differences in PERMDISP, the PCoA plot shows that the communities do not overlap along PCoA1. The dispersion of communities was greater in communities from the 5.0-µm fraction compared with the 0.22-µm fractions along PCoA2, indicating greater variability of communities in the 5.0 µm fraction. A variety of environmental factors (e.g., temperature, season, and water stratifications) (Cram et al., 2015; Sunagawa et al., 2015; Cheng et al., 2020) are known to contribute to differences in observed beta diversity. Filter pore size is another factor that can affect the beta diversity, especially when the bacterial communities in water samples include free-living and particle-associated fractions; in this scenario, the particle-associated communities are likely to be retained on the larger pore size filter (i.e., 5.0 µm in the current study) (Padilla et al., 2015; Mestre et al., 2017) as previously discussed. Additionally, bacterial size, shape, and flexibility are known to affect their passage through membrane filters (Wang et al., 2008), resulting in differential distribution of communities between the two pore size filters used in this study.

Taxonomic Representation

Phyla

Overall, Actinobacteriota, Bacteroidota (formerly, Bacteroidetes), and Proteobacteria collectively represented 96.9% and 93.9% (by relative proportion of all phyla) in 0.22- and 5.0-µm pore size filters, respectively (**Figure 4**). Individually, comparing 0.22- and 5.0-µm pore size filters, Actinobacteriota were 23.3% ± 8.24% (mean and SD) vs.

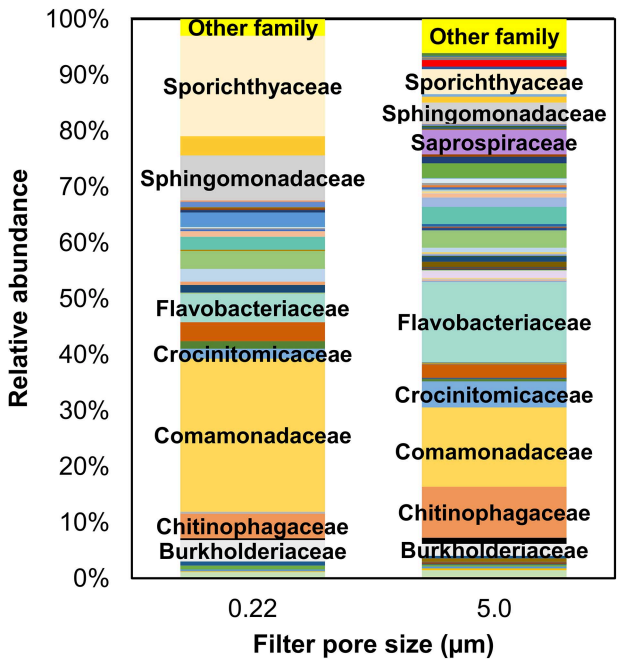


FIGURE 5 | Mean relative abundance of bacterial families identified in water samples from all sampling locations by filter pore size (i.e., 0.22- and 5.0-µm filters). Only families >0.1% by relative abundance are included in this illustration; the rest (<0.1% by relative abundance) are included under "Others."

9.4% ± 5.2%, Bacteroidota 22.2% ± 5.1% vs. 43.5% ± 11.5%, and Proteobacteria 51.3% ± 6.4% vs. 41.0% ± 8.9%. Although the dominance of these phyla is consistent with our recent studies in Lake Michigan (Nakatsu et al., 2019; Kinzelman et al., 2020) and other similar studies elsewhere (Newton et al., 2011; Mou et al., 2013), their differential relative abundances between the two filter fractions are noteworthy. Because these phyla are equally abundant in soil (Janssen, 2006) and aquatic vegetation (e.g., algae) (Miranda et al., 2013; Chun et al., 2017), their relative abundances in coastal waters may reflect the possibility of more than one source of these bacteria in interconnecting riparian (streams, rivers) and shoreline waters (**Figure 1**).

Other phyla with mean relative abundances ≥0.1% included Bdellovibrionota (previously Bdellovibrio), Campilobacterota, and Verrucomicrobiota (all 0.22 µm) and Acidobacteriota, Chloroflexi, Cyanobacteria, Desulfobacterota, and Firmicutes (5.0 µm), among others (**Figure 4**). In general, 5.0 µm filter fraction had more taxa than the 0.22 µm fraction. Cyanobacteria represented only a fraction of the sequences in both filter fractions (0.04% ± 0.05% and 0.25% ± 0.39% in 0.22 µm and 5.0 µm filters, respectively). Even though the two filter fractions were primarily dominated by Actinobacteriota, Bacteroidota, and Proteobacteria, albeit in different relative abundances, a differential distribution of several less abundant phyla was evident between the two filters. Notably, Cyanobacteria, Desulfobacterota, Firmicutes, Gemmatimonadota, Myxococcota, and Nitrospirota were mainly found in the 5.0-µm fraction, implying that analyzing both filters were required to capture the relative abundance of the measured (identified) taxa in this study.

Families

Like phyla, there were more families identified in the 5.0-µm pore size filter (368) than 0.22-µm (228). Only data with relative abundances >0.1% are graphically depicted in **Figure 5**; families with <0.1% in relative abundance are included under "other." Several families were common in both 0.22- and 5.0-µm filter fractions, albeit in varying proportions: *Burkholderiaceae* (3.8% vs. 2.1%), *Chitinophagaceae* (4.3% vs. 9.1%), *Comamonadaceae* (27.3% vs. 14.2%), *Crocinitomicaceae* (1.7% vs. 4.6%), *Flavobacteriaceae* (5.1% vs. 14.7%), *Sphingomonadaceae* (8.1% vs. 3.8%), and *Sporichthyaceae* (18.0% vs. 4.5%).

Interestingly, families comprising the traditional fecal indicators, such as *Enterobacteriaceae* (*E. coli*), *Enterococcaceae* (*Enterococcus* sp.), *Bacteroidaceae* (*Bacteroides* sp.), *Clostridiaceae* (*Clostridium*, especially *C. perfringens*), and *Lachnospiraceae* (Harwood et al., 2014), had sequence reads either too low (i.e., <0.1 by relative abundance)—*Bacteroidaceae, Enterobacteriaceae,* and *Lachnospiraceae*—or not detected in the samples (*Enterococcaceae* and *Clostridiaceae*) in both filter fractions. Thus, the 16S rRNA sequencing method might not be sensitive enough to detect these indicators in certain situations due to low sequence reads. However, any comparison of results related to the bacterial communities (noted above) between the two methods evaluated in this study (i.e., 16S rRNA gene sequencing and MST markers by qPCR) is tenuous because the data were not collected at the same time and qPCR detection is more sensitive than 16S rRNA sequencing.

Taxa Differences Between Filter Fractions

Linear discriminant analysis effect size analysis identified significant differences in taxa between the two filter fractions at all levels of taxonomic classification. There were 986 taxa with linear discriminant analysis (LDA) effect sizes greater than 2.0 (**Supplementary Table 1**). Those that could be classified to the genus level with the highest LDA effect sizes (>4.0, **Figure 6**) included *Limnohabitans*, *Sporichthyaceae* hgcl clade, *Sphigorhabdus*, *Pseudarcicella, Sediminibacterium, Rhodoferax*, and *Polynucleobacter* in the 0.22-µm fraction and *Flavobacterium, Fluviicola*, and Candidatus *Aquirestis* in the 5.0-µm fraction. Seven of these taxa were also significantly different using ANCOM. None of the taxa in either fraction represented bacteria of relevance to traditional indicators of water quality. In previous studies from the same study locations, there were several taxa that differed significantly among water sources (river, river mouth, stormwater wetland, lake near shore, and lake offshore) and collection sites (nearshore and river mouth); however, only 0.22-µm filters were used in those studies (Nakatsu et al., 2019; Kinzelman et al., 2020).

Collectively, noticeable differences were evident between the two pore size filters across all taxa (phyla, family, and genus) evaluated in this study. Such differences might be attributed to (a) cells and/or their DNA attaching to smaller particulate matter, such as clay minerals and organic fractions (Cai et al., 2006;

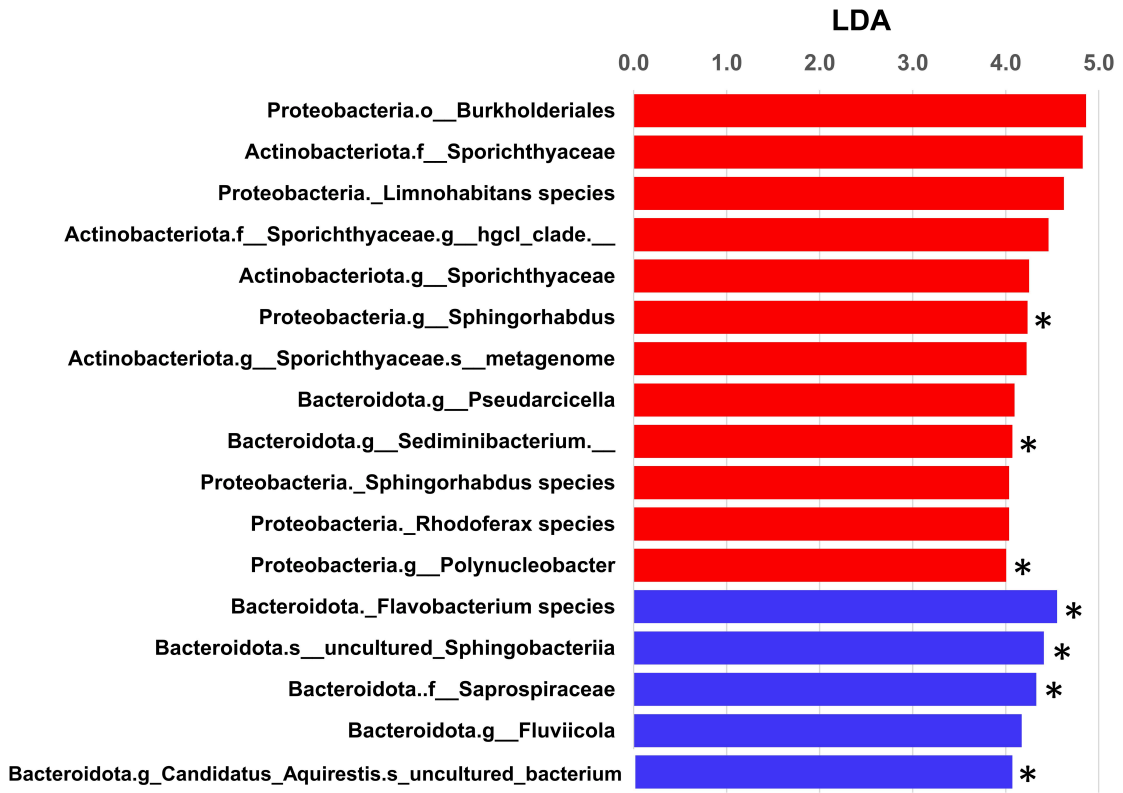


FIGURE 6 | Linear discriminate analysis effect size (LEfSe) analysis of bacterial communities from the two filter fractions analyzed (i.e., 0.22- and 5.0-μm filters). Taxa differed significantly at all levels of taxonomic classification. Only taxa that could be assigned to the genus level with LDA effect sizes >4.0 are shown in this graphical illustration (red and blue bars represent observed taxa in 0.22- and 5.0-μm filters, respectively; * denotes taxa that were significant by ANCOM). Overall, there were 986 taxa with LDA effect sizes greater than 2.0 (**Supplementary Table 1**).

TABLE 2 | Detection rates of Gull2 and HF183 host-specific MST markers in samples by filter pore size (0.22- and 5.0-μm) and water source [% detected includes samples within range of quantification (ROQ) and detected but not quantifiable (DNQ)].

Source	Percent detected (N)			
	Gull2		HF183	
	0.22 μm	5.0 μm	0.22 μm	5.0 μm
Overall	61 (87/143)	71 (102/143)	48 (69/143)	31 (45/143)
River	0 (0/26)	12 (3/26)	73 (19/26)	73 (19/26)
River mouth	39 (5/13)	69 (9/13)	77 (10/13)	46 (6/13)
Shoreline	79 (82/104)	87 (90/104)	39 (40/104)	19 (20/104)

Cuadros, 2018; Harrison et al., 2019; Sirois and Buckley, 2019), and passing through the larger pore size prefilter and (b) potentially unbound (i.e., free) bacterial cells primarily captured by the smaller pore size filters used (0.22 μm in the current study). Previous studies show that bacterial cells (e.g., *Escherichia coli, Hylemonella gracilis, Klebsiella pneumoniae, Pseudomonas aeruginosa*, and *Serratia marcescens*), can pass through pore size filters ranging from 0.1 to 0.45 μm (Hasegawa et al., 2003; Wang et al., 2007).

Host-Specific MST Markers

In general, the Gull2 marker was found in higher frequency in the 5.0-μm filter than the 0.22-μm filter (71% vs. 61%); however, the pattern was opposite for the HF183 marker: 31% vs. 48% (**Table 2**). These two markers were found in all water types, except for the upstream river samples from 0.22-μm filters (0/26). A low incidence of the Gull2 marker (12%, 3/26; 5.0-μm filters) indicates that contamination of rivers from gull feces is minimal or sporadic—most likely attributed to urban runoff from parking lots or other sources adjacent to streams (Green et al., 2019). Conversely, a relatively higher frequency of this marker in the river mouth samples (in both 0.22- and 5.0-μm filters from RRM), suggests higher activity of shoreline birds at public parks or recreational areas adjacent to beaches (Whitman and Nevers, 2003; Byappanahalli et al., 2015).

The two MST markers evaluated in this study represent two different fecal sources of distinct origin, namely wastewater (HF183) and environmental or wildlife (Gull2). Both markers were detected in both 0.22- and 5.0-μm filters but at different frequencies. One explanation might be that the bacteria (*C. marimammalium*) and/or DNA from gull feces is localized; likely associated with fecal droppings and/or deposited into the sand/sediment matrix underneath because a significantly higher incidence of the Gull2 marker was previously detected

in sand and sediment than water at the same study locations (Nevers et al., 2020). Thus, particle-bound cells/DNA from gull feces are likely to get trapped (in the larger pore size prefilter), and the unbound cells/DNA pass through the prefilter. The HF183 marker (from Bacteroidota), on the other hand, primarily comes from human fecal sources (e.g., treated sewage), which undergoes rigorous processing (e.g., screening, agitation or churning, and sedimentation) in wastewater treatment plants, leaving a portion of cells and/or their DNA unbound or attached to much smaller particulate matter such as clays and colloidal particles (Cai et al., 2006; Cuadros, 2018). Further, fecal contamination, originating from failed infrastructure or septic fields, which contributes to overall abundance (HF183 and other Bacteroidetes markers), more likely to be dispersed in the receiving water bodies vs. concentrated cells/DNA on particulate matter (e.g., MST markers in sand in the case of direct deposition from gulls) (Sauer et al., 2011; Nevers et al., 2020). Nonetheless, additional research is needed to better understand the reasons for differential detection of host-specific MST markers between pore size filters, including those not tested in this study as well as exploring other combinations of filter sizes for capturing capacities. In addition, future research should target samples from the same events for multiple analyses because any comparison of results, such as 16S rRNA gene sequencing and MST markers (current study), is tenuous.

SUMMARY AND CONCLUSION

DNA-based methods are increasingly applied in a wide range of environmental programs: from monitoring shoreline water quality to the assessment of biotic communities in aquatic systems. Membrane filtration remains a foundational step in the processing of water *via* a wide range of molecular assays employed for the identification of microbial or species-specific DNA. A prefiltration step is often included to overcome clogging and to ensure that sufficient cells/DNA are collected for the measured endpoints; however, there is little consistency in processing the pre- and final filters. In the current study, we analyzed the pre- and final filters independently to evaluate their influence on composition and abundance of bacterial communities and two MST markers routinely used in water quality monitoring programs.

Collectively, our findings show that (a) there were both qualitative and quantitative differences in bacterial communities between the two filter sizes with significant differences in number of taxa at all levels of taxonomic representation and (b) the two MST markers, Gull2 and HF183, showed different detection rates between the two filter sizes. Such differential detection could potentially result in an inaccurate or underrepresented pollution source profile when selectively analyzing a single pore size filter (either pre- or the final filter).

In summary, analyzing both pre- and final filters increases our confidence in the results. The molecular target and environmental substrate from which DNA is extracted play a role in the successful recovery and estimation of relative abundance. Targets extracted from hard substrates, such as sand, or highly turbid waters may be more difficult to recover due their adherence to particulate matter (e.g., sediments, clays), which can foul filters employed in the extraction process. When conducting investigative studies, it is important that the full complement of organisms be represented to render the greatest insight. Although analysis of multiple filters may increase costs, it provides more complete genomic data *via* increased sample volume for characterizing microbial communities in coastal waters.

DATA AVAILABILITY STATEMENT

Nucleotide sequences related to the results from 0.22- and 0.5-mm pore size filters are available in the National Center for Biotechnology Information-Sequence Read Archive (NCBISRA) database under study PRJNA725142 accession numbers SAMN18870789–SAMN18870931 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA725142). Data generated during this study are available as a USGS data release: Byappanahalli et al. (2021).

AUTHOR CONTRIBUTIONS

MB and MN contributed to conception and design of the study, and acquired funding for this research. MB, MN, DS, JK, and CN designed the experiments and conducted the research. CN and DS performed the statistical analyses. MB provided supervision and wrote the first draft. CN, MN, DS, JK, and MP reviewed the first draft. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was funded by the United States Environmental Protection Agency and Great Lakes Restoration Initiative. CN (Purdue University) received partial support from USDA National Institute of Food and Agriculture Hatch Grant IND010825.

ACKNOWLEDGMENTS

We thank Ashley Spoljaric, Kasia Kelly, Jennifer Creekmur, and Stephan Kurdas for their assistance with sample collection and analysis. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The Genomics core facility at Purdue University performed sequencing and the Purdue Bioinformatics core kindly provided servers to run the bioinformatics analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.665664/full#supplementary-material

Supplementary Figure 1 | Volcano plot of ANCOM results, *y*-axis *W*-statistic, and *x*-axis centered log ratio (clr).

Supplementary Table 1 | LEfSe results showing taxa differences between 0.22- and 5.0-μm pore size filters at all taxa levels with LDA > 2.0.

REFERENCES

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253. doi: 10.1111/j.1541-0420.2005.00440.x

Bej, A. K., Steffan, R. J., Dicesare, J. L., Haff, L., and Atlas, R. M. (1990). Detection of coliform bacteria in water by using PCR and gene probes. *Appl. Environ. Microbiol.* 56, 307–314.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.

Borrego, C., Sabater, S., and Proia, L. (2020). Lifestyle preferences drive the structure and diversity of bacterial and archaeal communities in a small riverine reservoir. *Sci. Rep*. 10:11288.

Brinkman, N. E., Haugland, R. A., Wymer, L. J., Byappanahalli, M., Whitman, R. L., and Vesper, S. J. (2003). Evaluation of a rapid, quantitative real-time PCR method for enumeration of pathogenic Candida cells in water. *Appl. Environ. Microbiol.* 69, 1775–1782. doi: 10.1128/aem.69.3.1775-1782.2003

Byappanahalli, M., Nevers, M. B., and Shively, D. A. (2021). Influence of filter pore size on microbial communities and microbial source tracking (MST) markers on water in Racine, Wisconsin; Chicago, Illinois; East Chicago, Indiana, 2015–2017. *Reston, VA: U.S. Geological Survey.* doi: 10.5066/P9UK7G0Z

Byappanahalli, M. N., Nevers, M. B., Whitman, R. L., Ge, Z., Shively, D., Spoljaric, A., et al. (2015). Wildlife, urban inputs, and landscape configuration are responsible for degraded swimming water quality at an embayed beach. *J. Great Lakes Res.* 41, 156–163. doi: 10.1016/j.jglr.2014.11.027

Cai, P., Huang, Q., Zhang, X., and Chen, H. (2006). Adsorption of DNA on clay minerals and various colloidal particles from an Alfisol. *Soil Biol. Biochem.* 38, 471–476. doi: 10.1016/j.soilbio.2005.05.019

Callahan, B. J., Mcmurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cheng, W.-H., Lu, H.-P., Chen, C.-C., Jan, S., and Hsieh, C.-H. (2020). Vertical beta-diversity of bacterial communities depending on water stratification. *Front. Microbiol.* 11:449. doi: 10.3389/fmicb.2020.00449

Chun, C. L., Peller, J. R., Shively, D., Byappanahalli, M. N., Whitman, R. L., Staley, C., et al. (2017). Virulence and biodegradation potential of dynamic microbial communities associated with decaying Cladophora in Great Lakes. *Sci. Total Environ.* 574, 872–880. doi: 10.1016/j.scitotenv.2016.09.107

Cram, J. A., Chow, C.-E. T., Sachdeva, R., Needham, D. M., Parada, A. E., Steele, J. A., et al. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* 9, 563–580. doi: 10.1038/ismej.2014.153

Cuadros, J. (2018). Clay minerals interaction with microorganisms: a review. *Clay Miner.* 52, 235–261. doi: 10.1180/claymin.2017.052.2.05

Francy, D. S., Stelzer, E. A., Brady, A. M. G., Huitger, C., Bushon, R. N., Ip, H. S., et al. (2013). Comparison of filters for concentrating microbial indicators and pathogens in lake water samples. *Appl. Environ. Microbiol.* 79, 1342–1352. doi: 10.1128/aem.03117-12

Fuhrman, J. A., Comeau, D. E., Hagström, A., and Chan, A. M. (1988). Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Appl. Environ. Microbiol.* 54, 1426–1429. doi: 10.1128/aem.54.6.1426-1429.1988

Gloor, G. B., Hummelen, R., Macklaim, J. M., Dickson, R. J., Fernandes, A. D., Macphee, R., et al. (2010). Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* 5:e15406. doi: 10.1371/journal.pone.0015406

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., et al. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* 7, 1299–1307. doi: 10.1111/2041-210x.12595

Green, H., Weller, D., 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., 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

Harrison, J. B., Sunday, J. M., and Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. R. Soc.* 286, 20191409–20191409. doi: 10.1098/rspb.2019.1409

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 to pathogens and human health outcomes. *FEMS Microbiol. Rev.* 38, 1–40. doi: 10.1111/1574-6976.12031

Hasegawa, H., Naganuma, K., Nakagawa, Y., and Matsuyama, T. (2003). Membrane filter (pore size, 0.22-0.45 micro m; thickness, 150 micro m) passing-through activity of *Pseudomonas aeruginosa* and other bacterial species with indigenous infiltration ability. *FEMS Microbiol. Lett.* 223, 41–46.

Jain, A., and Krishnan, K. P. (2017). Differences in free-living and particle-associated bacterial communities and their spatial variation in Kongsfjorden, Arctic. *J. Basic Microbiol.* 57, 827–838. doi: 10.1002/jobm.201700216

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi: 10.1128/aem.72.3.1719-1728.2006

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kinzelman, J., Byappanahalli, M. N., Nevers, M. B., Shively, D., Kurdas, S., and Nakatsu, C. (2020). Utilization of multiple microbial tools to evaluate efficacy of restoration strategies to improve recreational water quality at a Lake Michigan Beach (Racine. WI). *J. Microbiol. Methods* 178:106049. doi: 10.1016/j.mimet.2020.106049

Lacoursière-Roussel, A., Côté, G., Leclerc, V., and Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *J. Appl. Ecol.* 53, 1148–1157. doi: 10.1111/1365-2664.12598

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133

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

Mahbubani, M. H., Bej, A. K., Perlin, M., Schafer Iii, F. W., Jakubowski, W., and Atlas, R. M. (1991). Detection of Giardia cysts by using the polymerase chain reaction and distinguishing live from dead cysts. *Appl. Environ. Microbiol.* 57, 3456–3461. doi: 10.1128/aem.57.12.3456-3461.1991

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* 26:27663.

McLellan, S. L., Huse, S. M., Mueller-Spitz, S. R., Andreishcheva, E. N., and Sogin, M. L. (2009). Diversity and population structure of sewage-derived microrganisms in wastewater treatment influent. *Environ. Microbiol.* 12, 378–392. doi: 10.1111/j.1462-2920.2009.02075.x

Mestre, M., Borrull, E., Sala, M. M., and Gasol, J. M. (2017). Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *ISME J.* 11, 999–1010. doi: 10.1038/ismej.2016.166

Miranda, L. N., Hutchison, K., Grossman, A. R., and Brawley, S. H. (2013). Diversity and abundance of the bacterial community of the red macroalga porphyra umbilicalis: did bacterial farmers produce macroalgae? *PLoS One* 8:e58269. doi: 10.1371/journal.pone.0058269

Mou, X., Jacob, J., Lu, X., Robbins, S., Sun, S., and Ortiz, J. D. (2013). Diversity and distribution of free-living and particle-associated bacterioplankton in Sandusky Bay and adjacent waters of Lake Erie Western Basin. *J. Great Lakes Res.* 39, 352–357. doi: 10.1016/j.jglr.2013.03.014

Mull, B., and Hill, V. R. (2012). Recovery of diverse microbes in high turbidity surface water samples using dead-end ultrafiltration. *J. Microbiol. Methods* 91, 429–433. doi: 10.1016/j.mimet.2012.10.001

Nakatsu, C. H., Byappanahalli, M. N., and Nevers, M. B. (2019). Bacterial community 16S rRNA gene sequencing characterizes riverine microbial impact on Lake Michigan. *Front. Microbiol.* 10:996. doi: 10.3389/fmicb.2019.00996

Nevers, M. B., Byappanahalli, M. N., Nakatsu, C. H., Kinzelman, J. L., Phanikumar, M. S., Shively, D. A., et al. (2020). Interaction of bacterial communities and indicators of water quality in shoreline sand, sediment, and water of Lake Michigan. *Water Res.* 178, 115671. doi: 10.1016/j.watres.2020.115671

Nevers, M. B., Byappanahalli, M. N., Shively, D., Buszka, P. M., Jackson, P. R., and Phanikumar, M. S. (2018). Identifying and eliminating sources of recreational water quality degradation along an Urban Coast. *J. Environ. Qual.* 47, 1042–1050. doi: 10.2134/jeq2017.11.0461

Newton, R., Bootsma, M., Morrison, H., Sogin, M., and Mclellan, S. (2013). A Microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microb. Ecol.* 65, 1011–1023. doi: 10.1007/s00248-013-0200-9

Newton, R. J., Jones, S. E., Eiler, A., Mcmahon, K. D., and Bertilsson, S. (2011). A Guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* 75, 14–49. doi: 10.1128/mmbr.00028-10

Padilla, C. C., Ganesh, S., Gantt, S., Huhman, A., Parris, D. J., Sarode, N., et al. (2015). Standard filtration practices may significantly distort planktonic microbial diversity estimates. *Front. Microbiol.* 6:547. doi: 10.3389/fmicb.2015.00547

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

Sauer, E. P., Vandewalle, J. L., Bootsma, M. J., and Mclellan, S. L. (2011). Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. *Water Res.* 45, 4081–4091. doi: 10.1016/j.watres.2011.04.049

Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R., and Lukasik, J. (2002). Microbial source tracking: current methodology and future directions. *Appl. Environ. Microbiol.* 68, 5796–5803. doi: 10.1128/aem.68.12.5796-5803.2002

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60.

Sirois, S. H., and Buckley, D. H. (2019). Factors governing extracellular DNA degradation dynamics in soil. *Environ. Microbiol. Rep.* 11, 173–184. doi: 10.1111/1758-2229.12725

Staley, C., and Sadowsky, M. J. (2016). Application of metagenomics to assess microbial communities in water and other environmental matrices. *J. Mar. Biol.* 96, 121–129. doi: 10.1017/s0025315415001496

Staley, C., Unno, T., Gould, T. J., Jarvis, B., Phillips, J., Cotner, J. B., et al. (2013). Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J. Appl. Microbiol.* 115, 1147–1158. doi: 10.1111/jam.12323

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi: 10.1126/science.1093857

Wade, T. J., Calderon, R. L., Sams, E., Beach, M., Brenner, K. P., Williams, A. H., et al. (2006). Rapidly measured indicators of recreational water quality are predictive of swimming associated gastrointestinal illness. *Environ. Health Perspect.* 114, 24–28. doi: 10.1289/ehp.8273

Walsh, D. A., Zaikova, E., and Hallam, S. J. (2009). Large volume (20L+) filtration of coastal seawater samples. *J. Vis. Exp.* 28:1161.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm oore size filters and shape-dependent enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi: 10.1021/es0707198

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* 42, 6749–6754. doi: 10.1021/es800720n

Whitman, R. L., and Nevers, M. B. (2003). Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. *Appl. Environ. Microbiol.* 69, 5555–5562. doi: 10.1128/aem.69.9.5555-5562.2003

Xu, H., and Guo, L. (2017). Molecular size-dependent abundance and composition of dissolved organic matter in river, lake and sea waters. *Water Res.* 117, 115–126. doi: 10.1016/j.watres.2017.04.006

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648.

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.

At least a portion of this work is authored by Mureuleedhara Byappanahalli, Meredith B. Nevers, and Dawn Shively on behalf of the U.S Government and as regards, Dr. Byappanahalli, Dr. Nevers, and Dr. Shively and the U.S Government, is not subject to copyright protection in the United States. Foreign and other copyrights may apply. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 August 2021
doi: 10.3389/fmicb.2021.674214

A Combined Digital PCR and Next Generation DNA-Sequencing Based Approach for Tracking Nearshore Pollutant Dynamics Along the Southwest United States/Mexico Border

Amity G. Zimmer-Faust[1], Joshua A. Steele[1], Xianyi Xiong[2], Christopher Staley[2], Madison Griffith[1], Michael J. Sadowsky[3], Margarita Diaz[4] and John F. Griffith[1]*

[1] Southern California Coastal Water Research Project, Costa Mesa, CA, United States, [2] BioTechnology Institute, University of Minnesota Twin Cities, Saint Paul, MN, United States, [3] Department of Soil, Water, and Climate, University of Minnesota Twin Cities, Saint Paul, MN, United States, [4] Proyecto Fronterizo de Educación Ambiental, A.C., Tijuana, Mexico

OPEN ACCESS

Edited by:
Michèle Gourmelon, Laboratoire Santé, Environnement et Microbiologie, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Reviewed by:
Timothy Joe Wade, United States Environmental Protection Agency (EPA), United States
Ye Deng, Research Center for Eco-Environmental Sciences (CAS), China
Jeffrey Wayne Turner, Texas A&M University-Corpus Christi, United States

***Correspondence:**
Amity G. Zimmer-Faust
amyzf@sccwrp.org

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *28 February 2021*
Accepted: *25 May 2021*
Published: *06 August 2021*

Citation:
Zimmer-Faust AG, Steele JA, Xiong X, Staley C, Griffith M, Sadowsky MJ, Diaz M and Griffith JF (2021) A Combined Digital PCR and Next Generation DNA-Sequencing Based Approach for Tracking Nearshore Pollutant Dynamics Along the Southwest United States/Mexico Border. Front. Microbiol. 12:674214. doi: 10.3389/fmicb.2021.674214

Ocean currents, multiple fecal bacteria input sources, and jurisdictional boundaries can complicate pollution source tracking and associated mitigation and management efforts within the nearshore coastal environment. In this study, multiple microbial source tracking tools were employed to characterize the impact and reach of an ocean wastewater treatment facility discharge in Mexico northward along the coast and across the Southwest United States- Mexico Border. Water samples were evaluated for fecal indicator bacteria (FIB), *Enterococcus* by culture-based methods, and human-associated genetic marker (HF183) and *Enterococcus* by droplet digital polymerase chain reaction (ddPCR). In addition, 16S rRNA gene sequence analysis was performed and the SourceTracker algorithm was used to characterize the bacterial community of the wastewater treatment plume and its contribution to beach waters. Sampling dates were chosen based on ocean conditions associated with northern currents. Evidence of a gradient in human fecal pollution that extended north from the wastewater discharge across the United States/Mexico border from the point source was observed using human-associated genetic markers and microbial community analysis. The spatial extent of fecal contamination observed was largely dependent on swell and ocean conditions. These findings demonstrate the utility of a combination of molecular tools for understanding and tracking specific pollutant sources in dynamic coastal water environments.

Keywords: microbial source tracking, 16S ribosomal DNA analysis, wastewater, droplet digital PCR, coastal water

INTRODUCTION

Once bacterial and chemical contaminants enter the nearshore coastal environment, complicated mixing, dilution, and transport processes make it increasingly difficult to identify their origin. Moreover, jurisdictional boundaries, including international borders, can further complicate investigations aimed at identifying specific sources of contamination along coastlines. Despite this,

fecal pollution represents a leading cause of water quality impairments in coastal waters worldwide, and it is critical to identify its origin for successful management and mitigation of associated public health and economic consequences (DeFlorio-Barker et al., 2018).

In recent years, microbial source tracking (MST) methods that differentiate amongst different specific animal and human fecal sources have been developed and their application has become more widespread. Non-human sources of FIB do not carry the same pathogenic load compared to human point sources such as sewage (Soller et al., 2010, 2014; Schoen et al., 2011), underscoring the need to understand where fecal contamination is coming from. Among the different MST-based approaches, the marker gene approach, which relies on measurement of host source-associated DNA sequences by polymerase chain reaction (PCR)-based technologies, is among the most utilized (Harwood et al., 2014). Source-associated markers have been developed for common fecal sources including cow (Shanks et al., 2006, 2008; Kildare et al., 2007), dog (Green et al., 2014b), bird (Shanks et al., 2008; Lee et al., 2013), and human (Seurinck et al., 2005; Green et al., 2014a). Recent advances in digital PCR provide enhanced sensitivity and robustness to inhibitory substances (Cao et al., 2015), with select quantitative PCR (qPCR)-based MST assays readily adapted to digital PCR (Cao et al., 2015; Coudray-Meunier et al., 2015; Staley Z. R. et al., 2018). The MST marker gene approach has been used previously to effectively identify and differentiate between specific fecal sources in a variety of matrices, including estuarine (Riedel et al., 2015), fresh waters (Li et al., 2019), marine waters (Ervin et al., 2013), and sediments (Zimmer-Faust et al., 2017), resulting in recommendations for application of best management practices, specific infrastructure improvements, and regulatory support (USEPA, 2011; Verhougstraete et al., 2015; Goodwin et al., 2017).

More recently, technological advances in next generation sequencing (NGS) technology have led to source identification based on the comparative characterization of the entire microbial communities of environmental samples and pollution sources (Ahmed et al., 2015; Roguet et al., 2018; Staley C. et al., 2018), providing an additional avenue for identifying the origins of fecal contamination. These methods differ from the single marker MST-based approaches by their ability to characterize thousands of sequences in each sample (Sogin et al., 2006), in theory making them able to characterize any relevant source. Computational tools utilizing sequencing data that allow for community-based microbial source tracking have been developed and include the SourceTracker program (Mathai et al., 2020; Roguet et al., 2020). The SourceTracker program uses a Bayesian algorithm to estimate the proportion of each source in a set of samples (Knights et al., 2013). Previous studies have successfully applied the SourceTracker program to characterize bacterial contamination sources in recreational fresh (Baral et al., 2018) and estuarine (McCarthy et al., 2017) waters, with success of these methods reliant on microbial community profiles that are unique to each source (Brown et al., 2017; Staley C. et al., 2018). However, application of these MST tools for tracking specific pollutant sources within the nearshore coastal environment has not been explored fully, with previous efforts primarily focused on identifying the source of pollution to impacted water and ignoring the geographic reach of contamination. McCarthy et al. (2017) used microbial community-based source tracking to validate a 3-dimensional estuarine hydrodynamic model, finding this approach successful at identifying the primary water sources contributing to an urban estuary. However, in the nearshore coastal environment, mixing is further complicated by surf zone processes that can impact pollutant transport significantly.

In this study, the extent to which effluent discharged by a wastewater treatment plant in Tijuana, Mexico is transported within the nearshore coastal environment was examined. It has been hypothesized that when ocean currents are flowing south-to-north, wastewater from the outfall of the San Antonio de los Buenos (SADB) wastewater treatment plant (WTP) at Punta Bandera in Tijuana, Mexico contributes to elevated bacterial levels and introduces human fecal contamination to nearshore waters from its origin, across the United States/Mexico border, and as far north as the City of San Diego, CA, United States (Orozco-Borbón et al., 2006; Sassoubre et al., 2012; Thulsiraj et al., 2017). Previous spatial modeling efforts have suggested that south-to-north coastal currents have the potential to transport discharges from the SADB plant and impact water quality in southern San Diego County (CH2M Hill, 2009; Feddersen et al., 2020). However, the extent and impact of the plume has never been verified, in part due to logistical challenges associated with sampling across an international border. In this study, a combination of MST technologies consisting of human-associated marker genes and NGS was used to detect, quantify, and track human fecal contamination emanating from the outfall of the SABD WTP during specific south-to-north swell conditions from its origin in Tijuana, Mexico to San Diego, CA, United States.

MATERIALS AND METHODS

Study Site

The Southwest United States border region has a long history of poor water quality, with elevated levels of fecal indicator bacteria (FIB) leading to frequent beach postings and closures. At popular surf beaches near the United States/Mexico border in San Diego, CA, United States, sources of FIB causing poor beach water quality are unknown and of concern to public health officials, given the proximity of multiple potential inputs of human fecal contamination, which include the outfall of the SADB WTP in Tijuana, Mexico (Kim et al., 2009). The SADB WTP has capacity to treat up to 35 million gallons per day (MGD) and receives a mixture of wastewater from the cities of Tijuana and Rosarito, Mexico (USEPA, 2014). The plant is located approximately 7.7 km south of the United States/Mexico border and discharges minimally treated effluent to a coastal stream that terminates at the ocean at Punta Bandera, approximately 5 km south of the treatment plant and 11 km south of the United States/Mexico border (**Figure 1**).

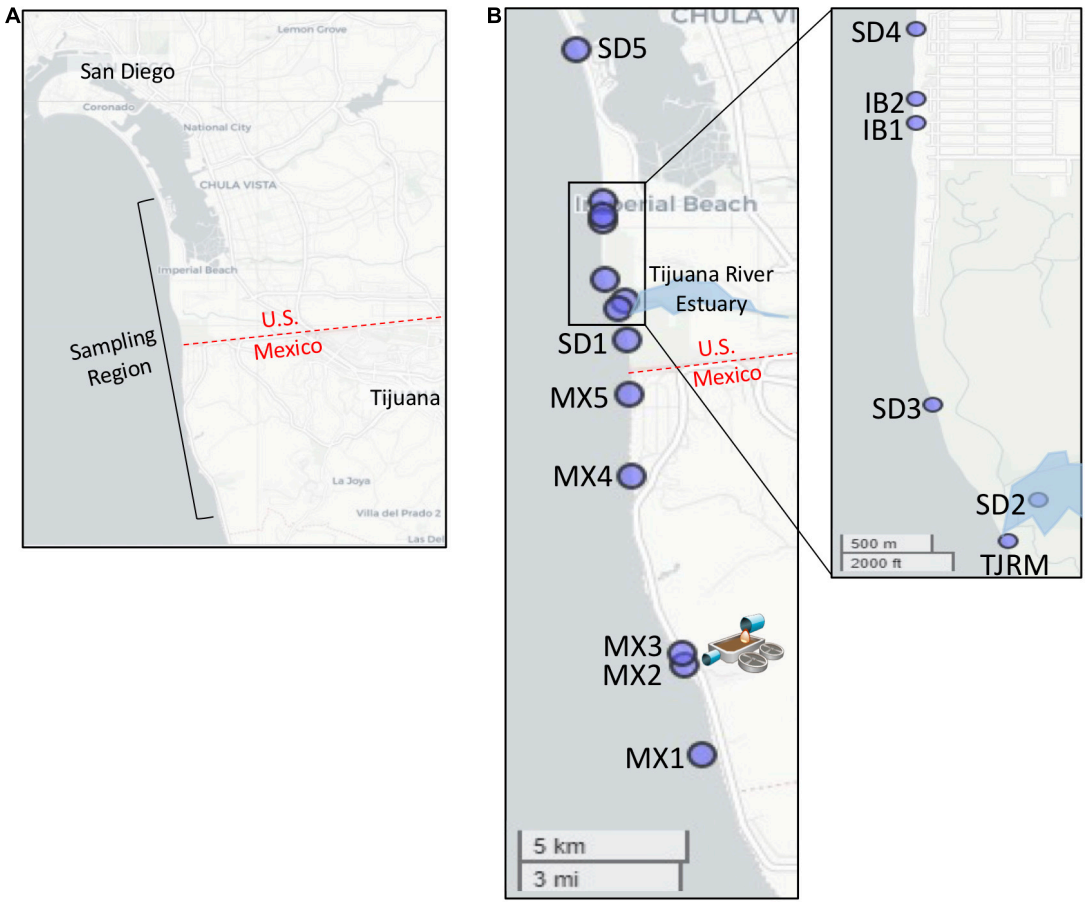


FIGURE 1 | Map of sampling locations. **(A)** Sampling region targeted. **(B)** Location of specific sites sampled. Map of Imperial Beach region inset. SAB WTP outfall located at site MX2.

During events when swells are coming from the south, the SADB source generally travels at 8–14 km/day (Feddersen et al., 2020), suggesting that the plume can reach the San Diego, CA, United States beaches targeted within this study within a 24–48 h period.

A second potential source of human fecal contamination in the region is the Tijuana River, which terminates into the Tijuana River estuary. The Tijuana River is 193 km long, flowing north through Mexico, before entering the United States. During dry weather, flows in the Tijuana River are largely intercepted by diversion structures in Mexico and pumped to either the SADB WTP or the South Bay International WTP before they cross into the United States. At times, however, pumps in the diversion structures can become clogged or malfunction, allowing potentially contaminated water to flow across the border into the United States (USEPA, 2014).

In the present study, sampling was designed to characterize impacts specifically from the SADB WTP. An approximately 23 km stretch of coastline was targeted, extending from just south of the SADB WTP discharge at Punta Bandera, Mexico to Silver Strand State Beach, CA, United States. Sampling took place during dry weather, to reduce the expected impact from the Tijuana River and other runoff-related sources. South swells dominate the coastal ocean in this region during the dry season and specific south swell events were targeted for sampling to increase the likelihood of capturing a gradient of human fecal pollution extending north from the SADB WTP outfall.

Sample Collection

Sites were sampled during dry weather along the coastline from just south of the SADB WTP outfall at Punta Bandera to Silver Strand State Beach, CA, United States (**Figure 1** and **Table 1**). Four dry weather, south swell events were targeted for sampling between October 2018 and October 2019. Two to three consecutive days were sampled per south swell event (**Supplementary Table 1**).

Grab water samples (2 L) were collected from each site. At all beach sites, ankle-to-knee deep water (swash zone) samples were taken on each sampling date. At site SD2, located just inside the mouth of the Tijuana River Estuary, water was collected with a pole sampler just below the water surface.

TABLE 1 | Description of sampling locations.

Site ID	Site	Latitude	Longitude	Distance to SADB WTP outfall	Description
MX1	SADB South	32.43786	−117.1019	1.2 km (South)	San Antonio del Mar, South of SADB WTP outfall; Ocean Site
MX2	SADB WTP Discharge	32.44804	−117.1053		SADB WTP outfall at Punta Bandera
MX3	SADB North	32.447652	−117.1086	0.3 km	North of SADB WTP (mixing zone); Ocean Site
MX4	El Vigia	32.502769	−117.1236	6.3 km	Southern end of Playas; Ocean site
MX5	Parque Mexico	32.527417	−117.1245	9.0 km	Parque Mexico; Ocean Site
SD1	Border field	32.5434	−117.125	10.8 km	Border Field State Park; Ocean site
SD2	TJ Slough	32.55299	−117.128	12.6 km	Tijuana River Estuary; Estuary site
TJRM	TJ River Mouth	32.55235	−117.127	11.7 km	Tijuana River Mouth; mixing zone
SD3	TJ Slough beach	32.561	−117.132	12.9 km	Tijuana Slough National Wildlife Refuge; Ocean site
IB1	Elder Ave	32.5788	−117.133	14.8 km	Imperial Beach; Elder Ave, Ocean site
IB2	Elm Ave	32.5803	−117.133	15.0 km	Imperial Beach Pier; Elm Ave, Ocean Site
SD4	Carnation	32.5847	−117.133	15.48 km	Imperial Beach municipal beach; Carnation Ave, Ocean Site
SD5	Silver Strand	32.6296	−117.141	20.4 km	Silver Strand State Beach; Ocean Site

In addition, paired surf zone samples (approximately 100 m offshore) were also collected from personal watercraft at all beach sites north of the United States/Mexico border (n = 7) during two of the four south swell events (Event 3 and Event 4). Surf zone samples were collected on all 3 days during Event 3 and during the first 2 days of Event 4 (**Supplementary Table 1**). Grab water samples were also collected from the two potential sources: the SADB WTP outfall at Punta Bandera (site MX2) and upstream in the Tijuana River, before it crosses into the United States. Samples were collected from the SADB WTP outfall on each sampling date to characterize effluent quality. Due to logistical challenges associated with access to the Tijuana River, upstream of the United States/Mexico border and the diversion structures, four grab samples total were collected on two sampling dates (August 1, 2019 and August 9, 2019). These samples were collected to characterize potential inputs to the Tijuana River that are theoretically diverted during dry weather.

All environmental sites were sampled in the morning to limit degradation of the bacterial signal. Water samples were stored on ice and transported to either laboratories at the Southern California Coastal Water Research Project (SCCWRP) in Costa Mesa, CA, United States (if collected north of the United States/Mexico border) or to laboratories at the Proyecto Fronterizo de Educación Ambiental, A.C. in Tijuana, Mexico (if collected south of the United States/Mexico border) for sample processing, which included analyzing for culturable enterococci and filtering for bacterial DNA. Filters were then transported to SCCWRP laboratories for analysis by digital PCR.

The four sampling events differed in swell direction and magnitude and tidal and wind conditions. Conditions per event are summarized in more detail in the **Supplementary Material** (**Supplementary Figures 1**–**4**). During the first sampling event (Event 1), samples were collected on two consecutive days with swell direction primarily from the southwest (**Supplementary Figure 1**). During events 2 and 3, similar conditions were sampled- the beginning of a south swell event. Samples were collected on three consecutive days with swell direction primarily from the west on Day 1 and from the south on Days 2 and 3 (**Supplementary Figures 2**, **3**). During event 4, the end of a south swell was targeted. Samples were collected on three consecutive days, with swell direction switching to more mixed by day 3 (**Supplementary Figure 4**).

Sample Processing

FIB Cultivation

Cultivable *Enterococcus* was quantified by using Enterolert in conjunction with the Quanti-Tray 2000^{TM} system (IDEXX, Westbrook, ME, United States), as per the manufacturer's instructions. Quantification was done using three dilutions covering a 100,000-fold range of concentration. Field and equipment blanks were collected and tested for FIB contamination in the same manner as regular samples. Laboratory blanks were performed using sterile phosphate buffered saline solution.

Filtration for Bacterial DNA

Fifty to 250 mL aliquots of sample water were filtered through 47 mm, 0.4 μm pore size, HTTP polycarbonate filters (EMD Millipore, Billerica, MA, United States). Each filter was placed in an individual 2 mL polypropylene screw cap tube, containing 0.7 mL dry volume of ZR BashingBead lysis matrix high density beads and 1 mL DNA/RNA Shield solution (Zymo Research, Costa Mesa, CA, United States). Bead tubes were stored at 4°C until transport to SCCWRP, and then stored at −80°C until processed. Filter blanks, consisting of sterile phosphate buffered saline passed through the polycarbonate filter, were also generated with each set of processed samples.

Microbial Source Tracking Marker Analysis

DNA was recovered from filters using the ZymoBiomics DNA Kit (Zymo Research, Costa Mesa, CA, United States), according to manufacturer's guidelines. Extracted DNA was eluted into 100 μL of buffer and aliquots were stored at −20°C until analyzed with droplet digital PCR (ddPCR) on the Bio-Rad QX200 platform

(Bio-Rad, Hercules, CA, United States). DNA from a halophilic, alkaliphilic archaeon (*Natronomonas pharaonis*) was added to the lysis buffer prior to extraction as an external extraction and inhibition control following USEPA method 1611 guidelines (USEPA, 2012). Negative extraction controls (NEC) containing only lysis buffer and halophile DNA were processed alongside each set of extracted samples.

Human-associated genetic markers and *Enterococcus* were quantified using digital PCR. Primer and probe sequences were those from published methods for detection of the HF183 (Cao et al., 2015) and Lachno3 (Feng et al., 2018) human-associated markers and for detection of *Enterococcus* (Cao et al., 2015). Primer and probe sequences are included in the **Supplementary Material** (**Supplementary Table 2**) and detailed methods employed have been described previously (Steele et al., 2018).

16S rRNA Gene Sequence Analysis

Next-generation DNA sequencing was used to characterize microbial community structure at coastal sites along the United States/Mexico border region and at sites potentially contributing non-indigenous bacteria to coastal waters. The DNeasy PowerSoil Pro DNA extraction kit (QIAGEN, Hilden, Germany) was used to extract DNA from water filters for next generation sequencing, according to manufacturer's guidelines.

The V5 and V6 regions of the bacterial 16S rRNA gene were amplified using primers F784 (5′-RGGATTAGATACCC-3′) and 1046R (5′-CGACRRCCATGCANCACCT-3′ (Claesson et al., 2010). Subsequently, the amplicons were pair-ended sequenced on the Illumina MiSeq platform (San Diego, CA, United States) using the dual-index method (Gohl et al., 2016) at the University of Minnesota Genomics Center (UMGC, Minneapolis, MN, United States).

Demultiplexed reads were analyzed via two different bioinformatic pipelines. The first was an OTU clustering-based approach and the second employed DADA2. For the OTU-based approach, the mothur pipeline v1.45.1 (Schloss et al., 2009) was utilized. Sequences were aligned against the SILVA Ribosomal RNA database v.132 and subjected to a 2% pre-cluster to remove likely sequence errors (Huse et al., 2010). Chimeric sequences were identified and removed using UCHIME ver. 4.2.40 (Edgar et al., 2011). Within each sample, operational taxonomic units (OTUs) were binned at 97% sequence similarity using the furthest-neighbor algorithm and were classified using the Ribosomal Database Project database v16 (Cole et al., 2009). For the DADA2 approach (Callahan et al., 2016), error rates were calculated and used for further quality filtering. Chimeras were removed using the removeChimeraDenovo function and taxonomy was assigned using DADA2's RDP Bayesian classifier against the Silva v.132 database (Yilmaz et al., 2014; Glöckner et al., 2017). Sequence data was submitted to the NCBI SRA database with accession number SRP250574.

SourceTracker Analysis

SourceTracker (ver. 0.9.8) was used to assign sources of fecal bacterial contamination at all 11 sink sites in **Figure 1** using default parameters in Knights et al. (2013). Separate SourceTracker assignment was run for each day to identify the compositional similarity in bacterial communities between sink sites and contamination "sources." The source assignment was completed for duplicate samples from each site for each sampling date, with source assignments (%) averaged.

SourceTracker was run two ways for each site, which differed based on assignment of the contamination sources. SourceTracker was run initially with the following two sources categorized as follows:

(1) Pristine/marine: composition calculated from samples collected in Silver Strand Beach, CA, United States on days when the plume was not thought to impact sites this far to the north (n = 6: Event 1 Day 1 and Day 2, Event 2 Day 1, Event 3 Day 1 and Day 2, Event 4 Day 1).
(2) SADB WTP outfall site: composition calculated by using samples collected from the SADB WTP outfall at Punta Bandera (site MX2) on the same day as the sink samples.

In addition, SourceTracker was re-run and potential inputs to the upstream Tijuana River were included as a third potential source.

(3) Tijuana River: composition calculated using samples collected from upstream on the Tijuana River, before the diversion structures. This source was included as a proxy for potential inputs to the Tijuana River that are theoretically diverted during dry weather but may be making it across the border, into the Tijuana River Estuary, and contributing to water quality impacts at beach sites.

Additional Measurements

Salinity and water temperature were measured at all sites using a portable meter (YSI Model Pro30). Swell direction, wave height, and tidal cycle information were downloaded from publicly available databases (swell direction/wave height; tidal conditions)[1,2] for all sampling periods.

Statistical Analyses

All statistical analyses were done using R-studio (R-studio version 1.1.463, Boston, MA, United States, R version 4.0.1). Correlation tests were completed to compare indicator measurements (HF183, Lachno3, and *Enterococcus* by ddPCR and culture) to each other and to SourceTracker analysis source assignments. Multiple linear regression models were run comparing the relationship between human marker measurements (Lachno3 and HF183) and distance from the SADB outfall at Punta Bandera as a function of event differences.

For statistical analyses of microbial communities, samples were rarefied to a depth of 11,000 reads/sample. Differences between bacterial communities were visualized by principal

[1]http://cdip.ucsd.edu/
[2]https://tidesandcurrents.noaa.gov/

TABLE 2 | Frequency of marker detection and concentration range detected for each marker by site.

Site	HF183			Lachno3			dENT			Frequency cENT > 104 MPN/100 mL
	Detection frequency	Concentration range		Detection frequency	Concentration range		Detection frequency	Concentration range		
		Lower	Upper		Lower	Upper		Lower	Upper	
MX1	100%	1.1E + 02	4.0E + 05	100%	3.2E + 02	5.0E + 05	100%	3.6E + 03	3.7E + 05	54%
MX2	100%	3.8E + 06	1.0E + 07	100.00%	5.4E + 05	2.5E + 07	100%	1.8E + 06	5.4E + 07	100%
MX3	100%	5.8E + 04	2.3E + 06	100.00%	8.5E + 04	6.3E + 06	100%	2.7E + 05	1.8E + 07	90%
MX4	82%	ND	1.2E + 04	90.90%	ND	3.0E + 04	100%	4.0E + 02	9.8E + 04	9%
MX5	89%	ND	2.6E + 04	100.00%	1.4E + 01	6.1E + 04	100%	4.5E + 02	1.3E + 05	11%
SD1	82%	ND	8.0E + 03	100.00%	4.3E + 01	2.0E + 04	100%	1.3E + 02	6.0E + 04	18%
SD2	60%	ND	2.2E + 04	100.00%	2.1E + 01	7.4E + 04	100%	5.4E + 02	1.1E + 05	20%
TJRM	67%	ND	8.4E + 03	100.00%	2.9E + 01	2.1E + 04	100%	8.0E + 02	7.4E + 04	33%
SD3	64%	ND	4.8E + 03	100.00%	1.8E + 01	1.5E + 04	100%	2.8E + 02	5.4E + 04	9%
IB1	83%	ND	7.5E + 03	100.00%	1.5E + 01	2.6E + 04	100%	3.3E + 02	9.2E + 04	17%
IB2	83%	ND	7.4E + 03	83.30%	ND	2.3E + 04	100%	5.5E + 02	8.4E + 04	0%
SD4	55%	ND	8.0E + 03	90.90%	ND	2.4E + 04	100%	1.1E + 02	7.9E + 04	0%
SD5	27%	ND	8.2E + 01	54.60%	ND	2.1E + 02	100%	6.4E + 01	2.5E + 03	9%

For cultivable enterococci (cENT), frequency of exceedance of the single sample beach water quality public health threshold listed. Data combined across all four sampling events and sites are ordered from south to north.

coordinate analysis (PCoA) using Bray–Curtis dissimilarity matrices. Vegan (version 2.5.6) and phyloseq (version 1.32) packages were used for visualizing taxonomic trends.

RESULTS

Human Fecal Marker and *Enterococcus* Alongshore Trends

Human-associated fecal marker (HF183 and Lachno3) and *Enterococcus* gene copies were evaluated in relation to increasing distance from the SADB WTP outfall at Punta Bandera (**Table 2**). There was a significant negative relationship between distance from the outfall and levels of human marker (HF183: $R^2 = -0.46$, $p < 0.01$; Lachno3: $R^2 = -0.49$, $p < 0.01$), evident during all four events (**Figure 2**). The slope of the relationship was significantly lower for Event 4 ($p < 0.01$), illustrating less dilution and likely faster transport of the plume when compared to Events 1–3. Estimated dilution of the SADB WTP plume was also calculated by comparing human-associated marker levels at each site relative to levels measured at the SADB WTP outfall, for estimated dilution see **Supplementary Figure 5**.

During Event 1, a gradient in human-associated marker levels was observed that extended to the United States/Mexico border. HF183 was not detected north of site IB1, while Lachno3 levels were detected at concentrations near the detection limit at the northernmost site (Silver Strand, SD5) on both sampling days. HF183 concentrations ranged from 547,250 copies/100 mL to not detected, with the highest concentrations at the sites bracketing the SADB WTP outfall at Punta Bandera (sites MX1 and MX3). Enterococci concentrations exceeded the single sample public health code regulatory limit (>104 MPN per 100 mL) at sites MX1 and MX3 on Day 1 and at Site MX3 (just North of Punta Bandera) on Day 2 only (**Figure 3A**).

During event 2, a sewage spill within a tributary of the Tijuana River confounded any along-coast human marker gradient observed. The spill bypassed the cross-border collector, leading to 0.86 million gallons of untreated sewage crossing from Mexico into the United States and being transported to the ocean via the Tijuana River Estuary the night before the first day of sampling. Water quality trends observed suggested that human marker detections were related to both the discharge from the SADB WTP outfall at Punta Bandera and the spill into the Tijuana River. Higher levels of both human markers were observed surrounding the two potential input sources, both bracketing the SADB WTP outfall at Punta Bandera and at sites located in the Tijuana River estuary (SD2) and just south of the terminus of the Tijuana River estuary (site SD1) (**Figure 3B**). Human marker levels decreased at sites near the Tijuana River Estuary by Day 3, likely due to dilution and degradation of the spilled sewage.

During Event 3, the swell switched from coming from the west to from the south by Day 3. Trends in human marker observed reflected the changing swell conditions, with a stronger south swell by Day 3 corresponding to an increasing gradient in human marker levels from the SADB WTP outfall moving toward the north. Human marker levels (HF183 and Lachno3) were similar on Day 1 and Day 2, with HF183 levels detected only at sites bracketing the SADB WTP outfall. By Day 3, human marker (HF183 and Lachno3) was detected extending from the SADB WTP outfall at Punta Bandera to Imperial Beach, CA, United States (sites IB1/IB2) (**Figure 3**). *Enterococcus* concentrations exceeded the single sample public health code

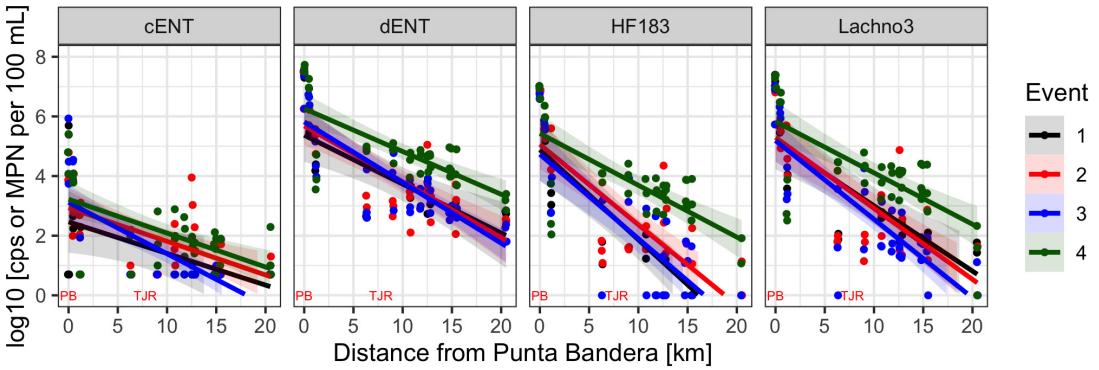


FIGURE 2 | Enterococci by culture (cENT) and ddPCR (dENT) and human marker (HF183 and Lachno3) levels versus distance from the SADB WTP outfall at Punta Bandera. Each event is represented by a different color, with all dates sampled combined per event. On the *x*-axis location of the SSADB WTP outfall is noted (PB) as is the location of the Tijuana River Estuary (TJR).

regulatory limit (>104 MPN per 100 mL) at sites bracketing Punta Bandera on Day 1 and at the site just north of Punta Bandera (MX3) on Day 2 and Day 3. Otherwise, culturable *Enterococci* levels at all ocean sites were well below the single sample regulatory limit.

The strongest gradient in human marker levels occurred during Event 4, with quantifiable human marker levels observed at the site sampled farthest from the SADB WTP outfall, site SD5, by Day 3. Human marker levels were similar between sites MX5 and SD4, suggesting minimal dilution of the plume as it traveled north. Culturable enterococci concentrations exceeded, or were near the single sample regulatory limit of >104 MPN per 100 mL, at most sites between Punta Bandera, MX and San Diego, CA, United States on Day 1 and Day 2 of Event 4, with enterococci levels below the public health code regulatory limit at all sites, with the exception of the Tijuana River Estuary site (site SD2), on Day 3.

Surf Zone Sampling

There were no significant differences observed between nearshore and surf zone *Enterococcus* by digital PCR and culture or human marker (HF183 and Lachno3) concentrations (all p-values > 0.05) (**Supplementary Figure 7**) from samples collected during events 3 and 4. In addition, measurements made in the surf zone and nearshore were highly correlated for all four targets measured (all r-values > 0.85; $p < 0.01$, **Supplementary Figure 8**).

Contamination Source Determined by 16S rRNA and SourceTracker

The SourceTracker program has been previously validated in coastal waters using OTU clustered sequencing data (Henry et al., 2016; Staley C. et al., 2018). For consistency, the SourceTracker results discussed below reflect data processed using the OTU-clustering based approach. SourceTracker was also applied to ASVs (from DADA2). Results were compared, and compositional data was highly correlated using either method ($r = 0.81$, $p < 0.01$). Trends in SourceTracker defined contributions using DADA2 are included in the **Supplementary Material** (**Supplementary Figure 6**).

Source assignment profiles differed by event, with results paralleling human-associated marker results when samples were evaluated with the SADB WTP as the only contamination source (**Figure 4A**). Estimated contributions of the SADB WTP outfall to beach sites north of the United States/Mexico border were highest during Event 4 with ∼3% of ocean water attributed to the SADB WTP. Although there was still evidence of impact from the SADB WTP plume to San Diego beach sites during the other events, total contributions to ocean waters were estimated to be much lower (<1%).

During Event 2, there was evidence of contamination contributed from the Tijuana River, which corresponded to the sewage spill event that occurred. When SourceTracker was run with both the SADB outfall and the Tijuana River included as potential sources, contributions were attributed to the Tijuana River at the Tijuana River Estuary site (site SD2) and at sites located adjacent to the estuary mouth (**Figure 4B**).

Samples generally clustered by graphical location, with differences noted between events (**Figure 5**). Samples collected at the SADB WTP outfall at Punta Bandera, clustered together regardless of when they were collected, suggesting a homogeneous microbial community in the discharge. In contrast, samples collected from SD2 (the Tijuana River Estuary site), varied dependent on Event, clustering more closely to the SADB WTP outfall during Event 2, when there was a sewage spill into the Tijuana River the night preceding sampling. The beach samples clustered most closely to MX2 (the SADB WTP outfall site) during Event 4.

The geographic extent of the SADB WTP plume was also evidenced by the presence of abundant taxa associated with the SADB WTP plume in marine water samples. ASVs (from DADA2) associated with the SADB WTP plume were resolved to a lower taxonomic level (genus or species) when compared to OTUs; thus, taxonomic trends described below reflect assignments made by the DADA2 pipeline.

Bacterial taxa that were dominant in the SADB WTP plume included several wastewater-associated genera,

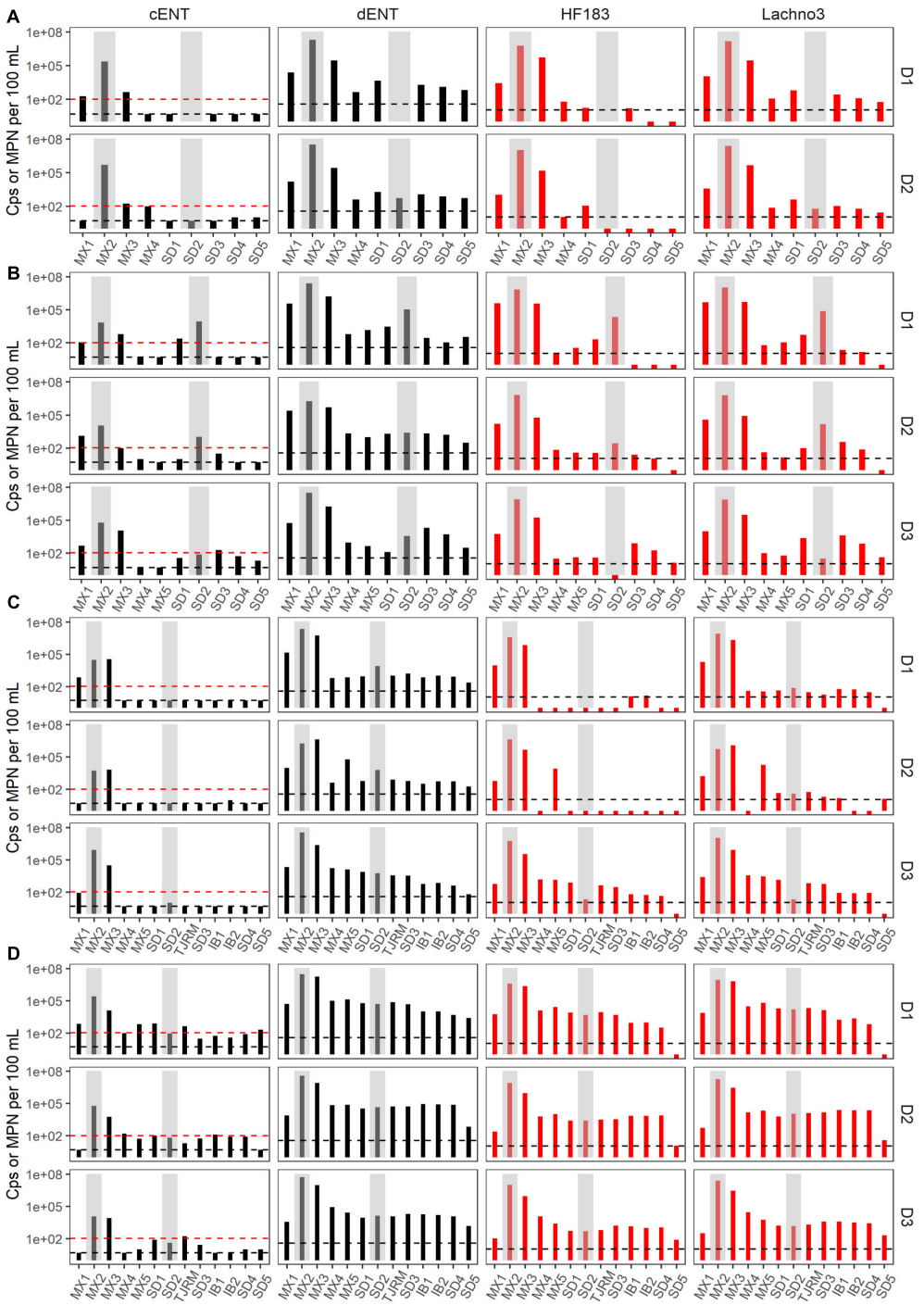


FIGURE 3 | Enterococci (black bars) by culture (cENT) and ddPCR (dENT) and human marker (red bars), HF183 and Lachno3, results for Event 1–4 **(A–D)** for each day sampled (D1–D3). Gray bars are overlayed on top of sites MX2 (the SADB WTP outfall at Punta Bandera) and SD2 (the Tijuana River Estuary), representing the two non-beach sites. Black dashed lines represent the limit of detection and the red dashed line indicates the single sample beach water quality public health threshold for culturable enterococci of 104 MPN per 100 mL.

Cloacibacterium, *Comamonas*, *Acidovorax*, *Bacteroides*, and *Macellibacteroides*. There were also several potentially pathogenic genera commonly associated with wastewater that were present in high abundance including *Acinetobacter*, *Aeromonas*, and *Arcobacter*. Taxa associated with the SADB WTP (Site MX2) were generally present in decreasing abundances at sites moving to the north (**Figure 6**), with abundances dependent on event. The highest abundances of

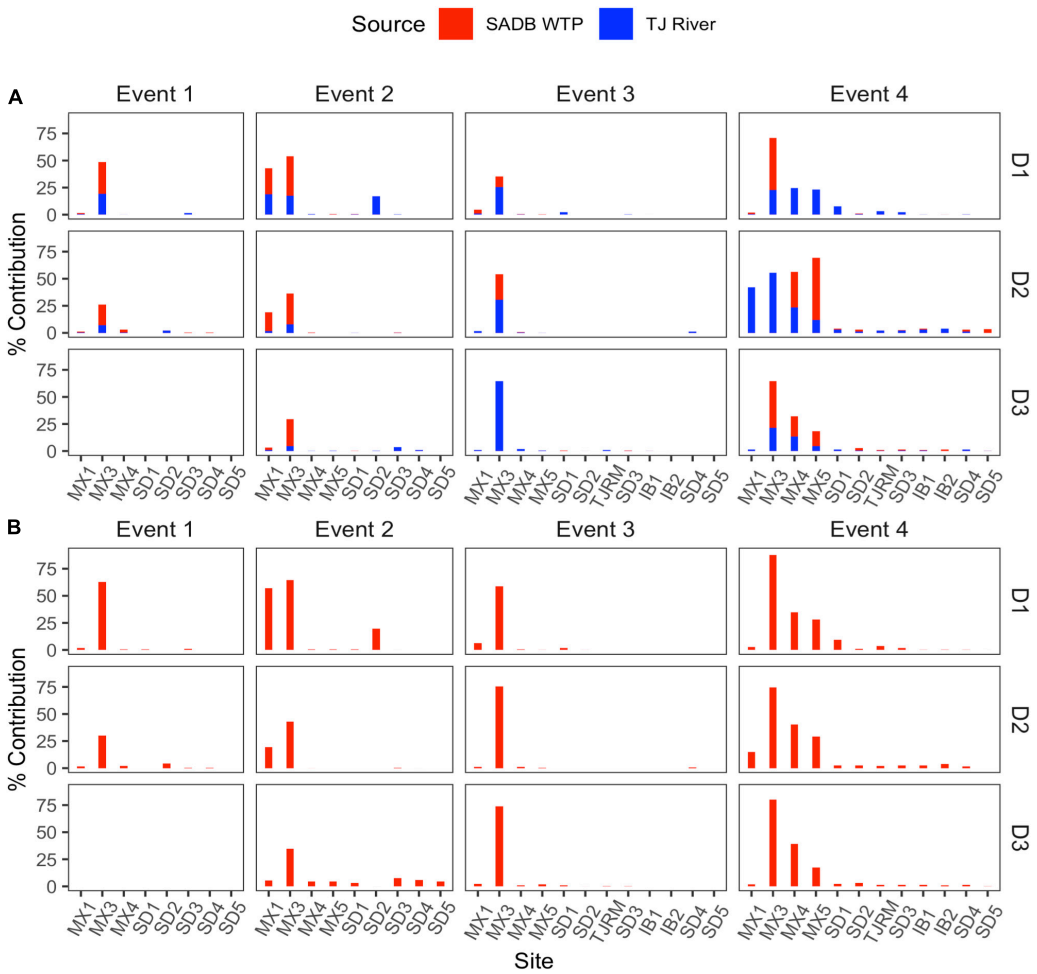


FIGURE 4 | Contributions from potential sources (SADB WTP and Tijuana River) in each sink site using SourceTracker. Sites are arranged from a north to south along the *x*-axis. **(A)** SADB WTP included as the only source. **(B)** SADB WTP and Tijuana River included as potential sources.

taxa associated with the SADB WTP were identified at beach sites during Event 4. During Event 2, high abundances of taxa associated with the SADB WTP were also observed at site SD2 (the mouth of the Tijuana River Estuary), reflecting the sewage spill event.

For comparison, abundant taxa associated with samples collected from Silver Strand State Park, CA, United States located 20 km north of the SADB WTP outfall at Punta Bandera, were used to characterize the unimpacted marine microbial community. Samples were evaluated from days when the plume was not thought to impact the Silver Strand State park site (n = 6: Event 1 Day 1 and Day 2, Event 2 Day 1, Event 3 Day 1 and Day 2, Event 4 Day 1). Marine-associated taxa were present in high abundance in those samples and included members of the marine Rhodobacteraceae family (*Amylibacter*, *Planktomarina*, *Planktotalea*, and from the coastal HIMB11 *Roseobacter* group), members of the SAR11 clade of *Alphaproteobacteria*, the *Flavobacteriaceae* NS4 and NS5 marine groups, and marine cyanobacterium *Synechococcus* CC9902 (**Figure 6**). Marine-associated taxa generally decreased in relative abundance at sites located nearest the SADB WTP outfall.

Relationship Between SourceTracker, Human Marker, and *Enterococcus* Results

There were statistically significant positive correlations ($p < 0.05$) between the magnitude of the SADB WTP contribution determined by the SourceTracker algorithm and human marker measurements, when SADB WTP was included as the only contamination source. Human marker levels, Lachno3, and HF183, were strongly correlated with estimated percent contribution from the SADB WTP outfall, $r = 0.80$ and $r = 0.70$, respectively (**Figure 7**).

Significant correlations were present between enterococci and human marker measurements (HF183 and Lachno3), regardless of measurement method (all $r > 0.6$, all p-values < 0.01; **Figure 7**). Human marker measurements, Lachno3 versus HF183, were also highly correlated to each other ($r = 0.96$,

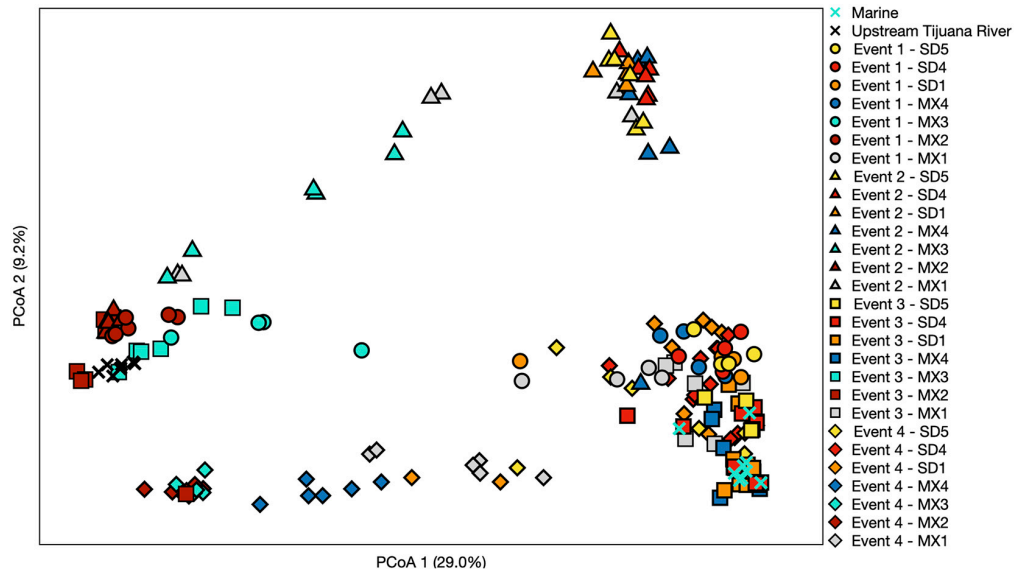


FIGURE 5 | Principal coordinate analysis performed using a Bray–Curtis dissimilarity matrix. Each symbol represents a different sampling event, while each color represents a different site. Three sites north of the United States/Mexico border were chosen to include that represent trends observed. Numbers in parentheses denote % of variation explained by each axis.

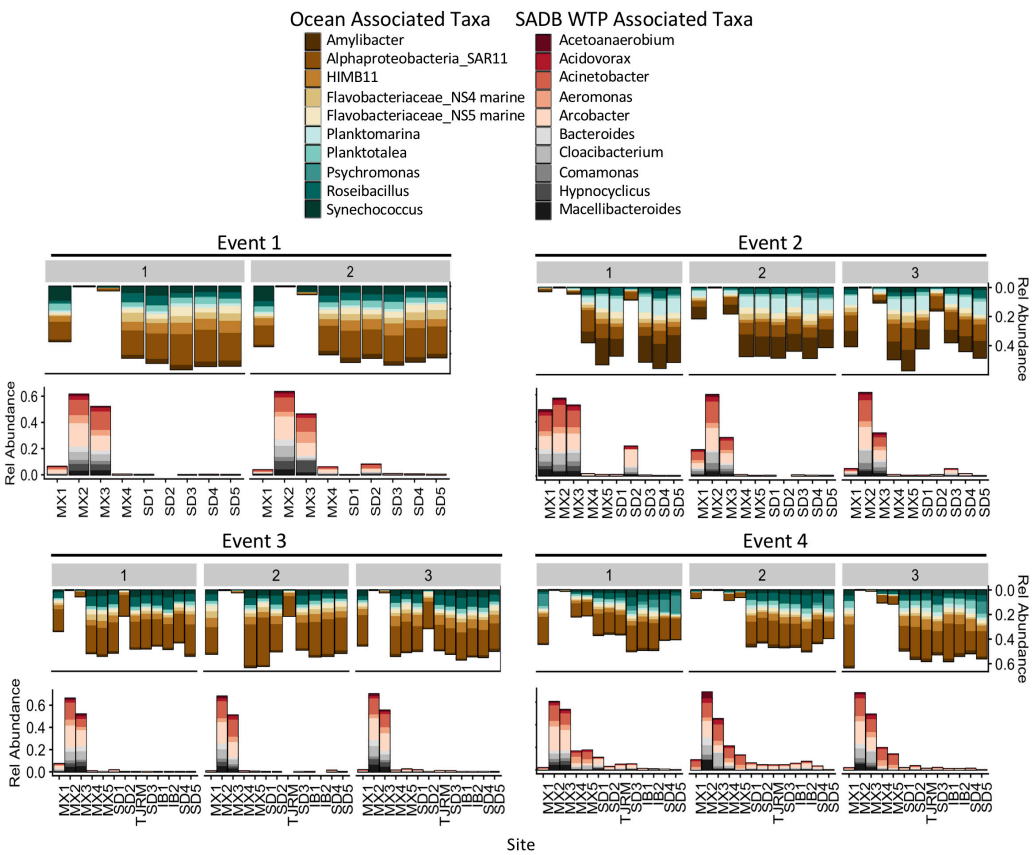


FIGURE 6 | Relative abundances for the 10 most abundant families and genera associated with SADB WTP effluent stream (MX2) and Silver Strand (SD5) samples-representing the unimpacted marine microbial community. Relative abundances are shown at all sites moving south to north along the *x*-axis. Marine associated taxa are presented on the top and SADB WTP associated taxa are presented on the bottom.

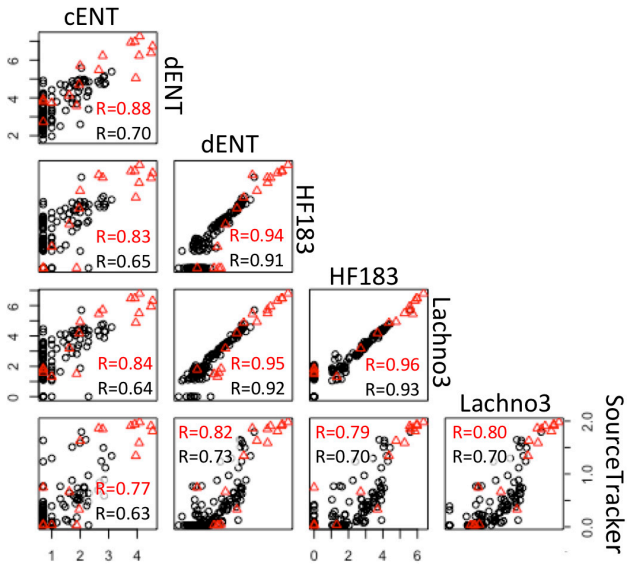


FIGURE 7 | Correlations between log-normalized microbial indicator (enterococci by culture and ddPCR and human markers, HF183 and Lachno3) results and log-normalized (x + 1) SADB WTP contributions estimated by SourceTracker. Red *r*-values refer to correlations between data collected at all sites. Black *r*-values refer to correlations between environmental beach sites only (with the SADB WTP outfall and mixing zone excluded).

$p < 0.001$; **Figure 7**), as were *Enterococcus* levels by culture and digital PCR ($r = 0.88$, $p < 0.001$; **Figure 7**).

DISCUSSION

Alongshore Human Fecal Contamination Trends

The SADB WTP discharge at Punta Bandera represents a consistent point source of human fecal contamination to the nearshore environment, with the geographic extent of its impact dependent on dilution and dispersion processes (Kim et al., 2009). In this study, the use of a molecular toolbox approach allowed for an increased understanding of how the SADB plume impacts water quality in the United States/Mexico border region. Digital PCR analysis of human-associated markers and 16S rRNA bacterial community sequencing both identified a significant gradient from the SADB WTP outfall at Punta Bandera moving south to north, when current direction was from the south, with sites closer to the discharge impacted more heavily.

Previous tracer studies have found that nearshore freshwater sources can either be rapidly transported offshore or become entrained within the surf zone, depending on wave conditions and source volume and flow rates (Rodriguez et al., 2018). During this study, conditions observed during Event 3 and Event 4 reflect the significant impact changing ocean condition can have on the impact of the SADB plume within the nearshore coastal zone. The strongest gradient in human fecal contamination (by SourceTracker and human-associated marker) was observed during Event 4, which targeted a longer south swell event. Swell direction was coming from the south for several days preceding sampling, which likely contributed to the higher human-associated marker concentrations and the more limited dilution of the plume observed. During this event, SourceTracker analysis attributed between 1 and 5% of total beach water at sites in San Diego, CA, United States and between 5 and 50% of beach water at popular beach sites located in Tijuana, Mexico to the SADB WTP. Considering that the SADB WTP outfall consists of minimally treated sewage, exposure to the plume has the potential for significant health risk. The results observed during Event 4 suggest that the impact of the SADB WTP plume can extend to beaches far upcoast, with evidence of the plume detected up to 20 km north of the SADB WTP outfall at Punta Bandera. These results support previous modeling efforts conducted that have illustrated that the SADB WTP plume can propagate northward during south swell events, leading to minimally treated sewage transported to San Diego beaches as far upcoast as 32 km, under the right conditions (Feddersen et al., 2020).

In contrast to the far-reaching extent of the plume during Event 4, during the first 2 days of Event 3, when currents were coming from the west, human associated markers were either not detected or were only present near the limit of detection at sites upcoast of the SADB WTP outfall. SourceTracker analysis attributed 1–2% of total beach water to the SADB WTP at sites located to the north of the plant in Tijuana, Mexico and 1–6% of total beach water to the plume at site MX1, located <1 mile south of the SADB WTP outfall. Although not the focus of this study, these results suggest that under certain ocean conditions the plume may be quickly transported offshore, having a lesser impact on beach water quality.

Further bolstering the SourceTracker results, the abundant taxa in the microbial communities revealed a distinct sewage plume assemblage that contrasted with the coastal ocean assemblage. *Acidovorax, Bacteroides, Cloacibacterium*, *Comamonas*, *Macellibacteroides,* and potentially pathogenic genera *Acinetobacter*, *Aeromonas*, and *Arcobacter,* the most abundant taxa in this sewage plume assemblage, have been consistently reported to be abundant taxa in sewage 16S rRNA microbiomes throughout the world (e.g., McLellan et al., 2010; Vandewalle et al., 2012; Shanks et al., 2013; Cai et al., 2014; Newton et al., 2015; Ahmed et al., 2017; Numberger et al., 2019). In contrast, the most abundant taxa identified in the coastal ocean assemblage were typical of the environmental marine microbiome and have all been identified previously as abundant in southern California coastal ocean communities: *Alphaproteobacteria* including the ubiquitous SAR11 clade, *Amylibacter*, *Planktomarina*, *Planktotalea*, and the coastal HIMB11 *Roseobacter* group, *Flavobacteriaceae* marine groups NS4 and NS5, and *Synechococcus* CC9902, a ubiquitous marine cyanobacterium (Morris et al., 2002; Fuhrman et al., 2006; Dufresne et al., 2008; Giebel et al., 2009; Steele et al., 2011; Cram et al., 2015). The relative abundances of taxa associated with

the sewage plume assemblage decreased with increasing distance from the SADB WTP outfall and at the same time, the coastal ocean assemblage became much more abundant, matching the SourceTracker results. The clear difference between the communities, along with the spatial gradient from the outfall, provides another line of evidence illustrating the influence of the SADB WTP plume during south swells.

Paired surf zone samples were also collected for comparison to the shoreline samples during the last two south swell events. There were no significant differences between the surf zone and shoreline samples collected and results from the two sampling areas were highly correlated, suggesting that under the water quality conditions measured in this study, shoreline samples collected as part of routine monitoring efforts are likely to reflect water quality conditions in the surf zone. This is contrast to previous efforts that found higher concentrations of FIB in the shoreline versus surf zone (SCCWRP, 2007). However, locations tested during those efforts were all located in close proximity to flowing freshwater creek inputs.

Human Source Identification Limitations

It was originally hypothesized that the effect of treatment processes at the SADB WTP may be large enough to alter the bacterial community of the wastewater (Hu et al., 2012), allowing for discrimination between contamination from the SADB WTP and wastewater traveling to beach sites from the Tijuana River. However, neither detection of human-associated markers by ddPCR or SourceTracker analysis of 16S rRNA sequences were able to definitively differentiate between human fecal contamination from these two sources, despite treatment and transport differences. These two primary sources harbored similar bacterial communities, with variations observed only for bacterial taxa that were present at low abundances. This is perhaps not surprising, as it is well documented that the majority of fecal contamination to the Tijuana River emanates from faulty infrastructure in the City of Tijuana and that a portion of the sewage that spills into the river is also diverted directly to the SADB WTP (USEPA, 2014). Moving forward, taxa more closely associated with the SADB outfall could be targeted for specific marker analysis that may provide better discrimination between the SADB WTP and similar human sources, like the Tijuana River.

Overall, results obtained using human-associated markers and SourceTracker analysis of 16S rRNA sequences were generally in agreement and positively correlated with only slight differences observed in estimated dilution of the SADB WTP plume by the two methods. These differences may be attributed to differences in environmental fate and transport. The SourceTracker algorithm utilizes a combination of different sequences to create a unique fingerprint for each source. In contrast, PCR marker-based methods target a specific sequence associated with a particular source. Previous efforts have evaluated the persistence of both human markers (Bae and Wuertz, 2009; Ahmed et al., 2014; Mattioli et al., 2017) and wastewater associated bacterial communities in marine waters (Sassoubre et al., 2015; Zhang et al., 2015; Ahmed et al., 2018). Side-by-side comparisons are also needed that evaluate decay of these signals under relevant environmental conditions and to ensure the stability of bacterial community-based pollution assignments.

Implications for Pollutant Tracking

The tools applied in this study represent a promising approach for further understanding pollution fate and transport in complex environmental settings. Modeling efforts have demonstrated the potential for the discharge plume from the SADB WTP to track north, across the United States/Mexico border under specific ocean conditions (CH2M Hill, 2009; Feddersen et al., 2020). However, this hypothesis has not been confirmed due to logistical challenges associated with both sampling in this region and with teasing apart multiple fecal pollution sources in the nearshore environment. In this study, more sensitive tools coupled with a targeted sampling approach that took advantage of specific ocean conditions helped characterize the potential impact of the SADB WTP plume. Although additional sources within this region could also be contributing to the elevated human marker levels; the trends observed in both the SourceTracker analysis and human-associated marker results tracked each other and swell conditions closely, pointing to the SADB WTP plume as a potential source of human fecal contamination to beach waters extending north, across the United States/Mexico border, under south swell conditions.

This is one of the first studies to effectively use microbial source tracking tools for nearshore plume characterization, which unlike other tracers, offer biological measurements that are a function of both decay and dilution. Hydrodynamic modeling-based approaches commonly used to predict plume transport typically rely on dyes, or other conservative tracers, to estimate how specific water sources move within the coastal environment (McCarthy et al., 2017). However, as pollutant sources travel in the environment, physical, biological, and chemical factors all govern the fate and transport of fecal-borne microbes (Sinton et al., 2002; Boehm et al., 2005, 2009; Rippy et al., 2013). FIB measurements have also been used to develop nearshore (Stark et al., 2016) and estuarine hydrodynamic models (Gao et al., 2015). However, FIB can come from multiple animal and human sources, confounding the ability to track specific pollutant sources. Advances in microbial source tracking tools may be a useful addition to hydrodynamic simulations, helping to contextualize, validate, and calibrate measurements made.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRP250574.

AUTHOR CONTRIBUTIONS

AZ-F, JS, JG, and MD contributed to conception and design of the study. AZ-F, MG, XX, CS, and MS contributed to sample and statistical analyses. CS, MS, XX, and AZ-F contributed to data interpretation. AZ-F wrote the first draft of the manuscript. All authors contributed to revision, read, and approved the submitted version.

FUNDING

This study was funded by the County of San Diego and the U.S. Environmental Protection Agency. The authors thank the following for their invaluable contribution to the project: Jeff Chokry, David Wanless, Jason Lindquist, Robert Stabenow, and the entire Proyecto Fronterizo de Educación Ambiental Team.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.674214/full#supplementary-material

Supplementary Table 1 | Description of number of samples collected at each site and events when sampling occurred at each site.

Supplementary Table 2 | Primer and probe sequences utilized for microbial source tracking assays.

REFERENCES

Ahmed, W., Gyawali, P., Sidhu, J. P. S., and Toze, S. (2014). Relative inactivation of faecal indicator bacteria and sewage markers in freshwater and seawater microcosms. *Lett. Appl. Microbiol.* 59, 348–354. doi: 10.1111/lam.12285

Ahmed, W., Staley, C., Kaiser, T., Sadowsky, M. J., Kozak, S., Beale, D., et al. (2018). Decay of sewage-associated bacterial communities in fresh and marine environmental waters and sediment. *Appl. Microbiol. Biotechnol.* 102, 7159–7170. doi: 10.1007/s00253-018-9112-4

Ahmed, W., Staley, C., Sadowsky, M. J., Gyawali, P., Sidhu, J., Palmer, A., et al. (2015). Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water. *Appl. Environ. Microbiol.* 81, 7067–7077. doi: 10.1128/AEM.02032-15

Ahmed, W., Staley, C., Sidhu, J., Sadowsky, M., and Toze, S. (2017). Amplicon-based profiling of bacteria in raw and secondary treated wastewater from treatment plants across Australia. *Appl. Microbiol. Biotechnol.* 101, 1253–1266. doi: 10.1007/s00253-016-7959-9

Bae, S., and Wuertz, S. (2009). Rapid decay of host-specific fecal Bacteroidales cells in seawater as measured by quantitative PCR with propidium monoazide. *Water Res.* 43, 4850–4859. doi: 10.1016/j.watres.2009.06.053

Baral, D., Speicher, A., Dvorak, B., Admiraal, D., and Li, X. (2018). Quantifying the relative contributions of environmental sources to the microbial community in an urban stream under dry and wet weather conditions. *Appl. Environ. Microbiol.* 84, 1–36. doi: 10.1128/AEM.00896-18

Boehm, A. B., Keymer, D. P., and Shellenbarger, G. G. (2005). An analytical model of enterococci inactivation, grazing, and transport in the surf zone of a marine beach. *Water Res.* 39, 3565–3578. doi: 10.1016/j.watres.2005.06.026

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

Brown, C. M., Staley, C., Wang, P., Dalzell, B., Chun, C. L., and Sadowsky, M. J. (2017). A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a lake superior estuary. *Environ. Sci. Technol.* 51, 8263–8271. doi: 10.1021/acs.est.7b01353

Cai, L., Ju, F., and Zhang, T. (2014). Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl. Microbiol. Biotechnol.* 98, 3317–3326. doi: 10.1007/s00253-013-5402-z

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cao, Y., Raith, M. R., and Griffith, J. F. (2015). Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water Res.* 70, 337–349. doi: 10.1016/j.watres.2014.12.008

CH2M Hill (2009). *Coastal observations and monitoring in South Bay San Diego IBWC/Surfrider Consent Decree: Final Report prepared for the International Boundary and Water Commission*. El Paso, TX: International Boundary Water Commission (IBWC).

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38, e200. doi: 10.1093/nar/gkq873

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, 141–145. doi: 10.1093/nar/gkn879

Coudray-Meunier, C., Fraisse, A., Martin-Latil, S., Guillier, L., Delannoy, S., Fach, P., et al. (2015). A comparative study of digital RT-PCR and RT-qPCR for quantification of Hepatitis A virus and Norovirus in lettuce and water samples. *Int. J. Food Microbiol.* 201, 17–26. doi: 10.1016/j.ijfoodmicro.2015.02.006

Cram, J. A., Chow, C. E. T., Sachdeva, R., Needham, D. M., Parada, A. E., Steele, J. A., et al. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* 9, 563–580. doi: 10.1038/ismej.2014.153

DeFlorio-Barker, S., Wing, C., Jones, R. M., and Dorevitch, S. (2018). Estimate of incidence and cost of recreational waterborne illness on United States surface waters. *Environ. Heal. A Glob. Access Sci. Source* 17, 1–10. doi: 10.1186/s12940-017-0347-9

Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 9:R90. doi: 10.1186/gb-2008-9-5-r90

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Ervin, J. S., Russell, T. L., Layton, B. A., Yamahara, K. M., Wang, D., Sassoubre, L. M., et al. (2013). Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods. *Water Res.* 47, 6873–6882. doi: 10.1016/j.watres.2013.02.060

Feddersen, F., Wu, X., and Giddings, S. (2020). *Modeling Impacts of Various Wastewater and Stromwater flow Scenarios on San Diego South Bay and Tijuana beaches.* Final report prepared for the North American Development Bank. San Antoni, TX: North American Development Bank.

Feng, S., Bootsma, M., and McLellan, S. L. (2018). Human-associated Lachnospiraceae genetic markers improve detection of fecal pollution sources in urban waters. *Appl. Environ. Microbiol.* 84, 1–14. doi: 10.1128/AEM.00309-18

Fuhrman, J. A., Hewson, I., Schwalbach, M. S., Steele, J. A., Brown, M. V., and Naeem, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13104–13109. doi: 10.1073/pnas.0602399103

Gao, G., Falconer, R. A., and Lin, B. (2015). Modelling the fate and transport of faecal bacteria in estuarine and coastal waters. *Mar. Pollut. Bull.* 100, 162–168. doi: 10.1016/j.marpolbul.2015.09.011

Giebel, H. A., Brinkhoff, T., Zwisler, W., Selje, N., and Simon, M. (2009). Distribution of roseobacter RCA and SAR11 lineages and distinct bacterial communities from the subtropics to the Southern Ocean. *Environ. Microbiol.* 11, 2164–2178. doi: 10.1111/j.1462-2920.2009.01942.x

Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., et al. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi: 10.1016/j.jbiotec.2017.06.1198

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., et al. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* 34, 942–949. doi: 10.1038/nbt.3601

Goodwin, K. D., Schriewer, A., Jirik, A., Curtis, K., and Crumpacker, A. (2017). Consideration of natural sources in a bacteria TMDL – lines of evidence, including beach microbial source tracking. *Environ. Sci. Technol.* 51, 7775–7784. doi: 10.1021/acs.est.6b05886

Green, H. C., Haugland, R. A., Varma, M., Millen, H. T., Borchardt, M. A., Field, K. G., et al. (2014a). 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

Green, H. C., White, K. M., Kelty, C. A., and Shanks, O. C. (2014b). Development of rapid canine fecal source identification PCR-based assays. *Environ. Sci. Technol.* 48, 11453–11461. doi: 10.1021/es502637b

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

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., et al. (2016). Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Res.* 93, 242–253. doi: 10.1016/j.watres.2016.02.029

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

Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a bayesian approach. *Water Res.* 41, 3701–3715. doi: 10.1016/j.watres.2007.06.037

Kim, S. Y., Terrill, E. J., and Cornuelle, B. D. (2009). Assessing coastal plumes in a region of multiple discharges: the U.S.-mexico border. *Environ. Sci. Technol.* 43, 7450–7457. doi: 10.1021/es900775p

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., et al. (2013). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763. doi: 10.1038/nmeth.1650.Bayesian

Lee, C., Marion, J. W., and Lee, J. (2013). Development and application of a quantitative PCR assay targeting *Catellicoccus marimammalium* for assessing gull-associated fecal contamination at Lake Erie beaches. *Sci. Total Environ.* 45, 1–8. doi: 10.1016/j.scitotenv.2013.03.003

Li, X., Sivaganesan, M., Kelty, C. A., Zimmer-Faust, A., Clinton, P., Reichman, J. R., et al. (2019). Large-scale implementation of standardized quantitative real-time PCR fecal source identification procedures in the Tillamook Bay Watershed. *PLoS One* 14:e0216827. doi: 10.1371/journal.pone.0216827

Mathai, P. P., Staley, C., and Sadowsky, M. J. (2020). Sequence-enabled community-based microbial source tracking in surface waters using machine learning classification: A review. *J. Microbiol. Methods* 177:106050. doi: 10.1016/j.mimet.2020.106050

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

McCarthy, D. T., Jovanovic, D., Lintern, A., Teakle, I., Barnes, M., Deletic, A., et al. (2017). Source tracking using microbial community fingerprints: method comparison with hydrodynamic modelling. *Water Res.* 109, 253–265. doi: 10.1016/j.watres.2016.11.043

McLellan, S. L., Huse, S. M., Mueller-Spitz, S. R., Andreishcheva, E. N., and Sogin, M. L. (2010). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* 12, 378–392. doi: 10.1111/j.1462-2920.2009.02075.x

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806–810. doi: 10.1038/nature01240

Newton, R. J., McLellan, S. L., Dila, D. K., Vineis, J. H., Morrison, H. G., Murat Eren, A., et al. (2015). Sewage reflects the microbiomes of human populations. *MBio* 6, e02574. doi: 10.1128/mBio.02574-14

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, 1–14. doi: 10.1038/s41598-019-46015-z

Orozco-Borbón, M. V., Rico-Mora, R., Weisberg, S. B., Noble, R. T., Dorsey, J. H., Leecaster, M. K., et al. (2006). Bacteriological water quality along the Tijuana-Ensenada, Baja California, México shoreline. *Mar. Pollut. Bull.* 52, 1190–1196. doi: 10.1016/j.marpolbul.2006.02.005

Riedel, T. E., Thulsiraj, V., Zimmer-Faust, A. G., Dagit, R., Krug, J., Hanley, K. T., et al. (2015). Long-term monitoring of molecular markers can distinguish different seasonal patterns of fecal indicating bacteria sources. *Water Res.* 71, 227–243. doi: 10.1016/j.watres.2014.12.037

Rippy, M. A., Franks, P. J. S., Feddersen, F., Guza, R. T., and Moore, D. F. (2013). Factors controlling variability in nearshore fecal pollution: The effects of mortality. *Mar. Pollut. Bull.* 66, 191–198. doi: 10.1016/j.marpolbul.2012.09.003

Rodriguez, A. R., Giddings, S. N., and Kumar, N. (2018). Impacts of nearshore wave-current interaction on transport and mixing of small-scale buoyant plumes. *Geophys. Res. Lett.* 45, 8379–8389. doi: 10.1029/2018GL078328

Roguet, A., Eren, A. M., Newton, R. J., and McLellan, S. L. (2018). Fecal source identification using random forest. *Microbiome* 6:185. doi: 10.1186/s40168-018-0568-3

Roguet, A., Esen, ÖC., Eren, A. M., Newton, R. J., and McLellan, S. L. (2020). FORENSIC: an online platform for fecal source identification. *mSystems* 5, 1–14. doi: 10.1128/msystems.00869-19

Sassoubre, L. M., Love, D. C., Silverman, A. I., Nelson, K. L., and Boehm, A. B. (2012). Comparison of enterovirus and adenovirus concentration and enumeration methods in seawater from Southern California, USA and Baja Malibu, Mexico. *J. Water Health* 10, 419–430. doi: 10.2166/wh.2012.011

Sassoubre, L. M., Yamahara, K. M., and Boehm, A. B. (2015). Temporal stability of the microbial community in sewage-polluted seawater exposed to natural sunlight cycles and marine microbiota. *Appl. Environ. Microbiol.* 81, 2107–2116. doi: 10.1128/AEM.03950-14

SCCWRP (2007). *Southern California Bight 2003 Regional Monitoring Program: VI Shoreline Microbiology.* SCCWRP Technical Report. Costa mesa, CA: SCCWRP.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schoen, M. E., Soller, J. A., and Ashbolt, N. J. (2011). Evaluating the importance of faecal sources in human-impacted waters. *Water Res.* 45, 2670–2680. doi: 10.1016/j.watres.2011.02.025

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* 7, 249–259. doi: 10.1111/j.1462-2920.2004.00702.x

Shanks, O. C., Atikovic, E., Blackwood, A. D., Lu, J., Noble, R. T., Domingo, J. S., et al. (2008). Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 74, 745–752. doi: 10.1128/AEM.01843-07

Shanks, O. C., Newton, R. J., Kelty, C. A., Huse, S. M., Sogin, M. L., and McLellan, S. L. (2013). Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl. Environ. Microbiol.* 79, 2906–2913. doi: 10.1128/AEM.03448-12

Shanks, O. C., Santo Domingo, J. W., Lamendella, R., Kelty, C. A., and Graham, J. E. (2006). Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. *Appl. Environ. Microbiol.* 72, 4054–4060. doi: 10.1128/AEM.00023-06

Sinton, L. W., Hall, C. H., Lynch, P. A., Davies-colley, R. J., Sinton, L. W., Hall, C. H., et al. (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

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103

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

Soller, J. A., Schoen, M. E., Varghese, A., Ichida, A. M., Boehm, A. B., Eftim, S., et al. (2014). Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. *Water Res.* 66, 254–264. doi: 10.1016/j.watres.2014.08.026

Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V. J., Brown, C. M., et al. (2018). Application of sourcetracker for accurate identification of fecal pollution in recreational freshwater: a double-blinded study. *Environ. Sci. Technol.* 52, 4207–4217. doi: 10.1021/acs.est.7b05401

Staley, Z. R., Boyd, R. J., Shum, P., and Edge, T. A. (2018). Microbial source tracking using quantitative and digital PCR to identify sources of fecal contamination in stormwater, river water, and beach water in a Great Lakes area of concern. *Appl. Environ. Microbiol.* 84, 1–12. doi: 10.1128/AEM.01634-18

Stark, J. S., Bridgen, P., Dunshea, G., Galton-Fenzi, B., Hunter, J., Johnstone, G., et al. (2016). Dispersal and dilution of wastewater from an ocean outfall at Davis Station, Antarctica, and resulting environmental contamination. *Chemosphere* 152, 142–157. doi: 10.1016/j.chemosphere.2016.02.053

Steele, J. A., Blackwood, A. D., Griffith, J. F., Noble, R. T., and Schiff, K. C. (2018). Quantification of pathogens and markers of fecal contamination during storm events along popular surfing beaches in San Diego. *California. Water Res.* 136, 137–149. doi: 10.1016/j.watres.2018.01.056

Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* 5, 1414–1425. doi: 10.1038/ismej.2011.24

Thulsiraj, V., Zimmer-Faust, A. G., and Jay, J. A. (2017). Use of viability-based methods for improved detection of recent fecal contamination in a microbial source tracking study near Tijuana, Mexico. *Water. Air. Soil Pollut.* 228:63. doi: 10.1007/s11270-016-3204-5

USEPA (2011). *Using Microbial Source Tracking to Support TMDL Development and Implementation Acknowledgments*. Washington, DC: USEPA.

USEPA (2012). *Method 1611: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) Assay. Environment Prot*. Washington, DC: USEPAL, 56. doi: 10.1017/CBO9781107415324.004

USEPA (2014). *Wastewater collection rehabilitation for mainlines in Tijuana, Mexico: Environmental Assessment. February, 2014*. Washington, DC: USEPA.

Vandewalle, J. L., Goetz, G. W., Huse, S. M., Morrison, H. G., Sogin, M. L., Hoffmann, R. G., et al. (2012). Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. *Environ. Microbiol.* 14, 2538–2552. doi: 10.1111/j.1462-2920.2012.02757.x

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

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, 643–648. doi: 10.1093/nar/gkt1209

Zhang, Q., He, X., and Yan, T. (2015). Differential decay of wastewater bacteria and change of microbial communities in beach sand and seawater microcosms. *Environ. Sci. Technol.* 49, 8531–8540. doi: 10.1021/acs.est.5b01879

Zimmer-Faust, A. G., Thulsiraj, V., Marambio-Jones, C., Cao, Y., Griffith, J. F., Holden, P. A., et al. (2017). Effect of freshwater sediment characteristics on the persistence of fecal indicator bacteria and genetic markers within a Southern California watershed. *Water Res.* 119, 1–11. doi: 10.1016/j.watres.2017.04.028

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.

Copyright © 2021 Zimmer-Faust, Steele, Xiong, Staley, Griffith, Sadowsky, Diaz and Griffith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/fmicb.2021.684533

Integrative Survey of 68 Non-overlapping Upstate New York Watersheds Reveals Stream Features Associated With Aquatic Fecal Contamination

Hyatt Green[1], Maxwell Wilder[1], Martin Wiedmann[2] and Daniel Weller[1]*

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

OPEN ACCESS

Edited by:
Georg H. Reischer, Vienna University of Technology, Austria

Reviewed by:
David McCarthy, Monash University, Australia
Arnau Casanovas-Massana, Tangen Biosciences, Inc., United States

***Correspondence:**
Hyatt Green
hgreen@esf.edu

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *23 March 2021*
Accepted: *05 July 2021*
Published: *12 August 2021*

Citation:
Green H, Wilder M, Wiedmann M and Weller D (2021) Integrative Survey of 68 Non-overlapping Upstate New York Watersheds Reveals Stream Features Associated With Aquatic Fecal Contamination. Front. Microbiol. 12:684533. doi: 10.3389/fmicb.2021.684533

Aquatic fecal contamination poses human health risks by introducing pathogens in water that may be used for recreation, consumption, or agriculture. Identifying fecal contaminant sources, as well as the factors that affect their transport, storage, and decay, is essential for protecting human health. However, identifying these factors is often difficult when using fecal indicator bacteria (FIB) because FIB levels in surface water are often the product of multiple contaminant sources. In contrast, microbial source-tracking (MST) techniques allow not only the identification of predominant contaminant sources but also the quantification of factors affecting the transport, storage, and decay of fecal contaminants from specific hosts. We visited 68 streams in the Finger Lakes region of Upstate New York, United States, between April and October 2018 and collected water quality data (i.e., *Escherichia coli*, MST markers, and physical–chemical parameters) and weather and land-use data, as well as data on other stream features (e.g., stream bed composition), to identify factors that were associated with fecal contamination at a regional scale. We then applied both generalized linear mixed models and conditional inference trees to identify factors and combinations of factors that were significantly associated with human and ruminant fecal contamination. We found that human contaminants were more likely to be identified when the developed area within the 60 m stream buffer exceeded 3.4%, the total developed area in the watershed exceeded 41%, or if stormwater outfalls were present immediately upstream of the sampling site. When these features were not present, human MST markers were more likely to be found when rainfall during the preceding day exceeded 1.5 cm. The presence of upstream campgrounds was also significantly associated with human MST marker detection. In addition to rainfall and water quality parameters associated with rainfall (e.g., turbidity), the minimum distance to upstream cattle operations, the proportion of the 60 m buffer used for cropland, and the presence of submerged aquatic vegetation

at the sampling site were all associated based on univariable regression with elevated levels of ruminant markers. The identification of specific features associated with host-specific fecal contaminants may support the development of broader recommendations or policies aimed at reducing levels of aquatic fecal contamination.

Keywords: microbial water quality, *Escherichia coli*, land use, Finger Lakes, microbial source-tracking

INTRODUCTION

Aquatic fecal contamination poses risks to human and environmental health through the introduction of pathogens (Craun et al., 2006; Arnone and Walling, 2007), nutrients (Sharpley et al., 2003), antimicrobials (Chee-Sanford et al., 2009; Karkman et al., 2019), and hormones (Boxall et al., 2003; Hanselman et al., 2003; Combalbert and Hernandez-Raquet, 2010; Bartikova et al., 2016). From 1978 to 2014, contact with contaminated untreated recreational water (e.g., lakes) resulted in 184 outbreaks associated with acute gastrointestinal illness, acute respiratory illness, and skin-related illness among other symptoms (CDC, 2019). Food-borne outbreaks can also be caused by using contaminated water for various agricultural purposes, such as frost protection and irrigation during produce production (FDA, 2018). Harmful algal blooms that pose both human and environmental health risks are also caused by, at least in part, phosphorus-rich runoff from fertilizer or manure applied to agricultural fields (Kane et al., 2014; Vadas et al., 2017; Kumaragamage and Akinremi, 2018). Therefore, the identification of not only the origins of fecal contaminants but also the factors that affect their transport, storage, and decay is essential to protecting ambient water quality and human and environmental health.

The protection and remediation of surface water quality often begins with assessing potential hazards, including sources of fecal contamination (WHO, 2016). A multitude of factors that affect levels of observed aquatic fecal contamination have been identified. Examples of well-known factors include, but are not limited to, recent rainfall and stormwater runoff (VanWormer et al., 2016), livestock density in the upstream catchment (Oliver et al., 2018), and poorly maintained wastewater infrastructure (Ahmed et al., 2018). However, many sources are diffuse and/or intermittent and may be difficult to pinpoint. Previous studies have assessed the association between the presence and levels of fecal indicator bacteria (FIB), such as *Escherichia coli* or fecal coliforms, and various factors thought to affect their distribution. However, it is well-established that FIB are ubiquitous in mammalian hosts, including some wildlife, and represent the total level of fecal contamination in a water body (Dufour, 1984). Therefore, it is difficult to gage the importance of the different factors controlling the input of a single fecal type (e.g., human and ruminant) using FIB as a response because measured FIB levels are a composite of FIB from different sources that may have different origins, routes of introduction, and other factors that control their distribution.

Over the last two decades, microbial source-tracking (MST) methods that target host-specific bacteria (Bernhard and Field, 2000; Mieszkin et al., 2010; Green et al., 2012) or viruses (Zhang et al., 2005; Stachler et al., 2017) have been used to distinguish and quantify levels of source-specific fecal contaminants in water samples. MST has most often been used to facilitate remediation of specific waterbodies, but could be used as a model input to identify and rank the major factors controlling different types of fecal contaminants. Such information could inform site-specific, watershed-scale, or regional management plans to help avoid sourcing or using water for agriculture or recreation that presents a risk to human health, such as water contaminated with human sewage or septage.

In this study, we visited 68 streams over a single growing season in the Finger Lakes region of Upstate New York, United States, and collected weather, land-use patterns, and water quality data (e.g., *E. coli*, MST markers, and physical–chemical parameters), as well as other stream features with the goal of identifying factors that affect fecal contamination at a regional scale. Our specific objectives were to (i) identify the associations between levels of *E. coli*, the detection of MST markers (i.e., avian, canid, human, and ruminant), and the presence of potential sources of upstream fecal contamination, (ii) characterize the relationship between *E. coli* levels and MST marker detection, and (iii) identify key spatial, weather, and physical–chemical water quality factors associated with an increased or decreased likelihood of fecal contamination.

MATERIALS AND METHODS

Spatial Analysis

Land Cover Characterization

Inverse-distance weights (IDW) were used to characterize land cover as described previously (Weller D.L. et al., 2020). Land cover percentages were calculated for the following distance intervals around each sampling site: 0–100, 100–250, 250–500, 500–1,000, 1,000–2,000, 2,000–5,000, 5,000–10,000, 10,000–20,000, and >20,000 m. The IDW proportion under each land cover class was then calculated for the total watershed and stream corridor (60 m buffer from the stream channel) using a modified version of the equation from King et al. (2005) (example R code available at https://github.com/wellerd2).

Feature Density

The presence, density, and flow path distance (minimum and median) from the sampling site to features that could affect measured water quality parameters were also determined. Flow path distances were estimated by creating flow networks using the National Hydrology Dataset (NHD) flow accumulation and flow direction rasters, as well as NHD flowline data, and analyzed using the riverdist package version 0.15.3 (Tyers, 2017) in R as

done previously (Weller et al., 2019). Flowline data were used for features that intersected streams (e.g., roads), while flow networks were used for all other features. The only data available for septic systems were aggregated to census tracts; therefore, septic density in each watershed was interpolated based on the percentage overlap between each census tract and the watershed as described previously (Weller et al., 2019).

Sample Site Selection

Streams and possible access points were identified using ArcGIS 10.2. Stream access points that represented an upstream watershed area $\geq$10 km^2 with streams adjacent (<400 m) to fields used to grow produce covered by the Food Safety Modernization Act over 4 of the past 8 years were identified as candidate sampling sites. From the candidate sampling sites, 68 sampling sites were randomly selected to ensure that the sampled watersheds did not overlap. While funding for this work was contingent on the confidentiality of precise sampling locations, all samples were collected near commercial farms from waterways used for irrigation in the Great Lakes or Finger Lakes watersheds. A map of approximate sampling locations is provided in a previous publication (Weller et al., 2019).

Meteorological Data Acquisition

Meteorological data were obtained from the nearest Network for Environment and Weather Applications (NEWA) weather station.[1] Total rainfall, average air temperature, and average solar radiation were then calculated using non-overlapping periods [i.e., 0–1, 1–2, 2–3, 3–4, 4–5, 5–10, 10–20, and 20–30 days before sample collection (d BSC)] accounting for the precise time of day the sample was collected.

Sample Collection

Between April and October 2018, streams were sampled either two (n = 8 streams) or three (n = 60 streams) times each resulting in a total of 196 1-liter samples. Samples were stored on ice during transit and filtered within 6 h of sample collection. To create split samples for MST and enumeration of fecal indicators, 1-liter samples were shaken by hand before filtration of two separate 100 ml aliquots.

Microbiological Analysis and MST

Enumeration of *E. coli*

Escherichia coli were enumerated using the Colilert Quanti-Tray 2000 kit (IDEXX, Westbrook, ME, United States) following the manufacturer's instructions. Sample dilution was not performed so the upper and lower limits of quantification were 2,419.6 Most Probable Number (MPN)/100 ml and 1 MPN/100 ml.

Sample Processing for MST

One hundred milliliter subsamples were filtered using sterile, pre-bagged, single-use vacuum filtration units with 0.45 μm pore-size 47 mm diameter polyethersulfone filters. Filters were transferred to Lysing Matrix E tubes (MPBio, Irvine, CA, United States) and stored at −80°C for between 55 and 218 days (mean = 197 days) until DNA extraction. Prior to extraction, filters were allowed to thaw to room temperature, and a 29.2 μl aliquot of prepared *Caenorhabditis elegans* lysate was added to each tube as described previously (Kirtane et al., 2019). Following thawing and lysate addition, filters were homogenized on a FastPrep-24-5G (MPBio), and DNA was extracted with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) with a final elution volume of 100 μl. Eluted DNA was stored at −20°C for between 1 and 27 days (mean = 21 days) until qPCR. Two DNA extraction blanks were included in each of the five extraction batches by conducting extractions as described above, but omitting the filters.

qPCR for MST

qPCR assays for human (HF183; Bernhard and Field, 2000; Green et al., 2014a), ruminant (Rum2Bac; Mieszkin et al., 2010), canine (DG3; Green et al., 2014b), avian (GFD; Green et al., 2012; Weller et al., 2019), and internal control *C. elegans* (CG4; Kirtane et al., 2019) molecular markers were performed as described previously (Green et al., 2019; Weller et al., 2019). Oligonucleotides were obtained from IDT (IA, United States), except for MGB probes that were obtained from Applied Biosystems (MA, United States). A version of the GFD assay that was modified for probe-based detection and that detects fecal contamination from gulls, geese, ducks, and chickens was used as previously reported (Weller et al., 2019). Gulls, geese, and ducks are common and are therefore likely sources of fecal contamination in the study area. Duplicate 25 μl qPCR reactions were prepared using 12.5 μl TaqMan Environmental Master Mix (ThermoFisher, Waltham, MA, United States), molecular grade water, primers and probes, and 2 μl DNA template. Reactions were run under default cycling parameters (50°C for 2 min; 95°C for 10 min; and 40 cycles of 95°C for 15 s, and 60°C for 1 min) on either a QuantStudio3 or QuantStudio5 (ThermoFisher). Each 96-well qPCR plate contained at least four no template control reactions (NTCs) for which molecular grade water was substituted in place of DNA template, as well as two wells of positive control consisting of 10^3 copies/reaction of a custom designed gBlock (IDT). The automatic baseline setting and a threshold value of 0.03 were used to estimate C_T values. qPCR assays used in this study and their performance metrics are reported elsewhere (Weller et al., 2019).

qPCR Quality Control and Data Analysis

Assay-specific standard curves generated using 10^0–10^6 copies/reaction of custom-designed gBlocks (IDT) were used to convert C_T values to copy number per reaction. Reactions with C_T values greater than the intercept of the standard curve were considered below limits of detection (LOD). For each assay, a sample was considered detectable for MST markers only if both C_T values were within the LOD. Amplification was not detected in any of the NTC (n = 100) or extraction blank reactions (n = 20). Recovery of spiked *C. elegans* DNA estimated using the CG4 assay was used as a proxy for total DNA recovery through the extraction process as done previously (Kirtane et al., 2019). DNA recoveries ranged from

[1] newa.cornell.edu

unmeasurable due to no recovery of spike (one sample) to >100% indicating a large variability in recovery (**Supplementary Figure 1**). Because samples with low recoveries still provide informative data (e.g., detection of MST markers), no samples were removed from analysis based on their estimated recovery values. Kinetic outlier detection for all CG4 reactions was used to detect the possible effects of qPCR inhibition as reported previously (Bar et al., 2011; Green and Field, 2012; Kirtane et al., 2019). Briefly, a sigmoidal model was fit to each CG4 amplification curve from which the first and second derivative maxima were calculated. A difference of ≥10 fractional cycle values between the derivative maxima was considered a sign of significant inhibition in the reaction well. Using these methods, no samples displayed signs of qPCR inhibition; therefore, no samples were excluded from analysis because of qPCR inhibition.

Data Analysis

Generalized Linear Mixed Models

General or generalized linear mixed models (GLMMs) were developed using the lme4 version 1.1 (Bates et al., 2015) and censReg version 0.5 (Henningsen, 2017) packages in R. Model outcomes were the presence or absence of the target MST marker or the log10 MPN of *E. coli*/100 ml. For models where the outcome was binary, a logit link was used. For models where the outcome was *E. coli* concentration (or "level"), censored regression was used to account for samples where the MPN was above the range of quantification (2,419.6 CFU/100 ml). No samples were below the limits of quantification. Week of the year was included as a fixed effect in all models to account for non-independence of samples due to the collection of samples in the same week. While stream ID was not itself a factor of interest, it was included as a random effect to account for pseudo-replication in all GLMMs.

Conditional CART Model Development

Conditional CART models [conditional inference trees (CTrees)] were included as an additional statistical technique that is robust to correlation and missingness and inherently considers hierarchical relationships within the data. CTrees were developed using the partykit package version 1.2 (Hothorn et al., 2006; Hothorn and Zeileis, 2015) using the following control parameters: mincriterion = 0.80, minbucket = 5, minsplit = 15, and maxsurrogate = 3. Separate models were developed for each outcome: (i) the presence or absence of the target MST marker or (ii) the log10 MPN of *E. coli*/100 ml. For models where *E. coli* levels were the outcome, the nine samples that were above the range of quantification maximum (2,419.6 CFU/100 ml) were assigned a value of 2,500 when developing the CTrees.

RESULTS

FIB and MST Summary

Escherichia coli was detectable in all streams and samples with a mean value of 212 MPN/100 ml (Standard Deviation [SD] = 637 MPN/100 ml); nine samples were above the range of quantification. MST markers were less prevalent than *E. coli*, occurring in only 68% of streams and only 38% of samples. HF183 (a human fecal marker) was the most prevalent MST marker, followed by Rum2Bac (a ruminant fecal marker), GFD (an avian fecal marker), and DG3 (a canid fecal marker; **Table 1**). Rum2Bac marker concentrations were higher than HF183, GFD, and DG3. GFD and DG3 marker levels were approximately one and two orders of magnitude below HF183 and Rum2Bac marker levels, respectively. The scarcity of GFD (found in only 4% of samples) and DG3 (found in only 0.5% of samples) markers greatly limited our ability to draw conclusions about drivers of avian and canid contamination, respectively. GLMMs indicated that the probability of finding both human (4.520, $p < 0.001$, 95% CI = 2.012–10.155) and ruminant (6.838, $p < 0.001$, 95% CI = 2.406–19.433) markers increased with elevated *E. coli* levels.

Factors Associated With *E. coli* Concentration

GLMMs

Results from GLMMs indicated that the density of upstream pig farms had a significant positive effect on *E. coli* concentrations (1.868, $p = 0.036$, 95% CI = 0.121–3.615; **Supplementary Table 1**), while the presence of upstream stormwater outfalls was also associated with higher levels of *E. coli* (0.309, $p = 0.008$, 95% CI = 0.083–0.536). Interestingly, the presence of upstream goat/sheep farms had a significant negative effect on *E. coli* concentrations (−1.049, $p = 0.018$, 95% CI = −1.919 to −0.180). Despite being one of the most common forms of livestock in the study area, there was no clear association between dairy/cattle farm density and *E. coli* concentrations (−0.161, $p = 0.071$, 95% CI = −0.336 to 0.014). Increased dissolved oxygen (−0.155, $p < 0.001$, 95% CI = −0.204 to −0.106) and pH (−0.468, $p < 0.001$, 95% CI = −0.718 to −0.219) were associated with lower levels of *E. coli*, while turbidity was associated with higher levels (0.580, $p < 0.001$, 95% CI = 0.376–0.784). Average rainfall, temperature, and solar radiation also had a significant effect on *E. coli* levels. Total rainfall within 0–2 d BSC was associated with higher levels of *E. coli* (0–1 d BSC, 0.375, $p < 0.001$, 95% CI = 0.252–0.497; 1–2 d BSC, 0.194, $p < 0.001$, 95% CI = 0.099–0.289), while total rainfall 5–10 d BSC was somewhat negatively associated with *E. coli* levels (−0.077, $p = 0.007$, 95% CI = −0.132 to −0.021). Average temperature 5–10 days prior to sample collection was more strongly associated with *E. coli* levels (0.046, $p < 0.001$, 95% CI = 0.029–0.062) than average temperature 0–5 days prior to sample collection (0.019, $p < 0.019$, 95% CI = 0.003–0.035). Average solar radiation 4–5 days prior to sampling was positively associated with *E. coli* (0.455, $p < 0.001$, 95% CI = 0.189–0.721), while average solar radiation 0–2 days prior was negatively associated with *E. coli* (0–1 day prior, −0.387, $p = 0.008$, 95% CI = −0.674 to −0.101; 1–2 days prior, −0.399, $p = 0.008$, 95% CI = −0.696 to −0.103). The presence of exposed rock in the form of cobble, boulder, or bedrock in the streambed was associated with lower levels of *E. coli* (−0.434, $p < 0.001$, 95% CI = −0.641 to −0.228), while the presence of organic matter was significantly associated with higher levels of *E. coli* (0.440, $p < 0.001$, 95% CI = 0.225–0.656).

TABLE 1 | Study-wide detection frequency and concentrations of molecular markers.

	Number of positive samples (%)	Number of positive streams (%)	Geometric mean concentration in positive samples (copies/100 ml)	Median concentration in positive samples (copies/100 ml)	Minimum concentration in positive samples (copies/100 ml)	Maximum concentration in positive samples (copies/100 ml)
DG3	1 (0.5%)	1 (1%)	76	76	76	76
GFD	8 (4%)	7 (10%)	410.7	466.5	64	7,040
HF183	49 (25%)	31 (46%)	1,640.1	1,204	48	320,448
Rum2Bac	34 (17%)	26 (38%)	1,971.6	1,315.5	145	117,605

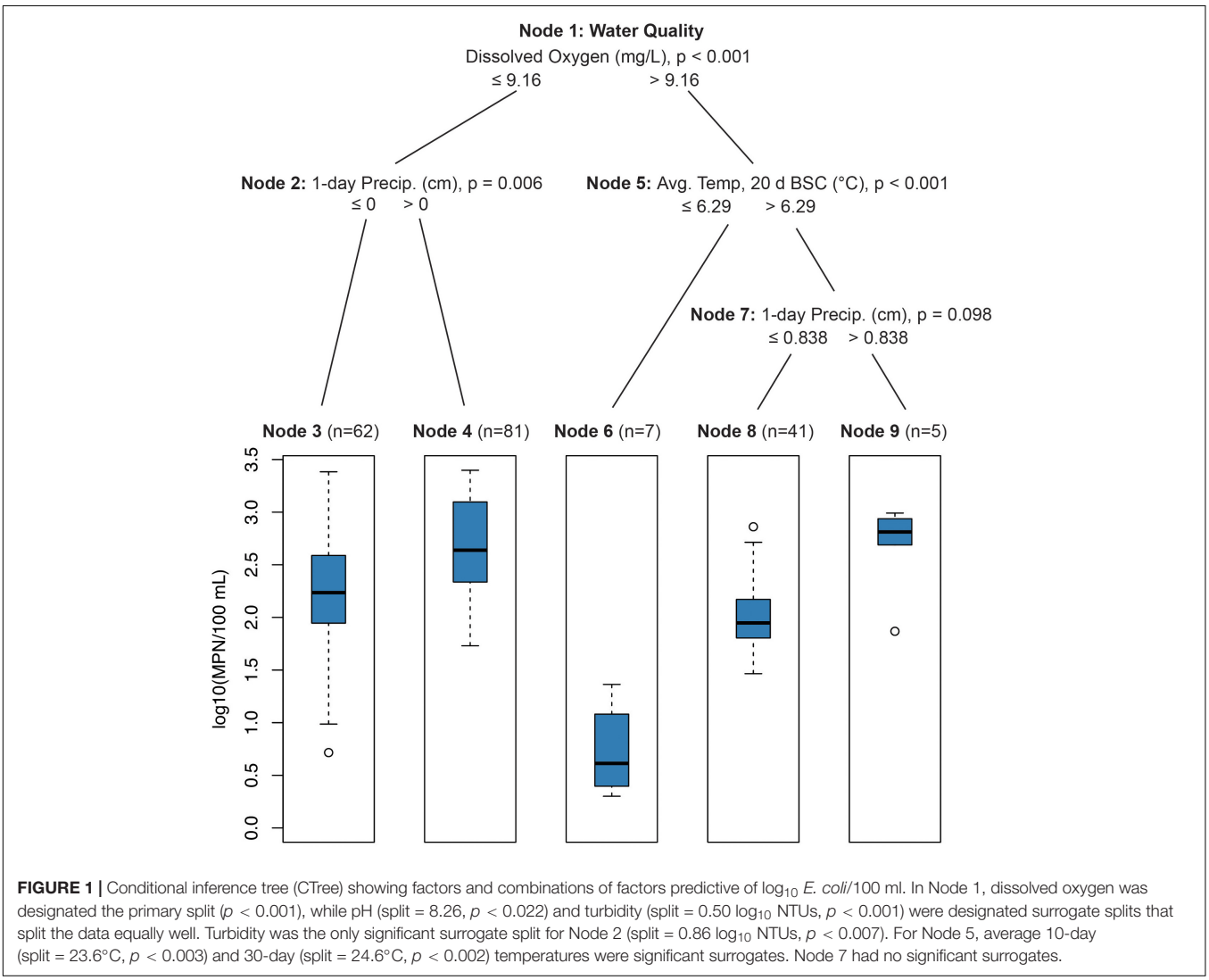


FIGURE 1 | Conditional inference tree (CTree) showing factors and combinations of factors predictive of $\log_{10}$ *E. coli*/100 ml. In Node 1, dissolved oxygen was designated the primary split ($p < 0.001$), while pH (split = 8.26, $p < 0.022$) and turbidity (split = 0.50 $\log_{10}$ NTUs, $p < 0.001$) were designated surrogate splits that split the data equally well. Turbidity was the only significant surrogate split for Node 2 (split = 0.86 $\log_{10}$ NTUs, $p < 0.007$). For Node 5, average 10-day (split = 23.6°C, $p < 0.003$) and 30-day (split = 24.6°C, $p < 0.002$) temperatures were significant surrogates. Node 7 had no significant surrogates.

CTrees

Conditional inference trees indicated that elevated *E. coli* levels were driven by physical–chemical water quality parameters (i.e., dissolved oxygen, turbidity, and pH) and precipitation 0–1 d BSC (**Figure 1**). *E. coli* levels were lowest (mean = 5.5 MPN/100 ml, SE = 12.5) when dissolved oxygen was above 9.16 mg/L and the average temperature 20–30 days before sampling was below 6.3°C (**Figure 1**, Node 6; $p < 0.001$). *E. coli* levels were highest (mean = 496.6 MPN/100 ml, SE = 89.9) when dissolved oxygen was less than 9.16 mg/L and total rainfall 0–1 d BSC was greater than zero centimeters (**Figure 1**, Node 2; $p = 0.006$).

Factors Associated With Human MST Markers

GLMMs

Somewhat surprisingly, rural features, such as the presence of horse stables upstream of sampling points (3.409, $p = 0.020$,

95% CI = 1.216–9.556; **Supplementary Table 2**) and IDW % of pasture land use (0.968, p = 0.039, 95% CI = 0.938–0.998), were significantly associated with the presence of human markers. Also, as distance from upstream goat/sheep (1.065, p = 0.038, 95% CI = 1.003–1.131) or pig farms (1.084, p = 0.024, 95% CI = 1.011–1.163) increased, so did the probability of finding human markers, potentially reflecting rural to urban transitions. Unsurprisingly, the presence of upstream stormwater outfalls (5.297, p = 0.002, 95% CI = 1.879–14.929), wastewater discharges (4.426, p = 0.003, 95% CI = 1.672–11.722), and campgrounds (3.549, p = 0.024, 95% CI = 1.184–10.637) was associated with the presence of human markers. Total watershed area was also significantly associated with human marker presence (1.041, p = 0.006, 95% CI = 1.012–1.071). Rainfall 0–1 d BSC (3.412, $p < 0.01$, 95% CI = 1.729–6.732) and average solar radiation 2–3 d BSC (0.097, p = 0.008, 95% CI = 0.017–0.540) were also positively associated with human marker presence.

CTrees

In areas where either the IDW developed area within a 60 m stream buffer exceeded 3.4%, the IDW developed watershed area exceeded 40.9%, or where there were stormwater outfalls present upstream, 45% of samples contained detectable levels of human markers ($p < 0.05$; **Figure 2**, Node 5). The probability of detecting human markers was lowest (13.6%) when none of the above conditions were met and rainfall 0–1 d BSC was less than 1.5 cm ($p < 0.05$; **Figure 2**, Node 3).

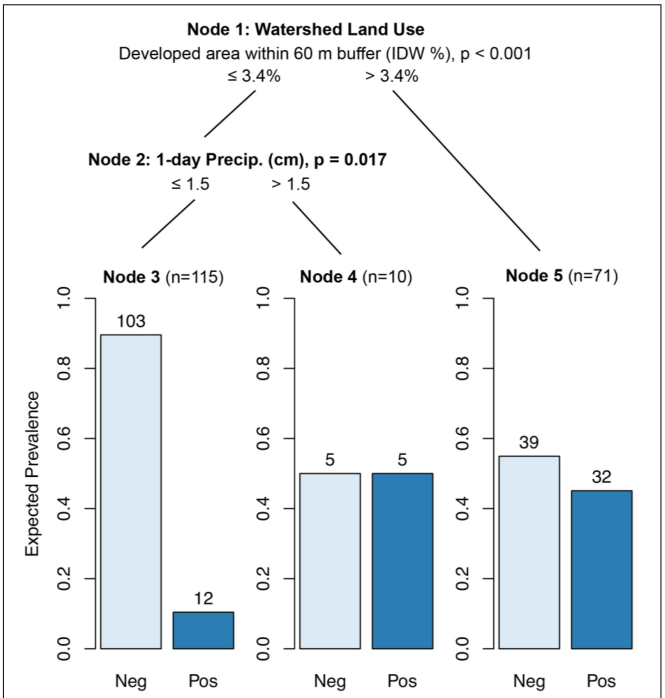


FIGURE 2 | Conditional inference tree (CTree) showing factors and combinations of factors predictive of HF183 (human) marker presence. In Node 1, IDW % developed area within 60 m buffer was designated the primary split ($p < 0.001$), while developed watershed area (split = 41%, $p < 0.002$) and the presence of stormwater outfalls ($p < 0.012$) were designated surrogate splits that split the data equally well. Node 2 had no significant surrogates.

Factors Associated With Ruminant MST Markers

GLMMs

The minimum distance to upstream cattle operations (0.842, p = 0.047, 95% CI = 0.710–0.997; **Supplementary Table 3**), IDW % of cropland within the 60 m buffer (1.054, p = 0.019, 95% CI = 1.009–1.101), and IDW % of forest/wetland (1.040, p = 0.036, 95% CI = 1.003–1.078) were all significantly associated with the presence of ruminant markers. Water quality parameters typically associated with poor water quality, such as high levels of *E. coli* (6.838, $p < 0.01$, 95% CI = 2.406–19.433) and turbidity (9.354, $p < 0.01$, 95% CI = 2.684–32.603), were also associated with the detection of ruminant markers as were average solar radiation 0–2 d BSC (0–1 d BSC, 0.041, p = 0.001, 95% CI = 0.007–0.261; 1–2 d BSC, 0.080, p = 0.005, 95% CI = 0.014–0.466) and total rainfall 0–1 d BSC (4.387, $p < 0.01$, 95% CI = 2.140–8.991) and 3–4 d BSC (2.770, p = 0.025, 95% CI = 1.139–6.734). Interestingly, the presence of exposed rock (0.257, p = 0.005, 95% CI = 0.099–0.669) and submerged aquatic vegetation (SAV) were also significantly associated with the presence of ruminant markers (0.278, p = 0.030, 95% CI = 0.087–0.885).

CTrees

While CTree results for ruminant marker presence/absence were more complex, they largely matched the results from the GLMMs. Ruminant markers were most likely to be found in samples with 0–1 d BSC precipitation greater than 1.17 cm ($p < 0.001$), *E. coli* levels greater than 631 MPN/100 ml (p = 0.008), average solar radiation greater than 0.43 MJ/m^2 0–1 d BSC (p = 0.008), or turbidity greater than 19.95 NTUs (p = 0.056; **Figure 3**, Node 7). When none of these conditions were met, ruminant markers were still more likely to be found if the pasture area within the 60 m buffer exceeded 35.8% ($p < 0.001$; **Figure 3**, Node 6). IDW % forested/wetland watershed area greater than 59.6% increased the probability of ruminant marker detection from 6.8% (**Figure 3**, Node 4) to 36.8% (**Figure 3**, Node 5; p = 0.038).

DISCUSSION

In this study, we used data collected from 68 streams to identify factors associated with fecal contamination across the Finger Lakes region of Upstate New York. Few studies have sampled this many streams multiple times to assess the causes of fecal contamination. While location-specific regulatory tools, such as total maximum daily loads (TMDLs) (U.S. Environmental Protection Agency, 1991), may offer pathways toward improved water quality for specific streams, they take years to develop and implement, and it is often resource-intensive to ensure compliance. On the other hand, federal

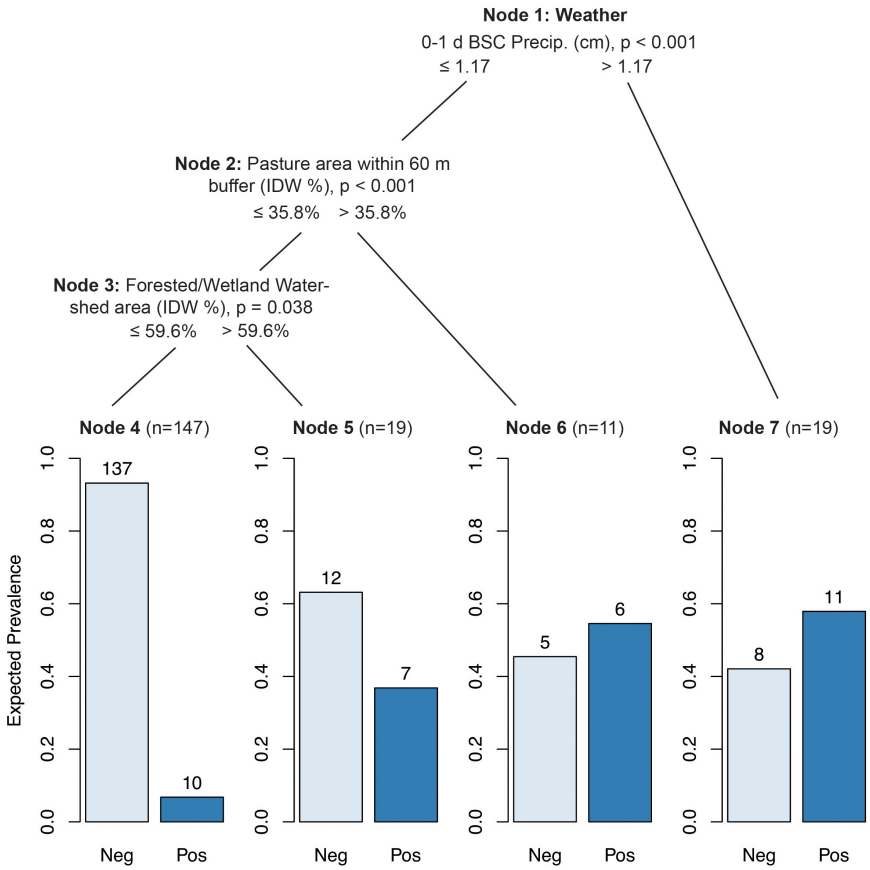


FIGURE 3 | Conditional inference tree (CTree) showing factors and combinations of factors predictive of Rum2Bac (ruminant) marker presence. In Node 1, 0–1 d BSC precipitation was designated the primary split ($p < 0.001$), while *E. coli* (split = 2.8 $\log_{10}$ MPN/100 ml, $p < 0.008$), average solar radiation in the 0–1 d BSC (split = 0.43 MJ/m^2, $p < 0.008$), and turbidity (split = 1.30 $\log_{10}$ NTUs, $p < 0.056$) were designated surrogate splits that split the data equally well. Nodes 2 and 3 had no significant surrogates.

standards for assessing human health hazard presence in recreational water (US EPA, 2012) and surface water used for produce production (Food and Drug Administration, 2015) may not fully account for spatiotemporal variation in microbial water quality, or the heterogeneity inherent to freshwater environments at local and regional scales (Truitt et al., 2018; Weller D. et al., 2020). In contrast, and intermediate in scale between the prior two approaches, the identification of factors associated with elevated levels of fecal contamination across a region may indicate effective mitigation strategies with relatively low overhead (Grayson et al., 1997; Verhougstraete et al., 2015). Although there were practical limitations that prevented an assessment of year-to-year variability on each of the 68 streams, follow-up studies should be conducted that assess this variability over a smaller number of streams.

Escherichia coli is an FIB and the target for existing agricultural and many recreational water quality standards and monitoring programs. In the present study, we primarily found associations between *E. coli* levels and meteorological or physical–chemical variables, although two land-use factors, upstream pig farm density and stormwater outfall presence, were also associated with elevated *E. coli* levels. More investigation is needed to determine why *E. coli* levels were negatively associated with goat/sheep farm density. It may be that goat/sheep farms may contribute lower levels of *E. coli* than larger livestock with higher amounts of waste, or that goat/sheep farms are associated with unmeasured factors that are also associated with low levels of *E. coli*. Rainfall is often associated with poor water quality (Pandey et al., 2012; Francy et al., 2013; Stocker et al., 2016; Weller et al., 2019) due to its propensity to promote the transport of contaminants into streams *via* stormwater runoff. The observed negative association between rainfall 5–10 d BSC and *E. coli* levels could be attributable to the rainfall-mediated removal of *E. coli* 5–10 BSC followed by limited terrestrial loading and/or limited transport due to little or no rainfall in the 0–5 d BSC window. The extended persistence and even growth of *E. coli* in the environment has been associated with elevated temperature previously (Porter et al., 2019) that is further supported by our observation of significant associations between *E. coli* and average temperature 0–10 days prior to sample collection.

Our observation that average solar radiation 4–10 days prior to sampling had a positive effect on *E. coli* levels could be explained by the strong correlation between sunlight and warmer temperatures. In contrast, our observation that higher levels of solar radiation 0–2 d BSC had a negative effect on *E. coli* could be due to the fact that UV light damages cells directly.

Human-derived aquatic fecal contamination is often considered more dangerous than other sources because of its association with a diverse range of human pathogens (Soller et al., 2014). Human sources are often associated with point sources, such as stormwater outfalls, and non-point sources, such as failing septic tanks, and can be detected using a wide range of source-tracking methods including bacterial, viral, and chemical methods. Based on the occurrence of HF183 markers, we found that agricultural (e.g., presence of upstream stables), engineered (e.g., stormwater outfalls), and residential features (e.g., presence of upstream campgrounds), as well as the amount of developed area in the watershed (IDW %) and in the 60 m stream buffer area, were highly associated with signs of human fecal contamination. While the associations of HF183 markers with upstream campgrounds, stormwater outfalls, and wastewater discharges were not surprising, the associations of HF183 markers with upstream stable presence and the proportion of pasture land (IDW %) in the watershed were somewhat unexpected given that there is no clear source of human contamination from stables or pasture land. It is possible that stable presence may be associated with the presence of septic systems that could serve as inputs, or that the HF183 assay cross-reacted with equine or other non-human fecal contaminants. Cross-reaction of other human MST markers, but not HF183, with equine contaminants has been observed previously (Feng et al., 2020). Further investigation of the sites in question is needed to fully explain these associations.

A number of factors associated with high levels of *E. coli* were also significantly associated with the presence of ruminant contaminants. The associations of Rum2Bac markers with recent rainfall (GLMMs, 0–1 and 3–4 d BSC; CTree, 0–1 d BSC) suggest that a major driver of transport of ruminant contaminants to streams in the study region is stormwater runoff. Based on CTree analysis, even when rainfall was low, the probability of finding ruminant markers increased with pasture area within the stream buffer or forested/wetland area in the watershed (IDW %). It is currently unclear if the associations between Rum2Bac markers and forests or wetlands (both land-use types were collapsed in this study) were due to bovine sources or to other ruminants (e.g., deer) that can also be detected by the Rum2Bac assay (Mieszkin et al., 2010).

The observation that different factors are associated with *E. coli* than human markers could not only be due to the fact that *E. coli* is a composite measure of all fecal inputs, but also could be due to the extended persistence or resilience of molecular markers compared with cultivable indicators (Green et al., 2011). For example, the presence of upstream wastewater discharges was significantly associated with human markers but not *E. coli* levels. We attribute this to the observation that wastewater treatment processes are usually much more effective at reducing levels of cultivable *E. coli* than they are at reducing concentrations of DNA-based molecular markers (Wu et al., 2020), which are relatively stable. A similar situation may have occurred downstream of campgrounds where molecular markers may have persisted through the duration of transport from, presumably, septage, whereas the rapid decay of *E. coli* may have limited its detection in these areas. Human molecular markers may not be good indicators of some pathogens in scenarios where contaminant introduction to the waterbody is preceded by treatment or storage processes.

CONCLUSION

The identification of a wide range of factors significantly associated with fecal contamination in the Finger Lakes region of Upstate New York streams points to the complex dynamics of fecal loading of streams in this area. Our observation that the presence of human contamination may be driven in some cases by watershed land use and specific features, such as stormwater outfalls, more so than meteorological factors suggests that limiting human fecal contamination of streams may be best confronted by management actions that prioritize spatial aspects of watersheds versus stream monitoring. However, the opposite may be true for mitigating ruminant contamination that appears to be highly associated with factors that are frequently monitored for in-stream (e.g., *E. coli*, turbidity, conductivity, temperature, and pH). Although the precise factors that control fecal contamination vary within and between watersheds, the factors identified herein could be useful for informing regional best management practices for reducing fecal contamination of waterbodies.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MWe obtained funding for the collection of samples. HG obtained funding for the microbial source-tracking analysis and wrote the first draft of the manuscript and all authors edited subsequent drafts. MWe and DW conceived and designed the study. DW coordinated the collection of samples, water quality data, and spatial data including land use features and, conducted the statistical analysis. MWl collected the microbial source-tracking data. All authors approved the submitted version of the manuscript.

FUNDING

Data collection and cleaning were partially funded by a grant from the Center for Produce Safety under award number 2017CPS09. Data analysis was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (NIH) under award number T32ES007271.

ACKNOWLEDGMENTS

We are grateful for the technical assistance of Sherry Roof, Maureen Gunderson, Tanzy Love, Sriya Sunil, Ahmed Gaballa, Kayla Ferris, and Julia Muuse.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.684533/full#supplementary-material

Supplementary Figure 1 | Histogram of DNA recoveries (%) estimated by quantifying spiked *C. elegans*.

Supplementary Table 1 | Parameters significantly associated with log-transformed *E. coli* concentrations using univariate GLMMs.

Supplementary Table 2 | Parameters significantly associated with log-transformed HF183 marker presence/absence using univariate GLMMs.

Supplementary Table 3 | Parameters significantly associated with log-transformed Rum2Bac marker presence/absence using univariate GLMMs.

REFERENCES

Ahmed, W., Zhang, Q., Lobos, A., Senkbeil, J., Sadowsky, M. J., Harwood, V. J., et al. (2018). Precipitation influences pathogenic bacteria and antibiotic resistance gene abundance in storm drain outfalls in coastal sub-tropical waters. *Environ. Int.* 116, 308–318. doi: 10.1016/j.envint.2018.04.005

Arnone, R. D., and Walling, J. P. (2007). Waterborne pathogens in urban watersheds. *J. Water Health* 5, 149–162. doi: 10.2166/wh.2006.001

Bar, T., Kubista, M., and Tichopad, A. (2011). Validation of kinetics similarity in qPCR. *Nucleic Acids Res.* 40, 1395–1406. doi: 10.1093/nar/gkr778

Bartikova, H., Podlipna, R., and Skalova, L. (2016). Veterinary drugs in the environment and their toxicity to plants. *Chemosphere* 144, 2290–2301. doi: 10.1016/j.chemosphere.2015.10.137

Bates, D., Mäechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

Bernhard, A. E., and Field, K. G. (2000). 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

Boxall, A. B. A., Kolpin, D. W., Halling-Sørensen, B., and Tolls, J. (2003). Peer reviewed: are veterinary medicines causing environmental risks? *Environ. Sci. Technol.* 37, 286A–294A.

CDC (2019). *2013—2014 Recreational Water-Associated Outbreak Surveillance Report Supplemental Figures*. Atlanta, GA: CDC.

Chee-Sanford, J. C., Mackie, R. I., Koike, S., Krapac, I. G., Lin, Y. F., Yannarell, A. C., et al. (2009). Fate and transport of antibiotic residues and antibiotic resistance genes following land application of manure waste. *J. Environ. Qual.* 38, 1086–1108. doi: 10.2134/jeq2008.0128

Combalbert, S., and Hernandez-Raquet, G. (2010). Occurrence, fate, and biodegradation of estrogens in sewage and manure. *Appl. Microbiol. Biotechnol.* 86, 1671–1692. doi: 10.1007/s00253-010-2547-x

Craun, M. F., Craun, G. F., Calderon, R. L., and Beach, M. J. (2006). Waterborne outbreaks reported in the United States. *J. Water Health* 4(Suppl. 2), 19–30. doi: 10.2166/wh.2006.016

Dufour, A. P. (1984). *Health Effects Criteria for Fresh Recreational Waters*. Cincinnati, OH: US Environmental Protection Agency.

FDA (2018). *Environmental Assessment of Factors Potentially Contributing to the Contamination of Romaine Lettuce Implicated in a Multi-State Outbreak of E. coli O157:H7*. Silver Spring, MD: Food and Drug Administration.

Feng, S., Ahmed, W. A.-O., and McLellan, S. A.-O. (2020). Ecological and technical mechanisms for cross-reaction of human fecal indicators with animal hosts. *Appl. Environ. Microbiol.* 86:e02319-19.

Food and Drug Administration (2015). *Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption*. Silver Spring, MD: Food and Drug Administration.

Francy, D. S., Stelzer, E. A., Duris, J. W., Brady, A. M. 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. doi: 10.1128/aem.02995-12

Grayson, R. B., Gippel, C. J., Finlayson, B. L., and Hart, B. T. (1997). Catchment-wide impacts on water quality: the use of 'snapshot' sampling during stable flow. *J. Hydrol.* 199, 121–134. doi: 10.1016/s0022-1694(96)03275-1

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., 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., Haugland, R. A., Varma, M., Millen, H. T., Borchardt, M. A., Field, K. G., et al. (2014a). 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

Green, H. C., 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., 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

Green, H. C., White, K. M., Kelty, C. A., and Shanks, O. C. (2014b). Development of rapid canine fecal source identification PCR-based assays. *Environ. Sci. Technol.* 48, 11453–11461. doi: 10.1021/es502637b

Hanselman, T., Graetz, D., and Wilkie, A. (2003). Manure-borne estrogens as potential environmental contaminants: a review. *Environ. Sci. Technol.* 37, 5471–5478. doi: 10.1021/es034410\%2B

Henningsen, A. (2017). *censReg: Censored Regression (Tobit) Models. R package version 0.5*. Available online at: http://CRAN.R-Project.org/package=censReg

Hothorn, T., Hornick, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 652–674.

Hothorn, T., and Zeileis, A. (2015). partykit: a modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* 16, 3905–3909.

Kane, D. D., Conroy, J. D., Peter Richards, R., Baker, D. B., and Culver, D. A. (2014). Re-eutrophication of Lake Erie: correlations between tributary nutrient loads and phytoplankton biomass. *J. Great Lakes Res.* 40, 496–501. doi: 10.1016/j.jglr.2014.04.004

Karkman, A., Parnanen, K., and Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat. Commun.* 10:80.

King, R. S., Baker, M. E., Whigham, D. F., Weller, D. E., Jordan, T. E., Kazyak, P. F., et al. (2005). Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecol. Appl.* 15, 137–153. doi: 10.1890/04-0481

Kirtane, A., Wilder, M., and Green, H. C. (2019). Development and validation of rapid environmental DNA (eDNA) detection methods for bog turtle (*Glyptemys muhlenbergii*). *PLoS One* 14:e0222883. doi: 10.1371/journal.pone.0222883

Kumaragamage, D., and Akinremi, O. O. (2018). Manure phosphorus: mobility in soils and management strategies to minimize losses. *Curr. Pollut. Rep.* 4, 162–174. doi: 10.1007/s40726-018-0084-x

Mieszkin, S., Yala, J. F., Joubrel, R., and Gourmelon, M. (2010). Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J. Appl. Microbiol.* 108, 974–984. doi: 10.1111/j.1365-2672.2009.04499.x

Oliver, D. M., Bartie, P. J., Louise Heathwaite, A., Reaney, S. M., Parnell, J. A. Q., and Quilliam, R. S. (2018). A catchment-scale model to predict spatial and temporal burden of *E. coli* on pasture from grazing livestock. *Sci. Total Environ.* 616-617, 678–687. doi: 10.1016/j.scitotenv.2017.10.263

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

Porter, K. D. H., Quilliam, R. S., Reaney, S. M., and Oliver, D. M. (2019). High resolution characterisation of *E. coli* proliferation profiles in livestock faeces. *Waste Manag.* 87, 537–545. doi: 10.1016/j.wasman.2019.02.037

Ritz, C., and Spiess, A. (2008). qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* 24, 1549–1551. doi: 10.1093/bioinformatics/btn227

Sharpley, A. N., Daniel, T., Sims, T., Lemunyon, J., Stevens, R., and Parry, R. (2003). *Agricultural Phosphorus and Eutrophication*. Washington, DC: U.S. Department of Agriculture, 44.

Soller, J. A., Schoen, M. E., Varghese, A., Ichida, A. M., Boehm, A. B., Eftim, S., et al. (2014). Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. *Water Res* 66, 254–264. doi: 10.1016/j.watres.2014.08.026

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O. C. (2017). Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154. doi: 10.1021/acs.est.7b02703

Stocker, M. D., Rodriguez-Valentín, J. G., Pachepsky, Y. A., and Shelton, D. R. (2016). Spatial and temporal variation of fecal indicator organisms in two creeks in Beltsville, Maryland. *Water Qual. Res. J.* 51, 167–179. doi: 10.2166/wqrjc.2016.044

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 Prot.* 81, 1661–1672. doi: 10.4315/0362-028x.jfp-18-185

Tyers, M. (2017). *riverdist: River Network Distance Computation and Applications. R package version 0.15.3*. Available online at: https://CRAN.R-project.org/package=riverdist

U.S. Environmental Protection Agency (1991). *Guidance for Water Quality-Based Decisions: The TMDL Process*. Washington, DC: U.S. Environmental Protection Agency.

US EPA (2012). *Recreational Water Quality Criteria*. Washington, DC: United States Environmental Protection Agency.

Vadas, P. A., Good, L. W., Jokela, W. E., Karthikeyan, K. G., Arriaga, F. J., and Stock, M. (2017). Quantifying the impact of seasonal and short-term manure application decisions on phosphorus loss in surface runoff. *J. Environ. Qual.* 46, 1395–1402. doi: 10.2134/jeq2016.06.0220

VanWormer, E., Carpenter, T. E., Singh, P., Shapiro, K., Wallender, W. W., Conrad, P. A., et al. (2016). Coastal development and precipitation drive pathogen flow from land to sea: evidence from a *Toxoplasma gondii* and felid host system. *Sci. Rep.* 6:29252.

Verhougstraete, M. A.-O., 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

Weller, D., Belias, A., Green, H. C., Roof, S., and Wiedmann, M. (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

Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., et al. (2020). Complex interactions between weather, and microbial and physiochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water. *bioRxiv* [Preprint] bioRxiv: 2020.2001.2002.892851,

Weller, D. L., Love, T. M. T., Belias, A., and Wiedmann, M. (2020). 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. Sustain. Food Syst.* 4:561517. doi: 10.3389/fsufs.2020.561517

WHO (2016). *Protecting Surface Water for Health: Identifying, Assessing and Managing Drinking-Water Quality Risks in Surface-Water Catchments*. Geneva: WHO.

Wu, Z., Greaves, J., Arp, L., Stone, D., and Bibby, K. (2020). Comparative fate of CrAssphage with culturable and molecular fecal pollution indicators during activated sludge wastewater treatment. *Environ. Int.* 136, 105452. doi: 10.1016/j.envint.2019.105452

Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W. L., et al. (2005). RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4:e3. doi: 10.1371/journal.pbio.0040003

Author Disclaimer: The content is solely the responsibility of the authors and does not represent the official views of the NIH or any other government agency.

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.

Copyright © 2021 Green, Wilder, Wiedmann and Weller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fmicb.2021.710346

Monitoring Urban Beach Quality on a Summer Day: Determination of the Origin of Fecal Indicator Bacteria and Antimicrobial Resistance at Prophète Beach, Marseille (France)

*Mylène Toubiana[1], Christian Salles[1], Marie-George Tournoud[1], Patricia Licznar-Fajardo[1,2], Isabelle Zorgniotti[1], Marie-Laure Trémélo[3], Estelle Jumas-Bilak[1,2], Samuel Robert[3] and Patrick Monfort[1]**

[1] HydroSciences Montpellier, UMR 5151 Université de Montpellier, CNRS, IRD, Montpellier, France, [2] Département d'Hygiène Hospitalière, CHU Montpellier, Montpellier, France, [3] ESPACE, UMR 7300 Aix Marseille Université, Avignon Université, Université Côte d'Azur, CNRS, Aix-en-Provence, France

OPEN ACCESS

Edited by:
Anicet Ramon Blanch, University of Barcelona, Spain

Reviewed by:
Lorena Rodriguez-Rubio, University of Barcelona, Spain
Phil Giffard, Charles Darwin University, Australia

***Correspondence:**
Patrick Monfort
patrick.monfort@umontpellier.fr

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *16 May 2021*
Accepted: *26 July 2021*
Published: *25 August 2021*

Citation:
Toubiana M, Salles C, Tournoud M-G, Licznar-Fajardo P, Zorgniotti I, Trémélo M-L, Jumas-Bilak E, Robert S and Monfort P (2021) Monitoring Urban Beach Quality on a Summer Day: Determination of the Origin of Fecal Indicator Bacteria and Antimicrobial Resistance at Prophète Beach, Marseille (France). Front. Microbiol. 12:710346. doi: 10.3389/fmicb.2021.710346

A highly frequented beach in Marseille, France, was monitored on an hourly basis during a summer day in July 2018, to determine possible water and sand fecal pollution, in parallel with influx of beach users from 8 a.m. to 8 p.m. Fecal indicator bacteria were enumerated, together with four host-associated fecal molecular markers selected to discriminate human, dog, horse, or gull/seagull origins of the contamination. The antimicrobial resistance of bacteria in water and sand was evaluated by quantifying (i) the class 1, 2, and 3 integron integrase genes *intI*, and (ii) bla_{TEM}, bla_{CTX-M}, and bla_{SHV} genes encoding endemic beta-lactamase enzymes. The number of beach users entering and leaving per hour during the observation period was manually counted. Photographs of the beach and the bathing area were taken every hour and used to count the number of persons in the water and on the sand, using a photo-interpretation method. The number of beach users increased from early morning to a peak by mid-afternoon, totaling more than 1,800, a very large number of users for such a small beach (less than 1 ha). An increase in fecal contamination in the water corresponded to the increase in beach attendance and number of bathers, with maximum numbers observed in the mid-afternoon. The human-specific fecal molecular marker HF183 indicated the contamination was of human origin. In the water, the load of *IntI*2 and 3 genes was lower than *IntI*1 but these genes were detected only during peak attendance and highest fecal contamination. The dynamics of the genes encoding B-lactamases involved in B-lactams resistance notably was linked to beach attendance and human fecal contamination. Fecal indicator bacteria, integron integrase genes *intI*, and genes encoding B-lactamases were detected in the sand. This study shows that bathers and beach users can be significant contributors to contamination of seawater and beach sand with bacteria of fecal origin and with bacteria carrying integron-integrase genes and beta lactamase encoding genes. High influx of users to beaches is a significant factor to be considered in order to reduce contamination and manage public health risk.

Keywords: fecal indicator bacteria, microbial source tracking, antimicrobial resistance, beach water, sand, attendance

INTRODUCTION

In summer, the French Mediterranean coasts record very high attendance due to the existence of numerous beaches which attract tourists and locals for recreational activities related to relaxation (bathing and sunbathing, etc.) and water sports (swimming, snorkeling, kite-surfing, jet-skiing, and paddle boarding, etc.). However, these recreational activities may be reduced and/or prohibited, depending on the sanitary quality of the bathing water.

Those decisions are controlled by relevant authorities when the sanitary quality of the bathing waters does not comply with standards set by the European directive (2006/7/CE-02/2006). These standards rely on two microbiological parameters indicating fecal contamination, thermo-tolerant coliforms (TTC, including *Escherichia coli*) and intestinal enterococci (IE), whose abundance in the water is controlled under the responsibility of the Regional Health Agency (ARS), during the entire summer season but at a variable frequency, depending on the beach. Unlike water, the sanitary quality of the sand is neither controlled nor regulated, and its potential fecal contamination is not considered. When decided, a ban on swimming and nautical activities reduces the number of visitors to beaches and repeated bans may alter the reputation of the seaside resort or the city. It is therefore important to detect and control factors related to poor bathing water quality, to be able to prevent degradation and improve the sanitary condition of the beach.

It has been shown that crowded beaches have high bacterial abundances in both sand and water during hot and sunny summer conditions (Papadakis et al., 1997; Pereira et al., 2013), simply because of the presence of users on the beach. The behavior of users, as well as the sanitary equipment provided by municipalities (showers, toilets, and bins, etc.), can have an impact on the microbiological quality of beach water and sand. In addition, a sanitation network failure resulting in direct discharge of untreated domestic wastewater to coastal waters can be severely detrimental to the quality of bathing water, especially during summer thunderstorms with intense and localized rainfall (Laplace et al., 2009; Figueras et al., 2015). Finally, accidental discharge of sewage from yachts and wild or domestic animal feces are additional factors to consider (Brouwer and De Blois, 2008). Standard microbial indicators do not allow identification of human or animal origin of the fecal contamination. However, in the event of problems with water quality, it is important to determine the origin of the fecal input in order to make management decisions that reduce the sources of contamination. In this perspective, a number of microbial source tracking (MST) methods, were employed, including methodologies developed over the last two decades that target closely related host-specific microorganisms (Roslev and Bukh, 2011).

In this study, a well-recognized and highly attended beach in Marseille, France, was monitored hourly on a summer day to determine water and sand fecal pollution that could be correlated with influx of beach users. Standard microbial indicators were measured to assess the level of fecal pollution (TTC, IE), together with four host-associated fecal molecular markers, chosen in order to discriminate human, dog, horse or gull/seagull origin of the contamination. These trackers targeted the host-specific 16S rRNA gene from *Bacteroidales* that is specific to the fecal microbiota of humans (Bernhard and Field, 2000a,b; Seurinck et al., 2005), dogs (Dick et al., 2005b), and horses (Dick et al., 2005a), and from the bacterium *Catellicoccus marimammalium* specific to seagulls (Lu et al., 2008; Wu et al., 2017). However, because geographical variation in sensitivity and specificity of these tracers has been reported in the literature (Harwood et al., 2014), they were first tested on a regional panel of host feces. Antimicrobial resistance (AMR) is a serious public health concern, therefore the AMR of bacteria in water and sand was determined. This was done by determining: (i) class 1, 2, and 3 integron integrase genes *intI* (Barraud et al., 2010; Gillings et al., 2015) associated with transmission and dissemination of resistance genes in the bacterial community, and (ii) bla_{TEM}, $bla_{\mathrm{CTX-M}}$, and bla_{SHV} genes, encoding endemic beta-lactamase enzymes that lead to resistance to the beta-lactam antibiotic family (Xi et al., 2009; Salverda et al., 2010; Marti et al., 2013). Finally, the total bacterial load was evaluated by quantifying a conserved region of the 16S rRNA gene (Maeda et al., 2003).

MATERIALS AND METHODS

Study Area and Beach Attendance Assessment

Located in Marseille, the second largest town in France (nearly 900,000 inhabitants), the study area was Prophète beach. Backing onto the ledge, this sandy beach is open to the sea on either side of a dike running parallel to the coast and offers two bathing areas, only delineated as a supervised bathing zone (**Figure 1**). The beach area is approximately 7,000 m^2. It accommodates a beach volleyball court, a snack bar, toilets, showers, and a small office for bath supervisors. The locale also hosts a snack bar and a restaurant located under the coastal boulevard overlooking the beach.

On July 18, 2018, the number of users at the beach was determined using an observation protocol implemented between 8 a.m. and 8 p.m., i.e., 12 h of continuous observation. Because two stairways are the only entry to the beach by land (**Figure 1**), this characteristic served to provide a manual count of the number of beach users entering and leaving per hour during the observation period. In addition to this counting, photographic shots of the beach and the bathing area (**Figure 1**) were carried out every hour and were used to count the number of persons in the water and on the sand, using a photo-interpretation method back at the laboratory (Robert et al., 2008).

Sampling and Processing of Water and Sand Samples

Sampling and Numeration of Fecal Indicator Bacteria

Samples of approximately 200 mL of water collected at each of ten sampling points spaced evenly across the supervised bathing area were combined in a sterile bottle (**Figure 1**). Sand was collected at ten points located at the water edge in the area of the surf. These points were spaced evenly across the beach. Sand samples were obtained using a sterile container to a final weight of approximately 250 g. At the time of sampling, the bathing



FIGURE 1 | Aerial view of the Prophète Beach in Marseille, Mediterranean coast, South of France. The only two pedestrian accesses by stairs, Northern and Southern, are indicated. On each side of the beach, two bathing areas (BA) are present: a supervised one (SBA) and a non-supervised one (NSBA). In the SBA, ten sampling points separated by 5 m each were determined to collect seawater (white dots) and sand (white triangles). These 10 samples are pooled in one sample for analysis. Map data:© IGN BN Ortho HR 2017; HSM.

water temperature was measured. All samples were stored in refrigerated coolers for immediate processing. Analyses were done in a field laboratory installed in a prefab module provided by the City of Marseille and located at the Roucas Blanc Marina near Prophet Beach (approximately 1 km).

Fecal indicator bacteria (FIB) in the water were enumerated using reference methods ISO 9308-1 and ISO 7899-2 referenced in the European bathing water directive 2006/7/EC to enumerate thermo-tolerant coliforms (TTC) and IE, respectively (European Parliament, 2006). Water volumes (1, 10, and 100 mL) were filtered in duplicate on 0.45 μm pore size cellulose nitrate filters (Sartorius). TTC were quantified on triphenyl tetrazolium chloride and tergitol 7 lactose agar (Biokar Diagnostics) incubated for 24 h at 44.5°C, and IE were enumerated on Slanetz medium (Biokar Diagnostics) incubated for 48 h at 37°C. The results are expressed as colony forming units per 100 mL (CFU/100 mL).

For sand, the sample was thoroughly mixed and 100 g was suspended in a bottle with 400 mL of sterile water and shaken to release the bacteria. After a settling time (1 min) of the larger particles, 4 × 100 mL of the wash water was filtered on 0.45 μm pore size cellulose nitrate filters (Sartorius). Two filters were used to count TTC and two filters for IE. This step was repeated three times in order to extract the bacteria. After incubation, eight membranes were counted for TTC or for EI to obtain the CFU for 100 g of sand.

The extraction protocol for sand is time consuming and tedious. Also, measurement of FIB in sand was done three times at 8 a.m., midday, and 5 p.m., while the measurements for the water samples were hourly, from 8 a.m. to 8 p.m.

Total DNA Extraction

Seawater samples: 500 mL of each composite water sample was filtered in triplicate through 0.2 μm pore size cellulose acetate filters (Sartorius) and filters were stored immediately at −20°C. Sand samples: 3 × 50 g of each composite sample was suspended in 200 mL of sterile water in three different bottles and were shaken to release the bacteria. The washing water was filtered using a 0.2 μm pore size cellulose acetate filter (Sartorius). Three more washes of the same three biological replicates of sand were made under the same conditions, and the wash water was filtered on the same cellulose acetate filter. In total, each biological replicate of 50 g of sand was washed using 800 mL of sterile water. Filters were immediately stored at −20°C until processing.

At the laboratory, total genomic DNA was extracted from filters using DNeasy PowerWater kit (Qiagen) according to the manufacturer's instructions and eluted in a final volume of 100 μL. Concentrations were measured on a Qubit fluorometer (Invitrogen) using the Qubit dsDNA HS assay kit (Invitrogen) following the manufacturer's instructions, and the DNA was stored at −20°C until qPCR analysis.

Fecal Sample Collection to Test for Host-Associated Molecular Markers

Four host-associated molecular markers were selected to determine the origin of fecal contamination: the human-specific HF183 *Bacteroides*-related, the dog-specific DF475 *Bacteroides*-related, the horse-specific HoF597 *Prevotella*-related and the gull/seagull-specific bacterium *Catellicoccus marimammalium* (**Table 1**). These markers were first tested using local fecal samples collected in 2018 in the Occitanie region, South of France, to determine sensitivity and specificity. Amplification tests were done by Real-time quantitative PCR (qPCR) on DNA extracted from fecal samples from 51 birds (11 species), including 19 gulls and seagulls, collected by the "Ligue de Protection des Oiseaux" (Hérault), from six dogs, collected by owners (Montpellier, Hérault) and from 15 horses, collected at a riding stable (Saint-Affrique, Aveyron). For birds, feces were sampled using a sterile swab, suspended immediately in 1 mL of sterile water and stored at 4°C. For dogs and horses, feces were collected directly in sterile tubes and stored at −20°C. Total DNA was purified from 1 mL of a feces suspension (birds) or from 50 mg of wet feces (dogs and horses), using the MasterPure™ Gram positive DNA purification kit (Epicentre) according to the supplier instructions (starting with feces pellet instead of bacteria). A total of

TABLE 1 | Primers and conditions used for end-point and real-time PCR quantifications.

Target	Primers (5′–>3′)	Final concentration (μM)	Amplicon size (bp)	Annealing temperature (°C)	References
Total bacteria					
16S rRNA gene	16SrRNA-F GTGSTGCAYGGYTGTCGTCA	0,4	147*	60	Maeda et al., 2003
	16SrRNA-R ACGTCRTCCMCACCTTCCTC	0,4			
Host-associated fecal molecular markers					
Human-specific *Bacteroides*-related HF183	HF183F ATCATGAGTTCACATGTCCG	0,4	83	60	Seurinck et al., 2005
	HF183R TACCCCGCCTACTATCTAATG	0,4			
Dog-specific *Bacteroides*-related DF475	DF475F CGCTTGTATGTACCGGTACG	0,4	251	56	Dick et al., 2005b
	Bac708R CAATCGGAGTTCTTCGTG	0,4			
Horse-specific *Prevotella*-related HoF597	HoF597F CCAGCCGTAAAATAGTCGG	0,5	127	56	Dick et al., 2005a
	Bac708R CAATCGGAGTTCTTCGTG	0,5			
Gull/Seagull-specific Sg2	Sg2-F TGCTAATACCGCATAATACAGAG	0,4	306	60	Wu et al., 2017
	Sg2-R CTATCGCTCCTGTTCTTCTCTAA	0,4			
Integron integrase genes					
Class 1-IntI1	intI1-LC1 GCCTTGATGTTACCCGAGAG	0,4	196	60	Barraud et al., 2010
	intI1-LC5 GATCGGTCGAATGCGTGT	0,4			
Class 2–IntI2	intI2-LC2 TGCTTTTCCCACCCTTACC	0,4	195	62	Barraud et al., 2010
	intI2-LC3 GACGGCTACCCTCTGTTATCTC	0,4			
Class 3–IntI3	intI3-LC1 GCCACCACTTGTTTGAGGA	0,4	138	62	Barraud et al., 2010
	intI3-LC2 GGATGTCTGTGCCTGCTTG	0,4			
Beta-lactam resistance genes					
bla_{TEM}	bla-TEM-F GCKGCCAACTTACTTCTGACAACG	0,4	247	60	Xi et al., 2009
	bla-TEM-R CTTTATCCGCCTCCATCCAGTCTA	0,4			
bla_{SHV}	bla-SHV-F CGCTTTCCCATGATGAGCACCTT	0,4	110	60	Xi et al., 2009
	bla-SHV-R TCCTGCTGGCGATAGTGGATCTT	0,4			
bla_{CTX-M}	bla-CTX-M-F CTATGGCACCACCAACGATA	0,4	103	62	Marti et al., 2013
	bla-CTX-M-R ACGGCTTTCTGCCTTAGGTT	0,4			

**Length of the amplicon obtained with positive control bacterial strain.*

55 DNA extracts from human fecal samples (anonymous patients) were obtained from a university hospital center. DNA extractions were obtained using a Seegene Nimbus Automated Liquid Handling Workstation. Quantities and qualities of DNAs were measured by spectrophotometry using a NanoDrop One (Thermo Scientific). All DNA samples were stored at −20°C until used.

Gene Quantifications by Real-Time Quantitative PCR

Quantitative PCR Reactions

The absolute number of genes was determined by real-time PCR (qPCR) in 96-well plates using a LightCycler® 480 (Roche). Each qPCR reaction employed Syber Green, was run in duplicate,

and contained 1X Luna® Universal qPCR Master Mix (New England BioLabs), 0.4–0.5 μM of each specific primer (**Table 1**), 1 μL of sample DNA (extracted from a fecal sample, seawater, or sand), and sterile water to a final volume of 10 μL. Sterile water was used instead of DNA as the negative control. After heating for 10 min at 95°C for activation of the hot-start DNA polymerase and denaturation of DNA, reactions were carried out for 45 cycles of 95°C for 10 s, 56–62°C for 10 s, and 72°C for 10 or 15 s (depending on the amplicon size, **Table 1**). Finally, qPCR products were gradually heated from 65 to 95°C to obtain a dissociation curve in order to determine the temperature of melting and the specificity of the amplicons. Bacterial type strains *Bacteroides dorei* (DSM 17855, DSMZ) and *Catellicoccus marimammalium* (DSM 18331, DSMZ) were used as positive control for the human-specific and the gull/seagull-specific markers, respectively. Environmental and clinical multi-drug resistant *Klebsiella pneumoniae* and *E. coli* strains (isolated from previous studies in Montpellier city, France) were used as positive control for total 16S rRNA, *intl* and *bla* genes. Total DNA was extracted from fresh cultures using the MasterPure™ Complete or Gram-positive purification kit (Epicenter) according to the manufacturer instructions. For dog and horse associated markers, DNA from two positive fecal samples was used as positive control.

Standard Curves and Determination of the Limits of Quantification

All specific amplicons were sequenced to ensure their identity (Genewiz, Germany) and cloned using TOPO® TA cloning® kit (Invitrogen) with pCR™4-TOPO® cloning vector, according to the manufacturer's instructions. After cloning, plasmids were purified using NucleoSpin Plasmid kit (Marcherey-Nagel) and linearized by enzymatic digestion using *Pst*1 (New England Biolabs). Finally linearized plasmids were purified using Monarch PCR and DNA Cleanup Kit (New England BioLabs), and concentrations measured on spectrophotometer NanoDrop One (Thermo Scientific) in order to calculate the number of plasmid copies per microliter. Then, 10-fold serial dilutions of each plasmid were made in 10 μg/mL sonicated salmon sperm DNA (Sigma) and these dilutions were used in qPCR to determine the standard curves, to establish the amplification efficiencies and to calculate the absolute quantifications of the genes in environmental samples. Based on standard curve results, the limits of quantification (LOQ) of the assays were determined for each gene as the lowest concentration of marker within the linear range of quantification during amplification.

Calculations and Statistics

The sensitivity (percentage of true positives among fecal samples of the targeted hosts) and specificity (percentage of true negatives among fecal samples of non-targeted hosts) were calculated for each host-associated molecular marker according to qPCR results.

For gene quantifications, results were transformed in number of gene copies for 1 mL of water sample (GC/mL) (or for 1 g of wet sand) and the final result was calculated as the arithmetical mean of the three biological replicates. The relative quantities of targeted genes were calculated as the ratio between the number of copies of each gene to the total number of 16S rRNA gene copies, in order to control the reproducibility and quality of DNA extractions and normalize their abundance in the collected samples.

The normality of the data distributions was tested by using a Shapiro-Wilk test, and equality of variances tested by using a Fisher test. In order to determine the correlations between variables, the Spearman's rank correlation coefficients (rs) were calculated and their nullity tested. All statistics were done using the GraphPad Prism software V 5.03. Test results were considered significant when the associated p-value was at least ≤ 0.05.

RESULTS

Beach User Influx on July 18, 2018

Weather conditions on July 18, 2018, were typical still air and sunny summer day. The average daily temperature from the nearby Marseille Corniche weather station was 26°C, bathing water temperature in the morning was 22°C and warmed to close to 25°C in the afternoon.

The total number of persons counted at the two entrances of the beach between 8 a.m. and 8 p.m. was 5,514 (Robert and Trémélo, 2019). This attendance is the highest ever at the Prophète beach, where such counts have been made every summer in July since 2016. As shown in **Figure 2**, the number of persons counted from the hourly photographs was lower than the count at the two entrances, which is not surprising. In photos, users can be hidden by beach umbrellas, tents, or other users, in particular when the beach is overcrowded. In water, counts made from the photographs are more reliable as the bathing zone is at the foreground, improving identification of individuals.

From counts at the two entrances, the number continued to increase from 8 a.m. to 4 p.m., reaching a little approximately 1,800 persons between 4 p.m. and 5 p.m. In the late afternoon and early evening, attendance decreased but remained high, with nearly 1,200 persons still at the beach at 8 p.m. From the photographs, users in the water in both the supervised and non-supervised bathing areas (SBA/NSBA) were present at a rather high level also. Between 2 p.m. and 7 p.m., more than 150 bathers could be counted every hour, reaching a maximum of 322 persons at 4 p.m., with more than two thirds in the supervised bathing area. During the attendance peak, nearly one beach user of six was in the water.

Fecal Indicator Bacteria Levels in Seawater and Sand

Quantities of TTC and IE numerated in seawater and sand are shown in **Table 2**. Concerning TTC, levels varied throughout the day in seawater, with low quantities (60–160 CFU/100 mL) from 8 a.m. to 3 p.m., and high levels estimated at over 500 CFU/100 mL (uncountable) from 4 p.m. to 8 p.m. These TTC loads exceeded the limits determined by the European directive 2006/7/CE-02/2006 (500 *E. coli*/100 mL) for sufficient water quality for bathing. In the sand, quantities were 44 TTC/100 g at 8 a.m., 108 at midday, and 360 at to 5 p.m.

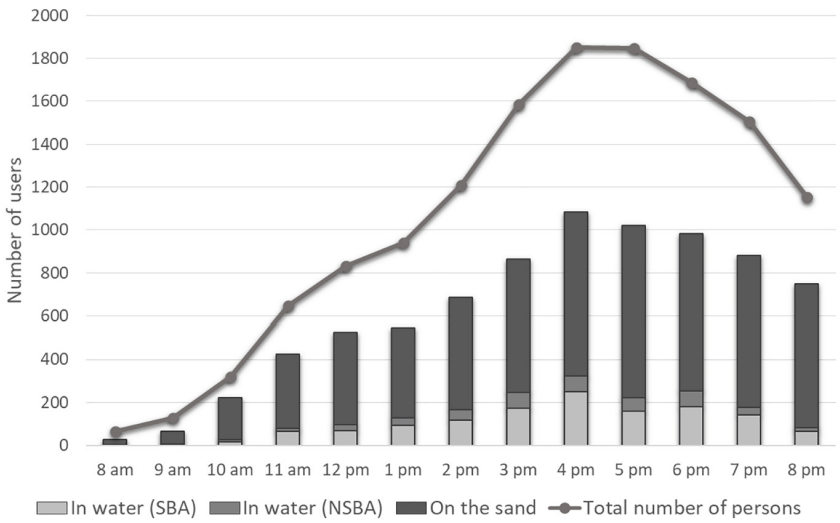


FIGURE 2 | Influx of users on July 18th 2020 on the Prophète beach. Number of users in the water in the supervised and non-supervised bathing areas (SBA and NSBA) and number of users on the sand were determined by photo-interpretation. Total number of persons was determined by the counts of entries and exits at the two only accesses (stairs).

TABLE 2 | Fecal indicator bacteria (FIB) levels in seawater and sand.

	Thermo-tolerant coliforms		Intestinal enterococci	
Hours of sampling	**Seawater (CFU/100 mL)**	**Wet sand (CFU/100 g)**	**Seawater (CFU/100 mL)**	**Wet sand (CFU/100 g)**
8 a.m.	160	44	67	326
9 a.m.	135	–	72	–
10 a.m.	N.A.	–	63	–
11 a.m.	N.A.	–	60	–
Midday	N.A.	108	54	270
1 p.m.	70	–	67	–
2 p.m.	65	–	93	–
3 p.m.	60	–	30	–
4 p.m.	*>500*	–	47	–
5 p.m.	*>500*	360	37	214
6 p.m.	*>500*	–	73	–
7 p.m.	*>500*	–	64	–
8 p.m.	*>500*	–	64	–

For seawater, quantities above the thresholds for sufficient quality for bathing are indicated in italics.
N.A., not accountable; –, not measured.

Concerning IE, quantities remained low in seawater during the day, with variations from 30 to 93 IE/100 mL. These results did not exceed 185 IE/100 mL which is the limit for sufficient water quality for bathing. In the sand, quantities were 326 IE/100 g at 8 a.m., 270 at midday, and 214 at to 5 p.m.

Origin of the Fecal Contamination

Sensitivity and Specificity of Host-Associated Fecal Molecular Markers

The detection results, obtained by qPCR, of host-associated fecal molecular markers in fecal samples are provided in **Table 3**. Of the 55 human fecal samples, 31 were positive for presence of the human-specific HF183 marker, which established its sensitivity at 56.36%. This marker was never detected in non-human fecal samples. Therefore, HF183 marker displayed moderate sensitivity of 56.36% and a high specificity of 100%. Dog-specific DF475 marker showed 83.33% sensitivity and 98.34% specificity due to its detection in two gull feces samples. Horse and gull/seagull-specific markers were both highly sensitive (100%) and highly specific (99.1 and 96.29%, respectively). Indeed, the horse-specific marker was found in one gull fecal sample, and the gull/seagull marker was detected in four fecal samples from other bird species. These tests were done on a small number of pure fecal samples. Nevertheless, due to their high sensitivity and specificity, the use

TABLE 3 | Sensitivity and specificity of host-associated fecal molecular markers.

	Host-associated fecal molecular markers			
Samples	**Human-specific HF183**	**Dog-specific DF475**	**Horse-specific HoF597**	**Gull/Seagull-specific Sg2**
Human ($n = 55$)	31/55 (56,36%)	0/55 (0%)	0/55 (0%)	0/55 (0%)
Dog ($n = 6$)	0/6 (0%)	5/6 (83,33%)	0/6 (0%)	0/6 (0%)
Horse ($n = 15$)	0/15 (0%)	0/15 (0%)	15/15 (100%)	0/15 (0%)
Gull/Seagull ($n = 19$)	0/19 (0%)	2/19 (10,53%)	1/19 (5,26%)	19/19 (100%)
Other birds ($n = 32$)	0/32 (0%)	0/32 (0%)	0/32 (0%)	4/32 (12,5%)
Sensitivity (r)	56,36%	83,33%	100,00%	100,00%
Specificity (s)	100,00%	98,34%	99,10%	96,29%

For each marker, the number of positive samples obtained by qPCR among each panel of samples are given, and corresponding percentages are indicated into brackets.

of these four markers was validated to quantify the origin of the fecal pollution in environmental samples in this study.

Quantification of Total 16S rRNA Gene and Host-Associated Fecal Molecular Markers in Seawater and Sand

The qPCR amplification efficiencies and LOQ obtained for the four host-associated fecal molecular markers and the total 16S rRNA gene are reported in **Supplementary Table 1**. Amplification efficiencies were excellent, ranging from 92 to 96.9%. The lowest LOQs were for the dog, horse and gull/seagull-specific markers, with a limit of 10 gene copy number (GC) per reaction, corresponding to 2 GC/mL of seawater and 20 GC/g of sand. The LOQ for the human-specific marker was a little higher, with 50 GC per reaction, corresponding to 10 GC/mL of seawater and 100 GC/g of sand. Finally, the highest LOQ was for total 16S rRNA gene, with a limit of 60 GC reaction, corresponding to 12 GC/mL and 120 GC/g of sand.

In seawater, the total number of 16S rRNA gene, representative of the total bacterial load, fluctuated during the day (**Figure 3**): levels first decreased during the morning from 8 a.m. (1.32×10^5 GC/mL) to 11 am (5.37×10^3 GC/mL), and then increased from 12 p.m. (4.59×10^4 GC/mL) to reach a maximum value of 3.05×10^5 GC/mL at 5 p.m. Levels then decreased until 8 p.m. (1.5×10^3 GC/mL). In the sand, the highest quantities were measured at 8 a.m., with 6.63×10^5 GC/g. The total load decreased at midday, with 3.87×10^4 GC/g, and increased at 5 p.m. to 2.69×10^5 GC/g (data not shown).

The human-specific fecal molecular marker HF183 (**Figure 3**) was first detected in seawater at 2 p.m. (10.6 GC/mL), and slowly increased until 5 p.m. (243 GC/mL). Levels then decreased from 5 p.m. to 7 p.m. (27 GC/mL) and the marker was not detected anymore at 8 p.m. Relative quantities of this marker (ratio on the total number of 16S rRNA gene copies), representative of its abundance in total bacterial community, increased in the same way during the afternoon from 2 p.m. to a maximum at 6 p.m., and finally decreased until 7 p.m. In the sand, the human-specific marker was only detected at 5 p.m. (<LOQ). The quantities of the human-specific marker in seawater were strongly positively correlated with the total number of persons at the beach ($rs = 0.908$, $p < 0.001$), in the water ($rs = 0.872$, $p < 0.001$) and on the sand ($rs = 0.89$, $p < 0.001$), with maximum values occurring between 4 p.m. and 6 p.m., when the number of persons at the beach was also maximum. This marker was also correlated with the total bacterial load ($rs = 0.603$, $p < 0.05$).

The presence of the dog and gull/seagull-specific markers were detected (<LOQ) in seawater approximately between 8 a.m. and 10 a.m., and between 2 p.m. and 6–7 p.m. (**Figure 3**). The dog-specific marker was also detected at midday. These two markers were absent in the sand. Finally, the horse-specific marker was not detected either in seawater or in the sand.

Antimicrobial Resistance of Bacteria in Seawater and Sand

The qPCR amplification efficiencies and LOQ for the integron-integrase genes (*Intl*1, 2, and 3) and beta lactamase encoding genes (bla_{TEM}, bla_{SHV}, and bla_{CTX-M}) are reported in **Supplementary Table 1**. The amplification efficiencies were also excellent, ranging from 96.1 to 102.4%. Efficiencies over 100% were surely due to minor pipetting errors but stayed in a range of good quality efficiencies. The LOQ for these six genes were 5 GC per reaction, which corresponded to 1 GC/mL of seawater and 10 GC/g of sand.

Quantification of Integron Integrase Genes (*Intl*)

In seawater, *Intl*1 gene was present almost all the day, except at 11 a.m. and at 8 p.m. (**Figure 4**). Measured quantities varied from 125 GC/mL at 8 a.m. to 27 GC/mL at 10 a.m., and increased gradually in the afternoon from 5 GC/mL at midday to a maximum of 114 GC/mL at 5 p.m. Densities then decreased at 6 p.m. and 7 p.m. (12 and 25 GC/mL), until no longer detected at 8 p.m. *Intl*1 gene density was higher in sand than in seawater. A maximum of 3,823 GC/g was detected at 8 a.m. Densities then decreased significantly at midday, with only 9 GC/g, and then increased again to 853 GC/g at 5 p.m. The relative amounts of *Intl*1 gene varied accordingly with the GC/mL in seawater (**Figure 4**) and with the GC/g in the sand (data not shown). Moreover, the quantities of *Intl*1 gene and of total 16S rRNA gene were strongly correlated ($rs = 0.927$, $p \leq 0.001$), which suggested that the variations in *Intl*1 gene quantities were proportional to the total bacterial concentration.

The density of *Intl*2 and *Intl*3 genes was significantly lower than *Intl*1 gene. The dynamics and load of both *Intl*2 and *Intl*3 genes in seawater were similar (**Figure 4**) and were significantly correlated ($rs = 0.903$, $p \leq 0.001$): both genes were detected (1 GC/mL or below the LOQ) from 8 to 11 a.m. and from 1 to 3 p.m. Both became quantifiable from 4 to 6 p.m., with quantities varying from 3 (*Intl*2) and 5 GC/ml (*Intl*3) to 1 GC/mL. These genes were still present at 7 p.m. but not at 8 p.m. In the sand, *Intl*2 and *Intl*3 genes were only slightly numerous (<LOQ) at 8 a.m. and in the afternoon, respectively. When regarding the relative abundance of both *Intl* 2 and 3 genes in the total bacterial community in seawater (**Figure 4**) a huge increase was observed at 6 p.m. but this could be artificial due to the very small quantities of *Intl* 2 and 3 genes measured. It is worth mentioning that the quantities of these two genes were significantly correlated with quantities of the human-specific marker HF183 ($rs = 0.796$ and 0.664, $p \leq 0.001$ and 0.05), and also for *Intl*2 gene to the number of persons in the water and on the sand ($rs = 0.684$ and 0.739, $p \leq 0.01$).

Quantification of Beta Lactamase Encoding Genes (*bla*)

The variations of the quantities of bla_{TEM} and bla_{CTX-M} genes in seawater were almost the same during the day (**Figure 5**): these genes were present (<LOQ) from 8 a.m. to 2 p.m. (bla_{TEM}) or to 1 p.m. (bla_{CTX-M}). At 3 p.m., bla_{CTX-M} was still present and bla_{TEM} began to be quantifiable, with almost 2 GC/mL. At 4 p.m., the measured quantities of both genes were maximum, with 157 and 110 GC/mL for bla_{TEM} and bla_{CTX-M} respectively, then the numbers slightly decreased to 39 and 84 GC/mL at 6 p.m., respectively. At 7 p.m., the numbers were much lower (1–3 GC/mL), and both genes were only detectable at 8 p.m. The proportion of these genes in the total flora (relative quantities,

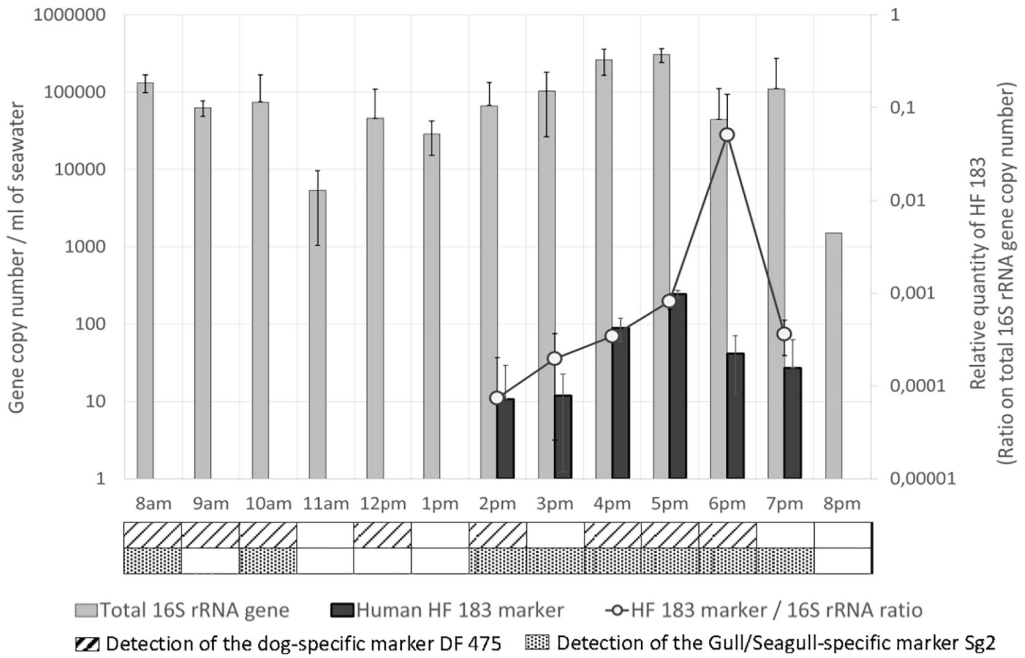


FIGURE 3 | Total 16S rRNA gene and host associated molecular markers levels in seawater. Quantities are given in gene copy number for 1 mL of seawater for the total 16S rRNA gene and the human-specific associated marker HF 183. Relative quantities of HF 183 are indicated on the secondary Y-axis. The detection (<Limit of quantification, LOQ) of the dog and the gull/seagull-specific associated markers are indicated in the table below the graph.

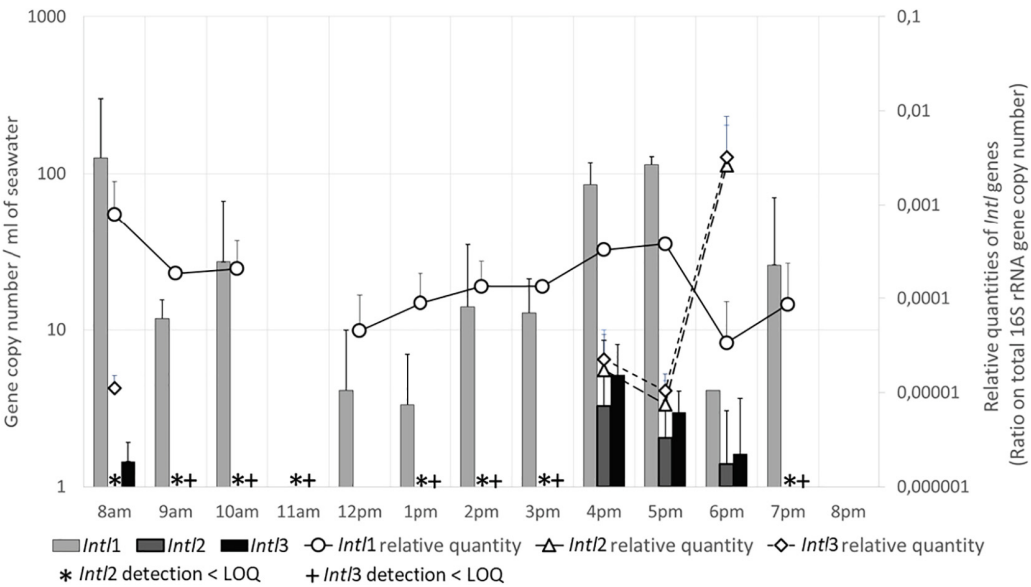


FIGURE 4 | Integron integrase genes (*Intl*) levels in seawater. Quantities are given in gene copy number for 1 mL of seawater. Relative quantities of *Intl* genes are indicated on the secondary Y-axis. LOQ, limit of quantification.

Figure 4) was increasing at 4 p.m., and was maximum at 6 p.m. In the sand, bla_{TEM} gene quantity was 14 GC/g at 8 a.m., and both genes were detected (<LOQ) all day.

The numbers of bla_{SHV} gene were much lower in seawater (**Figure 4**): it was detected only at 8 a.m., midday and 1 p.m., and was quantifiable between 3 p.m. to 7 p.m. (1 GC/mL) with a maximum value at 5 p.m. with 6 GC/mL. The relative number was maximum at 6 p.m. In the sand, this gene was detected only at 8 a.m.

The number of these three *bla* genes in seawater was positively correlated (rs = 0.781–0.926, $p \leq 0.01$–0.001). Maximum numbers were recorded between 4 and 6 p.m., when the number

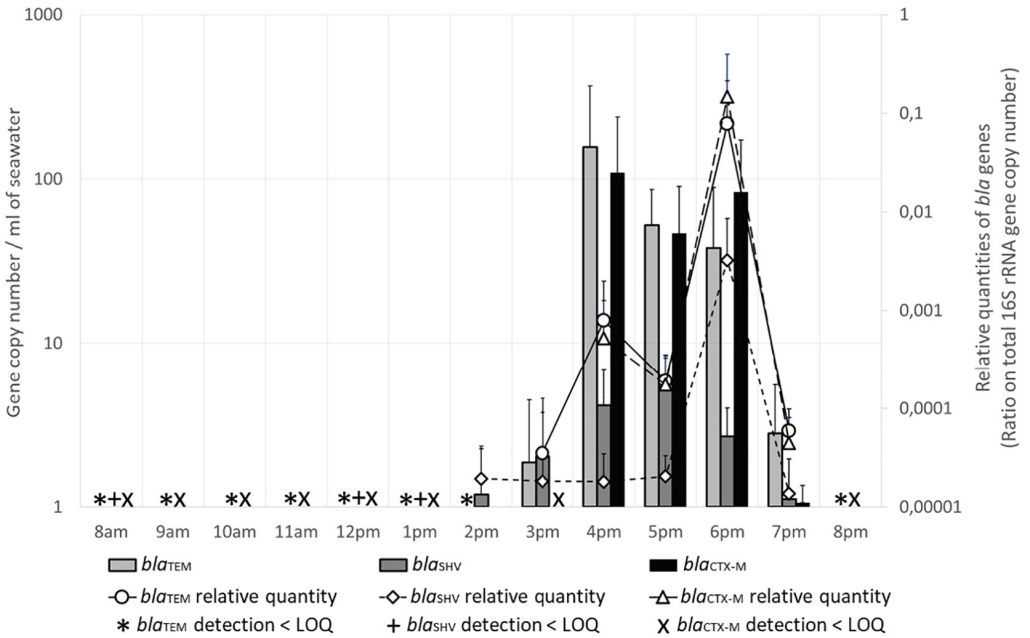


FIGURE 5 | Beta lactamase encoding genes (*bla*) levels in seawater. Quantities are given in gene copy number for 1 mL of seawater. Relative quantities of *bla* genes are indicated on the secondary Y-axis. LOQ, limit of quantification.

of users at the beach, the levels of the human-specific marker HF183, and the number of *Intl*2 and *Intl*3 genes were also at their maximum. Moreover, all these parameters were significantly correlated (rs = 0.62–0.947, $p \leq 0.05$ to ≤ 0.001).

DISCUSSION

The study was carried out during a very busy summer day at the Prophète Beach, an urban beach located in Marseille, a very large city. The number of beach users increased from early morning to a peak by mid-afternoon, as generally reported in other studies of beach attendance in summer time (Robert et al., 2008; Balouin et al., 2014). The peak of users in the middle of the afternoon was particularly significant (more than 1,800 people) for such a small beach (less than 1 ha).

The increase in fecal contamination (TTC) in the water followed the increase in beach attendance and bathers with maximum numbers of users observed mid-afternoon. However, the concentration of IE remains stable and does not increase with increasing TTC. The relationship between IE and TTC concentrations is not a rule. It has been shown that the concentrations of enterococci in the coastal estuarine/marine beach study area were largely controlled by particle-associated enterococci and mammal fecal source input (Rothenheber and Jones, 2018). The human-specific marker indicates that contamination was of human origin. The presence of bird- and dog-specifics markers was very low and did not contribute significantly to the fecal contamination observed in the water.

We demonstrated that monitoring fecal indicators and specific markers throughout the day and in presence of people using the beach provides valuable information for evaluating public health risk. Indeed, weekly to bi-weekly monitoring of water quality by health authorities is usually accomplished in the early morning when the beach is sparsely attended. The day this study was done, the Prophète beach was open, which means that surveillance showed the microbiological quality of the water was within the limits of the European directive. However, the mid-afternoon peak exceeded those limits and exposed bathers to potential health risk. The observed concentrations of *Bacteroides*-related HF183 marker represent a health risk for bathers as shown by Boehm et al. (2015). The dynamics of fecal indicators throughout the day could help define policies and/or remediation practices for more safe beach use.

Remediation efforts can be hindered because classical water quality surveillance does not take into account whether contamination originated from humans or animals. We determined the most likely origin of the fecal contamination by using markers previously described as efficient for origin tracing (Hughes et al., 2017). In particular, *Bacteroides*-related HF183 marker is considered the most sensitive measure of human fecal pollution (Hughes et al., 2017).

The analysis used host-associated markers of fecal samples and showed 100% specificity and 56.36% sensitivity for the human-specific *Bacteroides*-related HF183 marker. This sensitivity for the HF183 marker is consistent with other studies in France, with 54 and 62.5% sensitivity observed in Brittany (Mieszkin et al., 2009; Mauffret et al., 2012), and in other countries, e.g., in California, United States (Kildare et al., 2007), and in Kenya, Africa (Jenkins et al., 2009) with 61.1 and 58.3% sensitivity, respectively. For dog-specific *Bacteroides*-related DF475, horse-specific *Prevotella*-related HoF597, and Gull/Seagull-specific Sg2

markers, the results showed very high specificity and sensitivity. So, these four markers proved to be efficient and of great interest for fecal contamination tracking (Roslev and Bukh, 2011; Tran et al., 2015).

Various studies have shown that bathers contribute to contamination of water, namely with *Staphylococcus aureus*, enteric viruses, protozoan parasites, IE, and fecal coliforms (Breittmayer and Gautier, 1978; Gerba, 2000; Elmir et al., 2007). In particular, it has been shown that bathers may develop enteric infections from contamination with stools of other bathers (Keene et al., 1994). It has been shown that the behavior of toddlers can be a main cause of contamination. Indeed, during this study, we observed parents encouraging children under 5 to urinate and defecate in the water, and others collected baby diapers after wringing them out in seawater. Gerba (2000) showed in a review of the literature that bathers of all ages shed enteric microorganisms through normal recreational water contact or accidental fecal release. In our study of the Prophète beach, the same issues were highlighted. Overcrowding of such a small beach can lead to fecal contamination of bathing water, very likely caused by bathers.

Antibiotic resistance is a rising threat for human health and it is now recognized that a one-health based approach should be used to understand, prevent and treat AMR. In this context, spread in the marine environment is poorly understood. Therefore, besides the risk of contamination of recreational water by microorganisms able to cause enteric infections, the presence of human-originating bacteria raises the question of AMR and the potential role of recreational waters in diffusion of AMR. In this study, we tested a recognized marker of AMR in the environment, i.e., class 1 integron-integrase gene (*Intl*1) (Gillings et al., 2015). The *Intl*1 gene is widely used as a proxy for anthropogenic pollution (Gillings et al., 2015). Other genes, such as the sulphonamide resistance gene (*sul*2), are classical markers for AMR. It is noteworthy, that the intI1 gene is detected only in a small fraction of drug-resistant bacteria of health concern (Rodríguez et al., 2021) and the *sul*2 gene conferred resistance to a class of antimicrobial agents practically unused today in human medicine. For these reasons, we detected in this study the *Intl*2 and 3 genes, as well as three genes conferring resistance to B-lactams, the most used class of antibiotics in human and animal medicine. The *Intl*1, 2, and 3 genes are described as frequently occuring in the terrestrial environment and less frequently in the sea (Abella et al., 2015). In the water at Prophète Beach, the *Intl*1 load was not totally correlated with human presence or human fecal contamination. The load of *Intl*2 and 3 genes was lower than of *Intl*1 but these genes were detected only during peak attendance at the beach and with fecal contamination. This suggests the *Intl*2 and 3 genes may be more specific than *Intl*1 to detect human contamination in marine beach areas. *Intl*1 has been reported to be present in plankton-associated bacterial communities in the ocean (Di Cesare et al., 2018) and involved in bacterial adaptation not only to antimicrobial agents but also to hydrocarbon and metals (Gillings et al., 2015). Urban pollution of Prophète Beach may explain the high load of *Intl*1 and its dynamics, which is not fully correlated with human attendance.

The dynamics of genes encoding B-lactamases involved in B-lactam resistance was observed to be correlated with attendance and human fecal contamination. This important result suggests B-lactamass encoding genes may be valuable markers for AMR in the environment and also for human fecal contamination. B-lactamases encoding genes may also provide data for environment-based epidemiology because they are involved in resistance to the most used class of antibiotics. For instance, the load of $bla_{\mathrm{CTX-M}}$ observed mid-afternoon at the Prophète Beach is of particular concern because $bla_{\mathrm{CTX-M}}$ encodes an extended spectrum B-lactamase that confers multi-drug resistance to the enterobacteria. The spread of $bla_{\mathrm{CTX-M}}$, first in hospital and then in the community, is supposedly because extended spectrum B-lactamases are increasingly detected in human or animal infections caused by enterobacteria. A high load of $bla_{\mathrm{CTX-M}}$ in a water environment integrating contamination from numerous individuals demonstrated spread of $bla_{\mathrm{CTX-M}}$ within the entire population, including healthy individuals.

Fecal indicator bacteria, integron integrase genes *intI*, and genes encoding B-lactamases were present in the sand. There is some work on the presence and quantification of FIB in sand (Wheeler Alm et al., 2003; Yamahara et al., 2012; Sabino et al., 2014; Whitman et al., 2014). Although contamination in sand is of increasing interest, bathing regulations are based only on the quality of bathing water. In addition, there are no standard methods for monitoring and comparing results obtained around the world. The results of this study show that sand can contribute to contamination by beach users. However, more work needs to be done to understand this relationship between water and sand contamination.

This study shows that bathers and beach users are significant contributors to the contamination of seawater and beach sand with bacteria of fecal origin and also with bacteria carrying integron-integrase genes and beta lactamase encoding genes. This finding contributes to the understanding of an epidemiological risk that is not currently taken into account in sanitary approaches to the quality of bathing water and beaches. At present, the current rules and laws for managing the sanitary quality of bathing water are based on studies that consider possible sources of contamination as only from the beach watershed, particularly during heavy rains, as is the case for the urban beach of Marseille (Laplace et al., 2009). Our study concludes that high influx of users on beaches must be taken into account because beach users and bathers also may contribute to sea water and beach sand microbial contamination.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

M-GT, CS, SR, and PM led on conceptualization of the study and participated in the sampling of seawater and sand. M-LT and SR

carried out the counts of the users of the beach. MT, PL-F, IZ, and PM carried out the microbiological and molecular biology analyzes. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the LabEx DRIIHM, French Programme "Investissements d'Avenir" (ANR-11-LABX-0010) which is managed by the French ANR, under the Human-Environment Observatory "Mediterranean coast" (OHM Littoral Méditerranéen).

ACKNOWLEDGMENTS

The authors thank the City of Marseille (Directorate of Sea) for logistical support in providing technical rooms and access to sites. The authors would like to thank the master's students who contributed to the counting of beach users and to the microbiological analyzes: Brieuc Cabioch, Hadrien Di Costanzo, and Jules Grillot (Master of Geography at Aix Marseille University), and Emilie Nifaut and Arnaud Senegas (Master IMHE at Montpellier University). The authors would like to thank Philippe Clair, the manager of the qPCR facility of Université de Montpellier (Montpellier GenomiX) for his technical advice and for the lend of a LightCycler 480 during the study period. The authors also thank Rita Colwell, University of Maryland, for her critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.710346/full#supplementary-material

REFERENCES

Abella, J., Bielen, A., Huang, L., Delmont, T. O., Vujaklija, D., Duran, R., et al. (2015). Integron diversity in marine environments. *Environ. Sci. Pollut. Res. Int.* 22, 15360–15369. doi: 10.1007/s11356-015-5085-3

Balouin, Y., Rey-Valette, H., and Picand, P. A. (2014). Automatic assessment and analysis of beach attendance using video images at the Lido of Sète beach, France:*Ocean Coast. Manage.* 102, 114–122. doi: 10.1016/j.ocecoaman.2014.09.006

Barraud, O., Baclet, M. C., Denis, F., and Ploy, M. C. (2010). Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. *J. Antimicrob. Chemother.* 65, 1642–1645. doi: 10.1093/jac/dkq167

Bernhard, A. E., and Field, K. G. (2000a). 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

Bernhard, A. E., and Field, K. G. (2000b). 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

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

Breittmayer, J. P., and Gautier, M. J. (1978). Contamination bactérienne d'une zone balnéaire liée à sa fréquentation. *Water Res.* 12, 193–197. doi: 10.1016/0043-1354(78)90008-8

Brouwer, R., and De Blois, C. (2008). Integrated modeling of risk and uncertainty underlying the cost and effectiveness of water quality measures. *Environ. Model. Softw.* 23, 922–937. doi: 10.1016/j.envsoft.2007.10.006

Di Cesare, A., Petrin, S., Fontaneto, D., Losasso, C., Eckert, E. M., Tassistro, G., et al. (2018). ddPCR applied on archived continuous plankton recorder samples reveals long-term occurrence of class 1 integrons and a sulphonamide resistance gene in marine plankton communities. *Environ. Microbiol. Rep.* 10, 458–464. doi: 10.1111/1758-2229.12665

Dick, L. K., Bernhard, A. E., Brodeur, T. J., Santo Domingo, J. W., Simpson, J. M., Walters, S. P., et al. (2005a). Host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71, 3184–3191. doi: 10.1128/AEM.71.6.3184-3191.2005

Dick, L. K., Simonich, M. T., and Field, K. G. (2005b). Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. *Appl. Environ. Microbiol.* 71, 3179–3183. doi: 10.1128/AEM.71.6.3179-3183.2005

Elmir, S. M., Wright, M. E., Abdelzaher, A., Solo-Gabriele, H. M., Fleming, L. E., Miller, G., et al. (2007). Quantitative evaluation of bacteria released by bathers in a marine water. *Water Res.* 41, 3–10. doi: 10.1016/j.watres.2006.10.005

European Parliament (2006). *Council of the European Union. Directive 2006/7/EC of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC.*

Figueras, M. J., de Torres, M., Silvera, C., and Corrales, M. J. (2015). "Monitoring programs for bathing waters within the frame of the EU bathing water directive: the experience of Catalonia," in *Experiences from ground, coastal and transitional water quality monitoring. The Handbook of Environmental Chemistry, vol 43*, eds A. Munné, A. Ginebreda and N. Prat (Cham: Springer), doi: 10.1007/698_2015_425

Gerba, C. (2000). Assessment of enteric pathogen shedding by bathers during recreational activity and its impact on water quality. *Quant. Microbio.* 2, 55–68.

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M., and Zhu, Y.-G. (2015). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J.* 9, 1269–1279. doi: 10.1038/ismej.2014.226

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

Hughes, B., Beale, D. J., Dennis, P. G., Cook, S., and Ahmed, W. (2017). Cross-comparison of human wastewater-associated molecular markers in relation to fecal indicator bacteria and enteric viruses in recreational beach waters. *Appl. Environ. Microbiol.* 83, e00028–17. doi: 10.1128/AEM.00028-17

Jenkins, M. W., Tiwari, S., Lorente, M., Gichaba, C. M., and Wuertz, S. (2009). Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res.* 43, 4956–4966. doi: 10.1016/j.watres.2009.07.028

Keene, W. E., McAnulty, J. M., Hoesly, F. C., Williams, L. P., Hedberg, K., Oxman, G. L., et al. (1994). A swimming-associated outbreak of hemorrhagic colitis caused by Escherichia coli O157:H7 and Shigella sonnei. *N. Engl. J. Med.* 331, 579–584. doi: 10.1056/NEJM199409013310904

Kildare, B. J., Leutenegger, C. M., McSwain, B. S., Bambic, D. G., Rajal, V. B., and Wuertz, S. (2007). 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: a Bayesian approach. *Water Res.* 41, 3701–3715. doi: 10.1016/j.watres.2007.06.037

Laplace, D., Rovera, A., Guivarch, J. Y., and Giannetti, S. (2009). Gestion de l'assainissement et qualité des eaux de baignade à Marseille. *Cahiers de l'ASEES* 14, 39–44. doi: 10.1051/asees/2009003

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

Maeda, H., Fujimoto, C., Haruki, Y., Maeda, T., Kokeguchi, S., Petelin, M., et al. (2003). Quantitative real-time PCR using TaqMan and SYBR Green for Actinobacillus actinomycetemcomitans, Porphyromonas gingivalis, Prevotella intermedia, tetQ gene and total bacteria. *FEMS Immunol. Med. Microbiol.* 39, 81–86. doi: 10.1016/S0928-8244(03)00224-4

Marti, E., Jofre, J., and Balcazar, J. L. (2013). Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. *PLoS One* 8:e78906. doi: 10.1371/journal.pone.0078906

Mauffret, A., Caprais, M. P., and Gourmelon, M. (2012). Relevance of *Bacteroidales* and F-Specific RNA bacteriophages for efficient fecal contamination tracking at the level of a catchment in France. *Appl. Environ. Microbiol.* 78, 5143–5152. doi: 10.1128/aem.00315-12

Mieszkin, S., Furet, J. P., Corthier, G., and Gourmelon, M. (2009). Estimation of pig fecal contamination in a river catchment by Real-Time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl. Environ. Microbiol.* 75, 3045–3054. doi: 10.1128/aem.02343-08

Papadakis, J. A., Mavridou, A., Richardson, S. C., Lampiri, M., and Marcelou, U. (1997). Bather-related microbial and yeast populations in sand and seawater. *Water Res.* 31, 799–804. doi: 10.1016/S0043-1354(96)00377-6

Pereira, E., Figueira, C., Aguiar, N., Vasconcelos, R., Vasconcelos, S., Calado, G., et al. (2013). Microbiological and mycological beach sand quality in a volcanic environment: madeira archipelago, Portugal. *Sci. Total Environ.* 46, 469–479. doi: 10.1016/j.scitotenv.2013.05.025

Robert, S., Sillère, G., and Liziard, S. (2008). *Évaluer et représenter le nombre d'usagers sur une plage urbaine (Les Ponchettes, Nice), M@ppemonde, Maison de la géographie,* n°91, 3. Available Online at: http://mappemonde.mgm.fr/num19/articles/art08305.html

Robert, S., and Trémélo, M. L. (2019). *Data on Frequentation of the Prophète beach (Marseille) on July 18, 2018. [Research Report] CNRS ESPACE UMR 7300; Men-Environment Observatory Mediterranean coast. 2019. hal-02438256v2*. Available online at: https://hal.archives-ouvertes.fr/hal-02438256v2/document (accessed February 4, 2020).

Rodríguez, E. A., Ramirez, D., Balcázar, J. L., and Jiménez, J. N. (2021). Metagenomic analysis of urban wastewater resistome and mobilome: a support for antimicrobial resistance surveillance in an endemic country. *Environ. Pollut.* 276:116736. doi: 10.1016/j.envpol.2021.116736

Roslev, P., and Bukh, A. S. (2011). State of the art molecular markers for fecal pollution source tracking in water. *Appl. Microbiol. Biotechnol.* 89, 1341–1355. doi: 10.1007/s00253-010-3080-7

Rothenheber, D., and Jones, S. (2018). Enterococcal concentrations in a coastal ecosystem are a function of fecal source input, environmental conditions, and environmental sources. *Appl. Environ. Microbiol.* 84, e01038–18. doi: 10.1128/AEM.01038-18

Sabino, R., Rodrigues, R., Costa, I., Carneiro, C., Cunha, M., Duarte, A., et al. (2014). Routine screening of harmful microorganisms in beach sands: implications to public health. *Sci. Total Environ.* 472, 1062–1069. doi: 10.1016/j.scitotenv.2013.11.091

Salverda, M. L. M., Visser, J. A. G. M. D., and Barlow, M. (2010). Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* 34, 1015–1036. doi: 10.1111/j.1574-6976.2010.00222.x

Seurinck, S., Defoirdt, T., Verstraete, W., and Siciliano, S. D. (2005). Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environ. Microbiol.* 7, 249–259. doi: 10.1111/j.1462-2920.2004.00702.x

Tran, N. H., Gin, K. Y. H., and Ngoc, H. H. (2015). Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. *Sci. Total Environ.* 538, 38–57. doi: 10.1016/j.scitotenv.2015.07.155

Wheeler Alm, E., Burke, J., and Spain, A. (2003). Fecal indicator bacteria are abundant in wet sand at freshwater beaches. *Water Res.* 37, 3978–3982. doi: 10.1016/S0043-1354(03)00301-4

Whitman, R. L., Harwood, V. J., Edge, T. A., Nevers, M. B., Byappanahalli, M., Vijayavel, K., et al. (2014). Microbes in beach sands: integrating environment, ecology and public health. *Rev. Environ. Sci. BioTechnol.* 13, 329–368. doi: 10.1007/s11157-014-9340-8

Wu, B., Wang, X. C., and Dzakpasu, M. (2017). Genetic characterization of fecal impacts of seagull migration on an urban scenery lake. *Water Res.* 117, 27–36. doi: 10.1016/j.watres.2017.03.041

Xi, C., Zhang, Y., Marrs, C. F., Ye, W., Simon, C., Foxman, B., et al. (2009). Prevalence of Antibiotic Resistance in Drinking Water Treatment and Distribution Systems. *Appl. Environ. Microbiol.* 75, 5714–5718. doi: 10.1128/AEM.00382-09

Yamahara, K. M., Sassoubre, L. M., Goodwin, K. D., and Boehm, A. B. (2012). Occurrence and persistence of bacterial pathogens and indicator organisms in beach sand along the California coast. *Appl. Environ. Microbiol.* 78, 1733–1745. doi: 10.1128/AEM.06185-11

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.

Copyright © 2021 Toubiana, Salles, Tournoud, Licznar-Fajardo, Zorgniotti, Trémélo, Jumas-Bilak, Robert and Monfort. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

METHODS
published: 20 October 2021
doi: 10.3389/fmicb.2021.660566

Identifying Inorganic Turbidity in Water Samples as Potential Loss Factor During Nucleic Acid Extraction: Implications for Molecular Fecal Pollution Diagnostics and Source Tracking

Rita B. Linke[1†], Sibel Zeki[2], René Mayer[1†], Katharina Keiblinger[3], Domenico Savio[1,4†], Alexander K. T. Kirschner[4,5†], Georg H. Reischer[1,6†], Robert L. Mach[7], Regina Sommer[8†] and Andreas H. Farnleitner[1,4†]*

[1] Research Group Environmental Microbiology and Molecular Diagnostics 166/5/3, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Vienna, Austria, [2] Department of Marine Environment, Institute of Marine Sciences and Management, Istanbul University, Istanbul, Turkey, [3] Department of Forest and Soil Sciences, Institute of Soil Research, University of Natural Resources and Life Sciences Vienna, Vienna, Austria, [4] Division Water Quality and Health, Department Pharmacology, Physiology and Microbiology, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria, [5] Institute for Hygiene and Applied Immunology, Medical University of Vienna, Vienna, Austria, [6] Research Area Molecular Diagnostics, Department IFA-Tulln, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Tulln, Austria, [7] Research Division Biochemical Technology, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Vienna, Austria, [8] Unit of Water Microbiology, Institute for Hygiene and Applied Immunology, Medical University of Vienna, Vienna, Austria

Molecular diagnostic methods are increasingly applied for food and environmental analysis. Since several steps are involved in sample processing which can affect the outcome (e.g., adhesion of DNA to the sample matrix, inefficient precipitation of DNA, pipetting errors and (partial) loss of the DNA pellet during DNA isolation), quality control is essential at all processing levels. In soil microbiology, particular attention has been paid to the inorganic component of the sample matrix affecting DNA extractability. In water quality testing, however, this aspect has mostly been neglected so far, although it is conceivable that these mechanisms have a similar impact. The present study was therefore dedicated to investigate possible matrix effects on results of water quality analysis. Field testing in an aquatic environment with pronounced chemo-physical gradients [total suspended solids (TSS), inorganic turbidity, total organic carbon (TOC), and conductivity] indicated a negative association between DNA extractability (using a standard phenol/chloroform extraction procedure) and turbidity (spearman $\rho = -0.72$, $p < 0.001$, $n = 21$). Further detailed laboratory experiments on sediment suspensions confirmed the hypothesis of inorganic turbidity being the main driver for reduced DNA extractability. The observed effects, as known from soil samples, were also indicated to result from competitive effects for free charges on clay minerals, leading to adsorption of DNA to these inorganic particles. A protocol modification by supplementing the extraction buffer with salmon sperm DNA, to coat charged

OPEN ACCESS

Edited by:
Obulisamy Parthiba Karthikeyan, University of Houston, United States

Reviewed by:
Ashutosh Sharma, Escuela de Ingeniería y Ciencias, Instituto de Tecnología y Educación Superior de Monterrey (ITESM), Mexico
Gaurav Saxena, Shoolini University of Biotechnology and Management Sciences, India

***Correspondence:**
Rita B. Linke
rita.linke@tuwien.ac.at

† Interuniversity Cooperation Centre for Water and Health, Austria
www.waterandhealth.at

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: *29 January 2021*
Accepted: *13 August 2021*
Published: *20 October 2021*

Citation:
Linke RB, Zeki S, Mayer R, Keiblinger K, Savio D, Kirschner AKT, Reischer GH, Mach RL, Sommer R and Farnleitner AH (2021) Identifying Inorganic Turbidity in Water Samples as Potential Loss Factor During Nucleic Acid Extraction: Implications for Molecular Fecal Pollution Diagnostics and Source Tracking. Front. Microbiol. 12:660566. doi: 10.3389/fmicb.2021.660566

surfaces prior to cell lysis, was then applied on environmental water samples and compared to the standard protocol. At sites characterized by high inorganic turbidity, DNA extractability was significantly improved or made possible in the first place by applying the adapted protocol. This became apparent from intestinal enterococci and microbial source tracking (MST)-marker levels measured by quantitative polymerase chain reaction (qPCR) (100 to 10,000-fold median increase in target concentrations). The present study emphasizes the need to consider inorganic turbidity as a potential loss factor in DNA extraction from water-matrices. Negligence of these effects can lead to a massive bias, by up to several orders of magnitude, in the results of molecular MST and fecal pollution diagnostics.

Keywords: water quality, inorganic turbidity, DNA extraction, sample process control, microbial source tracking, fecal pollution diagnostics

INTRODUCTION

Methods for molecular fecal pollution diagnostics and microbial source tracking (MST) to quantify fecal pollution and allocate it to distinct pollution sources, are increasingly used worldwide (e.g., Wuertz et al., 2011; Reischer et al., 2013; Odagiri et al., 2015; Boehm et al., 2016; Mayer et al., 2018). Usually, these parameters are determined by applying a quantitative polymerase chain reaction (qPCR) to DNA extracted from environmental samples (e.g., water, soil, sediment). A variety of qPCR-based methods are suited for molecular diagnostics of fecal pollution and MST applications as well as the detection of microbial indicators and pathogens for monitoring water quality due to their achievable precision, specificity, and sensitivity. Yet, for the application to environmental samples, a number of possible sources of error have to be considered. In particular, it is the quantity and quality of extracted DNA that might be affected by the sample matrix or the co-extraction of PCR inhibiting substances (Green and Field, 2012; Hill et al., 2015; Lever et al., 2015; Rocha and Manaia, 2020). Moreover, genetic markers (e.g., for fecal indication, pathogenic microbes) are typically present at low to very low concentrations which is why they, in many cases, have to be enriched (filtration) or purified during sample processing. Therefore, the analysis of microbial samples remains a practical and technological challenge when it comes to confidence in the measurements, especially for measurements made at the point of interest or point of care where results are used to inform critical decision-making (Da Silva et al., 2016).

In contrast to molecular investigations of microbiological water quality, the aspect of the sample matrix influencing DNA extraction has attracted much attention in studies on soil microbiology (e.g., Herrera and Cockell, 2007; Saeki et al., 2010; Paulin et al., 2013; Lever et al., 2015). From these studies, it is known that the performance of DNA extraction can be directly related to the soil's composition (Frostegard et al., 1999; Miller et al., 1999; Lombard et al., 2011; Paulin et al., 2013). Among the various soil types studied it was shown that mainly those with a high clay content (Cai et al., 2006; Yu et al., 2013; Lever et al., 2015; Engel et al., 2019) are likely to adsorb and bind DNA. Taking into consideration that many rivers and lakes transport large quantities of fine sediment and clay minerals, an effect on DNA extraction efficiency is highly conceivable. In particular, under increased discharge conditions during flooding or in the presence of strong wind sediments are mobilized resulting in high inorganic turbidity, possibly interfering with DNA extraction efficiency in samples taken under those conditions. Also, the increased occurrence of heavy rainfall events as a result of climate change in combination with deforestation, clearance, and associated increased soil erosion can lead to an increased input of sediment into water bodies and temporarily high turbidity. Despite these aspects however, this issue for water quality analysis has only recently been the subject of first studies (Lebuhn et al., 2004; Haugland et al., 2012; Shanks et al., 2016; Li et al., 2019; Rocha and Manaia, 2020).

The present study aimed at testing the hypothesis that inorganic turbidity, if present, could affect the efficiency of DNA extraction, and, as a consequence, the reliability and reproducibility of quantitative molecular methods applied for water quality testing. The test strategy comprised three levels: (i) evaluation of the postulated detrimental effect on environmental samples from an aquatic habitat showing a strong gradient of chemo-physical parameters (inorganic turbidity, organic matter content, and salinity), (ii) experimental testing of the hypothesis in the laboratory on samples specifically generated to simulate an extreme gradient of inorganic turbidity and trying to understand the type of effects by applying competitive antagonists, and finally, (iii) adapting the routinely used DNA extraction protocol based on the formerly obtained findings and applying it on different environmental samples from the study area. As study area, a shallow steppe lake and a soda lake were chosen in order to cover a broad range of different sample matrices. A sample process control, employed on a sample-to-sample basis, was used to determine DNA extraction efficiency.

MATERIALS AND METHODS

Model Study Area

The study area was located in National Park Lake Neusiedl, which is situated in the lowlands of Eastern Austria. The area included a large steppe lake (Lake Neusiedl) as well as several smaller lakes.

These habitats are known for both high and fluctuating turbidity values which is why this area was chosen for the present study. Lake Neusiedl is surrounded by a reed belt, which is characterized by lower turbidity but higher contents of organic material. Therefore, this area in its complexity, offered a wide variety of different aquatic habitats according to their chemo-physical parameters [inorganic and organic turbidity, total suspended solids (TSS), total organic carbon (TOC), and conductivity] and water matrices making it ideally suited to test the hypothesis of the sample matrix interfering with sample processing efficiency. Samples were taken at six sampling sites: soda lake Oberer Stinker (OS), Lake Neusiedl (L), reed belt of Lake Neusiedl (RB), wastewater stabilization pond (P), un-treated (W-ut), and treated wastewater (W-t). The latter two were included in the study due to the high content of organic matter.

Lake Neusiedl (L) is a shallow alkaline brown-water steppe lake which is a nearly closed system without natural outlet, receiving 80% of its water input from precipitation, 20% from rivers and losing about 90% by evapotranspiration and 10% by an artificial channel regulated by a sluice gate (Kirschner et al., 2008; Soja et al., 2014). More than half (56%) of the lake area (320 km^2) is a covered by a *Phragmites australis*-stand (Hietz, 1992). Within the reed cover, water is not turbid due to reduced wind exposure and brown in color due to humic substances (Dokulil and Herzig, 2009). It is due to the extreme shallowness of the lake (mean depth 1.1 m, maximum depth 2.0 m) and wind exposure, that the lake is characterized by high concentrations of suspended solids (up to 800 mg dw L^{-1}) and large annual variations in temperature (28°C in summer and 1°C in winter) (Wolfram, 1996; Dokulil and Herzig, 2009). Chemically, the lake is characterized by elevated levels of alkalinity (5.0–14.6 meq L^{-1}), conductivity (1,100–3,100 μS) and pH (8.3–8.9) in comparison to typical freshwater lakes, with considerable spatial and temporal variability (Wolfram, 2006; Schauer et al., 2015).

The hypertrophic shallow soda lake (Oberer Stinker, OS) is characterized by high total salt concentrations and turbidity (Eiler et al., 2003; Kirschner et al., 2004). It was formed by mineral solutes ascending with the groundwater flux (Krachler et al., 2000) and is characterized by pH values ranging from 9.4 to 10 (Eiler et al., 2003). Na^+ is the dominating cation, and HCO_3^-, CO_3^{2-}, Cl^-, and SO_4^{2-} represent the major anions. Salinity of the soda lake varies strongly with seasons (Eiler et al., 2003).

Sampling for Test Series I—Evaluating Matrix Effects in Field Samples

Water and wastewater samples were collected over a period of 5 months in sterile 1 L sampling bottles (Nalgene, United Kingdom), stored in the dark in cooling boxes at 4°C during transport, and processed within 6 h after collection. A given volume of water (surface water: 100 mL, un-treated waste water: 10 mL, treated waste water: 50 mL) was filtered through Isopore 0.2 μm polycarbonate membrane filters (Millipore, Bedford, MA, United States). Immediately after filtration, the filters were frozen and stored at −80°C until nucleic acid extraction. Six independent filtrations were done for each sample. One replicate was used as an un-spiked control. The other five replicate samples were directly spiked with 5×10^7 cells of the defined target cell standard (DeTaCS; see section "Sample Process Control") during filtration. On each sampling occasion, an additional unused filter was placed directly into a 1.5 mL extraction vial as a blank filter control.

Sampling for Test Series II—Simulating an Extreme Gradient of Suspended Solids

To prepare samples mimicking an extreme gradient of suspended solid content, 20 L of water from Lake Neusiedl were sampled in sterile plastic sampling bottles (Nalgene, United Kingdom). In addition, a sediment sample from the bottom of the lake was taken by carefully dragging a sterile glass bottle (Duran Group, Germany) over the ground of the lake collecting the uppermost 2–3 cm of deposited sediment. Water and sediment samples were stored in the dark in cooling boxes at 4°C and were transported to the laboratory. Samples were stored over night at 4°C to ensure settling of the fine sediment fraction and were then processed. To evaluate the influence of suspended solids on DNA extractability, sediment-enriched samples containing 9–220 mg sediment L^{-1} water were prepared in the laboratory by suspending lake bottom sediment in water from Lake Neusiedl (for details regarding the production of sediment-enriched samples see **Supplementary Data 1**). From samples containing 9, 34, and 106 mg sediment L^{-1} water, 10 independent filtrations were performed which were used for further testing to modify and adapt the DNA extraction protocol. All so prepared sediment-enriched samples were filtered through Isopore 0.2 μm polycarbonate membrane filters (Millipore, Bedford, MA, United States) and filters were frozen immediately and stored at −80°C until nucleic acid extraction. An additional unused filter was placed directly into a 1.5 mL extraction vial as a blank filter control.

Sampling for Test Series III—Comparison of DNA Extraction Protocols on Environmental Samples

Samples from Lake Neusiedl (L), the stabilization pond (P) and the wastewater treatment plant (W-ut and W-t) were taken monthly from April to December 2015. Two independent filtrations were done for each sample (surface water: 100 mL, un-treated waste water: 10 mL, treated waste water: 50 mL) using Isopore 0.2 μm polycarbonate membrane filters (Millipore, Bedford, MA, United States). On each sampling occasion, an additional unused filter was placed directly into a 1.5 mL extraction vial as a blank filter control. All filters were frozen immediately and stored at −80°C until DNA extraction. One replicate sample was extracted according to the standard DNA extraction protocol, the other one with a protocol adapted to the sample matrix (see sections "DNA Extraction" and "DNA Extraction Protocol Modification"). Both, the sample process control, DeTaCS, and salmon sperm DNA as adsorption site blocker, were added directly at DNA extraction into the reaction mixture. DeTaCS was added at a concentration of 5×10^6 cells

per extraction and salmon sperm DNA was added at an amount of 250 μg per reaction. Since TSS values were not available on a sample-to-sample basis the salmon sperm DNA amount for protocol modification was chosen based on the results of test series II. The here used amount would have been sufficient to compensate for 106 mg sediments L^{-1} sample.

Chemo-Physical Water Analyses

Conductivity (LF 330, WTW, Germany), water temperature, pH (GHM, Seibold Vienna, Austria), oxygen (OXI 330i, WTW) and turbidity (calculated from Secchi depth) were measured *in situ*. For inorganic nutrients, TOC and TSS an extra water sample was collected in a clean 1 L plastic bottle and processed according to the methods published by Eiler et al. (2003). For the determination of TSS, a defined volume of sample water was filtered through pre-muffled glass-fiber filters (GF/C; Whatman, United Kingdom) and dried to constant weight at 120°C in a drying oven (Haereus, Germany). To obtain the inorganic and organic fraction of both suspended solids from water samples and lakebed sediments (uppermost 2–3 cm), the oven-dried samples were further combusted in a muffle furnace (480°C, 4 h; Nabertherm, Germany).

Sample Process Control

The DeTaCS used as sample process control was an *Escherichia coli* strain (DHB6501) carrying a single copy of the target sequence for a ruminant-associated source tracking marker (BacR; Reischer et al., 2006; Mayer et al., 2018) in its genome. The strain was constructed by applying an *E. coli* plasmid-chromosome shuttle system using a λ-phage (Boyd et al., 2000). Details on strain construction as well as strains and plasmids used therefore are given as Supplementary Material (**Supplementary Data 2**). Production of the DeTaCS strain was done by batch fermentation in a benchtop bioreactor (RALF Plus-System, Switzerland). Details on fermentation conditions are given as supplementary Material (**Supplementary Data 3**). Aliquots were prepared of the fermentation batch to be used as process control to determine DNA extraction efficiency. For this purpose, culture broth was supplemented with glycerol to a final concentration of 20% and aliquots of 100 μL were shock frosted in liquid nitrogen and stored at −80°C for further use. Cell numbers were determined using an epifluorescence microscope (Nicon Eclipse 8000, Japan; see **Supplementary Data 4**). Samples spiked during filtration were supplemented with 5×10^7 cells, samples spiked at DNA extraction directly were supplemented with 5×10^6 cells per extraction.

DNA Extraction

DNA extraction was performed using bead-beating and phenol/chloroform (Griffiths et al., 2000; Reischer et al., 2008; Mayer et al., 2018). In brief, cell lysis was achieved by addition of CTAB buffer and glass beads in a FastPrep 24 benchtop homogenizer for cell lysis (MP Biomedicals Inc., Irvine, CA, United States) at speed setting of 6 m s^{-1} for 30 s. Polycarbonate filters were completely dissolved at this step and the DNA was subsequently purified. Precipitation of the DNA was achieved by addition of isopropanol. The extracted DNA was eluted in 10 mmol L^{-1} TRIS buffer (pH 8.0) and stored at −80°C until further analysis. The detailed DNA extraction protocol is given as Supplementary Material (**Supplementary Data 5**).

DNA Extraction Protocol Modification

Modification of the DNA extraction protocol to enhance DNA yield were direct addition of either salmon sperm DNA (0–65 mg g^{-1} sediment) or Na-pyrophosphate (0–0.5 g g^{-1} sediment) to the extraction buffer. The following extraction controls were prepared: blank extraction control (reagents only), extraction control spiked with DeTaCS (5×10^6 cells per reaction), extraction control spiked with either salmon sperm DNA or sodium Na-pyrophosphate and an extraction control spiked with DeTaCS and either salmon sperm DNA or Na-pyrophosphate.

Salmon sperm DNA (Serva, Germany) added as adsorption site blocker was solubilized and purified using a standard phenol/chloroform extraction protocol. In brief, 500 mg salmon sperm DNA were solubilized in 50 mL sterile double-distilled water (4 h at room temperature on a rocking shaker; IKA, Germany) and NaCl concentration of the solution was adjusted to 0.1 M (Merck, Germany). DNA was extracted by addition of an equal volume of phenol (pH 7.5–8.0; Carl Roth, Germany). After centrifugation (5 min, RT, 13,000 rpm), the aqueous phase was transferred to a new reaction vial and extracted further by addition of an equal volume of phenol:chloroform (1:1). After centrifugation (5 min, RT, 13,000 rpm), the aqueous phase was again transferred to a new reaction vial containing an equal volume chloroform (Merck, Germany). After centrifugation (5 min, RT, 13,000 rpm), the aqueous phase was recovered and the DNA was sheared by passing it 10 times rapidly through a 20-gauge hypodermic needle (Braun, Germany). The recovered DNA was then precipitated by adding 2.5 volumes of ice-cold ethanol (96%; Merck, Germany). The solution was incubated at RT for 30 min. DNA was recovered by centrifugation (30 min, 4°C, 20.000 rpm). The obtained DNA pellet was finally dissolved in sterile double-distilled water and the final concentration was adjusted to 10 mg mL^{-1} salmon sperm DNA (Nanodrop 1000 spectrophotometer, Thermo Fisher Scientific). Before use as additive in the DNA extraction protocol, the DNA solution was tested for the absence of bacterial DNA contamination by 16S rRNA targeted qPCR (Klindworth et al., 2013). The obtained salmon sperm DNA solution was free of 16S rRNA gene targets.

Sample DNA concentration was determined with the QuantiFluor dsDNA Kit (Promega, United States) according to manufacturer's instructions and fluorescence readings were taken on an Anthos Multimode Fluorometer Zenyth 3100 (UK-Biochrom Ltd.).

Quantitative PCR

In addition to the sample process control DeTaCS (BacR assay; Reischer et al., 2006; Mayer et al., 2018), a general *Bacteroidetes* marker, AllBac (Layton et al., 2006), was run with the *ntb2* fragment as internal amplification control (IAC, non-competitive) in duplex to monitor for qPCR amplification

inhibition (Anderson et al., 2011). qPCR methods targeting an enterococcal-associated genetic marker (ENT; Haugland et al., 2005) and the human-associated MST assay HF183/BacR287 (Green et al., 2014) were applied to test for the performance of the modified DNA extraction protocol on environmental samples.

All qPCR reactions were performed in duplicate in a 15 μL volume on a Rotor-Gene Q thermocycler (Qiagen Inc.). The reaction mixture for the AllBac and IAC duplex assay was composed of 7.5 μL Rotor-Gene Multiplex PCR mastermix (Qiagen Inc.), 2.5 μL sample DNA dilution (1:4 and 1:16), 600 nM AllBac296f forward primer, 600 nM AllBac412r reverse primer, 25 nM AllBac375Bhqr TaqMan MGB probe, 500 nM ntb2-f forward primer, 500 nM ntb2-r reverse primer, 200 nM ntb2-p probe and 400 ng μL^{-1} bovine serum albumin. IAC template (plasmid containing the *ntb2* gene fragment) was spiked at a concentration of 10^3 copies per reaction. Cycling conditions were 3 min at 95°C for denaturation and 45 cycles of 30 s at 95°C followed by 45 s at 60°C. For the DeTaCS (BacR), HF183/BacR287 and the ENT assay the respective reaction mixture was composed of 7.5 μL Rotor-Gene Multiplex PCR mastermix (Qiagen Inc.), 2.5 μL sample DNA dilution (1:4) and 400 ng μL^{-1} bovine serum albumin, while the originally published primer and probe concentrations were maintained. Cycling parameters were 5 min at 95°C for denaturation and 45 cycles of 15 s at 95°C followed by 60 s at 60°C.

Data Analysis

All data analysis was done with either Microsoft Excel for Mac 16.17 or Sigma Plot 10 (Systat Software Inc., Chicago, IL, United States). Quality assessment of qPCR data was done as previously described (Reischer et al., 2006, 2011; Mayer et al., 2018). In brief, reaction efficiency of all qPCR runs ranged from 95 to 105%. All negative controls and no-template controls were consistently negative (i.e., fluorescence never exceeded the threshold). All samples were measured in duplicate in at least two 4-fold DNA dilution steps with the AllBac assay and the results were compared. Samples with matching concentrations (e.g., the ratio [(concentration 1:16)·4]/[(concentration1:4)] was between 0.5 and 2) in the 1:4 and 1:16 dilutions were judged free of PCR inhibiting substances in the 1:4 dilution. This dilution was then used for all further measurements.

Samples with replicate standard deviations of a Ct-value > 1 in the four-fold DNA extract dilutions were considered to be not quantifiable and were not considered for further analysis. Moreover, an inhibition of the qPCR reaction was assumed to be present if the threshold cycle (Ct value) of the IAC assay in a sample was shifted toward higher Ct values by more than one cycle in comparison to the mean of the Ct of the negative controls. qPCR standard dilutions ranging from 10^1 to 10^6 targets per reaction were used in a linear regression model for calculation of the qPCR calibration curve. Results are reported as marker equivalents per DNA extract (ME per extraction, cf. Reischer et al., 2006).

Spearman's rank-order correlation was used for calculation of the correlation coefficients among the parameters using SPSS Statistics Software version 25 (SPSS Inc., Chicago, IL, United States).

RESULTS

Turbidity and Chemo-Physical Properties of Samples From the Study Area

The amount of suspended particles in a water sample is regularly given as either turbidity (NTU) or as TSS (mg L^{-1}). In the present study, we provide values in both units to allow comparability with other studies. Turbidity and TSS values observed in the samples highlighted the high heterogeneity of the habitats within the study area in regard to the water matrix (**Table 1**). Turbidity values ranged between 3 and 341 NTU with the lowest values observed in treated wastewater and the highest ones in the soda lake (Oberer Stinker, OS > Lake Neusiedl, L > stabilization pond, P > reed belt, RB > treated wastewater, W-t). Values for TSS were highest in samples from the soda lake (OS) and lowest in samples from the reed belt (OS >> L ≥ P >> RB). The observed mean TSS values ranged from 4.6 mg L^{-1} (RB) to 3,015.5 mg L^{-1} (OS).

In water samples, the organic fraction of TSS ranged from 3.9 to 97.1% (P >> RB ≥ L >> OS) while the inorganic fraction ranged from 2.9 to 96.1% (OS >> L ≥ RB >> P; **Table 2**). In sediment samples, the mean organic fraction ranged from 0.6% in Lake Neusiedl to 16.2% in the reed belt (RB > P >> OS > L) and the mean inorganic fraction ranged from 83.8% in the reed belt to 99.4% in Lake Neusiedl (L > OS >> P > RB).

Total organic carbon values ranged from 6 to 39 mg L^{-1} with the lowest values observed in the stabilization pond and the highest ones in the Soda Lake (P < W-t < L < RB < OS). The open lake was therefore mainly characterized by high turbidity and lower TOC, while the reed belt was characterized by low turbidity and high TOC. The soda lake represented an extreme habitat in every respect (high turbidity and TOC levels).

TABLE 1 | Chemo-physical parameters of environmental water samples from the six different sampling sites within the study area.

		Turbidity (NTU)	TSS (mg L^{-1})	TOC (mg L^{-1})	Conductivity (μS cm^{-1})	pH
OS	**Mean**	**340.6**	**3,015.5**	**39.1**	**4,202.0**	**9.3**
	Range	70.0–650.0	1,739.7–3,307.8	31.2–47.0	2,980.0–6,000.0	9.1–9.5
L	**Mean**	**162.0**	**24.2**	**12.0**	**1,609.6**	**8.8**
	Range	31.0–455.0	22.3–26.4	11.2–12.5	1,456.0–1,842.0	8.7–8.9
RB	**Mean**	**3.2**	**4.6**	**22.3**	**1,878.6**	**8.6**
	Range	1.6–5.3	3.4–5.6	18.0–26.5	1,736.0–2,080.0	8.4–9.0
P	**Mean**	**11.2**	**21.2**	**5.8**	**994.5**	**7.9**
	Range	2.2–26.0	10.2–43.6	3.9–7.1	780.0–1,111.0	7.7–8.2
W-ut	**Mean**	n.a.	**286.5**	n.a.	n.a.	n.a.
	Range	n.a.	175.4–472.4	n.a.	n.a.	n.a.
W-t	**Mean**	**2.8**	**5.1**	**6.0**	**1,009.2**	**7.6**
	Range	1.6–4.2	5.0–5.2	4.4–10.6	811.0–1,138.0	7.3–7.8

OS, soda lake Oberer Stinker, L, Lake Neusiedl; RB, reed belt of Lake Neusiedl; P, wastewater stabilization pond; W-ut, untreated wastewater; W-t, treated wastewater; NTU, Nephelometric Turbidity Units; TOC, Total Organic Carbon; TSS, total suspended solids.
n = 5–11.

TABLE 2 | Percentage of the organic and inorganic fraction of suspended solids from water samples and lakebed sediment for environmental samples from Soda Lake (OS), Lake Neusiedl (L), the Reed Belt (RB), and the stabilization pond (P).

		OS		L		RB		P	
		Water	Sed.	Water	Sed.	Water	Sed.	Water	Sed.
Organic component (%)	**Mean**	**3.9**	**1.5**	**43.2**	**0.6**	**45.8**	**16.2**	**97.1**	**13.7**
	Median	4.1	1.5	41.6	0.5	47.8	16.5	100	9.9
	Min	2.7	1.1	30.8	0.4	28.8	13.8	91.4	7.2
	Max	5.0	1.8	63.9	0.8	62.2	17.8	100	35.2
Inorganic component (%)	**Mean**	**96.1**	**98.5**	**56.8**	**99.4**	**54.2**	**83.8**	**2.9**	**86.3**
	Median	95.9	98.5	58.4	99.5	52.2	83.6	0.0	90.1
	Min	95.0	98.2	36.1	99.2	37.8	82.2	0.0	64.8
	Max	97.3	98.9	69.2	99.6	71.2	86.2	8.6	92.8

n = 6. Min, minimum; max, maximum; Sed., sediment.

Electrical conductivity in the different sample types from the study area ranged between 1,000 and 4,200 $\mu S\ cm^{-1}$ (P < L < RB < OS) and the pH value was ≥8 (**Table 1**). The lowest pH values were found in the stabilization pond and in treated wastewater (pH 8), the highest pH values were observed in the soda lake (pH ≥ 9, **Table 1**).

Collecting Field Evidence That Chemo-Physical Parameters Influence DNA Extractability

As a measure of DNA extraction efficiency from environmental samples of the study area the recovery of the added process control (DeTaCS) was used. Losses observed during sample processing were most pronounced in samples from Lake Neusiedl (L) and the shallow soda lake (OS). In these two sample types, the marker concentrations retrieved were lower by a factor of 4 and 2 $\log_{10}$ units when compared to spiked controls, respectively. Both sample types were characterized by high turbidity values and a high percentage of inorganic compounds (**Tables 1**, **2**). In the other sample types (RB, P, W-t), the DeTaCS marker was detected in a concentration range similar to the extraction controls (**Figure 1**). A correlation analysis revealed a significant negative relationship of the DNA extraction efficiency with turbidity and TOC levels (turbidity: $\rho = -0.719$, $p < 0.001$, $n = 24$; TOC: $\rho = -0.407$, $p = 0.048$, $n = 24$, **Supplementary Table 3**). PCR inhibition, as would have been indicated by a shift of the IAC toward higher Ct values, could not be observed in any of the sample types. Observed Ct values in controls ranged from 30 to 31, those of samples ranged from 29 to 31.

Laboratory Testing of the Hypothesis

To further investigate the possible effects of the water sample matrix on DNA yield, samples were prepared by suspending lakebed sediment in lake water (equivalent to a TSS gradient covering a range of 9–220 mg L^{-1}). These sediment-enriched samples were then subjected to DNA extraction and the DNA yield and the AllBac marker concentration were subsequently determined to investigate possible effects of the suspended particles. The AllBac qPCR assay was chosen because it targets a broad range of bacteria of the *Bacteroidetes* phylum including target organisms of many MST assays.

The results (**Figure 2**) showed that sediment concentrations of 15 mg L^{-1} already reduced the DNA retrieval to approximately 43% (2 ng μL^{-1} DNA). Sediment concentrations of 40 mg L^{-1} and above resulted in a retrieval of approximately 20% (1.0–1.3 ng μL^{-1} DNA). AllBac marker concentrations obtained from this sample set reflected the same TSS dependent trend as observed for the DNA concentration. A spearman rank order correlation confirmed the significance of the negative correlation between the sediment content and the retrieved DNA concentration ($\rho = -0.916$, $p < 0.001$, $n = 11$, **Supplementary Table 4**) as well as the retrieved AllBac marker concentration ($\rho = -0.964$, $p < 0.001$; $n = 11$, **Supplementary Table 4**).

Optimization of the DNA Extraction Protocol by Addition of Adsorption Site Blockers

The results obtained from the previous experiments indicate that it is mainly those sample types with a high fraction of inorganic matter in which DNA extraction efficiency is strongly reduced. Since these particles are often characterized by high adsorptive capacities toward nucleic acids (see section "Discussion"), a DNA extraction protocol optimization was endeavored by testing two frequently used adsorption site blockers, salmon sperm DNA and Na-pyrophosphate.

In a first attempt Na-pyrophosphate was supplemented during DNA extraction as it would have the advantage not to bring in DNA into the sample. The amounts needed to compensate for adsorptive effects were high (>0.3 g g^{-1} sediment) but the DNA retrieval could be significantly increased (**Figure 3**). In samples with low sediment concentration Na-pyrophosphate addition did not affect DNA yield while for samples with higher sediment content (34 and 106 mg L^{-1}) no saturation plateau in the DNA increase could be observed. AllBac marker concentrations also increased after supplementing the reaction mixture with Na-pyrophosphate. However, in samples with the highest amounts of suspended sediments AllBac marker concentrations were not raised to the same extent as the DNA yield (**Figure 3**). In these samples, a strong inhibition of the qPCR reaction was observed as measured by application of the IAC. A shift in the Ct value of 2–4 cycles was observed in these samples for which reason Na-pyrophosphate was omitted from the spiking- and all further experiments.

The addition of salmon sperm DNA during the extraction process also strongly increased the DNA retrieval and the effect was greater the higher the sediment content was (**Figure 4**). Like for the DNA yield, the AllBac marker concentration also increased after salmon sperm DNA addition. The increase was most pronounced in samples with the highest sediment content tested (106 mg L^{-1}) while the marker concentrations in samples with the lowest sediment concentration (9 mg L^{-1}) remained merely unchanged. Further, a saturation plateau for the

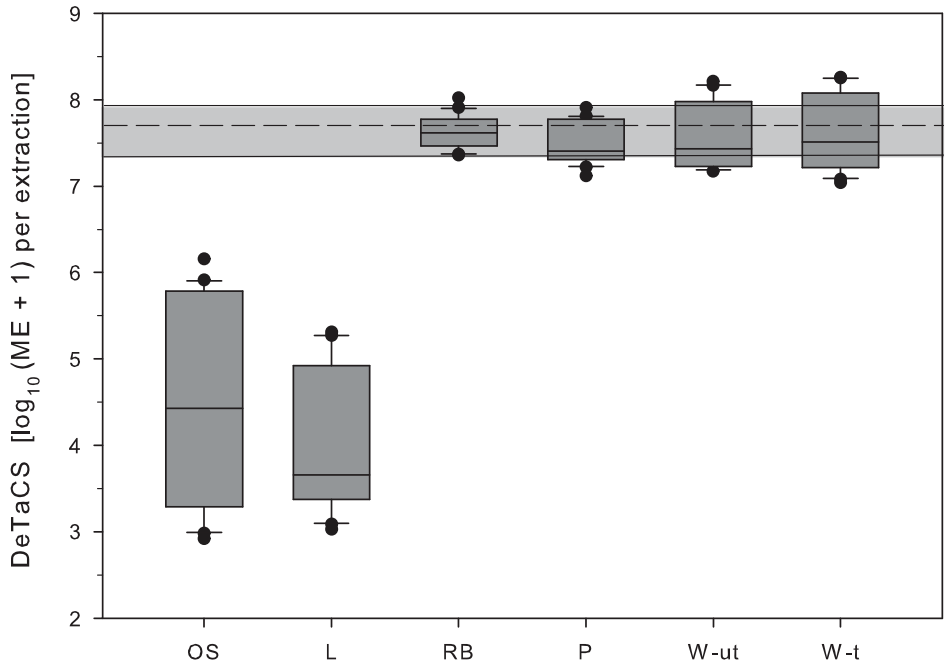


FIGURE 1 | Retrieval of DeTaCS marker concentrations from environmental samples of the study area to test the hypothesis that the sample matrix can affect DNA extraction efficiency. Environmental water samples were spiked with the reference cell standard during filtration directly into the sample. The given sample order reflects the gradient from high to low inorganic TSS fraction (OS > L >> RB > P > W-ut > W-t). OS, shallow soda lake Oberer Stinker; L, Lake Neusiedl; RB, reed belt; P, stabilization pond; W-ut, untreated wastewater; W-t, treated wastewater. Boxes, 25th and 75th percentile; lines within the boxes, median; whiskers, 10th and 90th percentile, respectively. The gray area spans the 10th–90th percentile of the DeTaCS marker concentration in the extraction controls. The dashed line marks the median. *n* = 22.

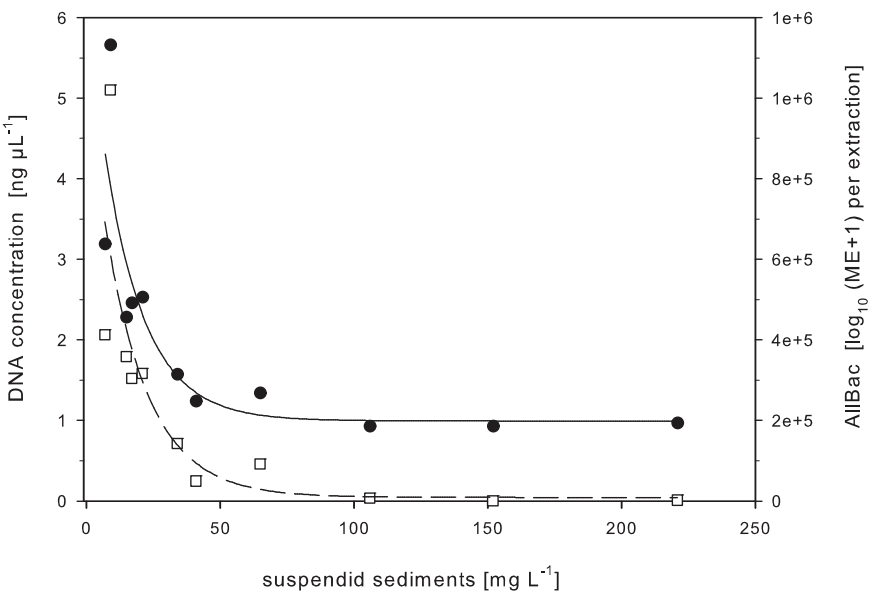


FIGURE 2 | DNA yield (black circles) and AllBac marker concentrations (white squares) in dependence of in lake water suspended lakebed sediments. ME, marker equivalents. *n* = 1 per sediment concentration.

AllBac marker concentration after salmon sperm addition was observed in samples with higher sediment concentrations (34 and 106 mg L^{-1}). The amount of salmon sperm DNA needed to reach this point, however, was different (**Figure 4**).

Supplementing the samples with salmon sperm DNA and additionally spiking with the DeTaCS revealed a picture comparable to that obtained for the AllBac marker. The DeTaCS marker concentrations could also be significantly increased due

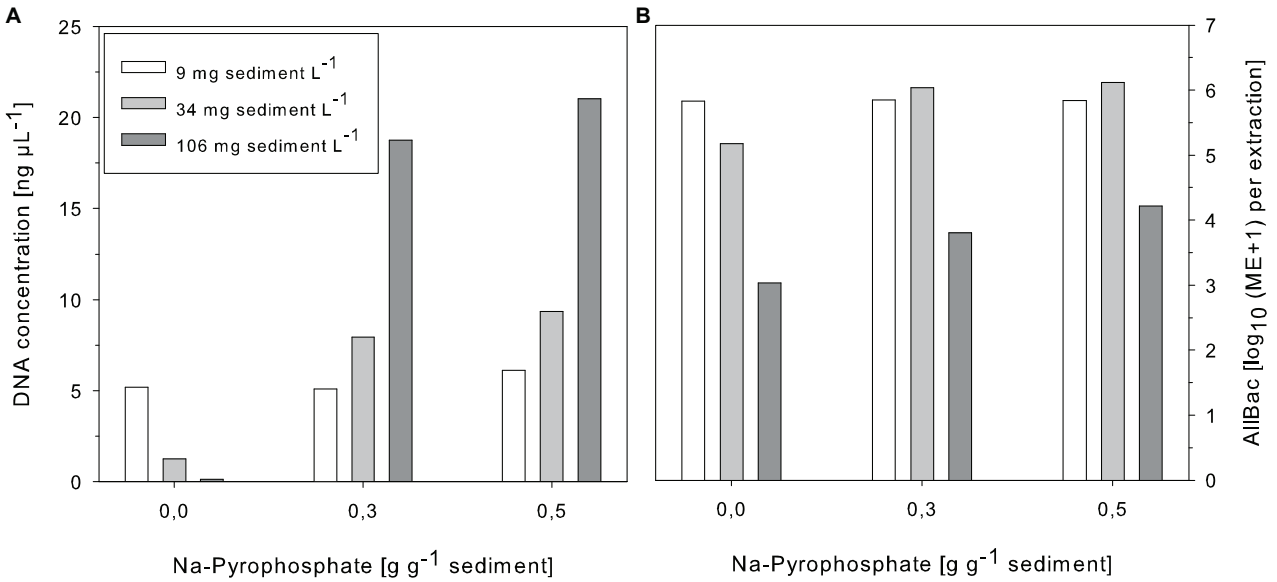


FIGURE 3 | Effects of Na-pyrophosphate supplementation during DNA extraction on the DNA yield **(A)** and AllBac marker concentrations **(B)** of sediment-enriched samples. ME, Marker equivalents, $n = 1$.

to supplementation with salmon sperm DNA and, as for the AllBac marker, a saturation plateau was reached after which further salmon sperm DNA addition did not affect the retrieved DeTaCS marker concentrations further. Significant correlations were obtained for the amount of added salmon sperm DNA with the DNA concentration retrieved ($\rho = 0.712$, $p < 0.001$, $n = 21$) as well as the DeTaCS marker concentration recovery ($\rho = 0.515$, $p = 0.017$ $n = 21$; **Supplementary Table 5**).

Impact of the Modified DNA Extraction Protocol on Molecular Fecal Pollution and MST Marker Results

Finally, the standard and the modified DNA extraction protocols were tested in comparison on environmental samples from the study area. The strongest effect of protocol modification became obvious in samples from Lake Neusiedl (L). In this sample type, the addition of salmon sperm DNA raised DNA extraction efficiency to close to control levels in the majority of the samples while the extraction efficiency of the other sample types tested (P, W-t, W-ut) remained merely unchanged. In samples from Lake Neusiedl, however, the DeTaCS marker concentrations retrieved with the modified protocol were, on average, 3 $\log_{10}$ units higher as with the standard procedure (**Figure 5A**).

When applied to molecular genetic fecal markers and MST markers, the benefit of using the modified protocol as opposed to the standard extraction protocol became especially clear. Again, the greatest differences were observed for samples from Lake Neusiedl. In this sample type, the median AllBac marker concentration increased by 4 $\log_{10}$ steps due to the addition of salmon sperm DNA as adsorption site blocker during DNA extraction (median standard protocol: $\log_{10}$ 1.4 ME 100 mL^{-1}, median modified protocol: $\log_{10}$ 6.0 ME 100 mL^{-1}, **Figure 5B**).

Particularly striking were the effects observed on the applied qPCR-based ENT marker and the human-associated MST marker HF183/BacR287. Since these markers are often present in the environment at very low concentrations, they were not detectable after extraction with the standard protocol. Only after application of the modified extraction protocol these genetic markers could also be detected in many samples (**Figures 5C,D**). The qPCR-based ENT marker, for example, which could not be detected at all after DNA extraction with the standard protocol, was found at median concentrations of $\log_{10}$ 1.7 ME 100 mL^{-1} (**Figure 5C**) after addition of salmon sperm DNA during extraction. The HF183/BacR287 marker was also not detected in samples extracted with the standard protocol but was found sporadically after extraction with the modified protocol (**Figure 5D**).

DISCUSSION

Many studies investigating microbiological water quality have addressed the issue of inhibitory effects of the sample matrix in the qPCR reaction which is routinely determined by the use of an internal amplification control (e.g., Gregory et al., 2006; Behets et al., 2007; Sen et al., 2007; Shanks et al., 2008; Haugland et al., 2010; Anderson et al., 2011; D'Agostino et al., 2011). Possible DNA losses during sample processing (filtration, DNA extraction) which might occur independently of PCR inhibition, remain undetected and unconsidered. This also became obvious from the results of the present study, where in neither of the sample types an inhibitory effect on the qPCR reactions was observed despite massive losses of DNA during the extraction process, highlighting the importance of sample process controls by which these effects could be uncovered in the present study.

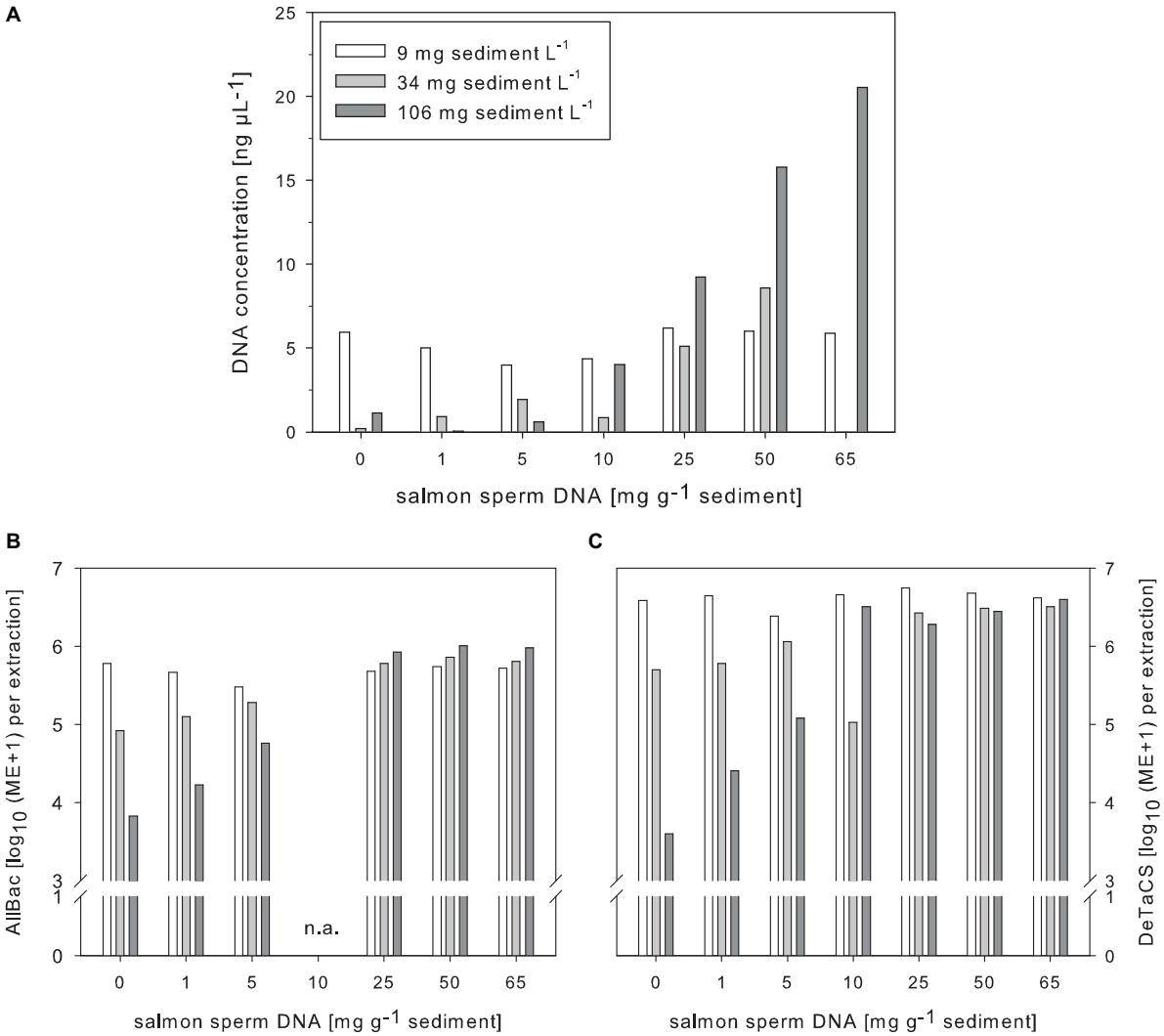


FIGURE 4 | Effects of salmon sperm DNA supplementation during DNA extraction on the DNA yield **(A)**, the AllBac marker concentration **(B)**, and the DeTaCS marker concentrations **(C)** of sediment-enriched samples. ME, Marker equivalents, $n = 1$.

However, from a large number of studies, mainly in the research field of soil microbiology, it has been shown that soil and sediment particles can interfere with DNA extraction efficiency (e.g., Lombard et al., 2011; Yu et al., 2013; Lever et al., 2015; Gardner and Gunsch, 2017; Xue and Feng, 2018). In studies on water quality analysis this aspect has only recently gained attention (Haugland et al., 2012; Shanks et al., 2016; Li et al., 2019; Rocha and Manaia, 2020) although it seems reasonable to assume that the mechanisms responsible for DNA losses in soil samples (organic and inorganic particles that have the capacity to adsorb nucleic acids during the process of DNA extraction) might also become active in water samples characterized by high amounts of suspended solids. In a first step this hypothesis was tested in the present study on environmental samples from an aquatic habitat covering a strong gradient of chemo-physical parameters (turbidity, TSS). The results clearly supported this hypothesis by showing a non-negligible effect of both, turbidity and TSS, on DNA extraction efficiency, with the greatest losses observed in samples with high inorganic content (L, OS; **Figure 1**) and less or no effect on other sample types (lower levels of inorganic particulate matter or higher levels of organic materials) as shown by application of the sample process control. This finding is in line with other studies that have shown that losses of DNA during the extraction process are directly related to the soil or sediment's composition (Frostegard et al., 1999; Miller et al., 1999; Lombard et al., 2011; Paulin et al., 2013). According to these studies, it is mainly those samples with a high clay content (Cai et al., 2006; Yu et al., 2013; Lever et al., 2015; Engel et al., 2019) to which DNA adsorbs and/or binds particularly well after its release from the cell. The therefore underlying mechanism is mainly explained as adsorption of DNA to inorganic particles by bridging effects exerted by divalent cations between phosphate groups of the DNA and the silicate anions of the minerals (Lorenz and Wackernagel, 1994; Saeki et al., 2010). Since the lakebed

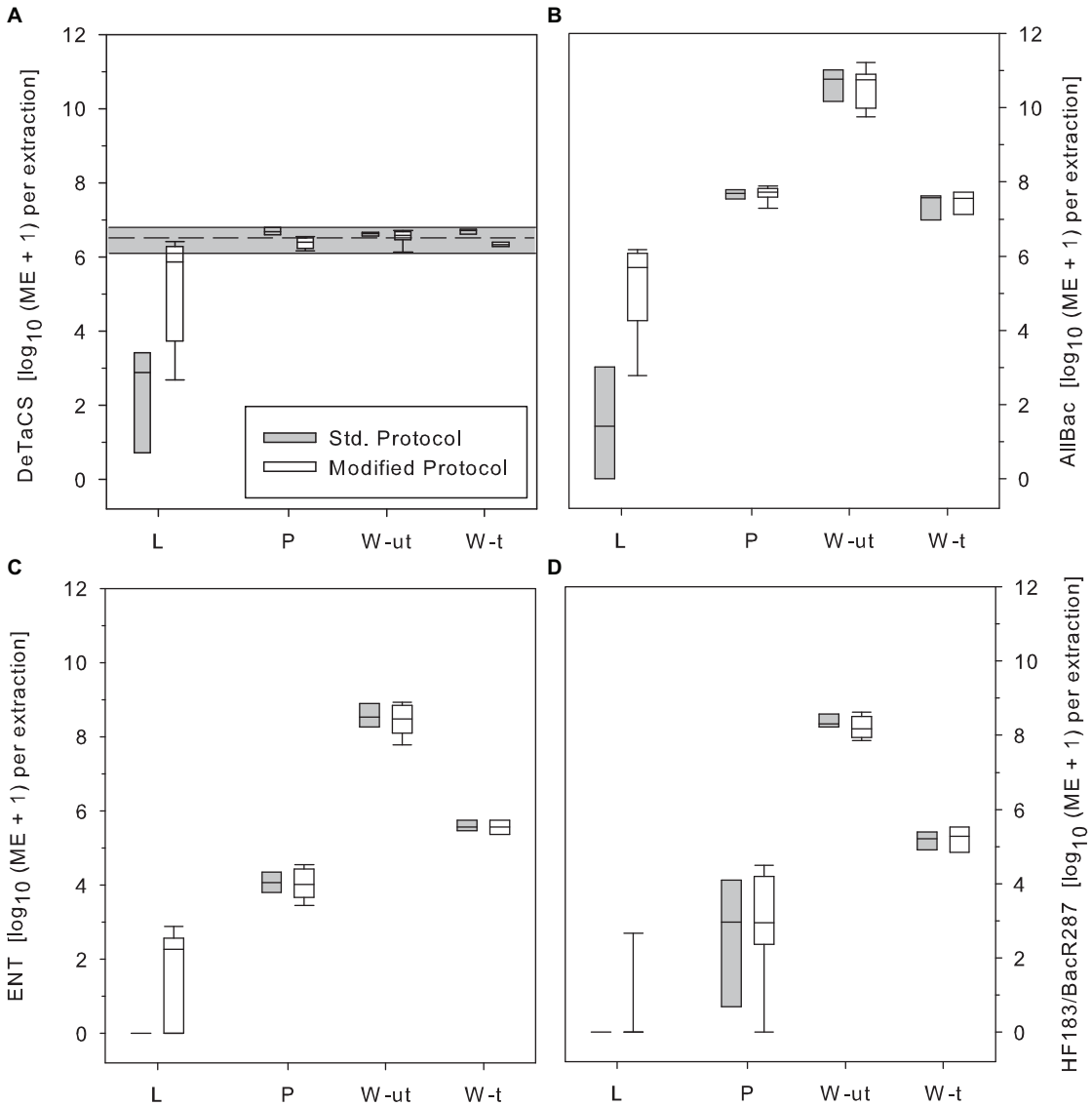


FIGURE 5 | Impact of the applied DNA extraction protocol (standard extraction protocol, gray; addition of salmon sperm DNA, white) on the retrieval of the sample process control (DeTaCS) **(A)**, the AllBac marker concentration **(B)**, the qPCR-based ENT marker **(C)**, and the human-asscoiated MST marker HF183/BacR287 results **(D)**. Std. Protocol, standard DNA extraction protocol; Modified Protocol, Supplementation of the reaction mixture with salmon sperm DNA during DNA extraction; L, Lake Neusiedl; P, stabilization pond; W-ut, untreated wastewater; W-t, treated wastewater. Boxes, 25th and 75th percentile; lines within the boxes, median; whiskers, 10th and 90th percentile, respectively. *n* = 8–9.

sediment of Lake Neusiedl consists mainly of fine-grained clays and carbonates interrupted by thin sand layers and small gravel (Löffler, 1979; Hedrich, 1983) an effect seems reasonable. Sand and silt contents of the lake sediment predominantly consist of dolomite, calcite, quartz, mica, and oligoklases while clay minerals are mainly composed of illite and, to a lesser extent, montmorillonite (Löffler, 1979). Further studies on laboratory-prepared samples with varying TSS content strengthened the hypothesis that inorganic turbidity strongly influences DNA extraction efficiency by showing a strict dependence between the amount of suspended lake bottom sediment and the recovered DNA concentration (presented as DNA concentration and AllBac marker equivalents, **Figure 2**).

While the assessment of DNA extraction efficiency is crucial for unraveling detrimental effects of the sample matrix, for some applications, especially when the results are used to inform critical decision-making, there may be a further need to adapt protocols to improve extraction efficiency. For this purpose, today a plethora of different protocols to extract DNA from difficult sample matrices is available (e.g., Paulin et al., 2013; Lever et al., 2015; Hinlo et al., 2017; Engel et al., 2019). Most protocol modifications were made by using additives such as

phosphate and inorganic phosphate types, nucleic acid building blocks, DNA, RNA, or skim milk powder (e.g., Frostegard et al., 1999; Hoshino and Matsumoto, 2005; Cai et al., 2006; Paulin et al., 2013; Lever et al., 2015) to coat charged surfaces before cell lysis. The choice of additives used, however, may depend on their availability and/or cost as well as on the research question posed. Using milk powder, for example, could raise the question whether traces of bacterial DNA could be introduced influencing the PCR reaction or affecting results in microbiome studies (Schrader et al., 2012; Paulin et al., 2013). In contrast, the use of salmon sperm DNA is only applicable if not simultaneously used as sample process control (e.g., Sketa Assay, U.S. EPA, 2010).

In the present study, salmon sperm DNA and Na-pyrophosphate were tested as additives. Both significantly increased DNA extraction efficiency. In the case of Na-pyrophosphate the high doses used were not able to compensate for the adsorptive capacities of the suspended matter since no saturation plateau was reached for the DNA concentration retrieved. However, in some samples negative effects of Na-pyrophosphate addition were observed in form of PCR inhibition (shift in the Ct value > 1 compared to controls), which is why the Na-pyrophosphate addition was no longer followed up in this study. This observation is also in line with that of Lever et al. (2015) who observed that DNA pellets retrieved after Na-pyrophosphate addition were not only larger but also darker in color, indicating an increased transfer of non-nucleic acid containing organic matter (e.g., humic substances) to the eluate which could further lead to PCR inhibition, as it was observed also in the present study.

The results obtained after salmon sperm DNA addition were somewhat different. As for Na-pyrophosphate addition no saturation plateau of the DNA concentration was retrieved. For the DeTaCS and AllBac marker concentrations, however, indeed a saturation plateau was reached, albeit the amounts of salmon sperm DNA needed to compensate for the adsorptive effects were very high compared to other studies. In general, the quantities of various DNAs described in the literature to be needed for adsorption site saturation vary greatly (e.g., Paulin et al., 2013; Lever et al., 2015; Gardner and Gunsch, 2017; Engel et al., 2019) as do the adsorption capacities described for different sediment types themselves. There seems no consensus conclusion regarding what type of clay exhibits the strongest adsorption capacity for nucleic acids and experimental setup of studies published vary greatly, making a direct comparison difficult. However, both illite and montmorillonite, which are present in the water bodies of Lake Neusiedl area (Löffler, 1979), are described as highly adsorptive for DNA (Ben-Hur et al., 1992; Poly et al., 2000; Cai et al., 2006; Saeki et al., 2010; Yu et al., 2013).

Finally, the established hypothesis was tested on environmental samples from the habitat. To evaluate possible and, based on the previous results, expected effects on quantitative results of selected molecular fecal and MST markers, DNA extraction was performed using the standard and a modified DNA extraction protocol, respectively. For the above mentioned reasons, in the comparative tests solely salmon sperm DNA was used as supplement. However, possible adverse effects on the measurements of DNA concentration should be considered as there might be competitive reactions of salmon sperm DNA and sample DNA. Both will compete for free charges, both will bind to them and it must not be assumed that all adsorption sites are fully saturated before cells are broken up and sample DNA is released from the cells. For this reason, the addition of salmon sperm DNA will not necessarily raise sample DNA retrieval to control levels since some sample DNA will also bind to free charges on particle surfaces. This was also observed in the present study. The DeTaCS retrieval was raised by three $\log_{10}$ units but still remained below the control level (**Figure 3A**). Furthermore, excess amounts of salmon sperm DNA could be co-precipitated alongside with sample DNA affecting the reliability of (sample) DNA concentrations determined (Paulin et al., 2013).

Despite all these considerations the comparative application of the two extraction protocols impressively showed the influence on the obtained quantitative results (**Figures 3A–D**). As in all previous tests the strongest influence of the extraction method was observed in samples from Lake Neusiedl. Both, the qPCR-based ENT marker and the human-associated MST marker HF183/BacR287, could not be detected in any sample subjected to standard DNA extraction. Only after application of the modified extraction protocol these markers could also be detected (**Figures 3C,D**), a result which underlines the importance of testing for matrix effects and possibly applying an adapted extraction protocol. These results, therefore, not only support the hypothesis that inorganic turbidity, when present, can affect the efficiency of DNA extraction, but also clearly demonstrate the influence on the ability of molecular methods to yield reliable quantitative data. Routine use of these markers to determine microbiological water quality would not have revealed, at least intermittently, fecal contamination in the study area if the standard extraction protocol had been used. In cases where such studies are used for risk assessment and targeting of management measures (e.g., guidance on closure of bathing sites), a proper assessment of the current situation would not be accurate.

CONCLUSION

The results of the present study clearly indicate that difficulties in the preparation of environmental samples due to matrix effects, as known from soil samples, should also be considered in the processing of water samples. Further, it seems that it is more the inorganic fraction of suspended solids which influences DNA extraction efficiency since in all experiments DNA retrieval from samples with high inorganic turbidity was affected most severely. Since these materials are present in many water bodies, this factor might become effective in a variety of sample types and even more if samples are taken under certain environmental conditions. By this, samples taken e.g., during flood events and/or strong wind events might be particularly vulnerable to the occurrence of such effects (whirling up of sediments). However, whether effects on DNA extraction are to be expected will not only depend on the amount of suspended material but also on its composition which in turn might affect the amount of additives required to improve the extraction efficiency. For this reason no universal

DNA extraction protocol or general rule of how to improve extraction efficiency can be given.

Apart from all considerations presented here, however, the main conclusion to be drawn is the immense importance of the use of controls, irrespective if further measures are taken in attempt to optimize an extraction protocol. The use of process controls (filtration and extraction controls) on a sample-to-sample basis, will allow direct conclusions on the extraction efficiency. As a more indirect measure, the monitoring of turbidity and TSS could be anticipated, whereby it should be borne in mind that different sediment types can differ strongly with their respect of influencing DNA extraction efficiency.

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

RL: investigation, laboratory work, writing—original draft, and writing—review and editing. SZ and ReM: investigation, laboratory work, and writing—original draft. KK: DeTaCS construction and writing—original draft. DS: investigation and writing—original draft. AK: conceptualization, funding acquisition, and writing—original draft. GR: conceptualization and writing—original draft. RoM: conceptualization. RS: conceptualization and funding acquisition. AF: conceptualization, writing—original draft, writing—review and editing, funding acquisition, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management/Province of Burgenland (grant number B401184 "Nachhaltiges Gesamtkonzept zum Management des mikrobiologisch-hygienischen Zustandes des Neusiedler Sees unter Berücksichtigung der Siedlungswasserwirtschaft der Umlandgemeinden Phase 1: Pilotstudie") and Gesellschaft für Forschungsförderung Niederösterreich (GFF) (grant number LSC 19-016 "Future Danube").

ACKNOWLEDGMENTS

We would like to thank Gerhard Linder for his supportive assistance during the sampling campaigns. Simone Ixenmaier and Stefan Jakwerth are thanked for their supportive assistance in the laboratory. This is a joint investigation within the Interuniversity Cooperation Centre for Water & Health (www.waterandhealth.at). We further acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.660566/full#supplementary-material

REFERENCES

Anderson, A., Pietsch, K., Zucker, R., Mayr, A., Müller-Hohe, E., Messelhäusser, U., et al. (2011). Validation of a duplex real-time PCR for the detection of *Salmonella* spp. in different food products. *Food Anal. Methods* 4, 259–267. doi: 10.1007/s12161-010-9142-8

Behets, J., Declerck, P., Delaedt, Y., Creemers, B., and Ollevier, F. (2007). Development and evaluation of a Taqman duplex real-time PCR quantification method for reliable enumeration of *Legionella pneumophila* in water samples. *J. Microbiol. Methods* 68, 137–144. doi: 10.1016/j.mimet.2006.07.002

Ben-Hur, M., Malik, M., Letey, J., and Mingelgrin, U. (1992). Adsorption of polymers on clays as affected by clay charge and structure, polymer properties, and water quality. *Soil Sci.* 153, 349–356.

Boehm, A. B., Wang, D., Ercumen, A., Shea, M., Harris, A. R., Shanks, O. C., et al. (2016). Occurrence of host-associated fecal markers on child hands, household soil, and drinking water in rural Bangladeshi households. *Env. Sci. Technol. Lett.* 3, 393–398. doi: 10.1021/acs.estlett.6b00382

Boyd, D., Weiss, D. S., Chen, J. C., and Beckwith, J. (2000). Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *J. Bacteriol.* 182, 842–847. doi: 10.1128/JB.182.3.842-847.2000

Cai, P., Huang, Q., Zhang, X., and Chen, H. (2006). Adsorption of DNA on clay minerals and various colloidal particles from an Alfisol. *Soil Biol. Biochem.* 38, 471–476. doi: 10.1016/j.soilbio.2005.05.019

Da Silva, S. M., Vang, L. K., Olson, N. D., Lund, S. P., Downey, A. S., Kelman, Z., et al. (2016). Evaluation of microbial qPCR workflows using engineered *Saccharomyces cerevisiae*. *Biomol. Detect. Quantif.* 7, 27–33. doi: 10.1016/j.bdq.2016.01.001

D'Agostino, M., Cook, N., Rodriguez-Lazaro, D., and Rutjes, S. (2011). Nucleic acid amplification-based methods for detection of enteric viruses: definition of controls and interpretation of results. *Food Environ. Virol.* 3, 55–60. doi: 10.1007/s12560-011-9063-8

Dokulil, M. T., and Herzig, A. (2009). An analysis of long-term winter data on phytoplankton and zooplankton in Neusiedler See, a shallow temperate lake, Austria. *Aquat. Ecol.* 43, 715–725. doi: 10.1007/s10452-009-9282-3

Eiler, A., Farnleitner, A. H., Zechmeister, T. C., Herzig, A., Hurban, C., Wesner, W., et al. (2003). Factors controlling extremely productive heterotrophic bacterial communities in shallow soda pools. *Microb. Ecol.* 46, 43–54. doi: 10.1007/s00248-002-2041-9

Engel, K., Coyotzi, S., Vachon, M. A., McKelvie, J. R., and Neufeld, J. D. (2019). Validating DNA extraction protocols for bentonite clay. *mSphere* 4, e00334–19. doi: 10.1128/mSphere.00334-19

Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., et al. (1999). Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* 65, 5409–5420. doi: 10.1128/AEM.65.12.5409-5420.1999

Gardner, C. M., and Gunsch, C. K. (2017). Adsorption capacity of multiple DNA sources to clay minerals and environmental soil matrices less than previously estimated. *Chemosphere* 175, 45–51. doi: 10.1016/j.chemosphere.2017.02.030

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., 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

Gregory, J. B., Litaker, R. W., and Noble, R. T. (2006). Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. *Appl. Environ. Microbiol.* 72, 3960–3967. doi: 10.1128/AEM.02291-05

Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., and Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* 66, 5488–5491. doi: 10.1128/AEM.66.12.5488-5491.2000

Haugland, R. A., Siefring, S., Lavender, J., and Varma, M. (2012). Influences of sample interference and interference controls on quantification of enterococci fecal indicator bacteria in surface water samples by the qPCR method. *Water Res.* 46, 5989–6001. doi: 10.1016/j.watres.2012.08.017

Haugland, R. A., Siefring, S. C., Wymer, L. J., Brenner, K. P., and Dufour, A. P. (2005). Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Res.* 39, 559–568. doi: 10.1016/j.watres.2004.11.011

Haugland, R. A., Varma, M., Sivaganesan, M., Kelty, C., Preed, L., and Shanks, O. C. (2010). Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected *Bacteroidales* species and human fecal waste by qPCR. *Syst. Appl. Microbiol.* 33, 348–357. doi: 10.1016/j.syapm.2010.06.001

Hedrich, E. (1983). Determination of layers in the sediment of Neusiedler See by neutron activation analysis. *Mikrochim. Acta* 79:22.

Herrera, A., and Cockell, C. S. (2007). Exploring microbial diversity in volcanic environments: a review of methods in DNA extraction. *J. Microbiol. Methods* 70, 1–12. doi: 10.1016/j.mimet.2007.04.005

Hietz, P. (1992). Die rolle des detritus im ökosystem des schilfgürtels am Neusiedler See. *Verhandlungen der Zoologisch-Botanischen Gesellschaft Österreich* 129, 35–66.

Hill, V. R., Narayanan, J., Gallen, R. R., Ferdinand, K. L., Cromeans, T., and Vinjé, J. (2015). Development of a nucleic acid extraction procedure for simultaneous recovery of DNA and RNA from diverse microbes in water. *Pathogens* 4, 335–354. doi: 10.3390/pathogens4020335

Hinlo, R., Gleeson, D., Lintermans, M., and Furlan, E. (2017). Methods to maximize recovery of environmental DNA from water samples. *PLoS One* 12:e0179251. doi: 10.1371/journal.pone.0179251

Hoshino, Y. T., and Matsumoto, N. (2005). Skim milk drastically improves the efficacy of DNA extraction from andisol, a volcanic ash soil. *Jpn. Agric. Res. Q.* 39, 247–252. doi: 10.6090/jarq.39.247

Kirschner, A. K., Schlesinger, J., Farnleitner, A. H., Hornek, R., Süß, B., Golda, B., et al. (2008). Rapid growth of planktonic *Vibrio cholerae* non-O1/non-O139 strains in a large alkaline lake in Austria: dependence on temperature and dissolved organic carbon quality. *Appl. Environ. Microbiol.* 74, 2004–2015. doi: 10.1128/AEM.01739-07

Kirschner, A. K. T., Zechmeister, T. C., Kavka, G. G., Beiwl, C., Herzig, A., Mach, R. L., et al. (2004). Integral strategy for evaluation of the fecal indicator performance in bird-influenced saline inland waters. *Appl. Environ. Microbiol.* 70, 7396–7403. doi: 10.1128/AEM.70.12.7396-7403.2004

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

Krachler, R., Krachler, R., Milleret, E., and Wesner, W. (2000). Limnochemische Untersuchungen zur aktuellen Situation der Salzlacken im burgenländischen Seewinkel. *Burgenländische Heimatblätter* 62, 3–49.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., and Sayler, G. (2006). Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* 72, 4214–4224. doi: 10.1128/AEM.01036-05

Lebuhn, M., Effenberger, M., Garces, G., Gronauer, A., and Wilderer, P. A. (2004). Evaluating real-time PCR for the quantification of distinct pathogens and indicator organisms in environmental samples. *Water Sci. Technol.* 50, 263–270. doi: 10.2166/wst.2004.0065

Lever, M. A., Torti, A., Eickenbusch, P., Michaud, A. B., Santl-Temkiv, T., and Jørgensen, B. B. (2015). A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front. Microbiol.* 6:476. doi: 10.3389/fmicb.2015.00476

Li, X., Sivaganesan, M., Kelty, C. A., Zimmer-Fuast, A., Clinton, P., Reichman, J. R., et al. (2019). Large-scale implementation of standardized quantitative real-time PCR fecal source identification procedures in the Tillamook Bay Watershed. *PLoS One* 14:e0216827. doi: 10.1371/journal.pone.0216827

Löffler, H. (1979). Neusiedlersee, the limnology of a shallow lake in central Europe. *Monogr. Biol.* 37:543.

Lombard, N., Prestat, E., van Elsas, J. D., and Simonet, P. (2011). Soilspecific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol.* 78, 31–49. doi: 10.1111/j.1574-6941.2011.01140.x

Lorenz, M. G., and Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* 58, 563–602.

Mayer, R. E., Reischer, G. H., Ixenmaier, S. K., Derx, J., Blaschke, A. P., Ebdon, J. E., et al. (2018). Global distribution of human-associated fecal genetic markers in reference samples from six continents. *Environ. Sci. Technol.* 52, 5076–5084. doi: 10.1021/acs.est.7b04438

Miller, D. N., Bryant, J. E., Madsen, E. L., and Ghiorse, W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* 65, 4715–4724. doi: 10.1128/AEM.65.11.4715-4724.1999

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., et al. (2015). Validation of bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci. Total Environ.* 502, 462–470. doi: 10.1016/j.scitotenv.2014.09.040

Paulin, M. M., Nicolaisen, M. H., Jacobsen, C. S., Gimsing, A. L., Sorensen, J., and Baelum, J. (2013). Improving Griffith's protocol for co-extraction of microbial DNA and RNA in adsorptive soils. *Soil Biol. Biochem.* 63, 37–49. doi: 10.1016/j.soilbio.2013.02.007

Poly, F., Chenu, C., Simonet, P., Rouiller, J., and Monrozier, L. J. (2000). Differences between linear chromosomal and supercoiled plasmid DNA in their mechanisms and extent of adsorption on clay minerals. *Langmuir* 16, 1233–1238. doi: 10.1021/la990506z

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Aström, J., et al. (2013). Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* 47, 8548–8556. doi: 10.1021/es304367t

Reischer, G. H., Haider, J. M., Sommer, R., Stadler, H., Keiblinger, K. M., Hornek, R., et al. (2008). Quantitative microbial faecal source tracking with sampling guided by hydrological catchment dynamics. *Environ. Microbiol.* 10, 2598–2608. doi: 10.1111/j.1462-2920.2008.01682.x

Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., and Farnleitner, A. H. (2006). Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Appl. Environ. Microbiol.* 72, 5610–5614. doi: 10.1128/AEM.00364-06

Reischer, G. H., Kollanur, D., Vierheilig, J., Wehrspaun, C., Mach, R. L., Sommer, R., et al. (2011). Hypothesis-driven approach for the identification of fecal pollution sources in water resources. *Environ. Sci. Technol.* 45, 4038–4045. doi: 10.1021/es103659s

Rocha, J., and Manaia, C. M. (2020). Cell-based internal standard for qPCR determinations of antibiotic resistance indicators in environmental water samples. *Ecol. Ind.* 113:106194. doi: 10.1016/j.ecolind.2020.106194

Saeki, K., Kunito, T., and Sakai, M. (2010). Effects of pH, ionic strength, and solutes on DNA adsorption by andosols. *Biol. Fertil. Soils* 46, 531–535.

Schauer, S., Jakwerth, S., Bliem, R., Baudart, J., Lebaron, P., Huhulescu, S., et al. (2015). Dynamics of *Vibrio cholerae* abundance in Austrian saline lakes, assessed with quantitative solid-phase cytometry. *Environ. Microbiol.* 17, 4366–4378. doi: 10.1111/1462-2920.12861

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* 113, 1014–1026. doi: 10.1111/j.1365-2672.2012.05384.x

Sen, K., Schable, N. A., and Lye, D. J. (2007). Development of an internal control for evaluation and standardization of a quantitative PCR assay for detection of *Helicobacter pylori* in drinking water. *Appl. Environ. Microbiol.* 73, 7380–7387. doi: 10.1128/AEM.00687-07

Shanks, O. C., Atikovic, E., Blackwood, A. D., Lu, J., Noble, R. T., Santo Domingo, J., et al. (2008). Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* 74, 745–752. doi: 10.1128/AEM.01843-07

Shanks, O. C., Kelty, C. A., Oshiro, R., Haugland, R. A., Madi, T., Brooks, L., et al. (2016). Data acceptance criteria for standardized human-associated fecal source identification quantitative real-time PCR methods. *Appl. Environ. Microbiol.* 82, 2773–2782. doi: 10.1128/AEM.03661-15

Soja, G., Kitzler, B., and Soja, A.-M. (2014). Emissions of greenhouse gases from Lake Neusiedl, a shallow steppe lake in Eastern Austria. *Hydrobiologia* 731, 125–138. doi: 10.1007/s10750-013-1681-8

U.S. EPA (2010). *Method A: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (QPCR) Assay*. Report NO. EPA-821-R-10e004. Washington, DC: US Environmental Protection Agency.

Wolfram, G. (1996). Distribution and production of chironomids (Diptera: Chironomidae) in a shallow, alkaline lake (Neusiedler See. Austria). *Hydrobiologia* 318, 103–115. doi: 10.1007/BF00014136

Wolfram, G. (2006). "Bedeutung und vorkommen von salzlebensräumen," in *Salzlebensräume in Österreich*, eds G. Wolfram, et al. (Wien: Umweltbundesamt), 13–26.

Wuertz, S., Wang, D., Reischer, G. H., and Farnleitner, A. H. (2011). "Library-independent source tracking methods," in *Microbial Source Tracking: Methods, Applications, and Case Studies*, eds C. Hagedorn, A. R. Blanch, and V. J. Harwood (New York, NY: Springer), 61–113. doi: 10.1007/978-1-4419-9386-1_4

Xue, J., and Feng, Y. (2018). Determination of adsorption and desorption of DNA molecules on freshwater and marine sediments. *J. Appl. Microbiol.* 124, 1480–1492. doi: 10.1111/jam.1373

Yu, W. H., Li, N., Tong, D. S., Zhou, C. H., Lin, C. X., and Xu, C. Y. (2013). Adsorption of proteins and nucleic acids on clay minerals and their interactions: a review. *Appl. Clay Sci.* 80–81, 443–452. doi: 10.1016/j.clay.2013.06.003

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.

Copyright © 2021 Linke, Zeki, Mayer, Keiblinger, Savio, Kirschner, Reischer, Mach, Sommer and Farnleitner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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