December 2013

Transcending Microbial Source Tracking Techniques Across Geographic Borders: An Examination of Human and Animal Microbiomes and the Integration of Molecular Approaches in Pathogen Surveillance in Brazil and the United States

Amber Mae Koskey
University of Wisconsin-Milwaukee

TRANSCENDING MICROBIAL SOURCE TRACKING TECHNIQUES ACROSS

GEOGRAPHIC BORDERS: AN EXAMINATION OF HUMAN AND ANIMAL

MICROBIOMES AND THE INTEGRATION OF MOLECULAR APPROACHES IN

PATHOGEN SURVEILLANCE IN BRAZIL AND THE UNITED STATES

by

Amber M. Koskey

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Freshwater Sciences and Technology

at

The University of Wisconsin- Milwaukee

December 2013

ABSTRACT
TRANSCENDING MICROBIAL SOURCE TRACKING TECHNIQUES ACROSS
GEOGRAPHIC BORDERS: AN EXAMINATION OF HUMAN AND ANIMAL
MICROBIOMES AND THE INTEGRATION OF MOLECULAR APPROACHES IN
PATHOGEN SURVEILLANCE IN BRAZIL AND THE UNITED STATES

by

Amber M. Koskey

The University of Wisconsin-Milwaukee, 2013
Under the Supervision of Professor Sandra McLellan

Waterborne illnesses, attributed to the ingestion or contact with contaminated water,

present a significant global health concern. Surface water sources can be impacted by

wide array of pollution inputs, but fecal pollution generates the most significant and acute

threat to human health. Therefore, the detection of fecal bacteria in surface water sources

remains an important public health objective. Current surface water monitoring employs

the use of fecal indicator bacteria (FIB) including *E. coli* and enterococci as proxies for

pathogenic organisms carried in fecal pollution. These traditional indicators, detected by

culture-based microbiological methods, do not discriminate fecal sources from another.

New molecular approaches in pathogen surveillance, such as microbial source tracking

(MST) and fecal-associated signatures, are culture-independent and are better suited for

both the detection and identification of fecal pollution sources. By identifying fecal

pollution sources, human health risks can be more accurately assessed and remediation

strategies can be effectively implemented.

This paper examines a variety of MST markers, and the basis for these by integrating in

host source microbiome studies. Chapter 2 describes work with *Catellicoccus*

marimammalium, where next generation sequencing demonstrates this marker is a dominant member of the gull microbiome. This work has important implications for reconciling high fecal indicator levels at beaches with health risk. Chapter 3 extends MST work to areas of poor sanitation in Jenipapo, Brazil. The distribution of human specific indicators in surface water fecal contamination and prevalence of the waterborne illness schistosomiasis is described. Lastly, Chapter 4 explores the microbial community of humans and animals across different geographic regions, Brazil and the United States, to evaluate the applicability of existing MST methods, assess host-specific organisms and fecal-associated bacterial groups, and investigate the potential to develop new and geographically-appropriate markers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

CHAPTER 1: Background and Significance

Fecal pollution in surface water sources as a public health threat

Worldwide, fecal pollution in surface water sources is a significant water quality impairment that constitutes a major public health threat. Waterborne illnesses, related to the inadequate provision of water and sanitation services, are responsible for 4 billion cases of diarrhea and 1.8 million deaths each year, mostly impacting children 5 years of age and younger living in the rural communities of developing countries (1-2). In these communities, drinking water supplies are often local waterbodies (i.e. streams, rivers or lakes) that are contaminated from non-point runoff and at times, the direct discharge of untreated sewage (3). Quite often, water collected from these sources - for household use (i.e. sanitation and hygiene purposes) and consumption - is used or consumed prior to treatment (3). The lack of sanitation and drinking water infrastructures within these rural communities creates a self-perpetuating cycle of diarrheal diseases, along with other waterborne infections associated with differing routes of exposure, resulting in: eye infections, skin irritations, ear, nose and throat infections and respiratory illness (4). Schistosomiasis, in particular, is a parasitic waterborne infection whose transmission is dependent on human fecal contamination of fresh water. Unlike most bacterial and viral waterborne illnesses that are acquired through ingestion, infection results through the contact of contaminated water (5).

In more developed regions of the world, drinking water sources are treated and piped directly into one's house, and sanitation services - sewage treatment facilities or septic systems - are readily abundant and utilized, so the risk of acquiring waterborne illnesses

from drinking water sources is minimized. However, waterborne illnesses attributed to recreational water exposures remains a risk factor, especially during summer months when bacterial levels increase with increasing water temperatures and when recreational water use is at its peak (6). Numerous studies have found correlative relationships between gastrointestinal (GI) illnesses and exposure to recreational waters impacted by fecal contamination (4, 7). As indicated through the Centers for Disease Control and Prevention (CDC) Surveillance for waterborne disease outbreaks Morbidity and Mortality Weekly Report (MMWR) Surveillance Summaries, the incidence of illness attributable to recreational water exposure appears to be increasing over time. There were 21 recreational water outbreaks reported in 2000, 78 from 2005 to 2006, and 134 from 2007 to 2008 (8-9). Beach closings and advisories have also been increasing in recent years; the Natural Resources Defense Council (NRDC) reported in 2008 that beach closings and advisories hit their fourth highest level in the 19 years that the NRDC has been tracking them (10). NRDC also reports that in 2008, the number of closings and advisory days at ocean, bay and Great Lakes Beaches topped 20,000 for the fourth year in row, confirm that our nation's beaches continue to suffer from serious water pollution that puts swimmers at risk (10). Whether one's waterborne illness is attributed to ingestion (i.e. contaminated drinking water) or through physical contact (i.e. general household use or recreational activities), understanding the exposure routes, transmission dynamics and sources of waterborne infections is imperative in preventing human illness.

Surface water sources and fecal pollution inputs

Surface water sources are susceptible to a wide range of pollution inputs, physical, chemical and microbiological (11). The protection from pathogenic microbes, such as those identified in fecal contamination, is designated by the Clean Water Act as the most important constituent of water used for recreation and public water supplies, along with other uses (11-12). Fecal contamination enters surface water sources via non-point runoff, domesticated animals, wildlife and agriculture, and through point sources, such as the direct discharge of untreated sewage. Whether from animals or humans, fecal contamination is significant water quality impairment and risk to human health; however, exposure to human feces poses a greater health risk. Although animals serve as reservoirs for a variety of enteric pathogens including: various serotypes of *Salmonella*, *Escherichia coli*, *Cryptosporidium* spp. *Campylobacter*, *Giardia* spp., human feces are more likely to contain human-specific enteric pathogens: *Shigella* spp., hepatitis A virus, Norwalk-group viruses, and *Salmonella enterica* serovar Thypi, as well as the previously identified pathogens (4, 12). Consequently, humans are at higher risk when exposed to water contaminated with human feces. In a comparison study that looked that the human health risks from exposure to recreational waters impacted by human and non-human sources of fecal contamination, recreational water contaminated by sewage and cattle posed a higher relative risk of acquiring GI illnesses than water sources contaminated by gulls, chickens and pigs (4); therefore, knowing the source of fecal contamination (human vs. animal) is important for risk assessment (12). The identification of fecal contamination sources also enables the development of appropriate and cost-effective mitigation strategies.

Water quality monitoring techniques

Current surface water monitoring employs the use of fecal indicator bacteria (FIB), such as *E. coli* and enterococci, to assess the presence of fecal pollution in water; however, these traditional monitoring methods lack the ability to discriminate fecal sources. Traditional indicator microorganisms are ubiquitous in nature and are found in the feces of humans and animals alike, which limits its usefulness as a method to assess human health risk (12-13). In light of the limitations of traditional monitoring methods, integrating new molecular approaches in pathogen surveillance are advantageous as they: provide a better indicator of human health risk, assess host-specificity and can subsequently be used for mitigation efforts. Molecular techniques, such as microbial source tracking (MST) and fecal-associated signatures, can be used to detect and identify pollution sources in surface water sources.

MST uses bacterial genetics to characterize host-specific microorganisms. Host-specific microorganisms are recognized as an organism's microbial signature and once identified, can be used to track pollutants to their source inputs (12). Using fecal-associated microbial signatures is another molecular technique that can be used to identify fecal pollution in surface water sources. Similar to MST, fecal signatures consist of a suite of organisms that are abundant in the microbial community of fecal pollution (13-14). The profiling of multiple markers, instead of a single host-specific organism, enhances the specificity of identifying multiple fecal sources in contaminated and complex environmental samples (13). Given the complexity of watershed systems and array of fecal contamination inputs (i.e. wildlife, agriculture, sewage), microbial source tracking

and fecal-associated microbial signatures, are both excellent tools for detecting fecal
contamination in surface water sources and identifying pollution sources. By identifying
the pollution source, efforts can be further made to quantify human health risk, develop
appropriate mitigation strategies and protect human health.

Geographic applicability of microbial source tracking

Geographic variability and host specificity can factor into the applicability of MST
markers, so it is important that the regional utility of a MST marker is tested and
validated across different watersheds (12). Further research is needed to evaluate if
makers developed in one region, state or country will transcend different geographical
regions, as there are a multitude of factors that have been shown to impact and diversify
the gut microbiota of animals and humans, including: diet, sanitation, genetics,
environment, age, antibiotic exposure, geography, etc. (13-16). The intestinal tract of
humans and animals is a complex ecological community that fluctuates overtime and
varies substantially between different individuals and organisms (15), so the
identification of MST markers requires: a comprehensive investigation of fecal microbial
community populations; microbial community comparisons, within and between different
organisms; an assessment of host-specific organisms and fecal-associated bacterial
groups; and an investigation of geographically-appropriate markers.

Purpose of the Study

This paper investigates the exposure routes and transmission dynamics of waterborne
infections, utilizing MST methods and fecal-associated microbial signatures for the

detection and identification of fecal contamination in surface water sources. This paper will also examine the relationship between traditional FIB measurements and MST techniques, and compare potential MST markers with host microbiome assemblages. It will also look at the distribution of surface water fecal contamination and prevalence of the waterborne illness schistosomiasis. Lastly, this paper will also explore the microbial community populations of humans and animals across different geographic regions, Brazil and the United States, to: evaluate the applicability of existing MST methods; assess host-specific organisms and fecal-associated bacterial groups and; investigate the potential to develop new and geographically-appropriate markers. The overall goal of this research is to become more proficient at integrating molecular techniques to detect and identify fecal contamination sources, to subsequently minimize public health exposure risks and prevent human illness.

CHAPTER 1: References

1. Clasen, T.F., Thao, D.H., Boisson, S., and Shipin, O. (2008) Microbiological Effectiveness and Cost of Boiling to Disinfect Drinking Water in Rural Vietnam. Environmental Science and Technology. 42(12):4255-4260.

2. World Health Organization (WHO). (2007) Combating waterborne disease as the household level. International Network to Promote Household Water Treatment and Safe Storage. Retrieved from http://www.who.int/household_water/advocacy/combating_disease/en/index.html

3. World Health Organization (WHO) and UNICEF. (2013) Progress on Sanitation and Drinking Water. Retrieved from http://apps.who.int/iris/bitstream/10665/81245/1/9789241505390_eng.pdf.

4. Soller, J.A., Schoen, M.E., Bartrand, T., Ravenscroft, J.E., and Ashbolt, J.N. (2010) Estimated human health risks from exposure to recreational water impacted by human and non-human sources of faecal contamination. Water Research. 44(16):4674-4691.

5. World Health Organization (WHO). (2001) Water-related Diseases: Schistosomiasis. Retrieved from http://www.who.int/water_sanitation_health/diseases/schisto/en/

6. Schets, F.M., Roda, D.E., Husman, A.M., and Havelaar, A.H. (2010) Disease outbreaks associated with untreated recreational water use. Epidemiological Infections. 10:1-12.

7. Pruss, A. (1998) Review of epidemiological studies on health effects from exposure to recreational water. International Journal of Epidemiology. 27(1):1-9.

8. Lee, S.H., Levy, D.A., Craun, G.F., Beach, M.J., and Calderon, R.L. (2002) Surveillance for waterborne disease outbreaks—United States, 1999–2000. MMWR Surveillance Summary. 51(8):1–47.

9. Hlavsa, M.C., Roberts, V.A., Anderson, A.R., Hill, V.R., Kahler, A.M., Orr, M., Garrison, L.E., Hicks, L.A., Newton, A., Hilborn, E.D., Wade, T.J., Beach, M.J., and Yoder, J.S. (2011) Surveillance for Waterborne Disease Outbreaks and Other Health Events Associated with Recreational Water --- United States, 2007—2008 and Surveillance for Waterborne Disease Outbreaks Associated with Drinking Water – United States, 2007-2008. Center for Disease Control and Prevention (CDC). Morbidity and Mortality Weekly Report (MMWR). Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6012a1.htm

10. Dorfman, M and Rosselot, K. (2009) Testing the Waters: A Guide to Water Quality at Vacation Beaches. New York. Natural Resources Defense Council. Retrieved from http://www.nrdc.org/water/oceans/ttw/ttw2009.pdf

11. Simpson, J.M., Santo Domingo, J.W., and Reasoner, D.J. (2002) Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36(24):5279-5288.

12. Environmental Protection Agency (EPA). (2012) Clean Water Act Section 303. Water Quality Standards and Implementation Plans. Retrieved from http://water.epa.gov/lawsregs/guidance/303.cfm

13. Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., and Lukasik, J. (2002) Microbial Source Tracking: Current Methodology and Future Directions. Applied and Environmental Microbiology. 68(12):5796-5803.

14. Orskov, F and Orskov, I. (1981) Enterobacteriaceae, p. 340–352. In A. I. Broude (ed.), Medical microbiology and infectious diseases. W. B. Saunders Co., Philadelphia, Pa.

15. Newton, R.L., Bootsma, M.J., Morrison, H.G., Sogin, M.L., and McLellan, S.L. (2013) A Microbial Signature Approach to Identifying Fecal Pollution in the Waters Off an Urbanized Coast of Lake Michigan. Microbial Ecology. 65:1011-1023.

16. 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 plan influent. Environmental Microbiology. 12(2)378-392.

17. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhim, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R., and Gordon, J.I. (2012) Human gut microbiome viewed across age and geography. Nature. 486:222-227.

18. Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. PNAS. doi: 10.1073/pnas.1005963107.

19. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012) Diversity, stability and resilience of the human gut microbiota. Nature. 489 (7415):220-230.

20.   Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D., and Lewis, J.D. (2011) Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science. 334(6052):105-108.

CHAPTER 2: Analysis of gull fecal microbial community reveals dominance of

***Catellicoccus marimammalium* in relation to culturable enterococci**

Abstract

Gulls are prevalent in beach environments and can be a major source of fecal contamination. Gulls have been shown to harbor a high abundance of fecal indicator bacteria (FIB), such as *E. coli* and enterococci, which can be detected readily as part of routine beach monitoring. Despite the ubiquitous presence of gull fecal material in beach environments, the associated microbial community is relatively poorly characterized. We generated comprehensive microbial community profiles of gull fecal samples using Roche 454 and Illumina MiSeq platforms to investigate the composition and variability of the gull fecal microbial community and to measure the proportion of FIB. *Enterococcaceae* and *Enterobacteriaceae* were the two most abundant families in our gull samples. Sequence comparisons between short-read data and near full-length 16S rRNA gene clones generated from the same samples revealed *Catellicoccus marimammalium* as the most numerous taxon among all samples. The identification of bacteria from gull fecal pellets cultured on membrane-Enterococcus indoxyl-β-D-glucoside (mEI) plates showed that the dominant sequences recovered in our sequence libraries did not represent organisms culturable on mEI. Based on 16S rRNA gene sequencing of gull fecal isolates cultured on mEI plates, 98.8% were identified as *Enterococcus* spp., 1.2% *Streptococcus* spp.; and none were identified as *C. marimammalium*. Illumina deep sequencing indicated that gull fecal samples harbor significantly higher proportions of *C. marimammalium* 16S rRNA gene sequences (>50-fold), relative to typical mEI culturable *Enterococcus* spp. *C. marimammalium* therefore can be confidently utilized as a genetic marker to identify gull fecal pollution in the beach environment.

Introduction

Gulls, and other shorebirds and waterfowl, are prevalent in beach environments and their feces are considered a major source of FIB in coastal and lake waters worldwide (1-5). *E. coli* and enterococci are commonly used to monitor the presence of pathogenic organisms in beach environments and assess the relative human health risks associated with recreational water use (6); however, these FIB are ubiquitous as they are found in the feces of humans and animals alike, limiting their utility to accurately predict human health risks (7). Although there is evidence to support the shedding of human pathogens (8-11), the human health risks associated with shorebirds and waterfowl feces are inherently lower than those from human fecal inputs (12, 13). Furthermore, bird droppings are a more concentrated, point source of contamination on beaches, while sewage is more readily dispersed, presenting a larger radius of contamination risk. Beach advisories are implemented to protect public health, but many unnecessary beach closings result from FIB seeding by gulls (2, 3). These closings can have huge economic implications, especially in coastal areas (14); therefore, the ability to discriminate the source of fecal contamination is critical for both human health risk assessments and mitigation of economic losses.

Currently, most health-related surface water monitoring programs use selective media to culture FIB, usually *E. coli* and enterococci. Gulls often shed high densities (10^5-10^9 CFUs of *E. coli* and 10^4-10^8 CFUs of *Enterococcus* spp. g^{-1} feces) of these organisms via their droppings to surface waters and beach environments (3, 15). Thus, gull contamination can confound the monitoring efforts of recreational water sources

proposed by the United States Environmental Protection Agency (16). Gull populations have steadily increased in urban areas over the past thirty years because of more readily abundant food supplies, long lifespans, and few natural predators (17-19). Recreational water quality monitoring efforts are, in turn, impacted by these growing gull populations.

Traditional culture-based monitoring methods have been successful at establishing human health risk thresholds; however, these methods generally lack the ability to identify contamination sources (7). The inability to distinguish sources of fecal contamination - human sewage, avian species, domestic pets or other urban animals - hinders management efforts and conveys an inexact measure of human health risks associated with recreational water use. Discriminating between contamination sources in beach and coastal environments can improve risk assessment and mitigation strategies, and help limit unnecessary beach advisories and closings caused by sources that do not carry pathogenic organisms.

Despite the ubiquitous presence of gulls and their impact on water quality monitoring efforts, analysis of the organisms harbored in gull feces has been limited to traditional cloning methods (20-22) and a comprehensive microbial community profile has yet to be investigated. Next generation sequencing platforms have gained popularity in recent years as a cost-effective way to deeply explore the microbial community compositions of many samples, while providing taxonomic resolution nearly equivalent to full-length 16S rRNA gene sequences (23, 24). Deeper resolution of microbial communities from gull feces can aid in discerning the gulls' contribution to the total fecal load in beach

environments. Furthermore, an increased understanding of the gull fecal bacterial composition could assist in the development and validation of source-specific assays to enhance beach-monitoring efforts.

Previous attempts to monitor gull pollution have resulted in the development of assays targeting specific bacteria as markers of fecal pollution. Gull-2 (PCR) and gull2, gull3, gull4 and LeeSeaGull (real-time PCR assays) target the 16S rRNA gene of *Catellicoccus marimammalium* (Gull-2, gull2 and gull4) and the genus *Streptococcus* (gull3) (21, 22, 25). The assays targeting *C. marimammalium* seem the most promising as this organism is commonly found in gull fecal samples, is host-specific when compared to non-avian animals, and is detectable in water sources with suspected gull contamination (21). Geographic variability can factor into the efficiency and breadth of microbial source tracking marker use, so the regional utility of a marker must be well tested and validated. A recent nationwide, multi-laboratory assessment of PCR methods targeting *C. marimammalium* discovered that MST methods designated for gull detection were cross-reactive with pigeon feces from California (26). The prevalence, geographic scope and ecology of *C. marimammalium* in host birds warrants further investigation, and future assessments should encompass environmental samples from diverse geographic regions (26). Given these findings, it will be important to note the abundance of *C. marimammalium* and general composition of gull fecal samples from Lake Michigan.

In this study, we used the Roche 454 and Illumina MiSeq deep sequencing platforms to examine the composition and relative abundance of bacteria in gull fecal samples. We

obtained 16S rRNA gene sequences for bacterial species cultured on mEI plates using Sanger sequencing and determined their proportion in the overall community. We explored the potential for gull fecal contamination to confound current water quality monitoring efforts through their contribution of commonly cultured FIB. Lastly, we evaluated the regional utility of the Gull-2 assay for beaches along Lake Michigan using gull fecal samples and environmental samples with presumed gull contamination.

Materials and Methods

Study area, sample collection, and DNA extraction of gull fecal samples

Fresh gull fecal samples were collected aseptically with sterile spatulas from parking lot surfaces at Bradford Beach (Milwaukee, WI) and Grant Park (South Milwaukee, WI). Samples were collected within 2-3 minutes after deposition to prevent overgrowth with non-fecal bacteria. In total, 57 gull fecal samples were collected at Bradford Beach (n=42) and Grant Park (n=13), and Point Beach (n=2) in Two Rivers, WI, between December 2011 and October 2012. The collected samples were stored in sterile 2 mL tubes, transported to the laboratory within two hours, and stored at -80 °C. DNA extractions were performed on 200 mg of the fecal pellets using the Qiagen Stool Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol, with minor modifications (doubling the amount of ProteinaseK, transferring 2X supernatant volumes, and eluting with both DES and AE Buffer). Extracted DNA was stored at -20 °C until further analysis. Four samples from Grant Park (2012) and four from Bradford Beach (two each from 2011 and 2012) were further purified using the MO BIO PowerClean DNA cleanup kit for sequencing (MoBio Laboratories, Carlsbad, CA). DNA

concentration and purity were measured using a NanoDrop ND-1000 (Thermo Fisher Scientific, Pittsburg, PA).

Next generation sequencing and 16S rRNA gene data set analysis

454 pyrosequencing targeting the V4-V6 hypervariable regions of the 16S rRNA gene generated ~7,000 quality reads per gull fecal sample. Primers 518F (5′-CCAGCAGCYGCGGTAAN-3′), 1064R (5′-CGACRRCCATGCANCACCT-3′) amplified DNA, and primer 565F (5′-TGGGCGTAAAG-3′) allowed for bioinformatic trimming (27) of raw sequence reads from the Roche genome sequencer GS-FLX. Illumina sequencing was also performed for six of eight gull samples (Gulls 1, 4-8). The V4-V5 hypervariable regions were amplified in these samples according to protocols developed in the Josephine and Bay Paul Center at the Marine Biological Laboratory, Woods Hole, MA (28). Sequences from both datasets were trimmed, quality controlled, and aligned; data was stored in the Visualization and Analysis of Microbial Population Structures (VAMPS) database (http://vamps.mbl.edu). Taxonomic assignments were made for all sequences using Global Alignment for Sequence Taxonomy (GAST) (23). VAMPS taxonomic count tables were normalized to maximum for microbial community comparisons.

***E. columbae* and *C. marimammalium* reference sequence comparisons and Gull-2 alignment**

Similarity entropy plots comparing full-length sequences of *E. columbae* (GAST annotation of the most abundant species of bacteria in gull samples) and *C.*

marimammalium (identified by Lu et al. (21) as the dominant species in gulls and target for the Gull-2 assay) were made in Vector NTI (Life Technologies, Grand Island, NY). NCBI reference sequences (*E. columbae* - AF0661006, *C. marimammalium* – AJ854484) were used for comparison.

16S rRNA clone library generation from gull fecal samples

Three of the eight gull fecal samples (Gulls 1, 2 and 5) were used to generate 16S rRNA gene clone libraries to enable a more definitive classification of the most abundant taxa in the gull microbial communities. PCR was performed with universal primers, 8F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′). Products were purified and cloned as described previously (20); sequencing was carried out with both the M13F and M13R primers using the ABI BigDye Terminator Kit (Applied Biosystems, Foster City, CA) on an ABI Prism 3700xi (Applied Biosystems, Foster City, CA). Nearly full-length (~1400 bp) contigs were constructed from forward and reverse sequence reads with sufficient overlap (CLC Genomics Workbench, Cambridge, MA). A total of 223 clones were analyzed; incomplete and partial sequences were removed from the dataset. Quality sequences were aligned in mothur (29) using the SILVA alignment as a reference. Chimeras were identified using DECIPHER (30) and removed. Quality sequences were uploaded into Ribosomal Database Project (RDP) for classification (31). BLAST (32) was used to obtain species-level identification of sequences identified by RDP as *Catellicoccus.* Sequence alignment comparisons were done in MEGA 5 (33) to evaluate the number of times the Gull-2 assay would match our nearly full-length *C. marimammalium* contigs (n=176).

Sequencing and phylogenetic analysis of gull fecal bacterial isolates cultured on mEI plates

Fresh fecal samples were collected aseptically, as described above, from Grant Park, South Milwaukee, WI on October 18, 2012. Fecal pellets were re-suspended in a 1:1 weight/volume of PBS; further dilutions (~1:10 and 1:100) were made using sterile water. The dilutions were filtered onto a 47-mm diameter, 0.45-μm-pore-size nitrocelluose filters (Millipore, Billerica, MA) and placed on membrane-Enterococcus indoxyl-β-D-glucoside (mEI) agar plates (34) and incubated for 24 hours at 41°C. Plates that generated enough colonies, without being too numerous to count (TNTC), were selected for colony PCR with universal primers 8F and 1492R to amplify the 16S rRNA gene. PCR products were cleaned up using ExoSAP-IT (Affymetrix, Santa Clara, CA) and sequencing was carried out as above, but with only the 8F primer.

Phylogenetic analyses of high quality forward reads from the mEI isolates, ~874 base pairs long, were conducted in MEGA 5 (33). Clones representing the 111 unique sequences were selected out of 342 total. Four isolates that were identified by BLAST (32) as *Streptococcus lutetiensis* were removed from the final phylogenetic tree. A neighbor-joining phylogenetic tree based on evolutionary distances estimated using the p-distance method was generated with the unique sequence representatives and reference sequences in MEGA 5 (33); distances are given as the number of base differences per site. A bootstrap test was performed with 1000 replicates. Isolates that clustered around a known reference sequence, with 99% sequence identity or greater, were grouped together

and annotated as such. Isolates that clustered with two or more reference sequences were left in the phylogenetic tree without classification.

An alignment and cluster analysis was performed in mothur (29) to verify the proportion of mEI culturable organisms within our dataset. Sequence reads from the Illumina dataset and *Enterococcus* spp. sequences obtained from the mEI isolates (albeit not the same gull samples as were sequenced, but from the same study area) were compared; only exact matches were counted.

Sequences within the Illumina dataset identified as *C. marimammalium* (n=224,082) were compared to sequences identified as an exact match to our cultured mEI isolates to generate the proportional relationship between *C. marimammalium* and culturable *Enterococcus* spp. Proportions of culturable *Enterococcus* identified to the species level (e.g., *E. faecium, E. faecalis, E. hirae, E. durans*; n=142), and sequences that may represent organisms that might be cultured on mEI, but not identified to the species level (i.e., *Enterococcus (species N/A)*, n=4,132), were also compared to *C. marimammalium*.

Detection and prevalence of *C. marimammalium* gull fecal and environmental samples from the Milwaukee region

Gull-2 assay (Gull-2F: 5′-TGCATCGACCTAAAGTTTTGAG-3′ and Gull-2R: 5′-GTCAAAGAGCGAGCAGTTACTA-3′) PCR amplifications were carried out as described in Lu et al. (21) on fecal samples collected between December 2011 and October 2012 (n=57). Preliminary tests revealed that the Gull-2 assay amplified with

better efficiency when BSA was added to the reaction. Sensitivity tests were also conducted on the four gull fecal samples with the highest concentrations of DNA to test the detection limit of the Gull-2 assay. PCR was done on serial dilutions (10^0 to 10^6) from a 1:1 wt:vol mixture of fecal matter and PBS.

Environmental samples collected during the 2010 and 2011 field seasons from Bradford, Atwater, Point, and McKinley Beaches in and near Milwaukee with elevated plate counts of both *E. coli* and enterococci were selected for Gull-2 screening. Sand and water samples were previously extracted using the MO Bio Soil Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA) and stored at -80 ˚C. PCR amplifications of the Gull-2 assay were carried out as described above using 5 μL of DNA in a final reaction volume of 25 μL.

16S rRNA gene sequences

Isolate and clone library 16S rRNA gene sequences were deposited into GenBank under the following accession numbers: KF250762 - KF250872 (mEI isolate sequences) and KF250873 - KF251018 (clone library sequences). The 454 pyrosequences and Illumina sequences are available through the VAMPS database (www.vamps.mbl.edu/).

Results

Microbial community structure in gull fecal samples

A total of 56,428 pyrosequences were generated from eight gull fecal samples. *Enterococcaceae* and *Enterobacteriaceae* were the two most abundant families,

comprising 58% (*Enterococcaceae*) and 29% (*Enterobacteriaceae*) of the samples (Figure 1). Nearly all samples contained ≥ 50% *Enterococcaceae* or *Enterobacteriaceae*, with the exception of Gull 6, which was also composed of *Pseudomonadaceae*, *Erysipelotrichaceae, Moraxellaceae, Lactobacillaceae, Leuconostocaceae, Clostridiaceae,* and *Staphylococcaceae* and had the highest diversity by far of any of the fecal samples.

Catellicoccus was the most abundant genus on average, accounting for 55% of the total pyrosequences recovered from the eight gull fecal samples (Table 1). *Enterococcus*, *Escherichia*/*Shigella*, *Clostridium, Klebsiella*, and *Pseudomonas* were also common genera, each accounting for >10% of the sequences obtained in at least one fecal sample. *Catellicoccus* and *Enterococcus* were the only genera present in all samples. Gull 6 had a more heterogeneous distribution of genera than the others, with relatively high proportions of *Pseudomonas*, *Klebsiella*, and *Psychrobacter*. A few typically non-fecal genera were also relatively common in our samples, including: *Serratia, Staphylococcus*, and *Comamonas*.

Not only was the general taxonomic diversity in V4-V6 reads from gull fecal samples low, but sequences within the *Catellicoccus* group shared high similarity. A single sequence accounted for over 16% of the *Catellicoccus* sequence reads. Furthermore, ~33% of the pyrosequences shared ≥99% sequence identity over a >400 bp read length.

Clone libraries constructed from a subset of the same DNA samples verified the pyrosequencing results. Over 94% of the 223 clone library-based sequences were classified as *Catellicoccus. Catellicoccus* accounted for >98% of clones in Gull 1 (1.4% as *Salmonella)*, 100% in Gull 2, and 84% in Gull 5 (9% *Escherichia/Shigella,* 5 % *Salmonella*, and 1% as *Streptococcus;* Table 1). Nearly all sequence reads identified as *Catellicoccus* (208/210) were ≥99% match to the *C. marimammalium* reference sequence NR042357. The two remaining clones were a 96% match to the same reference sequence. A comparison between the clone libraries and 454 pyrosequences revealed that more than 90% of the *Catellicoccus* sequences shared ≥99% sequence identity in the V5-V6 regions.

Taxonomic resolution of *Catellicoccus marimammalium*, the dominant taxonomic group in gull feces

"*Enterococcus columbae"* was the initial taxonomic assignment of the most numerous sequence group in the 454 pyrosequence dataset. However, comparison to the RDP and NCBI databases showed that this sequence group matched most closely to *Catellicoccus marimammalium* (Table 1). The majority of nearly full-length (~1400 bp) 16S rRNA gene sequences of these libraries exactly matched the pyrosequences in the V4-V6 region, confirming that *C. marimammalium* was the dominant species in our gull samples.

E. columbae and *C. marimammalium* reference sequences shared the highest identity (>97%) at the end of the 16S rRNA gene within the 499 bp hypervariable region targeted

for pyrosequencing (V4-V6). The highest variability occurred within the V1-V3 regions, which are targeted by the Gull-2 assay (21) (Figure 2). The high sequence identity in the V4-V6 regions of the two species explained the incorrect annotation of *E. columbae* within our 454 pyrosequence dataset.

Identification and comparison of bacteria isolated on mEI plates

The quantity of enterococci cultured on mEI from four gull fecal samples ranged from 1.1×10^4 to 4.9×10^4 CFU g^{-1} (Table 2). The majority (338/342) of these isolates were identified as *Enterococcus* spp. and four were identified as *Streptococcus* spp. Both *E. durans* and *E. faecalis* were the most abundant species identified in each gull sample, but the distribution and proportions varied slightly between individuals. The overall *Enterococcus* species distribution was as follows: *E. durans* (52%), *E. faecalis* (36%)*, E. faecium* (4%), *E. mundtii* (1%) and *E. hirae* (2%) (Figure 3). All isolates shared ≥99% sequence identity to a reference sequence, with the exception of 11 *Enterococcus* isolates whose identity could not be confidently resolved to the species level. No isolates were closely related to *C. marimammalium*, the most abundant species from our sequencing approaches.

Use of Illumina deep sequencing to explore rare occurrence of mEI culturable bacteria in the gull fecal microbial community

MiSeq Illumina sequencing provided greater sequencing depth (~65,000 reads per sample) and further confirmed *C. marimammalium* as the dominant taxonomic group within our gull fecal samples. The identity and abundance of microbial taxa were

generally in good agreement with the distributions seen in the pyrosequence dataset. This deeper sequencing method highlighted the disparity between *C. marimammalium* and the low-abundance mEI-culturable *Enterococcus* spp.: *E. faecium*, *E. faecalis*, *E. durans* and *E. hirae.* Of the 389,838 bacterial sequences generated by MiSeq Illumina sequencing, 57% were identified as *C. marimammalium*; <2.0% of sequences represented organisms commonly cultured on mEI; and 0.01% were an exact match to mEI-cultured isolates. Comparison of the mEI isolate sequences to the Illumina dataset revealed only four unique matches, representing 46 total sequences (45 *Enterococcus* species N/A, 1 *E. durans*). *Escherichia coli* made up 0.05% of the Illumina dataset, and sequences that could not be distinguished between *Escherichia* and *Shigella* made up 24% (Table 3).

Validation of Gull-2 assay with clone library contigs, gull fecal and environmental samples

The primers for the Gull-2 assay were an exact match to 170 of the 176 near full-length *C. marimammalium* sequences generated from gull fecal samples. Four of the contigs had a single base pair mismatch in either the forward or reverse primer regions, and two had base pair mismatches in both the forward and reverse primer regions. Fifty-seven gull fecal samples were tested with the Gull-2 PCR assay; 54 (95%) tested positive. Sensitivity tests on the four samples with the highest DNA concentrations revealed that the Gull-2 assay was detected down to the $1:1 \times 10^5$ dilution. However, only six (27%) of 22 environmental samples (sand and water) tested with the Gull-2 assay were positive (Table 4).

Discussion

***Enterococcaceae* and *Enterobacteriaceae* families dominant within gull fecal samples**

Enterococcaceae and/or *Enterobacteriaceae* dominated the 454 pyrosequencing libraries, with most *Enterococcaceae* sequences matching the *C. marimammalium* reference strain. With the exception of a few sequences, *Enterobacteriaceae* sequences could not be classified beyond the family level. There appeared to be a dichotomy between gull populations at the two sampling locations, Bradford Beach (Gulls 1-4) and Grant Park (Gulls 5-8). Samples collected at Bradford Beach consisted almost entirely of *Enterococcaceae*, while those collected at Grant Park were more generally more diverse and were dominated by *Enterobacteriaceae* species. Variations in gull gut microflora can occur even within small geographic regions of a city (20), and highlight the need for an assay with wide geographic applicability. While both Bradford Beach and Grant Park are beaches on Lake Michigan, differences between them may account for the observed differences in gull fecal communities. Grant Park is relatively isolated and minimally impacted by human activities. Bradford Beach, within the Milwaukee city limits, is a well-utilized and highly impacted tourist beach. Human activities affect the abundance and diversity of food sources available to gulls. Furthermore, when gulls inhabit areas where contact with human and domestic pet excrement may occur, host-enteric bacterial interactions and cosmopolitan population structures can develop (35).

***Catellicoccus marimammalium* is numerically dominant, but not culturable on mEI**

C. marimammalium has been identified in multiple studies (21, 22) as the dominant species in gull fecal microbial communities. Our study confirms those findings with a

comprehensive analysis using three separate sequencing platforms, each targeting a different region and length of sequence. Results were generally in good agreement, but did show taxonomic classification was influenced by amplification region and assignment method. Proper taxonomic classification is limited by the length of the sequence read, the region targeted, and reference sequences available in the taxonomic databases. Targeting different hypervariable regions within the same sequence can result in different taxonomic annotations (V4-V6 compared to the V1-V3 regions of same sequence) (36, 37). We were able to demonstrate the benefits of pairing smaller, full-length sequence libraries with high-abundance, shorter sequence read libraries to elucidate the abundance and proper classification of *C. marimammalium*. Sequence comparisons illustrated that the 5′ end of the 16S rRNA gene sequences, notably the V1-V3 regions, appear to be more discriminatory for taxonomic classification than the central portion of the gene (V4-V6 region).

We identified near clonal populations of *C. marimammalium* species in our gull fecal samples as the most highly abundant sequences. Lu et al. (21) also found predominance (26%) of high-identity (≥99%) *C. marimammalium* sequences in gulls from West Virginia. Collectively, these findings suggest that *C. marimammalium* has relatively low sequence diversity in gull fecal samples across large temporal and spatial gradients. High abundance of *C. marimammalium* in our gull fecal samples collected from Milwaukee beaches, combined with near clonal populations within the species, further supports the selection of *C. marimammalium* as the target for the microbial source tracking of gulls in this region.

The abundance of *C. marimammalium* in DNA samples suggests that the gull gut is good habitat for this organism; yet we failed to detect it on any of the mEI plates that we screened, despite the fact that *C. marimammalium* is within the *Enterococcaceae* family. Identification of gull fecal bacteria cultured on mEI plates revealed that the intestinal track of gulls harbors several other, lower abundance, *Enterococcaceae* spp. that are readily culturable. The mEI medium selects for the growth of enterococci, notably *E. faecalis, E. casseliflavus* and *E. faecium*, but other *Enterococcus* spp. (*E. mundtii*, *E. hirae*, *E. gallinarum*, *E. avium*, and *E. durans*) have also been isolated on mEI (38, 39). The gull fecal isolates cultured on our mEI plates were similar to those identified in a previous study by Layton et al. (40) in which *C. marimammalium* was notably absent as well, suggesting that it is not mEI culturable.

Use of deep sequencing to explore the rare occurrence of culturable *Enterococcus* spp. in relation to high abundance *C. marimammalium*

A noteworthy finding from our study is the rare occurrence of mEI-culturable *Enterococcus* spp. in relation to the high abundance of *C. marimammalium* within the Illumina dataset. Only a small proportion (46/180,888) of Illumina sequences assigned to the *Enterococcus* genus were an exact match to our mEI isolates. Using the relative abundances of taxonomic counts from our six Illumina samples, we can extrapolate the proportional relationship between *C. marimammalium* and culturable *Enterococcus spp.* to infer that for each colony of enterococci observed on mEI plates, there are ~4,900 *C. marimammalium* present in gull fecal samples. If we take into account all *Enterococcus* spp., and not just our exact sequenced isolate matches, the relative fold-abundance is ~50

C. marimammalium per *Enterococcus* CFU. *Enterococcus* spp. typically have 4-6 *rrnA* operons, whereas *C. marimammalium* have only one, so these are minimum estimates of the relative fold-enrichment (41, 42). Comparing the quantity of enterococci cultured from our four gull fecal samples, which averaged 2.3 x 10^4 CFU g^{-1}, we can estimate that there would be over three orders of magnitude greater density (~1.1 x 10^8 cells g^{-1}) of *C. marimammalium* also present. Our results may help to better understand the relationship between organisms identified by traditional culture methods versus culture-independent molecular techniques.

Implications for screening assays, source tracking, and human health

The prevalence of *C. marimammalium* in fecal samples (55/57; 95%) was higher than previously reported (71%) (21), based on screening with the Gull-2 assay. Detection of *C. marimammalium* in 27% of environmental samples was considerably lower than previously observed (21). Lu et al. (21) reported that 96% of 48 freshwater samples with presumed gull contamination, including eight samples from Grant Park Beach in South Milwaukee were positive for Gull-2 (21); however, our environmental samples were not necessarily presumed to have gull contamination. Factors affecting detection can include varying fecal loads at time of sampling, seasonal effects on bacterial survival, lower proportions of gull fecal pollution, or dilution of *C. marimammalium* DNA sequences to below the limit of detection of the assay. Our serial dilutions showed an average detection limit of 0.3 pg of gull fecal DNA per reaction, which is 20 times lower than reported by Lu et al. (21) indicating that the gulls we sampled contained a higher concentration of *C. marimammalium*, a possible explanation for the increased Gull-2

assay efficiency on our gull fecal samples. Although the Gull-2 assay is sensitive to gull fecal contamination (21, 22), the high volume of water in beach environments, particularity Lake Michigan beaches, may dilute *C. marimammalium* DNA to below detectable limits. Sand and water sources may additionally impact the persistence and survival of *C. marimammalium*, which would also affect assay detection.

Geographic variability and host specificity can factor into the applicability of microbial source tracking markers. The gull fecal samples used in the Lu et al. (21) study were from WV, GA, OH, FL and Ontario, Canada; and our gulls were collected on Lake Michigan in Wisconsin. A more recent study assessed the prevalence of gull makers in animal feces collected from CA, OH, AK, GA, DE and South Africa, and detected *C. marimammalium* in >85% of the gull fecal samples tested (22). *C. marimammalium* is specific to gulls, when compared to other poultry, waterfowl and non-avian species including pigs, dogs, cows and humans (21, 22, 25, 43). Aside from pigeons, which were detected with the same sensitivity and specificity to that of gulls, most non-target hosts have been near or below the lower limit of quantification or non-detectable (26). The detection of *C. marimammalium* in our gull fecal samples, the amplification of *C. marimammalium* with the Gull-2 assay in our environmental samples, and the alignment of the Gull-2 primers (21) with our nearly full-length clone library contigs demonstrates the regional utility of this microbial source tracking marker for gull contamination in the Great Lakes region.

Our research provides information on the diversity and variability of the gull fecal microbial community, including the dominance of *Enterococcaceae* or *Enterobacteriaceae* spp. in most samples analyzed. Using three separate sequencing platforms, we were able to demonstrate that although taxonomic assignments can be biased by primer choice, amplification region and sequence length, all methods clearly show *C. marimammalium* to be the dominant taxon within our gull fecal samples.

Plate count assays such as mEI agar are the industry standard for detection of FIB in sand and water samples. However, as is often the case with environmental samples, the most abundant organisms are not readily culturable (44). Although the gull gut appears to be a good habitat for organisms in the *Enterococcaceae* family, *Enterococcus* spp. that are readily cultured on mEI accounted for 0.04% of our sequence reads. Establishing this relationship between *C. marimammalium* and culturable enterococci allows better quantification of the total gull fecal load to beaches and coastal waters, which may help to understand the true impact gulls have on water quality monitoring efforts. These findings also highlight the advantages of using culture-independent detection techniques for environmental monitoring, as they are more encompassing, time efficient, and increasingly more cost effective. Although gulls harbor traditional FIB that can be observed in plate assays, non-mEI culturable *C. marimammalium* dominates the gull fecal microbial community and is a more suitable genetic marker for tracking gull fecal pollution.

FIGURE 1: Microbial community populations of eight gull fecal samples
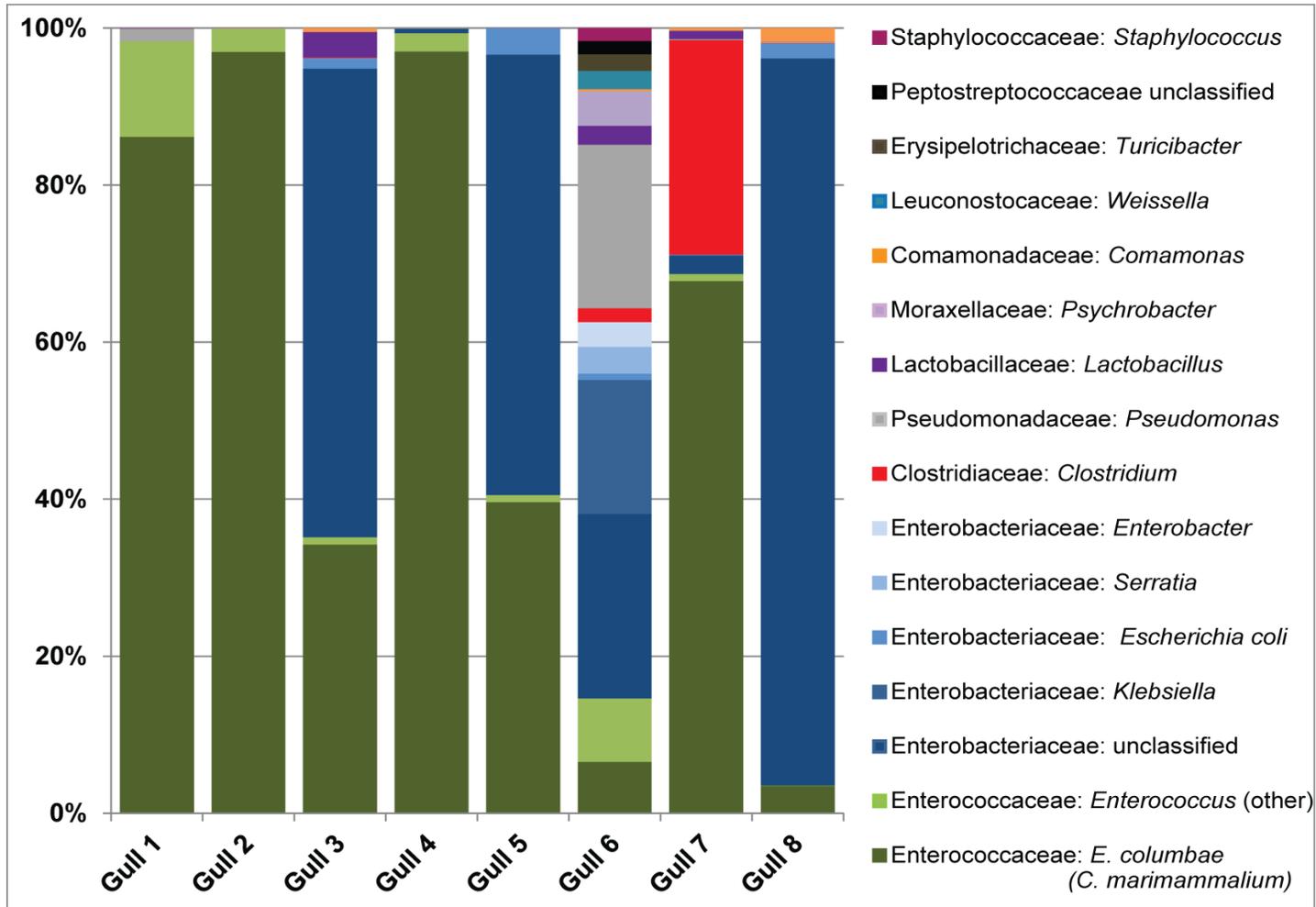
FIGURE 1: Microbial community populations of eight gull fecal samples. The V4-V6 hypervariable regions of the 16S rRNA gene from fecal genomic DNA were amplified and pyrosequenced. Taxonomy was assigned to sequences using GAST, and taxonomic counts were normalized to the maximum number of sequences. Only the most abundant genera ($\geq$1% of at least one sample) are presented. Fecal samples from Gulls 1-4 were collected from Bradford Beach, Milwaukee, WI on December 20, 2011 (1 and 2) and January 5, 2012 (3 and 4), and samples from Gulls 5-9 were collected from Grant Park, South Milwaukee, WI on January 1, 2012.

FIGURE 2: Entropy plot comparison of *E. columbae* and *C. marimammalium* reference sequences and Gull-2 assay alignment

Gull-2 Forward Assay (5' to 3')   Gull-2 Reverse Assay (5' to 3')
TGCTACACCGAAAGAAAGTGAG   GAGTGGAAAGTTCATCCTATGA   *E. columbae*
TGCATCGACCTAAAGTTTTGAG   TAGTAACTGCTCGCTCTTTGAC   *C. marimammalium*

Hypervariable regions for V4-V6 Pryosequencing: 565 – 1064
Base Pair Differences in V4-V6 regions: 12 in 499 base pairs

المنارة للاستشارات

www.manaraa.com
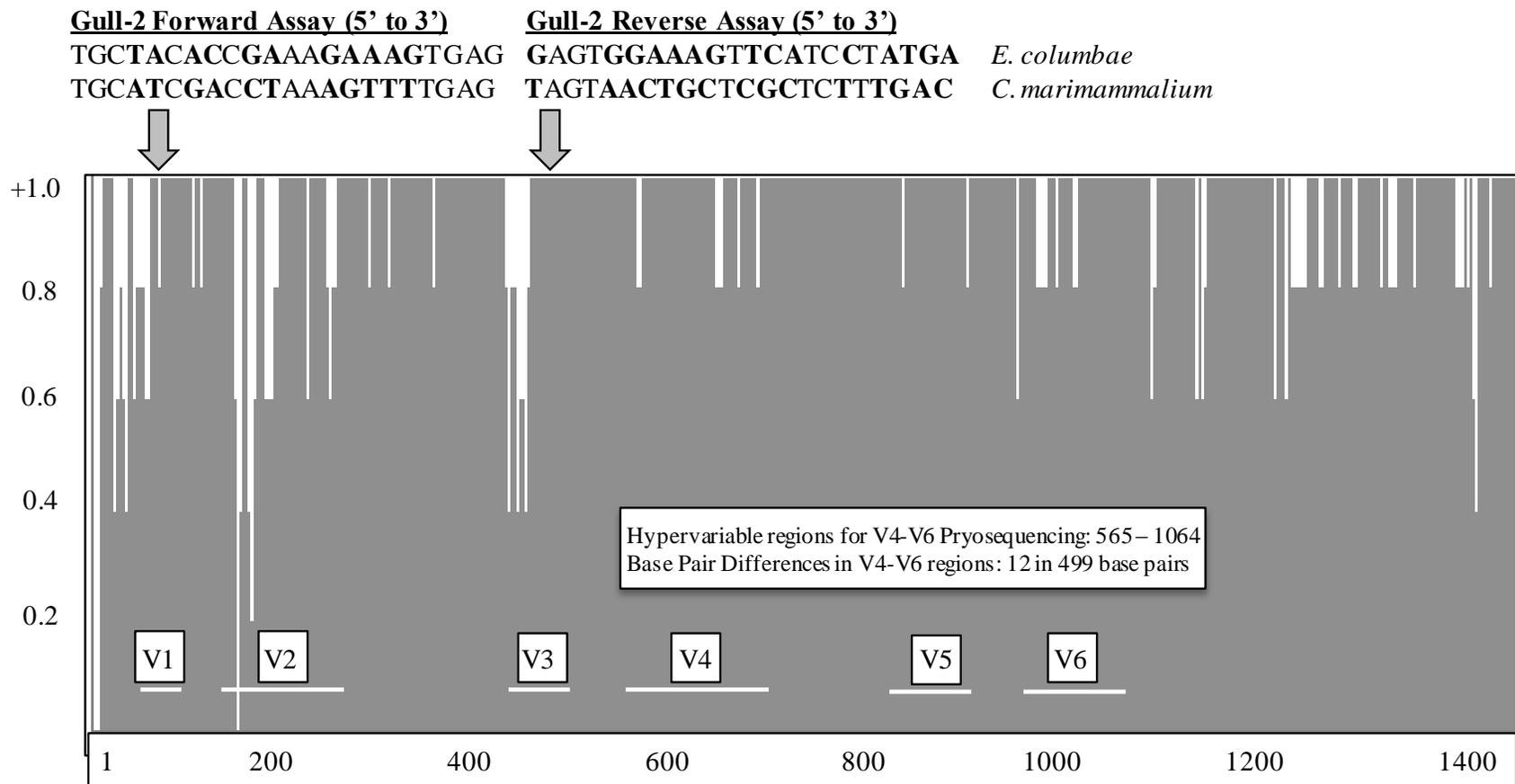
FIGURE 2: Entropy plot comparison of *E. columbae* and *C. marimammalium* reference sequences and Gull-2 assay alignment. The entropy plot illustrates base pair agreement and dissimilarity in hypervariable regions (V1-V6). The entropy plot was generated using Vector NTI; identical base pairs have a value of +1.0, similar base pairs a value of 0.5 and weakly similar base pairs a value of 0.2, as indicated on the y-axis. Variable regions, shown on the x-axis, were identified using the *E. coli* system of nomenclature noted in Chakravorty et al. (37). Variable regions targeted by the Gull-2 assay (21) include V1 and V3, and hypervariable regions selected for 454 pyrosequencing span the V4-V6 regions. Within the V1 and V3 regions targeted by the Gull-2 assay, there was 41% sequence identity within 44 base pairs, and 98% sequence identity with the V4-V6 regions, which spanned 499 base pairs. GenBank accession numbers of compared reference sequences are: *E. columbae* (AF061006) and *C. marimammalium* (AJ854484).

FIGURE 3: Phylogenetic tree of gull fecal samples cultured on mEI
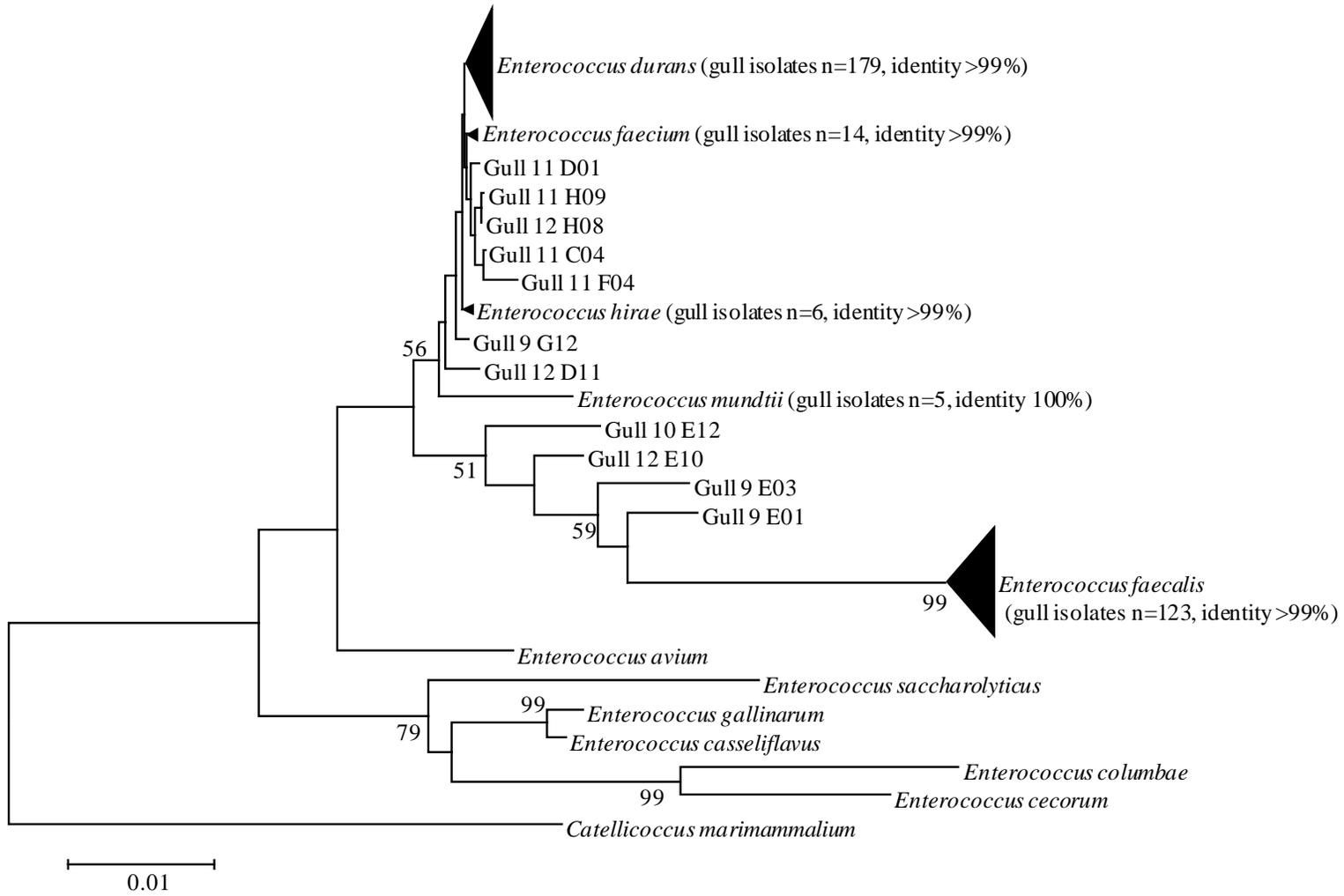
FIGURE 3: Phylogenetic tree of gull fecal samples cultured on mEI. The unrooted consensus phylogram from neighbor-joining phylogenetic analysis shows evolutionary relationships among bacterial isolates cultured on mEI from fecal pellets collected on October 18, 2012 from Grant Park, South Milwaukee, WI. *Enterococcus* spp. reference sequences (GenBank Accession numbers: AF061006, AB012212, AJ301830, AF039900, AF133535, AF061009, Y17302, AF061013, AF039903, AJ276354, and AF061004) were used for isolate clustering and classification. *Catellicoccus* reference sequences (NR042357 and AJ854484), which clustered together, were used as an outgroup. *Enterococcus* spp. reference sequences are listed next to the isolate name or cluster, along with the isolate's percent identity to the reference sequence. Gull fecal isolates were classified to the species level if they were >99% identical to GenBank reference sequences. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Only nodes with ≥50% confidence during bootstrapping are labeled. The scale bar indicates 1% sequence divergence.

TABLE 1: Comparison of taxonomic assignments for gull fecal samples

Taxonomic Assignment	Gull 1	Gull 2	Gull 3	Gull 4	Gull 5	Gull 6	Gull 7	Gull 8
Pyrosequencing								
Enterococcus	12.3%	3.4%	1.1%	2.5%	1.0%	7.5%	1.1%	0.1%
Staphylococcus	-	-	-	-	-	1.8%	-	-
Catellicoccus	85.4%	96.4%	33.6%	96.7%	39.3%	5.9%	67.0%	3.4%
Lactobacillus	0.1%	-	3.2%	-	-	2.5%	1.1%	0.1%
Leuconostoc	-	-	-	-	-	0.4%	-	-
Weissella	-	-	-	-	-	2.1%	-	-
Clostridium	-	-	-	-	-	1.6%	27.0%	0.1%
Turicibacter	-	-	-	-	-	1.9%	-	-
Comamonas	-	-	0.5%	-	-	0.3%	0.3%	1.7%
Enterobacter	-	-	-	-	-	3.0%	-	-
Escherichia	-	-	1.2%	-	3.3%	0.8%	0.1%	1.8%
Escherichia/Shigella	-	-	58.6%	0.6%	55.7%	21.1%	2.3%	90.7%
Klebsiella	-	-	-	-	-	15.4%	-	-
Serratia	-	-	-	-	-	3.1%	-	-
Pasteurella	-	-	-	-	-	0.7%	0.3%	0.1%
Psychrobacter	-	-	-	-	-	4.0%	-	-
Pseudomonas	1.6%	0.1%	-	0.1%	-	18.8%	0.2%	0.2%
Clone library sequences								
Catellicoccus[b]	98.6%	100.0%	n/a[b]	n/a	83.8%	n/a	n/a	n/a
Escherichia/Shigella	-	-	n/a	n/a	9.4%	n/a	n/a	n/a
Salmonella	1.4%	-	n/a	n/a	5.4%	n/a	n/a	n/a
Streptococcus	-	-	n/a	n/a	1.3%	n/a	n/a	n/a

[a] *Catellicoccus* sequences (n=210) uploaded to BLAST were a 99% (n=208) and 96% (n=2) match to *C. marimammalium* reference

sequence NR042357

[b] Not analyzed by clone library method

TABLE 2: Enterococci counts from gull fecal samples cultured on mEI

Sample	Fecal mass (g)	Enterococci counts (CFU g^{-1} feces)[a]
Gull 9	0.45	2.06×10^4 CFU g^{-1}
Gull 10	0.72	1.08×10^4 CFU g^{-1}
Gull 11	0.76	4.80×10^4 CFU g^{-1}
Gull 12	0.64	1.25×10^4 CFU g^{-1}

[a] CFU counts were obtained from 1:10 or 1:100 dilutions plated on mEI (see Methods).

Isolates for 16S rRNA sequencing were harvested from plates with well-dispersed colonies.

TABLE 3: Relative abundance comparisons of *C. marimammalium* and commonly cultured FIB

FIB type	Gull 1	Gull 4	Gull 5	Gull 6	Gull 7	Gull 8	TOTAL[a]
Enterobacteriaceae							
Escherichia/Shigella[b]	28	80	23,921	20,996	333	48,545	**93,903**
Escherichia coli	0	0	41	45	1	121	**208**
mEI culturable organisms[c]							
E. faecium	1	0	0	105	0	0	**105**
E. faecalis	0	0	0	28	0	0	**28**
E. hirae	0	0	0	0	0	0	**0**
E. durans	0	0	0	9	0	0	**9**
Enterococcus (species N/A)[d]	74	90	43	3854	59	12	**4,132**
Identical to mEI isolates[e]							
E. durans	0	0	0	1	0	0	**1**
Enterococcus (species N/A)	0	0	0	45	0	0	**45**
C. marimammalium	64,376	64,731	40,366	5,561	41,965	7,083	**224,082**
TOTAL Illumina sequences from gulls	**64,974**	**64,983**	**64,980**	**64,938**	**64,971**	**64,992**	**389,838**

[a] Gull samples 2 and 3 not analyzed with Illumina

[b] Sequences not distinguishable between *Escherichia* and *Shigella*

[c] Species previously observed to grow on mEI

[d] Sequences that may represent organisms that are typically cultured on mEI, but could not be assigned

[e] Subset of *Enterococcaceae* whose sequences exactly matched our mEI isolate sequences

TABLE 4: Detection of Gull-2 assay in gull fecal and environmental samples

Source	Samples Tested	Positive
Gull feces		
Grant Park	13	12
Bradford Beach	42	40
Point Beach	2	2
Lake Michigan[a]		
Beach water	9	4
Sand	13	2

[a] Bradford Beach, Atwater Beach, Point Beach and McKinley Beach

CHAPTER 2: References

1.   Alderisio KA, DeLuca N. 1999. Seasonal enumeration of fecal coliform bacteria from the feces of ring-billed gulls (*Larus delawarensis*) and Canada geese (*Branta canadensis*). Appl. Env. Microbiol. 65:5628-5630.

2.   Converse RR, Kinzelman JL, Sams EA, Hudgens E, Dufour AP, Ryu H, Santo-Domingo JW, Kelty CA, Shanks OC, Siefring SD, Haugland RA, Wade TJ. 2012. Dramatic improvements in beach water quality following gull removal. Environ. Sci. Technol. 46:10206-10213.

3.   Fogarty LR, Haack SK, Wolcott MJ, Whitman RL. 2003. Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull faeces. J. Appl. Microbiol. 94:865-878.

4.   Leevesque B, Brousseau P, Simard P, Dewailly E, Meisels M, Ramsay D, Joly J. 1993. Impact of the ring-billed gull (*Larus delawarensis*) on the microbiological quality of recreational water. Appl. Env. Microbiol. 59:1228-1230.

5.   Meerburg BG, Koene MG, Kleijn D. 2011. *Escherichia coli* concentrations in feces of geese, coots, and gulls residing on recreational water in The Netherlands. Vector Borne Zoonotic Dis. 11:601-603.

6.   USEPA. 1986. Ambient water quality criteria for bacteria. US Environmental Protection Agency Office of Water, Regulation and Standards, Washington DC.

7.   McLellan SL, Boehm AB, Shanks OC. 2013. Marine and freshwater fecal indicators and source identification. *In* Kanki P and Grimes DJ (ed.), Infectious Diseases : Selected Entries from the Encyclopedia of Sustainability Science and Technology. Springer Science + Business Media, New York.

8.   Albarnaz JD, Toso J, Correa AA, Simoes CMO, Barardi CRM. 2007. Relationship between the contamination of gulls (*Larus dominicanus*) and oysters (*Crassostrea gigas*) with *Salmonella* serovar *Typhimurium* by PCR-RFLP. Int. J. Environ. Heal. Res. 17:133-140.

9.   Kinzelman J, McLellan SL, Amick A, Preedit J, Scopel CO, Olapade O, Gradus S, Singh A, Sedmak G. 2008. Identification of human enteric pathogens in gull feces at Southwestern Lake Michigan bathing beaches. Can. J. Microbiol. 54:1006-1015.

10.   Lu JR, Ryu H, Domingo JWS, Griffith JF, Ashbolt N. 2011. Molecular detection of *Campylobacter* spp. in California gull (*Larus californicus*) excreta. Appl. Env. Microbiol. 77:5034-5039.

11.   Quesy S, Messier, S. 1992. Prevalence of *Salmonella* spp., *Campylobacter* spp. and *Listeria* spp. in ring-billed gulls (*Larus delawarensis*). J. Wildl. Dis. 28:526-531.

12.   Schoen ME, Ashbolt NJ. 2010. Assessing Pathogen Risk to Swimmers at Non-Sewage Impacted Recreational Beaches. Environ. Sci. Technol. 44:2286-2291.

13.   Soller JA, Schoen ME, Bartrand T, Ravenscroft JE, Ashbolt NJ. 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.

14.   Rabinovici SJM, Bernknopf R, Wein A. 2004. Economics and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Technol. 38:2737-2745.

15.   Edge TA, Hill S. 2007. Multiple lines of evidence to identify the sources of fecal pollution at a freshwater beach in Hamilton Harbour, Lake Ontario. Water Res. 41:3585-3594.

16.   USEPA. 2002. National beach guidance and required performance criteria for grants EPA-823-B-02-004. US Environmental Protection Agency Office of Water, Washington, DC.

17.   Engeman RM, Hartmann JW, Beckerman SF, Seamans TW, Abu-Absi S. 2012. Egg oiling to reduce hatch-year ring-billed gull numbers on Chicago's beaches during swim season and water quality test results. EcoHealth 9:195-204.

18.   Hartman J, Beckerman S, Seamans S, Engeman R, Abu-Absi S. 2010. Report to the City of Chicago on Conflicts with Ring-Billed Gulls and the 2009 Integrated Ring-Billed Gull Damage Management Project.

19.   Sauer J, Hines J, Fallon J. 2008. The North American Breeding Bird Survey, Results and Analysis 1966-2007. Center UPWR, Laurel, MD.

20.   Jeter SN, McDermott CM, Bower PA, Kinzelman JL, Bootsma MJ, Goetz GW, McLellan SL. 2009. Bacteroidales diversity in ring-billed gulls (*Laurus delawarensis*) residing at Lake Michigan beaches. Appl. Env. Microbiol. 75:1525-1533.

21.   Lu J, Santo Domingo JW, Lamendella R, Edge T, Hill S. 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl. Env. Microbiol. 74:3969-3976.

22.   Ryu H, Griffith JF, Khan IU, Hill S, Edge TA, Toledo-Hernandez C, Gonzalez-Nieves J, Santo Domingo JW. 2012. Comparison of gull feces-specific assays targeting the 16S rRNA genes of *Catellicoccus marimammalium* and *Streptococcus* spp. Appl. Env. Microbiol. 78:1909-1916.

23.   Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, Sogin ML. 2008. Exploring microbial diversity and taxonomy ssing SSU rRNA hypervariable tag sequencing. PLoS Genetics 4.

24.   Sogin ML MH, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. . 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Nat. Acad. Sci. USA 103:12115-12120.

25.   Lee C, Marion J, 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. 454-455:1-8.

26.   Sinigalliano CD EJ, Van De Werfhorst LC, Badgley BD, Balleste E, Bartkowiak J, Boehm AB, Byappanahalli M, Goodwin KD, Gourmelon M, Griffith J, Holden PA, Jay J, Layton B, Lee C, Lee J, Meijer WG, Noble R, Raith M, Ryu H, Sadowsky MJ, Schriewer A, Wang D, Wanless D, Whitman R, Wuertz S, Santo Domingo JW. 2013. Multi-laboratory evaluation of the performance of *Catellicoccus marimammalium* PCR assays developed to target gull fecal sources. Water Res.

27.   Marteinsson VT, Runarsson A, Stefansson A, Thorsteinsson T, Johannesson T, Magnusson SH, Reynisson E, Einarsson B, Wade N, Morrison HG, Gaidos E. 2013. Microbial communities in the subglacial waters of the Vatnajokull ice cap, Iceland. ISME J. 7:427-437.

28.   Morrison H, Grim S, Vineis J, Sogin M. 2013. 16S amplicon fusion primers and protocol for Illumina platform sequencing. figshare. http://dx.doi.org/10.6084/m9.figshare.833944 [Online.]

29.   Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Env. Microbiol. 75:7537-7541.

30.   Wright ES, Yilmaz LS, Noguera DR. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl. Env. Microbiol. 78:717-725.

31. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Env. Microbiol. 73:5261-5267.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Molec. Biol. 215:403-410.

33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary geneticsanalysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molec. Biol. Evol. 28:2731-2739.

34. USEPA. 2002. Method 1600: enterococci in water by membrane filtration using membrane-*Enterococcus* indoxyl-β-D-glucoside agar (mEI). US Environmental Protection Agency Office of Water, Washington DC.

35. Dick LK, Bernhard AE, Brodeur TJ, Santo Domingo JW, Simpson JM, Walters SP, Field KG. 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. Appl. Env. Microbiol. 71:3184-3191.

36. Kumar PS, Brooker MR, Dowd SE, Camerlengo T. 2011. Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. PLoS One 6.

37. Chakravorty S, Helb D, Burday M, Connell N, Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Meth. 69:330-339.

38. Bonilla TD NK, Esiobu N, McCorquodale DS, Rogerson A. 2006. Species assemblages of *Enterococcus* indicated potential sources of fecal bacteria at a South Florida recreational beach. Mar. Pollut. Bull. 52:800-815.

39. Mote BL, Turner JW, Lipp EK. 2012. Persistence and growth of the fecal indicator bacteria enterococci in detritus and natural estuarine plankton communities. Appl. Env. Microbiol. 78:2569-2577.

40. Layton BA, Walters SP, Lam LH, Boehm AB. 2010. *Enterococcus* species distribution among human and animal hosts using multiplex PCR. J. Appl. Microbiol. 109:539-547.

41. Lee ZM, Bussema C, Schmidt TM. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res 37:D489-493.

42.   Weigand MR, Ryu H, Bozcek L, Konstantinidis KT, Santo Domingo JW. 2013. Draft genome sequence of *Catellicoccus marimammalium*, a novel species commonly found in gull feces. Genome Announc 1:e00019-00012.

43.   Ryu H, Henson M, Elk M, Toledo-Hernandez C, Griffith J, Blackwood D, Noble R, Gourmelon M, Glassmeyer S, Santo Domingo JW. 2013. Development of quantitative PCR assays targeting the 16S rRNA genes of *Enterococcus* spp. and their application to the identification of *Enterococcus* species in environmental samples. Appl. Env. Microbiol. 79:196-204.

44.   Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143-169.

CHAPTER 3: Sources and distribution of surface water fecal contamination and prevalence of schistosomiasis in a Brazilian village

ABSTRACT

Background

The relationship between poor sanitation and the parasitic infection schistosomiasis is well-known, but still rarely investigated directly and quantitatively.

Methods

In a Brazilian village we studied the spatial relationship between human fecal contamination of its main river and the prevalence of schistosomiasis. We validated use of three bacterial markers of contamination in this population by 16S rRNA gene sequencing and qPCR of feces from local residents. The markers consisted of the *Bacteroides-Prevotella* group, human-specific *Bacteroides* HF8 cluster, and *Lachnospiraceae* Lachno2 cluster. We then quantified human fecal contamination and the distribution of *Schistosoma mansoni* infection along the river.

Findings

Sequence of total 16S rRNA DNA from stool samples validated the relative human specificity of the HF8 and Lachno 2 fecal indicators. The concentration of fecal contamination increased markedly along the river as it passed an increasing proportion of the population on its way downstream. An increase in multiple bacterial families associated with human feces was observed in the same distribution, with *Lachnospiraceae* the most robust human-specific signal. This was not due to a localized source of contamination since the spatial distribution of sewage draining directly into the river was random. The prevalence of schistosomiasis likewise increased downstream in

the community. The relationship between fecal contamination and schistosomiasis prevalence was explored in a spatial model using linear regression. A significant correlation was demonstrated between the prevalence of *S. mansoni* infection and local concentration of human fecal contamination based on the *Lachnospiraceae* Lachno2 cluster (r^2 0·53, p = 0·001, CI95% 0·48 – 0·58).

Interpretation

Fecal contamination in rivers has a downstream cumulative effect. The transmission of schistosomiasis correlates with very local factors probably resulting from the distribution of human fecal contamination, the limited movement of snails, and the frequency of water contact near the home. In endemic regions, the combined use of human associated bacterial markers and GIS analysis can quantitatively identify areas with risk for schistosomiasis as well as assess the efficacy of sanitation and environmental interventions for prevention.

KEY WORDS

Water quality, microbial source tracking, GIS, quantitative PCR, microbiome, Schistosoma mansoni, Bacteroides, Prevotella, Lachnospiraceae

Introduction

The culture of common fecal organisms such as coliforms and enterococci has historically been used as a proxy for the risk of infection with viral, bacterial, and parasitic pathogens (1). Despite the well-known association between fecal contamination of water and acute diseases, a correlation between these bacterial proxies and actual disease causing organisms has been difficult to demonstrate in the absence of a point-source such as sewage outflows (2). Known limitations that could explain this lack of association include the short survival of some fecal indicator organisms in water (3), their presence in environmental sources including soils and sediments (4, 5), contributions from non-human sources, and low sensitivity of detection methods for some pathogens (4, 5). The short incubation and shedding periods of these infections may also cause the pathogenic organism to no longer be present by the time an investigation is undertaken. Alternative indicators that are more specific to sources that carry pathogens offer a higher level of information than traditional fecal indicators and can improve disease surveillance and risk characterizations (6-10).

Schistosomiasis is a chronic parasitic infection that results from skin contact with water as opposed to ingestion, which is the route of transmission for most other waterborne diseases. It is a global disease that is transmitted in 78 countries with 240 million people infected (11). In Brazil, it is the second most common cause of morbidity and death due to parasitic infection (Ministry of Health, http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obtuf.def, accessed 09-29-13). Similar to other waterborne diseases, its transmission is dependent on human fecal

contamination of fresh water. Thus, in Brazil the distribution of schistosomiasis maps to areas with the poorest level of sanitation (National System of Sanitation Information; www.SNIS.gov.br, accessed 5-20-13).

The biology of *Schistosoma mansoni*, the only schistosome species found in Brazil, is very distinct from bacteria or protozoa. *S. mansoni* is a complex multicellular, multiorgan, sexually reproducing animal. Parasite eggs excreted in the stool of infected humans hatch after reaching fresh water and the released miracidium stage enters snails. In the snail they undergo asexual reproduction and develop into cercariae, the infectious stage for humans, which is released into the water. Cercariae quickly penetrate the skin of people who come into contact with these waters. After migrating to the lungs and then the liver, male and female worms mate and lodge in the intestinal mesenteric veins. Most of the eggs produced will reach the colon and then the stool, but a portion is swept back into the liver and produces fibrosis. The parasite is able to establish a long-term infection (5-40 years) if untreated that produces hundreds to thousands of eggs per day (12).

Given the complex life cycle of this parasite and its long-term survival in a community, bacterial indicators that track human sources of fecal contamination in water may contribute much to our understanding of the transmission dynamics of the parasite. Delineation of the spatial distribution of fecal contamination of surface water may have direct relevance for understanding the distribution of schistosomiasis at a fine spatial scale, i.e. household.

Methods

Study Site and Population

The village of Jenipapo in the state of Bahia, Brazil was selected for study because of its high prevalence of *S. mansoni* infection, the geographic distribution of its human population around surface waters, and its relative isolation from other settlements (13). The village is split north and south by the Jiquiriçá River and a two-lane highway. The Brejões River descends from the north, borders part of the village on the west, and enters the Jiquiriçá River at approximately the village midpoint (Figure 1a). Within Jenipapo, the Jiquiriçá River measures 5-10 meters across and less than 1 m deep, with areas of bare rock as well as thick aquatic vegetation. The Brejões is narrower and shallower, but still perennial. Most houses are located within 20 meters of these rivers. Topographically, the region is a narrow river valley with approximately equal elevations on both sides. Commercial activity is primarily devoted to raising livestock along with planting cassava, beans, and bananas. Demographic data and prevalence of schistosomiasis was obtained from interviews and a fecal survey of all residents of the village in 2009. The description of the community has been published previously (13).

The location of each home and human water contact sites in the community was established with a hand-held Trimble/Nomad GPS unit (Model 65220-11). The course of the river within the village was surveyed by walking along one bank. Data were imported into ESRI ArcGIS 10.0 (Redlands, CA) for mapping and analyses.

Sample collection and DNA extraction

Human fecal samples. Briefly, whole morning stools were collected and a slide was prepared by the Kato-Katz technique for microscopic examination. Samples positive for *S. mansoni* were processed to enrich for *S. mansoni* eggs and 5 ml of the remaining mass underwent total DNA extraction by a phenol and chloroform protocol (14). The DNA was stored at -20°C for future analysis.

Animal fecal samples. Animal samples consisted of fresh stool collected from three pigs, three dogs, two cows, and two horses residing in Jenipapo. Approximately 10 g were collected from each with a plastic scoop and stored in 2 ml screw cap tubes at -20° C. Efforts were made to collect samples within minutes after deposition to limit exposure and prevent environmental contamination. DNA was extracted from 200 mg of the sample using a Qiagen Stool Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Extracted DNA was stored at -20° C until further analysis.

Water samples. A total of eight water sample sites were chosen along the Jiquiriçá River. Six of the eight sites were most commonly reported for human contact with water (Figure 1a). Of the remaining two, one was collected 10 meters upstream from the first house and the other 55 meters downstream of the last house. An additional sample was collected from a small pond >4 km from the village. This was the source of the community's drinking water located distant from human habitation. Water samples were collected in August of 2012 using 500 ml plastic bottles. Where possible, a sample was taken from both banks of the river. Each bottle was stored at 4°C within 30 min of collection. One ml

of each sample was placed in culture media (3M™ Petrifilm™ E. coli/Coliform Count Plate, 3M, Saint Paul, MN) for 24 h at 37°C and counted for colony forming units (CFUs) of total coliforms and E. coli. The rest of the sample was filtered through a 0·22 µm nitrocellulose filter (EMD Millipore Corporation, Billerica, MA). Filters were folded and placed in 2 ml screw cap tubes and stored at -20ºC for one week in the field and -80ºC in the laboratory until DNA extraction. For DNA extraction, frozen filters were broken into small fragments with a sterile metal spatula and vortexed with a bead-beating matrix and buffers, according to the manufacturer's instructions for the Fast DNA SPIN Kit for Soil (MP Biomedicals, Solon, OH).

Quantification of traditional and alternative indicators

Six different qPCR assays were used for identification and quantification of fecal indicator bacteria, as well as human and ruminant specific fecal bacteria (Table 1). All qPCR assays were amplified in 25 µl reactions using 12·5 µl TaqMan Master mix, 1·0 µl 25 µM primer mixtures, 1·0 µl 2 µM probe mixtures, 5·5 µl water and 5·0 µl of DNA. Assays were carried out as previously described in the referenced literature in Table 1. All assays were run in duplicate.

Microbial community analysis using Hi-Seq Illumina Sequencing

Illumina deep sequencing was carried out at the Josephine Bay Paul Center of the Marine Biological Laboratory. A comprehensive microbial community profile was generated for five river samples, ten human fecal samples, and all collected animal fecal samples. The V6 hypervariable regions of the 16S rRNA gene were amplified in each of the samples

using previously described primers and protocols (15). Sequences were trimmed, controlled for low quality and contaminated reads, and then aligned. Nearly 27 million bacterial sequence reads were generated (~1 million reads per sample). The sequence data were further processed and stored in the Visualization and Analysis of Microbial Population Structures (VAMPS) database (http://vamps.mbl.edu). Taxonomic assignments were made for all sequences using Global Alignment for Sequence Taxonomy (GAST) (16).

To assess the proportion of bacterial community members that are potentially amplified by the human specific fecal indicator assays, a BLAST (17) search was performed against the Illumina sequence data sets with the HF8 and Lachno2 primers . Since the primers for the human-specific assays are in regions of the 16S rRNA gene different from the V6 sequences, the HF8 and Lachno2 primers were BLASTed against the complete reference sequences that encompassed the shorter V6 sequences. The V6 sequencing reads, each a proxy for a bacterial community member, were then binned within the HF8 cluster or Lachno2 cluster if their corresponding reference sequences contained both the forward and reverse human-specific primers.

Spatial Analysis

The basic spatial unit for analysis was the household, which were characterized in terms of presence of septic tank, number of household members, number of *S. mansoni* infected household members, and proximity to water contact sites. Kernel density estimation was used to assess and display the spatial density of the human population, schistosome

infection, and river use for sewage disposal. The Moran's I statistic was calculated using the Spatial Autocorrelation Tool in ArcGIS to assess spatial clustering.

To examine the association between proximity to fecally contaminated water and schistosomiasis, a linear regression model was created with SPSS version 19 and ArcGIS 10.1. For the model, we made the following simplifying assumptions: 1) infection occurs at the common water contact sites, 2) probability of infection depends only on proximity of place of residence to a water contact site, 3) distribution of snails along the river is homogeneous, and 4) prevalence of snail infection is proportional to degree of human fecal contamination in water.

To assign a value for fecal exposure for each household, spatial interpolation of two fecal marker DNA concentrations measured from the eight-water sample sites was performed using Inverse Distance Weighting (IDW). The village was divided into a two-dimensional grid of cells whose values were a function of their distance from a water contact site and the concentration of a fecal contamination marker at that site. A power of 2 was determined to be the best value for the weighting exponent by distance with a cell size of 20 m. Since *S. mansoni* is transmitted by human fecal water contamination, we hypothesized that a human specific fecal marker (Lachno2) would be a better predictor of schistosome infection compared to a general fecal marker, i.e. *E. coli*. Consequently, *E. coli* and Lachno2 estimated concentrations at each resident's location were extracted from the IDW generated surface to obtain an *E. coli*-IDW and Lachno2-IDW value for each home.

The spatial distribution of the prevalence of schistosomiasis was mapped to the residential area of Jenipapo as a grid of 200 m^2 blocks. The density of human population and number of cases of schistosomiasis per block was calculated using the point density tool in ArcGIS. The prevalence of schistosomiasis within each block was obtained by calculating the ratio of these two values using the Map Algebra tool. Prevalence of schistosomiasis per 200 m^2 was the dependent variable and the weighted *E. coli* and Lachno2 concentrations at each resident's home were independent variables. Since water samples were not taken from the Brejões River, the section of the community bordering the Brejões was not included in the analysis. The relationship of fecal contamination to prevalence of schistosomiasis was then assessed by standard linear regression. The model significance was determined by bootstrapping with 1000 resamples.

Results

Study site and Population

In 2009, Jenipapo consisted of 128 houses with 482 residents. Twenty-three residents had no house assigned, hence were not included in the analysis (Table 2). More than 98% of residents had tap water and an indoor flush toilet. There was access to adequate sanitation for 201 (43·8%) via home septic tanks, while sewage drained directly into the river for the remaining 258 (56·2%). Schistosomiasis was found in 209 individuals (45·5%) by examination of 3 stools collected on different days (13). The geometric mean of intensity (57 eggs per gram of feces) indicates generally light infections comparable to other studies in Brazil (18). Ten percent of the infections were heavy (>400 eggs per gram).

Spatial distribution of human population, sewage, and schistosomiasis

The Jiquiriçá River flows from west to east, and its course measures 1,542 meters from the upstream sampling site to the downstream site. There is one formal bridge across the river at the point where the Brejões River enters the Jiquiriçá. Human water contact sites were primarily used to cross the river as well as for bathing, washing clothes, or fishing (Figure 1a). The majority of houses are located on the south bank. Kernel density estimation shows clustering of human population density at both ends of the village along the south side of the Jiquiriçá River, but the greatest density clustered along the Brejões (Figure 1b). The distribution of houses with sewage draining directly into the river (Figure 1c), however, were not clustered based on Moran's I statistic, which fell within the 95% confidence interval of the null hypothesis (random spatial distribution) as indicated by a low z score (Moran's index = -0·013, p = 0·847, z = -0·19). In contrast, for the prevalence of schistosomiasis, kernel density estimation shows clusters located along the Brejões River and for people living along the most downstream segment of the village (Figure 1d). The positive Moran's index with a high z score and low p value indicated that the distribution of schistosomiasis was not random (Moran's index = 0·042, p = 0·006, z = 2·74).

Marker Specificity

The 16S rRNA sequencing reads of extracted DNA from fecal samples of ten humans and all collected animal fecal samples were normalized against their maximum number of reads and queried for the human-specific *Bacteroides* HF8 and Lachno2 clusters. Overall, humans had considerably lower amounts of *Bacteroides* in relation to *Lachnospiraceae*

or more specifically, *Blautia* (one genera within *Lachnospiraceae* from which the Lachno2 assay was designed). However, despite the low amount of overall *Bacteroides* in humans, the HF8 sequence was found in 28% of the *Bacteroides*. Overall, the proportion of sequence reads matching the HF8 cluster in humans was 10-fold higher than for pigs and dogs, and 100-fold higher than for horses and cows. The Lachno2 cluster showed even higher specificity with the proportion of reads in humans ~100-fold higher than three animal sources, but was only ~10-fold higher than for horses (Table 3).

Spatial distribution of the river's fecal contamination

Using qPCR, the concentration of the *Bacteroides-Prevotella* group was at its lowest ($<2.7 \times 10^5$ copies/100 ml) from site S1, located upstream of the first house of the village, through site S3 (Figure 2). There was a steep increase at S4 to 4.8×10^5 copies/100 ml. The highest concentration was found at S5 (5.4×10^5 copies/100 ml), where the Brejões River joins the Jiquiriçá. Its concentration then decreased gradually and by S8, located downstream of the last house, the concentration of fecal marker had returned to a value similar to S1 (2.7×10^5).

The human-specific markers (HF8 and Lachno2) followed a similar distribution; however, concentrations increased one site further downstream compared to the *Bacteroides-Prevotella* group marker. The HF8 marker was undetectable until site S5, at which point it also reached its peak ($0 \cdot 9 \times 10^4$ copies/100 ml), followed by a gradual decline. Lachno2 was detectable in minimal quantities at sites S1 to S4 (maximum concentration 684 copies/100 ml), and also had a marked increase by site S5. The peak

Lachno2 concentration was at site S7 ($1 \cdot 6 \times 10^4$ copies/100 ml), which is the last site downstream in Jenipapo that humans utilize to cross the river, and declined by S8. The ruminant specific marker was undetectable until S5 and remained in low concentrations without significant variation between sites thereafter. The *E. coli* marker showed a smaller degree of increase after S5. By contrast, the source of drinking water located $4 \cdot 8$ km north of the village had no copies of the HF8 human specific marker. Colony counts for coliforms, and less so for *E. coli*, also increased as the river moved down stream and declined sharply past the last house in the village (Figure 3).

Changes in bacterial communities across the river transect

Sequence data from the microbial communities found in river water was used to compare the relative abundance of >20 bacterial families. Consistent with the qPCR results, the proportion of *Prevotellaceae* and *Lachnospiraceae* increased significantly in the downstream portion of the village (Figure 4). *Ruminococcaceae* and *Enterobacteriaceae*, two other families associated with fecal communities, also increased. Combined, these fecal families increased from ~3% to ~9% of the community between upstream to downstream sites. Families associated with sewage- contaminated water - *Moraxellaceae* and *Aeromonadaceae*, specifically *Acinetobacter* spp. and *Aeromonas* spp. - also increased at sites six and seven (19-22). *Comamonadaceae*, a bacteria common to the environment and freshwater, was the most abundant family on average, accounting for over 40% of the microbial community populations at each sampling site (23). *Bacteroidaceae*, which includes the genera *Bacteroides*, were in low abundance and are not represented in Figure 4.

Site-specific risk model

The 200 m^2 prevalence grid for Jenipapo produced 7 blocks (Figure 5). Each house was assigned a value for exposure to fecal contamination based on proximity to a water sample site and the fecal marker concentration at site. The relationship of risk for infection with *S. mansoni* to the concentration and proximity to fecal contamination was modeled and tested statistically using the data from Jenipapo. Linear regression of prevalence of schistosome infection against fecal contamination yielded an r^2 of 0·28 for the *E. coli*-IDW value (two-tailed p<0·001, 95% CI 0·22–0·35) and 0·53 for the Lachno2-IDW value (two-tailed p<0·001, 95% CI 0·48–0·58). These results can be interpreted as local concentration of human fecal contamination explaining over 50% of the variance in risk for schistosomiasis.

Discussion

Although the village of Jenipapo is small, it is typical of many villages of Latin America. It also shares a pattern of development common with larger communities and even the great metropolises of Brazil. The village grew up along the two rivers that meet at its center, and most homes border these rivers in order to have access to a ready form of sewage removal. The community's drinking water supply is 4·8 km away where a dammed stream forms a small reservoir. Jenipapo's geometry is a simple, mostly linear distribution of residences and water contact sites, and this made it ideal for studying the dynamics of fecal contamination and its relationship to acquisition of schistosomiasis. Putting the degree of fecal contamination of the Jiquiriçá River within Jenipapo in context, the geometric mean CFUs for *E. coli* (113 CFU/100 ml) was at the upper limit of

the EPA's 2012 Recreational Water Quality Criteria value of 100 CFU/100 ml (24). This level of contamination was estimated to result in 32 gastrointestinal illnesses per 1,000 primary recreational contacts.

We were further able to identify human waste as the major contributor to this contamination. We validated both the HF8 and Lachno2 genetic markers as human-specific by directly assaying the resident population. Interestingly, both human specific markers were identified from humans in the US, but were also present in Brazilians. The Lachno2 marker in particular showed a high signal in the water sample and all human feces, but near absence in cows, the other major animal contributing to fecal contamination of the river. These markers indicated that human fecal waste was the major component of fecal contamination in this section of river. Overall, human-specific fecal indicators contribute important quantitative information on water quality that could be used for surveillance to gauge specific sanitation interventions.

The nearest community to Jenipapo is 8 km upstream with a population of 353 and similar level of sanitation, and there are few intervening houses, but many areas of pasture. Twelve km further upstream there is a town of 12,000. Despite nearby populations, quantitative tracking of human fecal contamination in this study suggests a predominance of local effects. The qPCR markers for human and other fecal contamination, as well as coliform colony counts, are very low at the entrance to the village and significantly increase as the river continues downstream. Inflow for the village has significant levels of the *Bacteroides-Prevotella* group, but is very low for

human fecal contamination indicating that most influence from communities upstream has dissipated. We presume this is not the result of the HF8 marker being sensitive to environmental degradation, since experimentally the duration of signals from *Bacteroides* ranges from days to several weeks (25). In addition, the other marker of human fecal contamination (Lachno2) shows a similar pattern. Within Jenipapo, the entry of sewage is not clustered to one area of the community and we noted the concentration of contamination is cumulative as the river moves downstream through the community. The analysis of bacterial communities was based on number of sequence reads and is consistent with the qPCR genetic marker data. The study is limited in the relatively small number of samples taken, sampling only ~50 m beyond the community's houses and a lack of water samples.

Human fecal contamination of water and the presence of snails are prerequisites for transmission of schistosomiasis. Snails are known to have a limited range of movement (26). Proximity to water bodies where there are infected snails is a known risk factor for schistosomiasis (27, 28). However, all inhabitants in Jenipapo are essentially equidistant from the river, and finding and determining which snails are infected can be laborious. In this study we show that, in a village with high prevalence of schistosomiasis, the risk of acquiring the infection is driven not only by proximity to surface water but also by its degree of human fecal contamination. The model explained a large amount of variation without including data on snail populations. To further support this, a study conducted in Jenipapo on *S. mansoni* population genetics found that parasite populations were more similar among infected members of the same household compared to parasite populations

of all infected individuals in the village (manuscript under revision). The variation not explained by our model was likely due to violations of our simplifying assumptions. Snails are not likely to be evenly distributed, and infection risk is influenced by more than distance to a contact site (age, type of activity, etc.). Some infection occurs outside of contact sites or not at the nearest contact site.

Although the human population disperses widely over this area, the local opportunities for exposure near the home may dominate the infection risk profile. Since awareness of schistosomiasis has been raised in the community and well before the analysis of fecal contamination, we have heard reports that teenage boys now prefer to enter the river upstream of the village. This may be a wise precaution, although the better solution will be to remove the contamination from the river rather than remove the boys and girls.

FIGURE 1: Spatial Analysis of the Jiquiriçá River
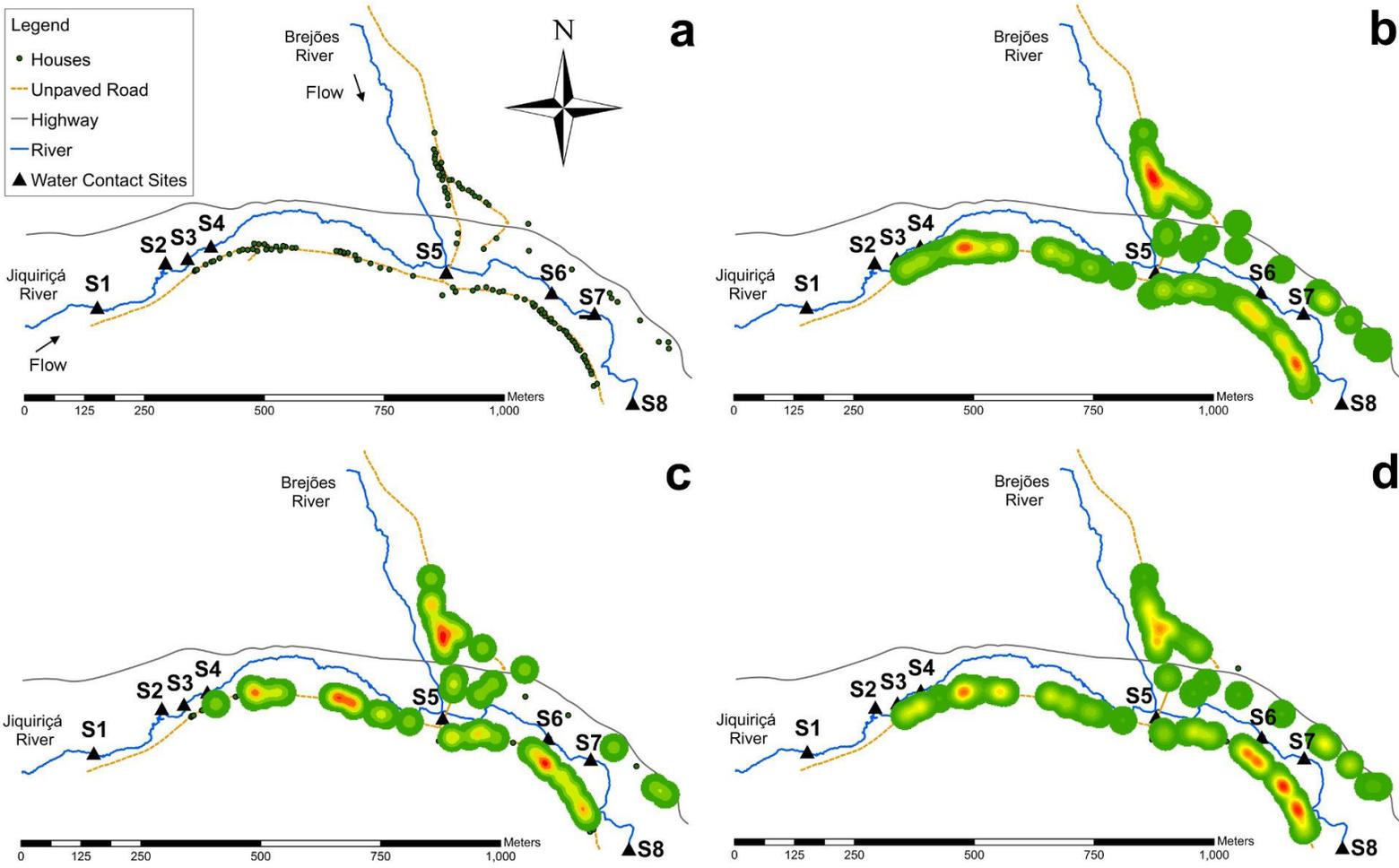
FIGURE 1: (a) Study area: village of Jenipapo in the state of Bahia, Brazil (b) Kernel density distribution of human population (c) Kernel density distribution of sewage draining directly in the river (d) Kernel density distribution of *S. mansoni* infection.

FIGURE 2: Bacterial concentration distribution along the Jiquirica River

FIGURE 2: Bacterial concentration distribution along the Jiquirica River.

Bac 32, *Bacteroides-Prevotella*; HF8, human-specific *Bacteriodes;* Lachno2, human-specific *Lachnospiraceae*.

FIGURE 3: CFU distribution along the Jiquiriçá River
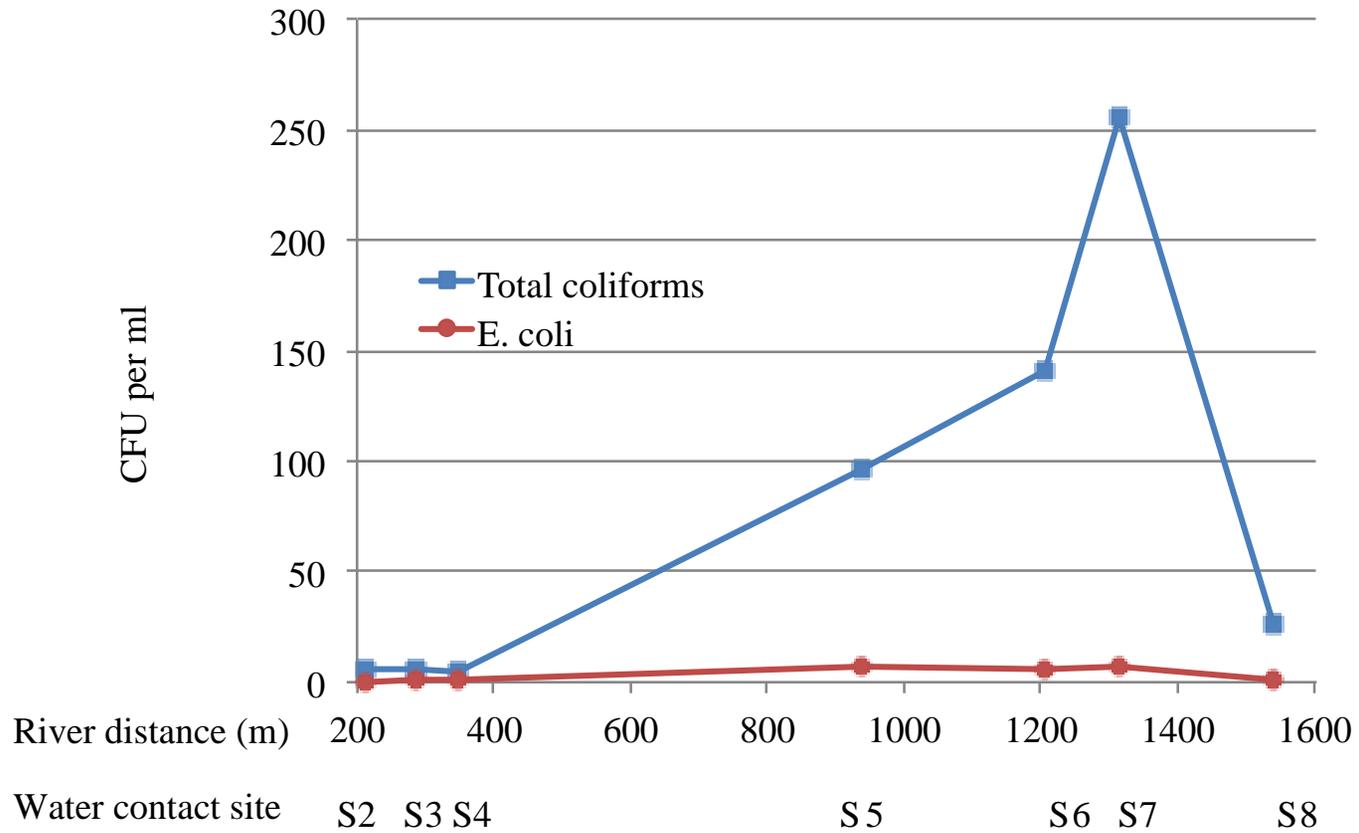
FIGURE 3: CFU distribution along the Jiquircia River.

CFU, colony forming units. No sample was sent for culture for site S1.

FIGURE 4: River microbial communities (Family level)

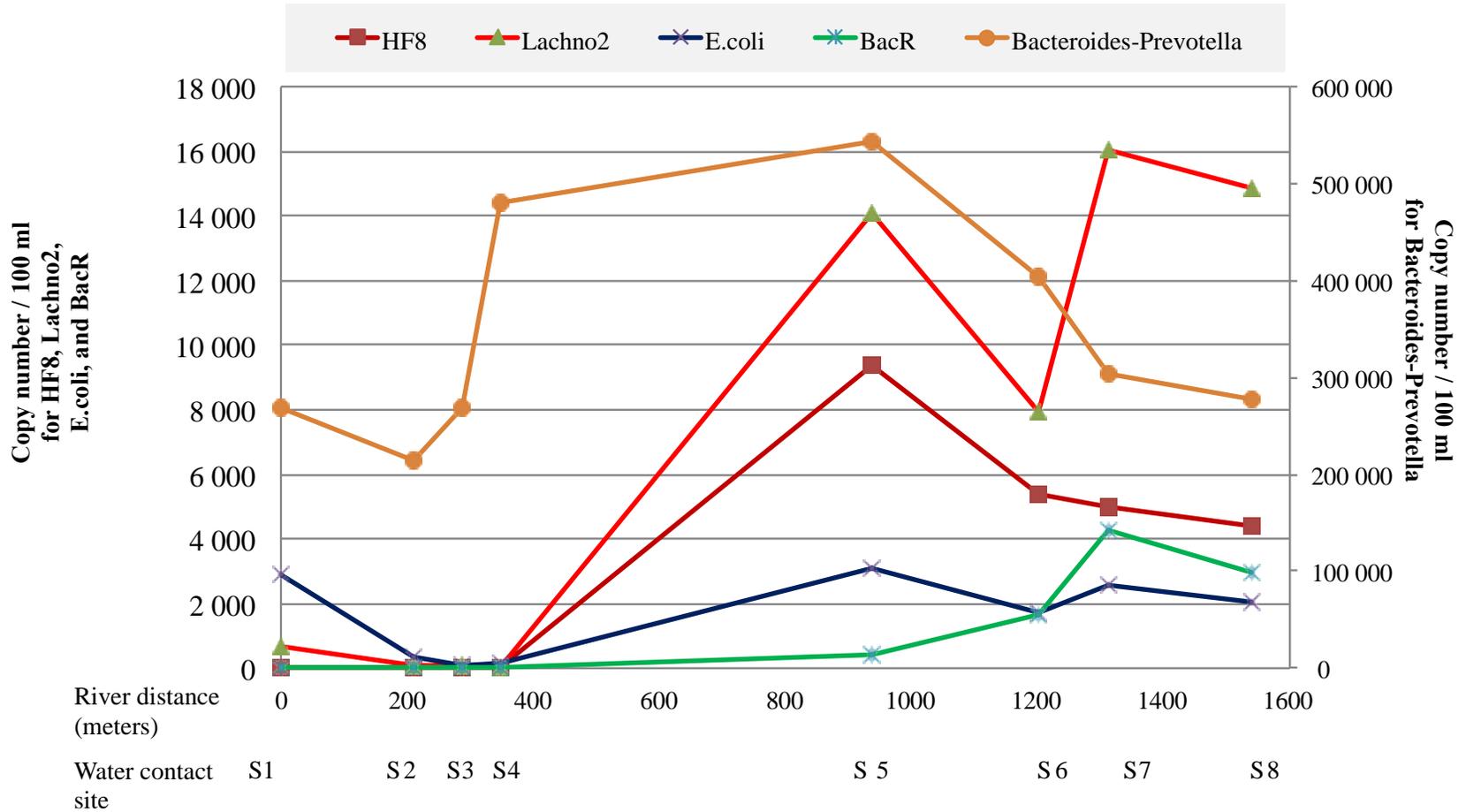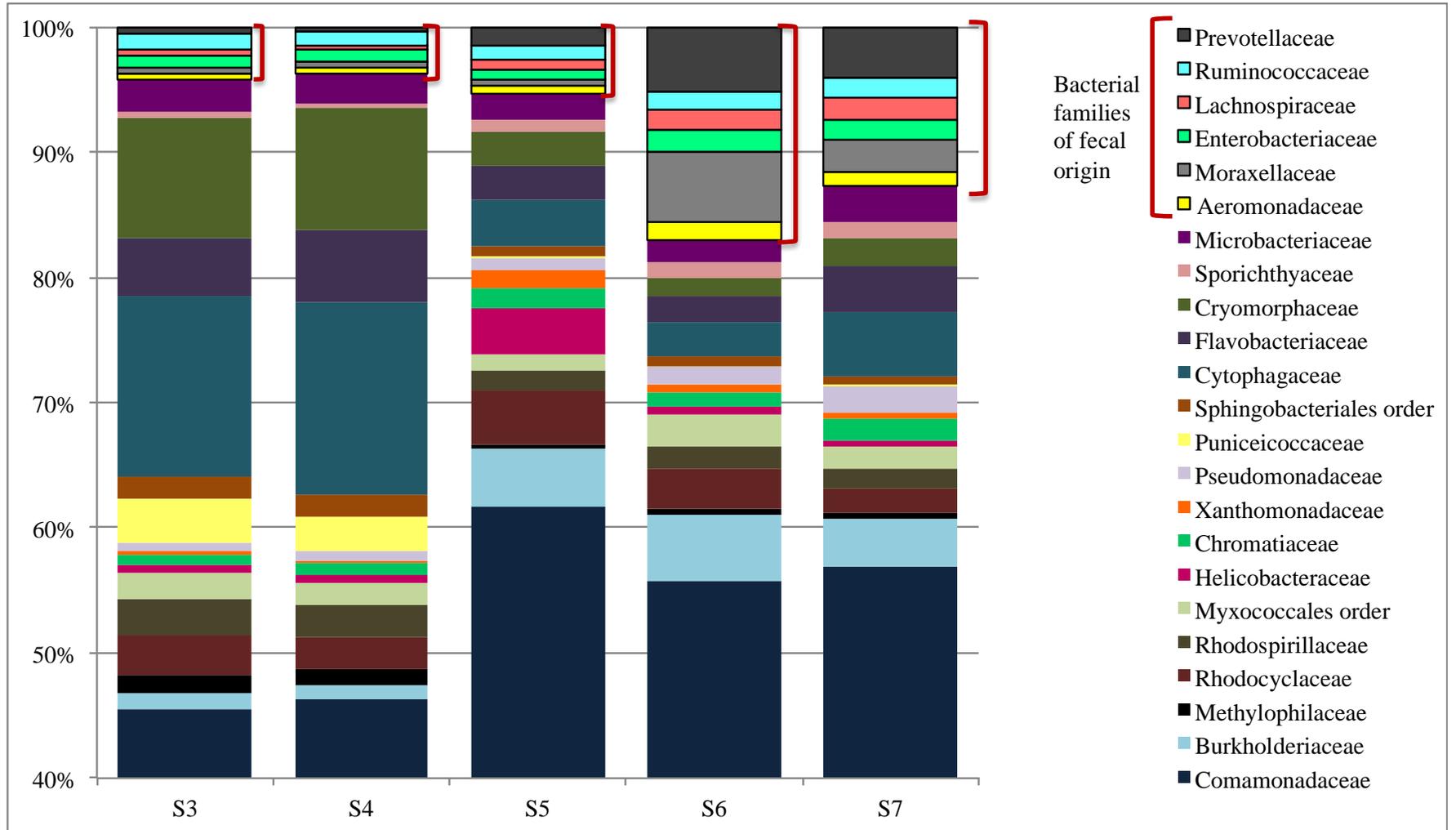**FIGURE 4:** River microbial communities (Family level). River samples were collected from five points along the Jiquiriçá River in Jenipapo, Brazil on August 18, 2012. Microbial community populations of five river samples. The V6 hypervariable regions of the 16S rRNA gene from community genomic DNA were amplified and sequenced using Hi-Seq Illumina Sequencing. Taxonomy was assigned to sequences using GAST and taxonomic counts were normalized to the maximum number of sequences. Only the most abundant genera (>1% of at least one sample) are presented. Families discussed in the text are outlined.

FIGURE 5: Map of Jenipapo with the area of houses divided in a 200 m^2 grid
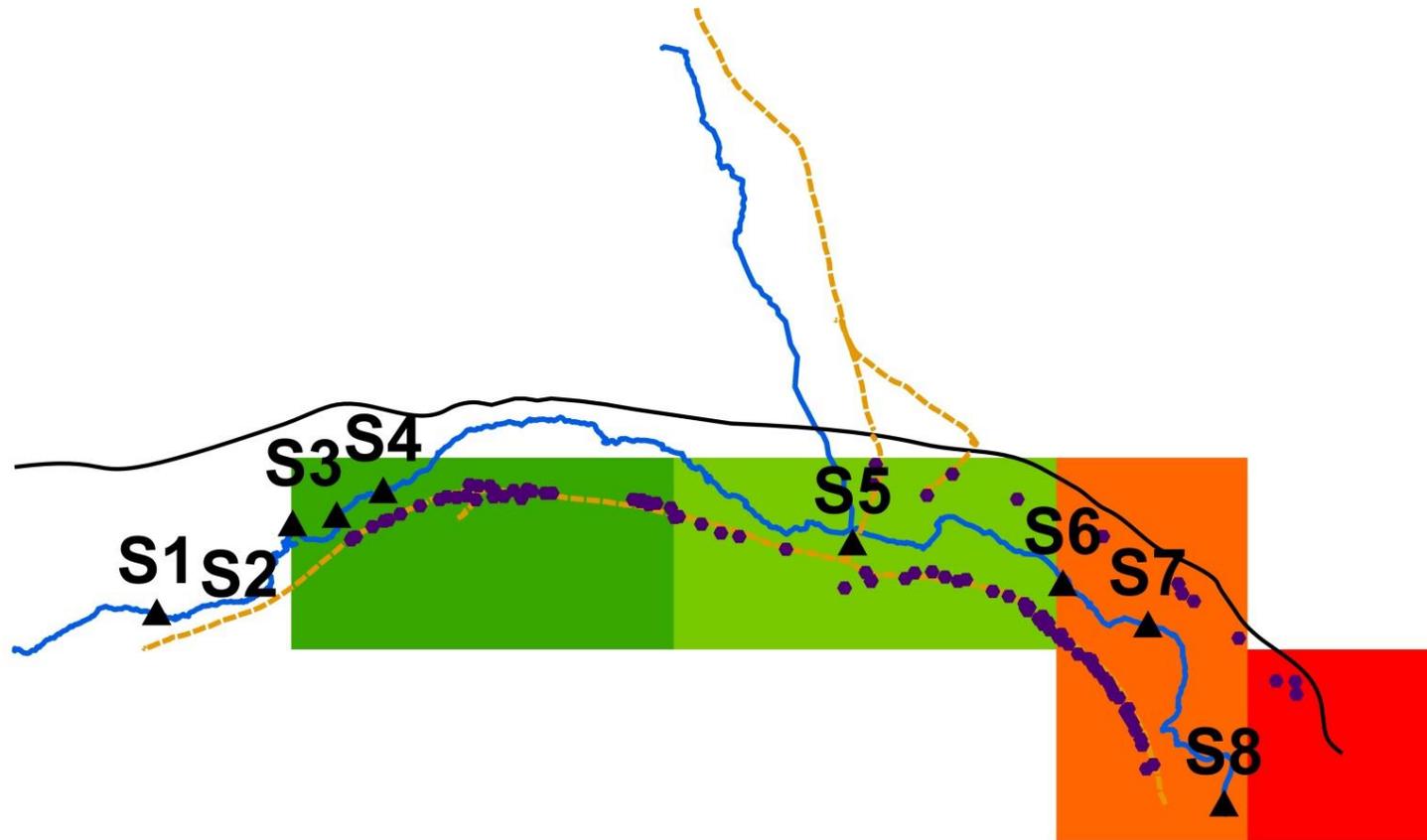
FIGURE 5: Map of Jenipapo with the area of houses divided in a 200 m^2 grid.

Schistosomiasis prevalence for each 200 m^2 block was estimated. Infection prevalence increased downstream.

TABLE 1: Primers used in this study

Bacterial target	Target Host	Primer	Sequence	Reference
Bacteroides-Prevotella group	Non specific	GenBac3F	GGGGTTCTGAGAGGAAGGT	(29)
		BacsppR	CCGTCATCCTTCACGCTACT	(29)
		Bacspp346p	[6FAM]-CAATATTCCTCACTGCTGCCTCCCGTA-[MGBNFQ]	(30)
Bacteroides HF8 cluster	Human specific	HF183F	ATCATGAGTTCACATGTCCG	(31)
		BacHum241R	CGTTACCCCGCCTACTATCTAATG	(31)
		BacHum193p	[6-FAM]-TCCGGTAGACGATGGGGATGCGTT-[MGBNFQ]	(32)
Lachnospiraceae Lacno2 cluster	Human specific	Lachno2F	TTCGCAAGAATGAAACTCAAAG	(33)
		Lachno2R	AAGGAAAGATCCGGTTAAGGATC	
		Lachno2p	[6FAM]-ACCAAGTCTTGACATCCG-[MGBNFQ]	
E. coli	Non specific	uidA1663F	GCGACCTCGCAAGGCATA	(34)
		uidA1790R	GATTCATTGTTTGCCTCCCTGCTGCG	
		uid1729p	[6FAM]-TGCAGCAGAAAAGCCGCCGACTTCGG-[MGBNFQ]	
Bacteroidetes	Ruminant specific	BacR_f	GCGTATCCAACCTTCCCG	(35)
		BacR_r	CATCCCCATCCGTTACCG	
		BacR_p	[6FAM]-CTTCCGAAAGGGAGATT-[MGBNFQ]	

TABLE 2: Population characteristics

Characteristics	
Population	459
Houses	128
Residents/house	3.6
Sex	
Male (%)	221 (48.1)
Mean Age (SD)	30.5 (21.6)
% Age < 15	28
Tap water %	
Yes	459 (100)
Flush toilet %	
Yes	453 (98.7)
Sewage destination by house (%)	
Septic tank	68 (53.2)
River	60 (46.8)
S. mansoni infection	
Prevalence (%)	209 (45.5)
Mean intensity epg (S.D.)	57 (4.1)

TABLE 3: Percent of sequences matching HF8 and Lachno2 clusters by fecal source

Average	Human	Cow	Pig	Dog	Horse
Bacteroides % of total	0.29	0.058	3.4	0.097	3.3
HF8 % of total	0.11	0.001	0.013	0.011	0.001
Lachnospiraceae % of total	17.74	0.070	12.5	0.14	7.5
Blautia % of total	1.2	1.00	0.011	1.3	0.35
Lachno2 % of total	0.082	0.001	0.001	0.002	0.009

[a] The 16S rRNA gene was sequences from total DNA extracted from stool or filtered fecal sediment (Human). There were ~1 million sequences per sample generated and matched to species or family. All fecal samples were collected from the village of Jenipapo, 10 humans 2 cows, 3 pigs, 3 dogs and 2 horses.

CHAPTER 3: References

1. Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res. 2006 Jul;40(13):2463-76.

2. Wu J, Long SC, Das D, Dorner SM. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. J Water Health. 2011 Jun;9(2):265-78.

3. McFeters GA, Bissonnette GK, Jezeski JJ, Thomson CA, Stuart DG. Comparative survival of indicator bacteria and enteric pathogens in well water. Appl Microbiol. 1974 May;27(5):823-9.

4. Horman A, Rimhanen-Finne R, Maunula L, von Bonsdorff CH, Torvela N, Heikinheimo A, et al. *Campylobacter* spp., *Giardia* spp., *Cryptosporidium* spp., noroviruses, and indicator organisms in surface water in southwestern Finland, 2000-2001. Appl Environ Microbiol. 2004 Jan;70(1):87-95.

5. Winfield MD, Groisman EA. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. Appl Environ Microbiol. 2003 Jul;69(7):3687-94.

6. Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiol Rev. 2013 Jul 1.

7. Santo Domingo JW, Bambic DG, Edge TA, Wuertz S. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. Water Res. 2007 Aug;41(16):3539-52.

8. Field KG, Bernhard AE, Brodeur TJ. Molecular approaches to microbiological monitoring: fecal source detection. Environ Monit Assess. 2003 Jan-Feb;81(1-3):313-26.

9. McLellan S, Boehm A, Shanks O. Marine and Freshwater Fecal Indicators and Source Identification. New York: Springer; 2013.

10. Roslev P, Bukh AS. State of the art molecular markers for fecal pollution source tracking in water. Appl Microbiol Biotechnol. 2011 Mar;89(5):1341-55.

11. WHO. Schistosomiasis: number of people treated in 2011. Wkly Epidemiol Rec. 2013 Feb 22;88(8):81-8.

12.   Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. Lancet. 2006 Sep 23;368(9541):1106-18.

13.   Blanton RE, Blank WA, Costa JM, Carmo TM, Reis EA, Silva LK, et al. *Schistosoma mansoni* population structure and persistence after praziquantel treatment in two villages of Bahia, Brazil. Int J Parasitol. 2011 Aug 15;41(10):1093-9.

14.   Maniatis T, Fritsch EF, Sambrook J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Press; 1982.

15.   Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One. 2013;8(6):e66643.

16.   Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet. 2008 Nov;4(11):e1000255.

17.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.

18.   Enk MJ, Lima AC, Drummond SC, Schall VT, Coelho PM. The effect of the number of stool samples on the observed prevalence and the infection intensity with *Schistosoma mansoni* among a population in an area of low transmission. Acta Trop. 2008 Nov-Dec;108(2-3):222-8.

19.   McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ Microbiol. 2010 Feb;12(2):378-92.

20.   Vandewalle JL, Goetz GW, Huse SM, Morrison HG, Sogin ML, Hoffmann RG, et al. Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. Environ Microbiol. 2012 Sep;14(9):2538-52.

21.   Collado L, Inza I, Guarro J, Figueras MJ. Presence of Arcobacter spp. in environmental waters correlates with high levels of fecal pollution. Environ Microbiol. 2008 Jun;10(6):1635-40.

22.   Khan IU, Loughborough A, Edge TA. DNA-based real-time detection and quantification of aeromonads from fresh water beaches on Lake Ontario. J Water Health. 2009 Jun;7(2):312-23.

23.   Willems A, De Ley J, Gillis M, Kersters K. Comamonadaceae, a New Family Encompassing the Acidovorans rRNA Complex, Including Variovorax paradoxus gen. nov.,comb. nov. for Alcaligenes paradoxus (Davis 1969). International Journal of Systematic Bacteriology. 1991;41(3):445-50.

24.   US Environmental Protection Agency: 2012 Recreational Water Quality Criteria OFFICE OF WATER 820-F-12-058 http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/RWQC2012.pdf

25.   Liang Z, He Z, Zhou X, Powell CA, Yang Y, Roberts MG, et al. High diversity and differential persistence of fecal *Bacteroidales* population spiked into freshwater microcosm. Water Res. 2012 Jan 1;46(1):247-57.

26.   Prugnolle F, Theron A, Pointier JP, Jabbour-Zahab R, Jarne P, Durand P, et al. Dispersal in a parasitic worm and its two hosts: consequence for local adaptation. Evolution. 2005 Feb;59(2):296-303.

27.   Clennon JA, King CH, Muchiri EM, Kariuki HC, Ouma JH, Mungai P, et al. Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. Am J Trop Med Hyg. 2004 Apr;70(4):443-8.

28.   Kloos H, Fulford AJ, Butterworth AE, Sturrock RF, Ouma JH, Kariuki HC, et al. Spatial patterns of human water contact and *Schistosoma mansoni* transmission and infection in four rural areas in Machakos District, Kenya. Soc Sci Med. 1997 Apr;44(7):949-68.

29.   Bernhard AE, Field KG. 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. 2000 Apr;66(4):1587-94.

30.   Seurinck S, Defoirdt T, Verstraete W, Siciliano SD. 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. 2005 Feb;7(2):249-59.

31.   Bernhard AE, Field KG. 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. 2000 Oct;66(10):4571-4.

32.   Kildare BJ, Leutenegger CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res. 2007 Aug;41(16):3701-15.

33. Newton RJ, Vandewalle JL, Borchardt MA, Gorelick MH, McLellan SL.
     Lachnospiraceae and *Bacteroidales* alternative fecal indicators reveal chronic
     human sewage contamination in an urban harbor. Appl Environ Microbiol. 2011
     Oct;77(19):6972-81.

34. Li J, McLellan S, Ogawa S. Accumulation and fate of green fluorescent labeled
     Escherichia coli in laboratory-scale drinking water biofilters. Water Res. 2006
     Sep;40(16):3023-8.

35. Reischer GH, Kasper DC, Steinborn R, Mach RL, Farnleitner AH. Quantitative
     PCR method for sensitive detection of ruminant fecal pollution in freshwater and
     evaluation of this method in alpine karstic regions. Appl Environ Microbiol. 2006
     Aug;72(8):5610-4.

CHAPTER 4: *Prevotella* and *Blautia* distinguish human and animal fecal pollution

in Brazil surface waters

Summary

Poor sewage handling and limited agricultural manure management practices contribute to fecal pollution in rural Brazilian waterways. Few microbial source tracking studies have tested host-specific indicators in underdeveloped regions such as this. Sequencing of sewage, human feces, and animal feces with Illumina HiSeq revealed Prevotella as the most abundant genus, with genera belonging to the families Lachnospiraceae and Ruminococcaceae comprising a large proportion of the microbiome as well. Bacteroides, the most commonly utilized human-specific genus in the US, was present only in low abundance. We used oligotyping to identify human-related Blautia and Prevotella sequences. Thirty-three Blautia oligotypes were differentially distributed in humans compared to animals, as were 13 Prevotella oligotypes; these sequences represent possible alternative indicators to Bacteroides. Application of these sequences to source tracking was tested in comparison to traditional fecal indicators along an increasing population gradient in a rural river. Prevotella and Blautia both increased considerably at sites downstream, but traditional FIB populations followed a steady or even decreasing trend. Human-specific and preferred oligotypes also increased downstream. While both genera were able to distinguish human and animal fecal pollution in Brazil surface waters, Blautia appears to contain more discriminatory and globally applicable markers for tracking sources of fecal pollution.

Introduction and Background

Fecal pollution in surface water sources constitutes a significant public health threat worldwide. Waterborne illnesses, related to the inadequate provision of water and sanitation services, are responsible for four billion cases of diarrhea and 1.8 million deaths each year, mostly impacting children five years of age and younger living in the rural communities of developing countries. The lack of sanitation and drinking water infrastructures within these rural communities creates a self-perpetuating cycle of waterborne diseases (33, 38). In Brazil these include a broad array of diarrheal illnesses from *Escherichia coli*, *Campylobacter*, *Giardia*, *Cryptosporidium,* rotavirus (15) and norovirus (37) to schistosomiasis, a systemic parasitic disease with a morbidity/mortality rate estimated at over 250,000 per year (39). While there has been significant progress in controlling schistosomiasis through repeated mass administration of drugs, it is still the second leading cause of death from parasitic infections in Brazil (4).

Data from legally-mandated fecal contamination surveillance in Brazil have shown that regions with the highest prevalence of diseases overlap with areas of low sanitation (1, National System of Sanitation Information; www.SNIS.gov.br, accession date 5-20-13). Although this relationship has been demonstrated (1), identifying human sources in waterways with fecal pollution is less certain. Poor sewage handling and limited agricultural manure management practices contribute to fecal pollution in Brazilian waterways (6). Exposure to human feces is generally a greater health risk than exposure to animal feces (29); therefore, identifying the source contamination is important for risk assessment (35, 13, 27, 30). Traditional fecal indicator bacteria (FIB), such as *E. coli* and

enterococci, are found in the feces of humans and animals alike, which limit their potential to accurately assess human health risk. Host specific alternative indicators may therefore be particularly useful in areas where there is ubiquitous fecal pollution in surface waters (24, 26).

Multiple studies have identified and tested various host specific indicators by detecting genetic markers (a specific sequence within the DNA of fecal bacteria), but few have explored the applicability of these markers to underdeveloped regions. Geographic variability and host specificity can factor into the efficacy of markers in discrimination of host sources (3, 17, 28). Recent human microbiome studies demonstrate a difference in the microbial community across geography, age and time (5, 22, 41). Differences in cultures and diets may also cause variation in the human gut microbiota, which manifest in the relative abundances of *Bacteroides* versus *Prevotella* (5, 40). Considering these findings, more research is needed to evaluate if the "human-specific" genetic markers developed in the US are relevant in other regions of the world.

With deep sequencing technologies becoming more commonplace, it is now possible to identify multiple microorganisms that provide a fecal signature within a sample (25). more reliably than a single marker (9, 25, 34). Given the complexity and depth of massively parallel, high-throughput sequencing datasets, the use of sensitive methods that can distinguish closely related but distinct organisms is critical for accurate identification of markers associated with certain hosts. Oligotyping is a recently described method that facilitates high-resolution partitioning of the 16S rRNA gene amplicon data into

"oligotypes" using Shannon entropy (12). By recovering subtle nucleotide variation among reads oligotyping can distinguish organisms with more than 99% identity over the sequenced region (11, 12). Previous studies identified oligotypes within the genus *Blautia* that distinguish human and animal sources with remarkable accuracy (23; A.M. Eren et al., unpublished).

In this study, we used Illumina HiSeq for ultra-deep sequencing of the bacterial communities associated with sewage, human feces, and animal feces. We applied an extremely stringent quality filtering on our sequencing reads to eliminate the vast majority of sequencing errors (10) to avoid false positives in our findings. Host-related *Blautia* and *Prevotella* oligotypes were determined and evaluated for environmental applicability by looking at the microbial community composition of river water data collected from the Jiquiriçá River in Jenipapo, Brazil and the Lucaia River in Salvador, Brazil. Previous work with these samples (Ponce Terashima et al., unpublished) used qPCR assays for quantification of human-associated *Bacteroides* (HF183;(20)) and *Blautia* (Lachno2 assay; (24)) combined with spatial analysis to correlate human fecal contamination in the Jiquiriçá River with risk of *Schistosomiasis mansoni* infection. An increasing gradient of human fecal contamination from upstream to downstream sites was correlated to a higher risk of Schistosomiasis and cumulative sewage discharge into river. We expand upon this study by describing the oligotype signature profile in human samples from Brazil and the imprint of this signature on contaminated surface water.

Results and Discussion

Microbial community structure of human fecal and sewage samples in Brazil

Brazilian human fecal samples and sewage shared common fecal taxa in similar proportions (Figure1), but sewage also contained taxa that have been previously associated with pipe infrastructure and surface water samples (22, 28, 36). *Prevotella,* within the family *Prevotellaceae,* was the most abundant genus, comprising >36% of the average Brazilian human fecal sample and >10% of the Brazilian sewage sample. Genera belonging to the families *Lachnospiraceae* (e.g., *Blautia, Roseburia*) and *Ruminococceae* (e.g., *Ruminococcus, Oscillibacter, Faecalibacterium*) also made up a large proportion of the sewage (20%) and human fecal (40%) samples when combined. The latter two families have been identified as dominant families in US human fecal and sewage samples (22, 31). Notably abundant in US human fecal samples, but less prevalent in Brazil human fecal samples, are the *Bacteroides,* which make up ~11% of US human fecal samples and ~2% of US sewage (7, 22, 31), but comprised <1% of the Brazil human and sewage samples.

Populations consuming carbohydrate-rich diets in Africa and South America have higher proportions of *Prevotella* their gut microbiomes, whereas *Bacteroides* tend to dominate the guts of individuals from Europe and North America who typically consume a higher protein diet (5, 40, 41). Our research further supports this geographical distinction, as the Brazilian human fecal and sewage samples had a higher abundance of *Prevotella* versus *Bacteroides.* In contrast to differences in the proportional abundance of Bacteroidetes genera between the US and Brazil, the genus *Blautia* had a consistent signal in both

populations. *Blautia* makes up 1.5% of US and 1.3% of Brazilian human fecal samples and 0.3% of US sewage and 2.9% of Brazilian sewage (31, 22, 24).

Microbial source tracking assays widely used throughout the US target *Bacteroides* found in animals (2, 3, 8, 20) and humans (3, 24). *Blautia* is another taxonomic group that shows host specificity that can be tracked by the sequence variation in 16S rRNA gene (23; A.M. Eren et al., unpublished). We recently described an assay for a *Blautia* spp. found in humans but nearly absent in cow and chicken samples (23, 24). These assays demonstrated their applicability in a previous study tracking human fecal inputs to surface waters in Brazil (Ponce Terashima et al., unpublished). Genetic markers based on *Lachnospiraceae*, and more specifically, *Blautia*, may prove more universally applicable than *Bacteroides*, given the higher proportion of this taxonomic group within the Brazilian fecal and sewage samples. The human-specific HF183 Bacteroides assay is a useful marker in certain regions (i.e., the United States and Europe); but *Prevotella* may serve as a better target in regions such as rural Brazil, where the standard diet favors its dominance in the gut microbiome. Future studies could examine Prevotella population structure in humans across a broad geographic area to develop a human *Prevotella* targeted assay, either alone or in conjunction with a *Blautia* assay.

Comparison of Brazil human and animal fecal microbial communities

All animal contained 14 common fecal families, but in different relative proportions; genera within these fecal families varied among samples as well (Table 1). A comparison of the 14 most abundant fecal-associated families in Brazilian humans,

animals, and sewage showed that *Ruminococcaceae* had the highest relative abundance, making up >12% of each sample, followed by *Prevotellaceae* and *Lachnospiraceae*. These families comprised >80% of the human sequence reads and >60% of the sewage and pig samples, but were less abundant in the cow, horse, and dog fecal samples. *Enterobacteriaceae,* a large family of facultative anaerobes that contain pathogenic members, made up >11% of the dog, horse and cow fecal samples, >5% of the sewage samples, but <2% of the human and pig fecal samples. *Bacteroidaceae,* the family containing the genus *Bacteroides,* comprised ~3% of the dog, horse and cow samples, but was considerably lower in the human and pig fecal samples. This particular suite of organisms, notably the top four fecal families, could be useful for tracking animal fecal contamination in the surface waters of Brazil. US studies have identified animal specific animal genetic markers (14, 16). These assays target the V2 to V4 regions, whereas our study targeted the V6 region. Therefore, additional full length sequencing and or deep sequencing of additional regions is needed to make direct comparisons.

Distribution of *Blautia* and *Prevotella* oligotype profiles among Brazil sewage, human and animal fecal microbial communities

We identified over 60 *Blautia* oligotypes in the ten human fecal samples, representing a total of 134,595 sequence reads (Figure 2). Thirty-three oligotypes (33,677 sequence reads; 25%) were human-specific (found in all human samples and only human samples; 6), human-associated (found only in some human samples; 26), or human-preferred (found at significantly higher abundance in humans than other sources; 3) when compared to the animals in our study (pigs, dogs, horses, and cows). A total of 31 of the

33 human oligotypes were also present in the sewage sample from Embasa. The top three human-specific oligotypes comprised 13% of the human and 3% of the sewage specific reads from the genus *Blautia*. Assays based on these human-specific *Blautia* oligotypes could be used in conjunction with the Lachno2 assay to enhance specificity of human source identification.

Although *Prevotella* was noted as the dominant organism in the Brazilian human fecal samples, the most dominant Prevotella oligotype in humans was also dominant in animals (Figure 3). Only 13 of the 108 *Prevotella* oligotypes identified in the ten human fecal were human-associated, while three were host-preferred. Sewage also contained eight of the 13 human-associated and all three human-preferred oligotypes. The top three human-preferred *Prevotella* oligotypes comprised 0.2% of the human and 0.2 % of sewage samples.

The relative abundance of human-associated and preferred *Prevotella* oligotypes was lower compared to *Blautia* oligotypes, and Prevotella had no strictly host-specific oligotypes. Thus, although *Prevotella* is the most abundant genus in terms of total sequence reads, the majority of the sequence reads are not specific to the mammals surveyed. A higher abundance of human-specific/preferred/associated *Blautia* oligotypes suggests that this genus may be a more discriminatory fecal source marker. It should be noted that our sequence reads are very short and that higher diversity may occur in other regions of the *Prevotella* 16S rRNA gene (14, 18). The combination of these two fecal

indicators could be very useful for simultaneously tracking total fecal loads as well as the relative contribution from humans.

Tracking human fecal contamination in surface water with oligotype versus FIB sequence abundance

The relative abundance of sequence reads identified as *Prevotella*, *Blautia*, *E. coli* and *Enterococcus* varied at sites along the Jiquiriçá River (Figure 4). *Prevotella* spp. and *Blautia* spp. both increased considerably at site 5 and steadily increased at subsequent sites downstream (sites 6 and 7), where the integrated impacts of sewage input are shown to be the greatest (Ponce Terashima et al., unpublished). In contrast to the increasing trend of our fecal signature organisms, traditional FIB populations followed a steady or even decreasing trend. *E. coli* steadily decreased from sites 3 to 5, with only a slight increase at sites 6 and 7; and *Enterococcus* spp. steadily decreased with each downstream site. The human *Blautia* and *Prevotella* oligotypes showed great facility in tracking the human fecal contamination along the Jiquiriçá River. Human-specific/preferred *Blautia* and *Prevotella* oligotypes were absent in sites 3 and 4, but appeared at sites 5-7 (Figure 4), consistent with the qPCR-based findings of the HF183 and Lachno2 markers (Ponce Terashima et al., unpublished).

We also examined the presence of human-specific *Blautia* and *Prevotella* oligotypes in the Lucaia River, a historic river in Salvador, Brazil, near the Embasa treatment facility, that now generally appears to be a sewer ditch. This sample had a similar oligotype composition to the Embasa sewage and Jiquiriçá River sites 5-7, demonstrating the

prevalence and traceability of *Blautia* and *Prevotella* oligotypes in both rural and urban environmental samples. Our cluster analyses illustrated that the human, Embasa sewage, downstream river and Lucaia River samples are similar in oligotype composition, while the animal (three dogs, three pigs, two cows, and one horse) and upstream river samples (sites 3 and 4) are similar (*Blautia* oligotypes; Figure 5). The *Prevotella* oligotypes also defined distinctions between the human, sewage (Embasa and the Lucaia River) and downstream river sites and the animal samples and upstream sites; however, the similarities were less pronounced than those of the *Blautia* oligotypes (Figure 6). These findings further confirm the distinction of human/sewage *Blautia* oligotypes from animals and demonstrate our ability to track human fecal contamination in the environment.

Conclusions

We identified *Prevotella* as the dominant genus in Brazilian human fecal samples and second most abundant in Brazilian sewage. The genus *Blautia* was also consistently present in human/sewage samples. While both genera have demonstrated their applicability in distinguishing human and animal fecal pollution in Brazil surface waters, *Blautia* appears to contain more discriminatory and globally applicable markers for human sources. *Blautia* and *Prevotella* oligotypes may be a useful combination for identifying source of fecal pollution: *Prevotella* might function as a "general" indicator to assess total fecal pollution, while the host-specific/preferred sequences identified in *Blautia* and *Prevotella* would assess the amount of human influence. This is particularly important in rural regions of the world, where sewage handling and animal waste are both

likely to impact waterways (33). Overall, the notable differences between Brazilian and US human fecal and sewage microbial communities reiterate the need to develop alternative indicators with global applicability for tracking of human fecal pollution in surface waters.

FIGURE 1: Microbial community populations of sewage and human fecal samples collected in Brazil
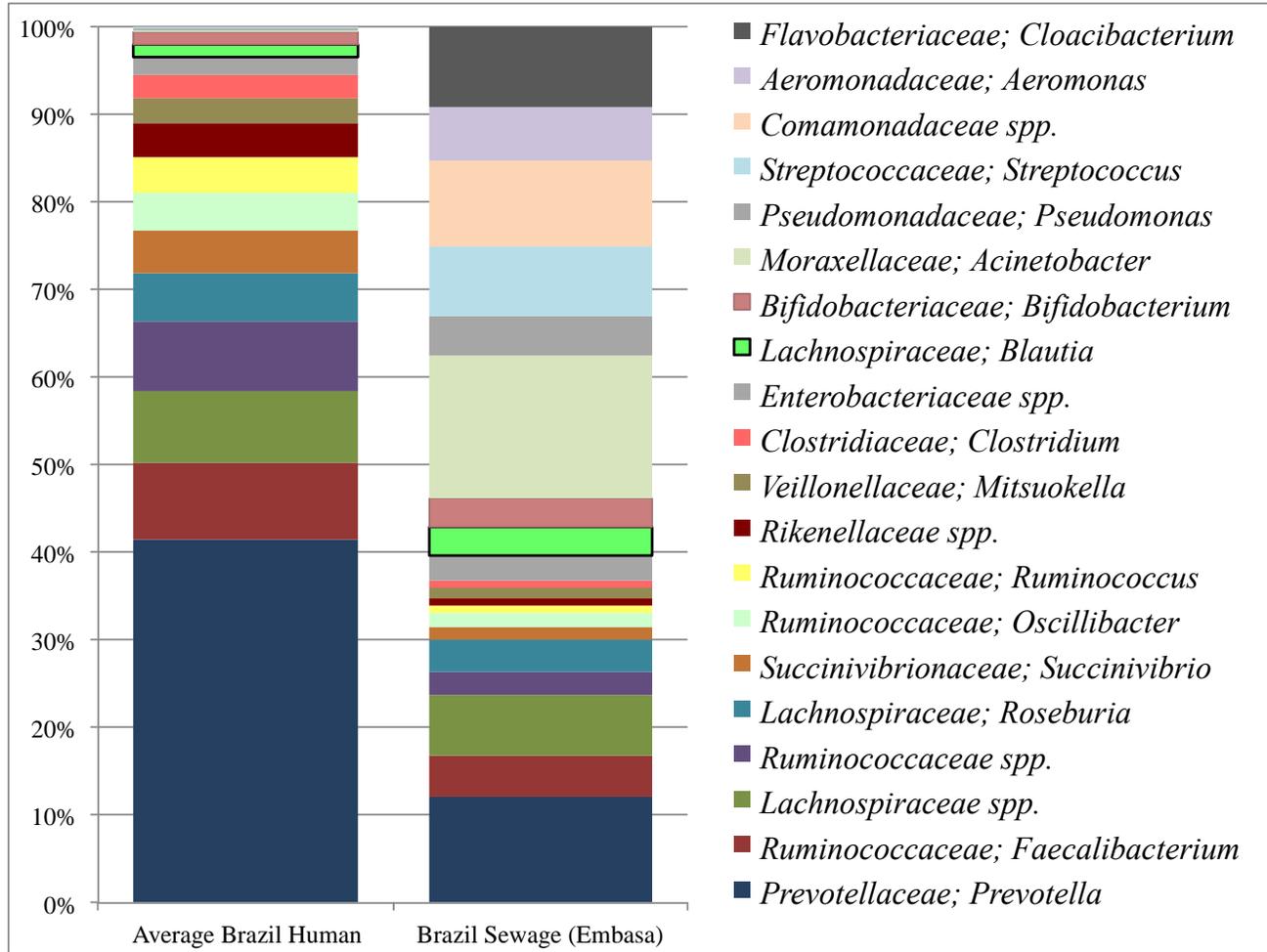
FIGURE 1: Microbial community populations of sewage and human fecal samples collected in Brazil. A 50 mL sewage sample was from Embasa (Empresa Baiana de Águas e Saneamento - The Bahian Water and Sanitation Company), a water treatment facility in Salvador, Brazil, on August 13, 2012. DNA extraction was carried out as described previously by our laboratory (22, 24). The human fecal samples (n=10) were collected as part of a 2009 schistosomiasis survey (4). The V6 hypervariable regions of the 16S rRNA gene from community genomic DNA were amplified and sequenced using Illumina HiSeq as described previously (10). Taxonomy was assigned to sequences using GAST (19), and taxonomic counts were normalized to the maximum number of sequences. Nearly 15 million bacterial sequence reads were generated (~1 million reads per sample). Sequence data is deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive. Taxonomic counts were normalized to maximum number of sequences and abundance parameters were set from 1 to 100%. The relative abundance of the top 20 genera occurring in sewage (n=1) and humans (average of n=10) are presented.

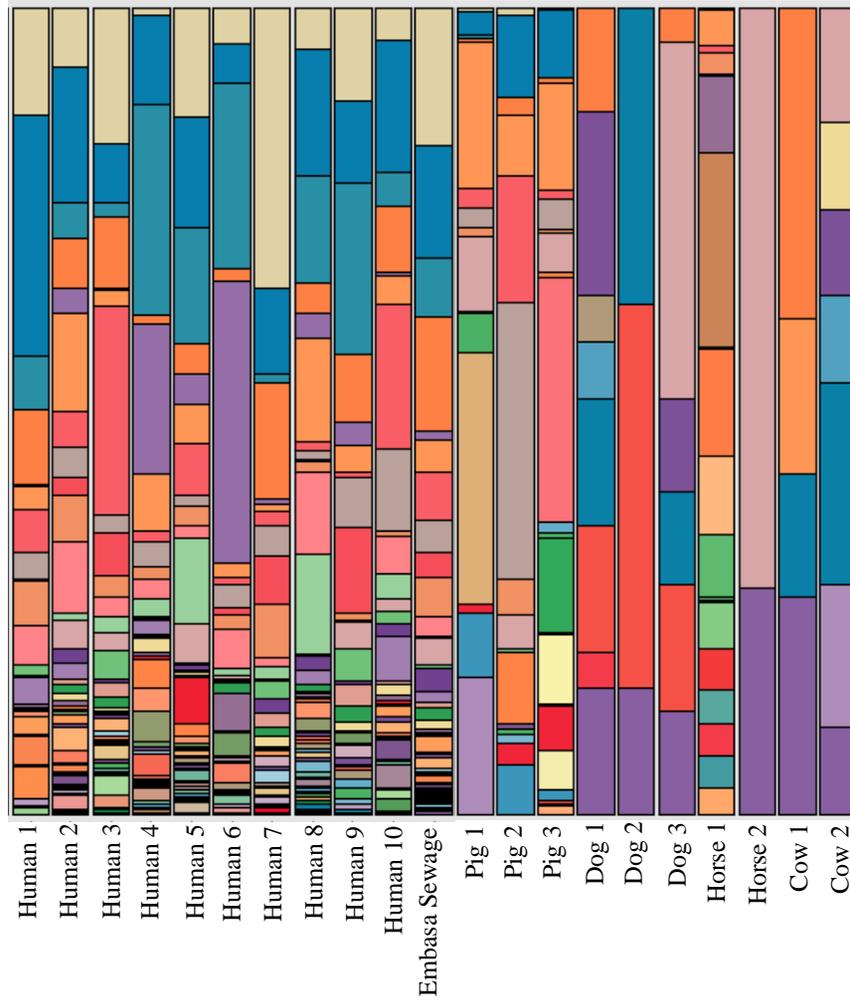**FIGURE 2: Distribution of *Blautia* oligotypes**

FIGURE 2: Distribution of *Blautia* oligotypes for samples from Brazil sewage, human and animal fecal microbial communities. Oligotyping analysis was performed on 175,985 reads that are identified as *Blautia* from 27 samples with the oligotyping pipeline version 0.96 (available from http://oligotyping.org) using 15 components following the initial entropy analysis. Entropy values that oligotype components were selected from (i.e. base locations of interest in the alignment) include: 1, 2, 14, 15, 16, 23, 24, 25, 31, 35, 36, 38, 40, 53, 56. To reduce the noise, each oligotype was required to have a most abundant unique sequence with a minimum abundance of 100. Oligotypes that did not meet this criterion were removed from the analysis. The final number of quality controlled oligotypes was 81, and they represented 169,295 reads (equivalent to 96.20% of all reads analyzed).

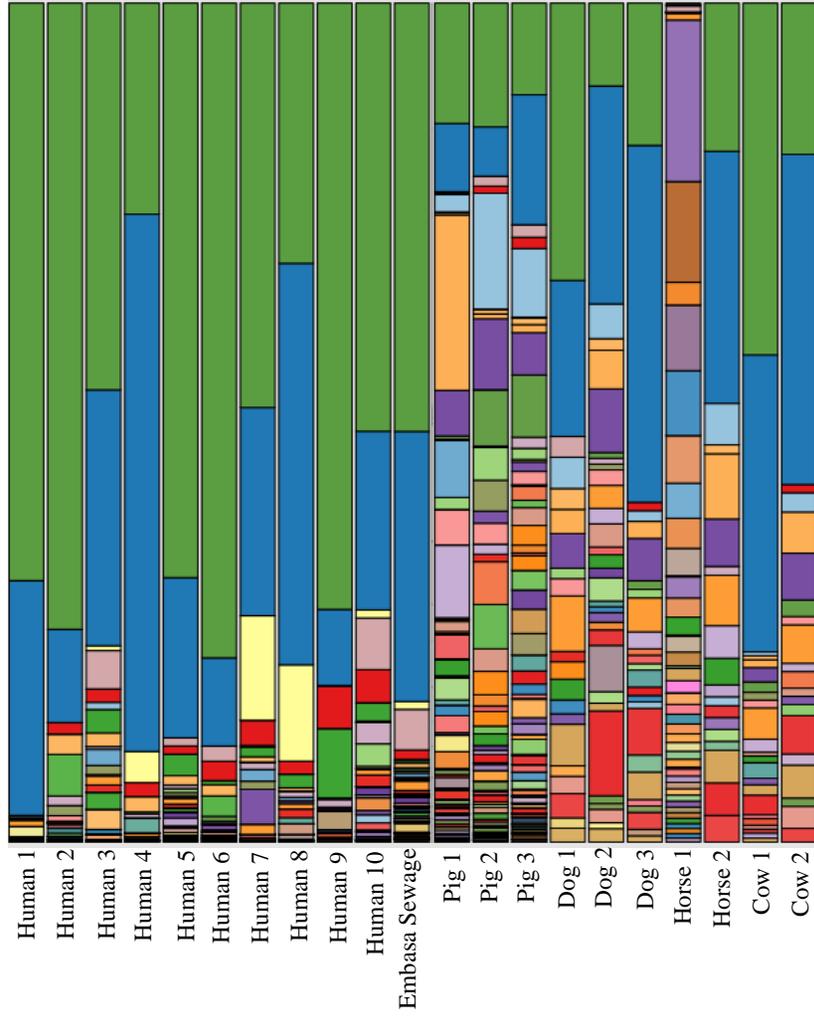**FIGURE 3: Distribution of *Prevotella* oligotypes**

FIGURE 3: Distribution of *Prevotella* oligotypes for samples from Brazil sewage, human and animal fecal microbial communities. Oligotyping analysis was performed on 4,676,785 reads that were identified as *Prevotella* from 27 samples with the oligotyping pipeline version 0.96 (available from http://oligotyping.org) using 19 components following the initial entropy analysis. Entropy values that oligotype components were selected from (i.e. base locations of interest in the alignment) include: 10, 13, 15, 17, 19, 20, 21, 23, 24, 25, 27, 34, 35, 37, 38, 40, 50, 54 and 55. To reduce the noise, each oligotype was required to have a most abundant unique sequence with a minimum abundance of 400. Oligotypes that did not meet this criterion were removed from the analysis. The final number of quality controlled oligotypes was 188, and they represented 4,641,000 reads (equivalent to 99.23% of all reads analyzed).

FIGURE 4: Abundance of human-specific/preferred *Blautia* and *Prevotella* oligotypes and Hi-Seq Illumina sequence reads
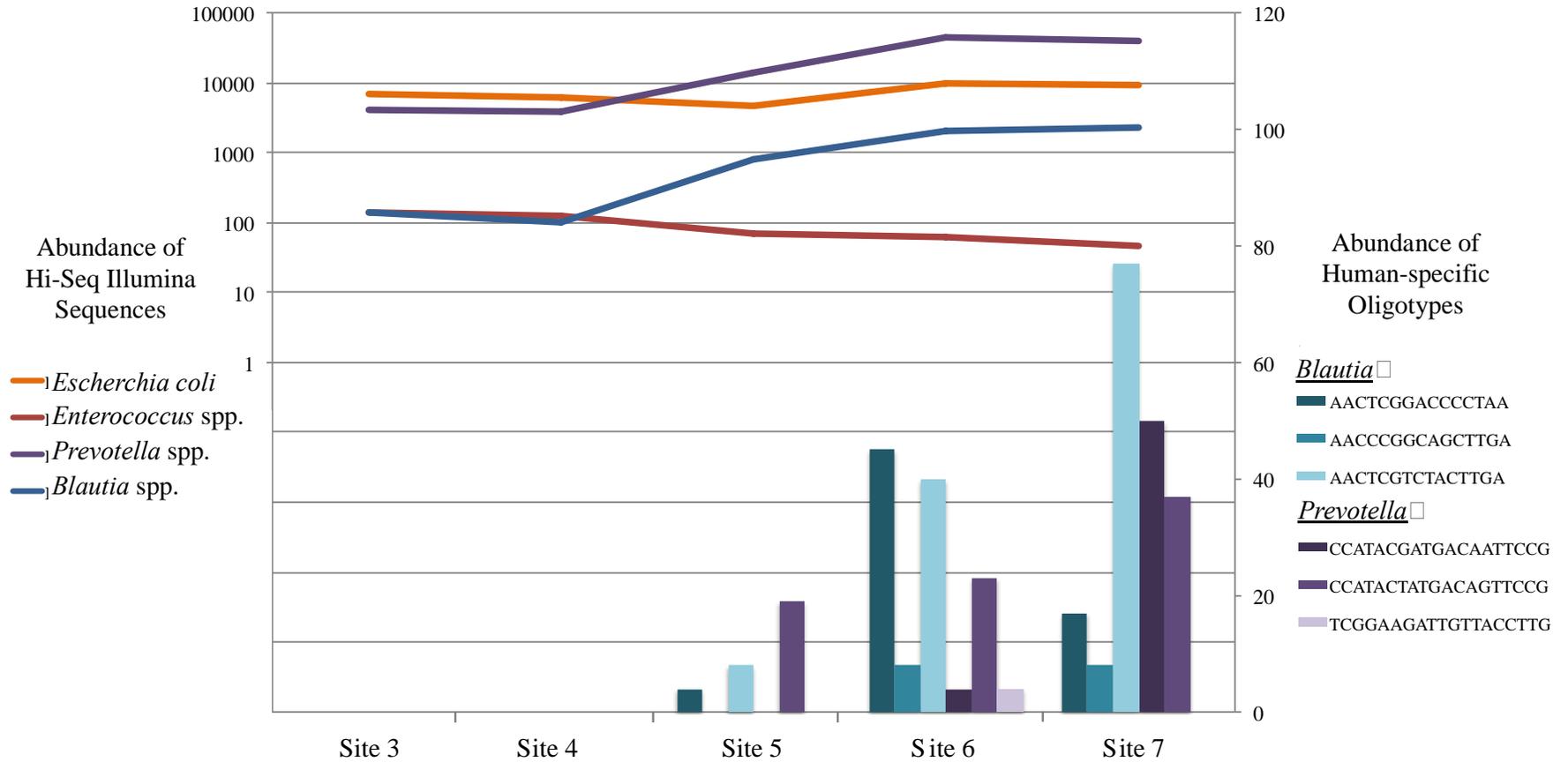
FIGURE 4: Human-specific/preferred *Blautia* and *Prevotella* oligotypes and HiSeq Illumina sequence reads identified as *Prevotella* spp., *Blautia* spp., and fecal indicator bacteria (*E. coli* and *Enterococcus* spp.) at five sites along the Jiquiriçá River in Jenipapo, Brazil. The water samples were collected on August 18, 2012 as part of a microbial source tracking and schistosomiasis survey (Ponce Terashima et al., unpublished). As with the sewage and fecal samples, sequence reads were generated using HiSeq Illumina sequencing, targeting the V6 hypervariable region of the 16S rRNA gene. Taxonomy was assigned to sequences using GAST, and taxonomic counts were normalized to the maximum number of sequences. Abundance parameters were set from 0 to 100%; sequence reads are reported on a log-scale. *Prevotella* spp. included *P. buccae, P. copri, P. species_NA*; *Blautia* spp., were *B. schinkii* and *B. species_NA*; and *Enterococcus* spp. included *Entero*coccus *casseliflavus, E. faecalis, E. faecium, E. ratti*, as well as *Enterococcus* species_NA. Sequences classified as *E. coli* and unclassifiable *Enterobacteriaceae* (non-distinguishable between *Escherichia* and *Shigella*) were added to *E. coli* counts. The top three human-specific *Blautia* and human-preferred *Prevotella* oligotypes are reported as sequence read counts.

FIGURE 5: Hierarchical cluster analysis of *Blautia* oligotypes

FIGURE 5: Hierarchical cluster analysis of 27 samples (ten human, ten animals, one sewage (Embasa), and six river samples) with respect to *Blautia* oligotypes they possess. Distances between samples were determined using Canberra distance metric. The fiver river oligotypes cluster more closely with animals upstream and with humans downstream after passing through the village.

FIGURE 6: Hierarchical cluster analysis of *Prevotella* oligotypes

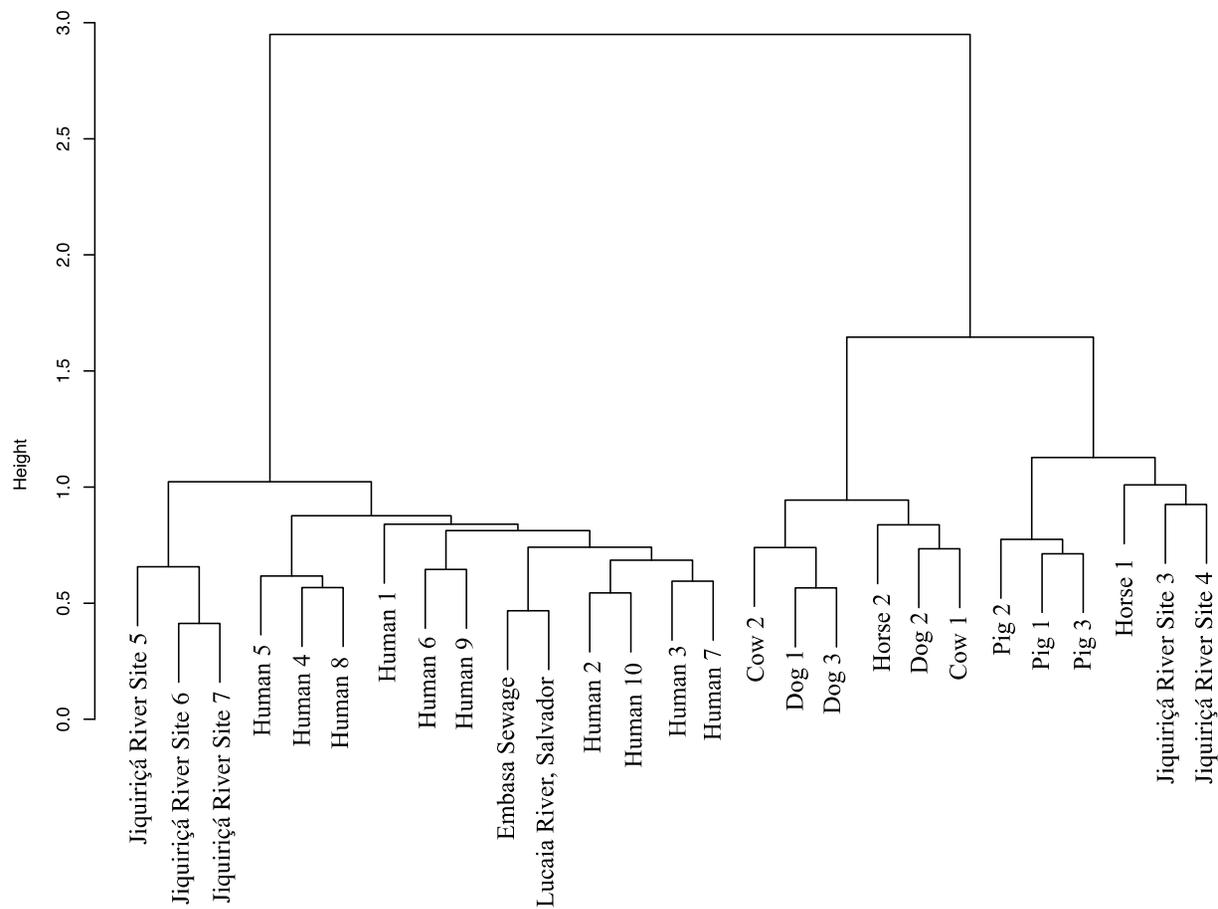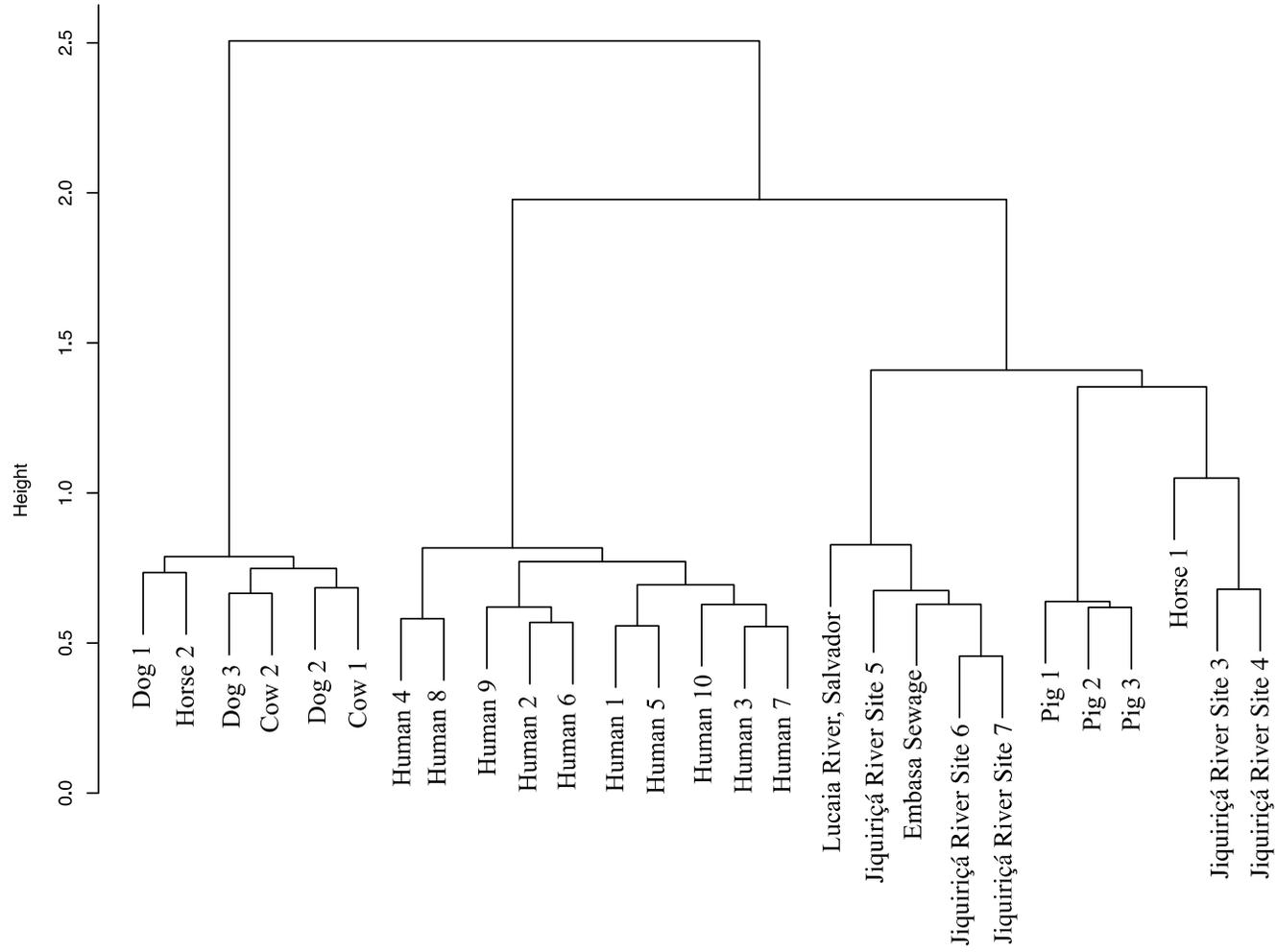**FIGURE 6:** Hierarchical cluster analysis of 27 samples (ten human, ten animals, one sewage (Embasa), and six river samples) with respect to *Prevotella* oligotypes they possess. Distances between samples were determined using Canberra distance metric.

TABLE 1: Average taxonomic composition of fecal-associated bacteria human fecal and sewage samples

Family	Emabsa Sewage %	Average Human %	Average Pig %	Average Dog %	Average Horse %	Average Cow %
Ruminococcaceae	18	25	30	13	23	18
Prevotellaceae	19	38	19	10	13	12
Lachnospiraceae	27	20	16	10	18	8.3
Enterobacteriaceae	5.4	2.0	1.4	36	11	20
Porphyromonadaceae	4.8	0.50	7.3	8.4	5.8	10
Veillonellaceae	7.9	4.4	11	3.2	3.8	2.7
Rikenellaceae	1.5	3.6	4.6	3.2	12	7.0
Clostridiaceae	1.3	2.3	1.7	3.5	4.1	8.1
Erysipelotrichaceae	4.4	2.7	3.9	3.5	2.0	2.1
Bacteroidaceae	4.9	0.30	0.30	3.8	3.1	3.4
Lactobacillaceae	0.70	0.10	1.5	3.8	1.7	4.2
Bifidobacteriaceae	4.9	1.4	0.30	0.90	1.0	0.60
Fusobacteriaceae	0.30	0	1.9	1.8	0.90	3.7
Enterococcaceae	0.10	0	0	0.10	0.10	0.10

TABLE 1: Average taxonomic composition of fecal-associated bacteria human fecal and sewage samples (from Figure 1), compared with ten Brazilian animal fecal samples (three pigs, three dogs, two cows and two horses) collected from Jenipapo, Brazil in August 2012. As with the previous samples, the V6 hypervariable regions of the 16S rRNA gene from community genomic DNA were amplified and sequenced using HiSeq Illumina sequencing. Taxonomy was assigned using GAST, stored in VAMPS and family level taxonomic counts were normalized to the maximum number of sequences. Data presented are based on the percentage of sequence reads associated with the taxonomic classification of these fourteen previously reported families of fecal bacterial (22, 25, 32).

TABLE 2: Abundance of top three *Blautia* and *Prevotella* human specific oligotypes

	Sequence read abundance (Human)	Sequence read abundance (Sewage)
Blautia		
Total Oligotypes (Among Ten Human Fecal Samples, n=63)	134,595	19,863
Human-Specific Oligotypes (Human Fecal, n=33; Sewage, n=31)	33,677	3,047
Top three human-specific *Blautia* oligotypes and tag sequences:		
AACTCGGACCCCTAA	9,910	205
AAATCTTGACATCCCTCTGACCGGGACTTAACCGTCCCCTTTCTTCGGAACAGAGGAGAC		
AACCCGGCAGCTTGA	4,070	59
AAATCTTGACATCCCCCTGACAGGGCATGTAATGTGCCTTTCCTTCGGGACAGGGGAGAC		
AACTCGTCTACTTGA	3,945	472
AAATCTTGACATCCCTCTGACCGGTCTTTAATCGGACCTTTCCTTCGGGACAGGGGAGAC		
Prevotella		
Total Oligotypes (Among Ten Human Fecal Samples, n=108)	3,787,929	76,261
Human-Preferred Oligotypes (Human Fecal, n=13; Sewage, n=8)	18,125	449
Top three human-specific *Prevotella* oligotypes and tag sequences:		
CCATACGATGACAATTCCG	4,635	7
CGGGCTTGAACTGCTAGTGACGTATGCAGAGACGCATATTTCTTCGGACACTAGCGG		
CCATACTATGACAGTTCCG	2,776	188
CGGGCTTGAACTGCCAGTGACTCATGCAGAGACGCATGTTTCTTCGGACACTGGCGG		
TCGGAAGATTGTTACCTTG	2,368	0
CGGGCTTGAATTGCAGAGGAAGGATTTGGAGACATTGACGCCCTTCGGGGTCTCTGT		

TABLE 2: Abundance of the top three *Blautia* and *Prevotella* human specific oligotypes in ten Brazilian human fecal samples and one Brazilian sewage sample. There were 33 human-specific, associated, or preferred *Blautia* oligotypes identified. Of those 33, 31 were also identified in the Embasa sewage sample. There were 13 human-associated or preferred *Prevotella* oligotypes identified. Of those 13, eight were also identified in the Embasa sewage sample. The top three human specific *Blautia* oligotypes comprised 13% of the human and 3% of the sewage specific reads and the top three human-associated/preferred *Prevotella* oligotypes comprised 0.2% of the human and 0.2% of the sewage specific reads.

CHAPTER 4: References

1.   Barreto, M.L., Genser, B., Strina, A., Teixeira, M.G., Assis, A.M.O., Rego, R.F. et al. (2007) Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies. *Lancet* **370**: 1622-1628.

2.   Bernhard, A.E., and Field, K.G. (2000a) Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl Environ Microbiol* **66**: 1587-1594.

3.   Bernhard, A.E., and Field, K.G. (2000b) A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Appl Environ Microbiol* **66**: 4571-4574.

4.   Blanton, R.E., Blank, W.A., Costa, J.M., Carmo, T.M., Reis, E.A., Silva, L.K. et al. (2011) Schistosoma mansoni population structure and persistence after praziquantel treatment in two villages of Bahia, Brazil. *International Journal for Parasitology* **41**: 1093-1099.

5.   De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S. et al. (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 14691-14696.

6.   de Oliveira, L.K., Fleck, J.D., Comerlato, J., Kluge, M., Bergamaschi, B., Fabres, R.B. et al. (2012) Enteric viruses in water samples from Brazilian dairy farms. *Agricultural Water Management* **111**: 34-39.

7.   Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A. (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.

8.   Dick, L.K., Bernhard, A.E., Brodeur, T.J., Santo Domingo, J.W., Simpson, J.M., Walters, S.P., and Field, K.G. (2005) Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. *Appl Environ Microbiol* **71**: 3184-3191.

9.   Dubinsky, E.A., Esmaili, L., Hulls, J.R., Cao, Y., Griffith, J.F., and Andersen, G.L. (2012) Application of Phylogenetic Microarray Analysis to Discriminate Sources of Fecal Pollution. *Environmental Science & Technology* **46**: 4340-4347.

10.   Eren, A.M., Vineis, J.H., Morrison, H.G., and Sogin, M.L. (2013a) A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS One* **8**.

11.   Eren, A.M., Zozaya, M., Taylor, C.M., Dowd, S.E., Martin, D.H., and Ferris, M.J. (2011) Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS One* **6**: e26732.

12.   Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013b) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution* **4**: 1111-1119.

13.   Field, K.G., and Samadpour, M. (2007) Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res* **41**: 3517-3538.

14.   Fogarty, L.R., Haack, S.K., Wolcott, M.J., and Whitman, R.L. (2003) Abundance and characteristics of the recreational water quality indicator bacteria Escherichia coli and enterococci in gull faeces. *J Appl Microbiol* **94**: 865-878.

15.   Franzolin, M.R., Alves, R.C.B., Keller, R., Gomes, T.A.T., Beutin, L., Barreto, M.L. et al. (2005) Prevalence of diarrheagenic Escherichia coli in children with diarrhea in Salvador, Bahia, Brazil. *Memorias Do Instituto Oswaldo Cruz* **100**: 359-363.

16.   Fremaux, B., Gritzfeld, J., Boa, T., and Yost, C.K. (2009) Evaluation of host-specific Bacteroidales 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Research* **43**: 4838-4849.

17.   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.

18.   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. *Systematic and Applied Microbiology* **33**: 348-357.

19.   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.

20.   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.

21.   Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* **489**: 220-230.

22.   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.

23.   McLellan, S.L., Newton, R.J., Vandewalle, J.L., Shanks, O.C., Huse, S.M., Eren, A.M., and Sogin, M.L. (2013) Sewage reflects the distribution of human faecal *Lachnospiraceae*. *Environmental Microbiology* **15**: 2213-2227.

24.   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.

25.   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. *Microbial Ecology* **65**: 1011-1023.

26.   Roslev, P., and Bukh, A.S. (2011) State of the art molecular markers for fecal pollution source tracking in water. *Applied Microbiology and Biotechnology* **89**: 1341-1355.

27.   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.

28.   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. *Applied and Environmental Microbiology* **79**: 2906-2913.

29.   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.

30.   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 Research* **47**: 6829-6838.

31.   Turnbaugh, P.J., and Gordon, J.I. (2009) The core gut microbiome, energy balance and obesity. *Journal of Physiology-London* **587**: 4153-4158.

32. Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E. et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480-484.

33. UNDP (2001) Closing the loop: Ecological sanitation for food security. In Publications of *Water Resources* **18**.

34. Unno, T., Jang, J., Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.S. et al. (2010) Use of Barcoded Pyrosequencing and Shared OTUs To Determine Sources of Fecal Bacteria in Watersheds. *Environmental Science & Technology* **44**: 7777-7782.

35. USEPA (2005) Microbial Source Tracking Guide Document. In. Washington, DC: Office of Research and Development.

36. 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. *Environmental Microbiology* **14**: 2538-2552.

37. Victoria, M., Guimaraes, F.R., Fumian, T.M., Martins Ferreira, F.F., Vieira, C.B., Shubo, T. et al. (2010) One year monitoring of norovirus in a sewage treatment plant in Rio de Janeiro, Brazil. *Journal of Water and Health* **8**: 158-165.

38. WHO (2013a) Progress on sanitation and drinking-water 2013 update. In: World Health Organization.

39. WHO (2013b) Schistosomiasis Fact Sheet. In. http://www.who.int/mediacentre/factsheets/fs115/en/index.html: World Health Organization

40. Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A. et al. (2011) Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science* **334**: 105-108.

41. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M. et al. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222-227.

CHAPTER 5: Concluding Comments

Overall, this research demonstrated the value of integrating molecular techniques for the detection and identification of fecal contamination in surface water sources. Traditional water monitoring efforts, using culture-based indicator organisms (i.e. enterococci and *E. coli*) as an index of human health risk, has its shortcomings, as recent studies have demonstrated weak correlations between FIB and bacterial pathogens (1). Although the association between fecal contaminated water and human illness is well established, the associations between traditional FIB and human health risk are less decisive; the survival and environmental persistence of FIB and pathogenic bacteria are variable and non-correlative, as are the techniques used to ascertain their presence (culture-dependent vs. culture–independent) (1). The ubiquity of FIB in humans and animals alike presents another challenge for pathogen detection and risk characterization (2-4). As demonstrated through the microbial community population profiles of the gull fecal samples, sometimes the most abundant bacteria (i.e. *Catellicoccus marimammalium*) are not readily culturable and culturable FIB are in low abundance. These findings highlight the accuracy and advantages of using culture-independent over culture-dependent detection techniques for surface water monitoring. Culture-independent techniques are more time efficient, cost effective, and encompassing as they are able to detect hard to culture organisms that might be more relevant to understanding the source of fecal pollution. Given the weak and confounding associations between culture-dependent FIB and human-specific bacterial pathogens, it is prudent that more decisive and accurate methods are utilized when available, such as culture-independent molecular approaches in pathogen surveillance.

Watersheds are complex systems and fecal pollution can be introduced from a variety of sources: sewage overflows, agricultural runoff, urban stormwater, wildlife, etc., making the identification and elimination of the contamination source obscure. Traditional microbiological methods used to detect the presence of pathogens provides no information as to the source of the fecal contamination and given the severity of some waterborne infections, source detection is paramount to prevent the further transmission of disease. Molecular pathogen surveillance, such as MST, has proven to be an effective method of determining sources of fecal pollution in water sources, showing considerable promise as a tool for public health prevention (2, 5-12). In surface water sources with ubiquitous fecal contamination, MST indicators are particularly useful for accurate surface water assessments, human health risk characterizations and pathogen surveillance (2, 5, 10, 12, 13). They can be further used to direct specific sanitation interventions. As applied in this research, MST techniques are also useful for tracking chronic parasitic infections such as schistosomiasis, in addition to acute gastrointestinal illnesses. The parasite causing schistosomiasis has a complex life cycle, long-term survival in the environment and can be difficult to detect in surface water sources; bacterial indicators that track human fecal contamination are a useful proxy for detecting and understanding schistosomiasis transmission.

As noted in the Brazil and US microbial community population comparisons, differences in diet, cultures, levels of sanitation, population densities, genetics, environment, age, antibiotic exposure and/or geography contribute to differences in the geographic applicability of MST marks (5-9). These findings highlight the need for the testing and

validation of different MST assays in across different watersheds to evaluate if makers developed in one region, state or country will transcend different geographical regions. Deep sequencing technologies - such as pyrosequencing, Mi-Seq and Hi-Seq Illumina sequencing – are becoming more commonplace, giving rise to techniques such as fecal-associated microbial signatures and oligotyping; both of which are additional tools that are useful for detecting fecal contamination in surface water sources (14, 15). Whereas MST assays traditionally track one host-specific organism, fecal signatures – a suite of fecally associated organisms – may offer more sensitivity in complex environments (16-18).

Waterborne illnesses continue to be a major public health problem throughout the world. Of the illnesses and outbreaks that are detected and reported, it is believed that there are many that go unreported and undiagnosed (19). Therefore, detection methods and MST that employ single source specific markers, fecal-signatures or oligotyping are important molecular techniques that can be useful for distinguishing sources of fecal contamination. Once established, efforts can be directed at minimizing public health exposures, which will consequently limit disease transmission and prevent further illness.

CHAPTER 5: References

1.   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. Applied and Environmental Microbiology. 78(8):1733-1746.

2.   Field, K.G. and Samadpour, M. (2007) Fecal source tracking, the indicator paradigm, and managing water quality. Water Research. 41:3517-3538.

3.   Field, K.G., Bernhard, A.E., and Brodeur, T.J. (2003) Molecular approaches to microbiological monitoring: fecal source detection. Environmental Monitoring and Assessment. 81(1-3):313-26.

4.   Orskov, F., and I. Orskov. (1981) Enterobacteriaceae. A. I. Broude (ed.), Medical microbiology and infectious diseases. W. B. Saunders Co., Philadelphia, Pa. 340–352.

5.   USEPA (U.S. Environmental Protection Agency). (2005) Microbial Source Tracking Guide Document. Washington, DC: Office of Research and Development, 131.

6.   Bernhard, A. E. and Field, K.G. (2000a) A PCR Assay to Discriminate Human andRuminant Feces on the Basis of Host Differences in Bacteroides-Prevotella Genes Encoding 16S rRNA. Applied and Environmental Microbiology. 66: 4571-4574.

7.   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. Applied and Environmental Microbiology. 66(4):1587-1594.

8.   Kildare, B.J., Leutengger, 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 Research. 41:3701-3715.

9.   Newton, R.L., Vandewalle, J.L., Borchardt, M.A., Gorelick, M.H., and McLellan, S.L. (2011) Lachnospiraceae and Bacteriodales Alternative Fecal Indicators Reveal Chronic Human Sewage Contamination in an Urban Environment. Applied and Environmental Microbiology. 77(19):6972-81. doi: 10.1128/AEM.05480-11.

10. Simpson, J.M., Santo Domingo, J.W., and Reasoner, D.J. (2002) Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36(24):5279-5288.

11. Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., and Lukasik, J. (2002) Microbial Source Tracking: Current Methodology and Future Directions. Applied and Environmental Microbiology. 68(12):5796-5803.

12. 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 Research. 41(16):3539-52.

13. Stewart, J.R., Boehm, A.B., Dubinsky, E.A., Fong, T.T., Goodwin, K.D., Griffith, J.F., Noble, R.T., Shanks, O.C., Vijayavel, K., Weisberg, S.B. (2013) Recommendations following a multi-laboratory comparison of microbial source tracking methods. Water Research. pii: S0043-1354(13)00544-7. doi: 10.1016/j.watres.2013.04.063

14. Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G. and Sogin, M.L. (2013) Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods in Ecology and Evolution. In press. doi: 10.1111/2041-210X.12114

15. Eren, A.M., Zozaya, M., Taylor, C.M., Dowd, S.E., Martin, D.H. and Ferris, M.J. (2011) Exploring the Diversity of Gardnerella vaginalis in the Genitourinary Tract Microbiota of Monogamous Couples Through Subtle Nucleotide Variation. PLoS ONE, 6, e26732. doi: 10.1371/journal.pone.0026732

16. McLellan, S., Boehm, A., and Shanks, O. (2013a) Marine and Freshwater Fecal Indicators and Source Identification. New York: Springer.

17. McLellan, S.L., Newton, R.J., Vandewalle, J.L., Shanks, O.C., Huse, S.M., Eren, A.M., and Sogin, M.L. (2013b) Sewage reflects the distribution of human faecal Lachnospiraceae. Environmental Microbiology. 15(8):2213-2227.

18. Newton, R.L., Bootsma, M.J., Morrison, H.G., Sogin, M.L., and McLellan, S.L. (2013) A Microbial Signature Approach to Identifying Fecal Pollution in the Waters Off an Urbanized Coast of Lake Michigan. Microbial Ecology. 65:1011-1023.

19. O'Brien, S.J., Rait, G., Hunter, P.R., Gray, J.J., Bolton, F.J., Tompkins, D.S., McLauchlin, J., Letley, L.H., Adak, G.K., Cowden, J.M., Evans, M.R., Neal, K.R., Smith, G.E., Smyth, B., Tam, C.C., and Rodrigues, L.C. (2010) Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). BMC Medical Research Methodology. 10(39):1-13.