Title
Household Reservoirs of Fecal Contamination in Rural Bangladesh

Author
Fuhrmeister, Erica

Publication Date
2019

Peer reviewed|Thesis/dissertation

Household Reservoirs of Fecal Contamination in Rural Bangladesh

by

Erica R Fuhrmeister

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kara L. Nelson, Chair
Professor Lisa Alvarez-Cohen
Professor John M. Colford Jr.

Summer 2019

Household Reservoirs of Fecal Contamination in Rural Bangladesh

Abstract

Household Reservoirs of Fecal Contamination in Rural Bangladesh

by

Erica R Fuhrmeister

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Berkeley

Professor Kara L. Nelson, Chair

Enteric pathogens are a leading cause of diarrheal illness in low-and middle-income countries and are transmitted through the fecal-oral route. Exposure routes can be blocked through adequate drinking water treatment, sanitation facilities and hygiene practices (WaSH). While these pathways are well understood conceptually, few studies have explored household transmission and the impacts of WaSH interventions with empirical measurements of enteric pathogens. In this dissertation, indicators and pathogens were measured in household reservoirs (stored drinking water, soil, and mothers and child hands) in rural Bangladesh to quantify the impact of a sanitation intervention on household level fecal contamination. Additionally, the reliability of indicator *Escherichia coli* to suggest risk of enteric pathogen infection was evaluated in this context. Fecal indicator organisms were also used to assess the importance of animal fecal sources in these households. Lastly, the association between bacterial communities in household reservoirs was quantified using 16S rRNA gene sequencing. Through this work I showed that the sanitation intervention in place had limited impact on household fecal contamination. In later chapters, indicator *E. coli* is shown to be a useful tool to indicate risk in this context as higher concentrations of indicator *E. coli* were associated with higher prevalence of pathogenic *E. coli* genes. I also determined that animals were potential sources of *Giardia* and pathogenic *E. coli* genes on mother hands. In Chapter 4, shared bacteria between mother and child hands, hands and soil, and mother hands and stored water suggest the potential for transmission of pathogens.

To my dad

Contents

List of Figures

List of Tables

Acknowledgments

I owe a debt of gratitude to the many individuals who contributed to this dissertation. First, I thank my advisor, Kara Nelson, for pushing me beyond what I thought capable. I also thank Kara for providing the space and support needed to instill a sense of confidence and never doubting my ability, even when I did. Kara also created a climate of acceptance and authenticity evident not just in her mentorship, but in the whole Nelson research group (both current students and alumni). I feel very fortunate to be a part of this network.

Thank you to my co-authors Kaitlyn Jeanis, Mahaa Ahmed, Sara Brown, Yoshika Crider, Ben Arnold, Alan Hubbard, Mahfuja Alam, Debashish Sen, Sharmin Islam, Mir Hamayet Kabir, Mahfuza Islam, Leanne Unicomb, Mahbubur Rahman, Laura Kwong, Jess Grembi, Ali Boehm, Steve Luby, and Jack Colford who contributed to all aspects of this work including study design, sample collection, molecular lab work, and helpful commentary on manuscripts. I would especially like to thank Ayse Ercumen and Amy Pickering for their integral role in this project. This work would not have been possible without their guidance.

The highlight of my PhD experience at Berkeley was the incredible network of peers to provide feedback, encouragement, and most importantly, friendship. I especially would like to thank Scott Miller, Lauren Kennedy, Karina Chavarria, Siva Rama Satyam Bandaru, Rachel Scholes, and Emily Marron for their support over the years and for always lending an ear. A thanks to my non CEE friends Yukiko Yano, Nishita Deka, Brian Gong, and Hannah Evans for providing an outside perspective.

A special thanks to my family Krista, Jon, Greg, Chrissie, Michael, Ma, and the Marchand Benmamans for your support during this endeavor. Lastly, I thank my partner Jorge Marchand for the unwavering encouragement and for being my champion every step of the way.

This dissertation was made possible by the National Institute of Health, the National Science Foundation Graduate Research Fellowship, and in part by the Bill and Melinda Gates Foundation.

Chapter 1

Introduction

1.1 Introduction

Every year, 6 million children under the age of five die around the world[1]. The majority of deaths occur in low-and middle-income countries (LMICs) and are preventable. Of these, diarrheal illnesses are a leading cause of communicable illness death, second to respiratory infections[1]. In 2015, one out of every ten children under five died from diarrheal illness, disproportionately impacting the most vulnerable age group, children under two[1, 2]. Diarrheal illnesses not only cause death, but can also initiate lifelong health effects, primarily associated with malnutrition such as growth stunting and impaired cognitive development[3–5]. Beyond the health impacts, diarrheal illnesses have economic consequences which include increased health care costs and lost wages due to illness[6].

There are a plethora of etiological agents of diarrhea spanning multiple types of microorganisms including viruses, bacteria, and protozoa. These three classes of microorganisms are all represented in the most common agents of moderate to severe diarrhea globally which are rotavirus (virus), *Cryptosporidium* (protozoa), enterotoxigenic *Escherichia coli*(ETEC) (bacteria), and *Shigella* (bacteria)[7]. The most deadly enteric pathogens include ETEC, enteropathogenic *E. coli* (EPEC), *Cryptosporidium*, and rotavirus, although leading agents of diarrheal death vary geographically[1, 7]. In Southeast Asia, other causes of moderate to severe diarrhea include *Aeromonas*(bacteria), adenovirus 40 and 41 (virus), and *Campylobacter jejuni* (bacteria)[7].

Enteric pathogens are transmitted from the feces of an infected host to a new host through environmentally mediated pathways (Figure 1.1)[8]. The F-Diagram was first adopted in 1958 to describe these fecal-oral pathways from feces to fluids, fields, flies, fingers, fomites and food. Enteric pathogens can originate from the feces of children and adults as well as animals. Until recently, animal and child feces have been largely ignored reservoirs of pathogens. These sources are especially important because neither are contained through conventional

sanitation (toilet or latrine). Animal and child feces are either left in place, buried, disposed of in a latrine, disposed of in the environment (river, drain, etc.), or disposed of with garbage[9, 10]. Animal feces can carry pathogens such as *Campylobacter*, *Cryptosporidium*, *Giardia* and pathogenic *E. coli*[11, 12]. Likewise, child feces are more likely to contain pathogens than adults due to a higher prevalence of diarrhea in children[13]. In a study of child feces disposal practices around the world, less than 50% of child feces were disposed of in a pit latrine or toilet in most countries[14]. Improper disposal of child feces has been linked to undernourishment and increased risk of diarrhea[9, 15].

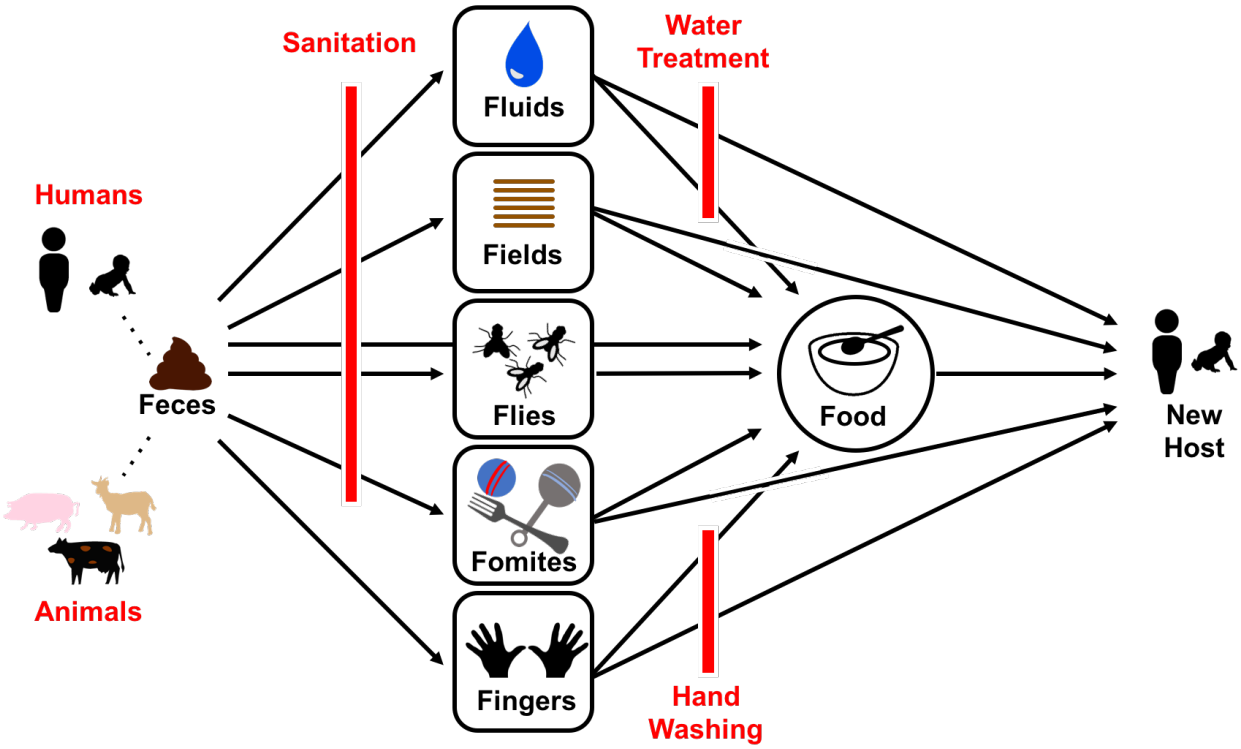


Figure 1.1: Fecal-oral transmission pathways through household reservoirs. Pathways blocked with saniation, water treatment, and hand washing interventions are shown in red.

Animal husbandry is widely practiced in LMICs and has been linked to increased prevalence of diarrhea[12]. In these settings, poultry and livestock are kept in close contact with human living areas due to socioeconomic factors, such as availability of land[16–18]. Genetic markers of animal fecal contamination have been found to be highly prevalent in the household environment (in hands, stored water, and soil) in Bangladesh and India[19–22]. Contact with animal feces is also introduced during cooking in places where animal feces are used as cooking fuel. These cooking practices result in direct contact between fecal piles and caretaker hands as they form fuel patties[23]. Pathogens on hands can then be ingested through direct oral contact, or indirectly through contact with food, fomites or drinking water.

Although the F-diagram was conceptualized over 60 years ago, household level transmission of enteric pathogens is still relevant today. Key pathways of transmission can be interrupted with water treatment, sanitation, and hygiene interventions to prevent further spread of pathogens downstream (Figure 1.1). In particular, sanitation is a crucial intervention that can block the transmission cycle at an early stage (feces to fluids, feces to flies, feces to fields, feces to fomites, feces to foods) and prevent subsequent transmission (fields, fluids, flies, fomites, food to a new host). Yet, 2.3 billion people still lack access to improved sanitation, facilities that provide barriers to separate humans from feces[24]. An additional 2.2 billion people do not have safely managed sanitation services, defined as unshared improved sanitation facilities with safe excreta disposal practices[24]. Although international development targets are implementing stricter definitions of safe sanitation, evident in the shift from improved to safely managed sanitation in the sustainable development goals, these targets focus on containment of primarily human feces. Animal feces and feces from young children are still overlooked in sanitation goals.

In order to improve the efficacy of sanitation interventions for both humans and animals it is important to understand how pathogen transmission is blocked, specifically which pathways in the environment are most important. This requires measuring enteric pathogens in household reservoirs such as hands, flies, stored drinking water, food, and soil. Measuring enteric pathogens in the environment is particularly challenging due to both the number of potential agents of disease and low number of organisms capable of causing disease[25, 26]. The infectious dose of pathogens can be lower than the detection limits of even the most sensitive detection methods. Therefore, fecal indicators are used to suggest the potential for enteric pathogens.

Ideal fecal indicators are microorganisms that are present when pathogens are present, do not grow in the environment, are part of the normal intestinal tract microbiome of warm blooded animals, and their concentration relates to the degree of contamination[27]. Coliform bacteria, specifically *E. coli*, are a common fecal indictor and are the basis for safe drinking water guidelines by the World Health Organization and recreational water quality criteria in the United States[28, 29]. Although *E. coli* exhibits many of the ideal characteristics of a fecal indicator, studies show that *E. coli* can grow in the environment, especially in warm tropical settings, and therefore concentration may not relate to the degree of fecal contamination[30, 31]. The degree to which *E. coli* correlates with enteric pathogens is highly variable depending on context[32]. Previous studies have found increased concentrations of thermotolerant coliforms in drinking water were associated with increased odds of diarrhea in LMICs[33], but few studies have investigated the association between concentration of indicators and presence of actual pathogens in drinking water and other household reservoirs.

Other fecal indicators include microbial source tracking markers, which are bacteria specific to the intestinal tract of warm blooded animals, used to determine the source of fecal contamination[27]. Bacteria from the order *Bacteroidales* exhibit host specificity and can dif-

ferentiate the presence of human versus animal feces, although the sensitivity and specificity of these markers varies geographically[34, 35]. Source tracking has been used most commonly to demonstrate the presence of human, avian, ruminant, and dog feces in environmental waters in developed countries, but have also been employed in household reservoirs in India, Bangladesh and Tanzania[19–22, 36, 37].

This dissertation contributes to the small but growing body of literature characterizing enteric pathogen transmission in the household environment in low-and middle-income countries. Through this research, I quantify the impact of a human and animal sanitation intervention, investigate the relationship between indicators and pathogens in the household setting, evaluate the contribution of animal fecal contamination, and elucidate sources and overlap in microbial communities in household reservoirs.

1.2 Study Site: Rural Bangladesh

The research presented in this dissertation was conducted within the control and sanitation arms of a randomized controlled trial investigating the impact of water, sanitation, hygiene and nutrition interventions on child health in rural Bangladesh (WASH Benefits)[38, 39] (Figure 1.2).

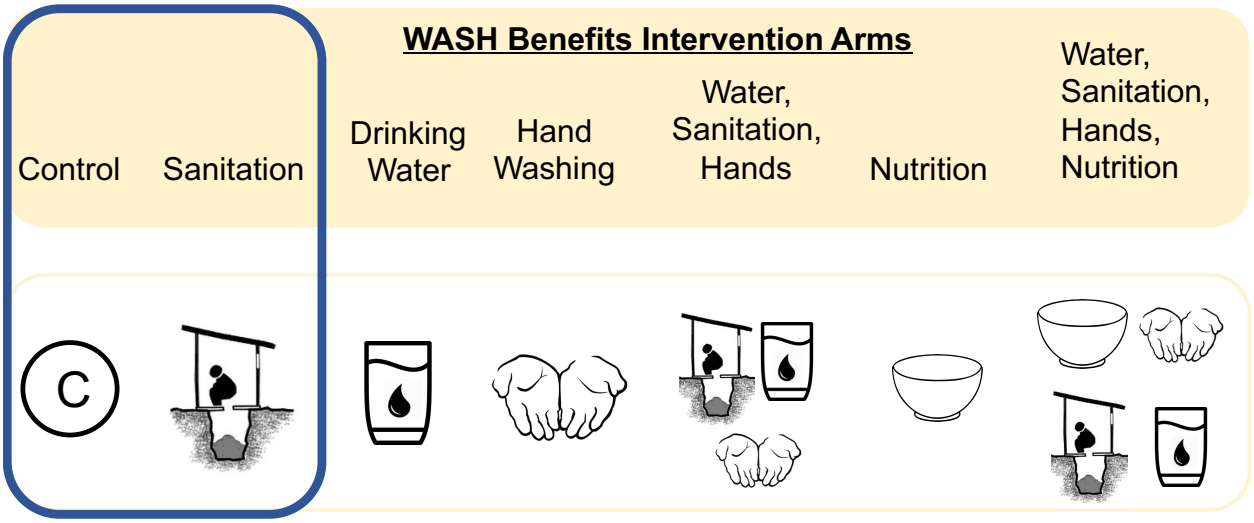


Figure 1.2: WASH Benefits treatment arms. This dissertation focuses on a subset of households in the sanitation and control groups.

Households were location in the Mymensingh, Tangail and Kishoreganj districts (Figure 1.3a). Households in Bangladesh are typically part of a larger compound, where homes of blood relatives surround a central courtyard (Figure 1.3b). For treatment assignment, compounds were clustered in groups of eight to reduce spillover effects and for the ease of community promoters delivering the interventions. Clusters were geographically pair

matched and then randomly assigned to a treatment arm. Samples for this dissertation were collected from one household per compound in a subset of households enrolled in the control and sanitation arms between March and October 2015. WASH Benefits enrolled pregnant women and followed their birth cohort. At the time of sample collection, each household contained a child between 9 and 44 months old, and any siblings of the birth cohort.

Figure 1.3: a.) Location of study cite in rural Bangladesh shown in black circle. b.) Example of a household compound c.) Goat present in a compound.

The sanitation intervention included dual pit latrine with a water seal, child potties, and a scoop for removal of animal feces (Figure 1.4). Community promoters provided an additional layer to the intervention by promoting behavior change and ensuring functionality of hardware. They visited enrolled households on average, six times per month. At the time of sample collection, it had been 16-35 months since the latrines were constructed.

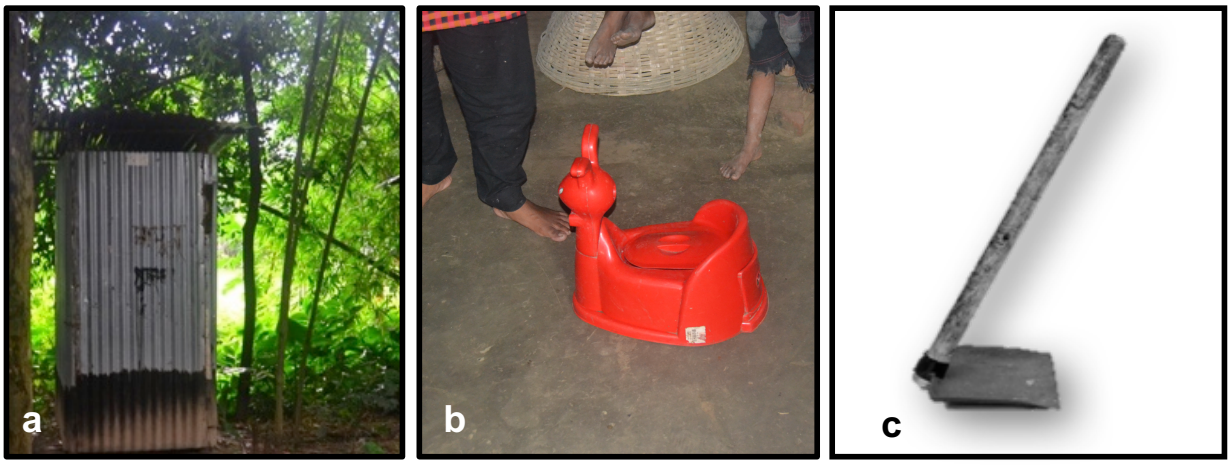


Figure 1.4: a.) Dual pit latrine b.) Child potty c.) Sani-scoop.

Study participants primarily obtained drinking water from shallow tubewells, as is common in Bangladesh. Households in this study consumed drinking water with low levels of iron

and arsenic determined by the Government of Bangladesh and British Geological Survey in order to avoid decreased efficacy of chlorine in the water treatment arm of the main study. There were no reported major WASH interventions from the government or international non-governmental organizations in the study area that would coincide with the study period. Bangladesh experiences a monsoon season from June to October, consequently households from hoar areas (areas completely submerged during the monsoon season) were excluded. In the study area, the main source of income was agriculture and common household animals included chickens, goats, cows, and ducks (Figure 1.3c).

1.2.1 Ethics

Participants provided written informed consent. The study protocol was approved by human subjects committees at the icddr,b (PR-11063), University of California, Berkeley (2011-09-3652), and Stanford University (25863).

1.3 Dissertation Outline

This dissertation contains three research chapters (Chapters 2-4) followed by a conclusion chapter (Chapter 5). In Chapter 2, the impact of a sanitation intervention on the prevalence of microbial source tracking markers and specific pathogen genes in household reservoirs is presented. The sanitation intervention consisted of dual pit latrines, scoops for removal of child and animal feces, and child potties. Reservoirs of contamination sampled within households included stored drinking water, mother and child hands, and soil. The prevalence, prevalence ratio, and prevalence difference were calculated for binary marker data (presence/absence). The concentration and change in concentration was calculated for continuous microbial source tracking marker data. Subgroup analyses were conducted to explore trends associated with season, number of individuals living in a compound, and number of animals present in a compound.

In Chapter 3 the relationships between indicator *E. coli*, microbial source tracking markers, select enteric pathogen genes, and potential sources of enteric pathogens in stored drinking water, soil, and on mother and child hands is evaluated. The association between indicator *E. coli* and select pathogen genes was used to assess the performance of indicator *E. coli* in the rural Bangladeshi household setting. Microbial source tracking markers as well as survey data on the number of fecal piles in compound courtyards were employed to determine specific sources of pathogen genes including those associated with pathogenic *E. coli*, *Giardia* and norovirus GII. These associations between indicators and pathogens were assessed by determining prevalence ratios and conditional probabilities such as negative predictive value, positive predictive value, sensitivity, and specificity.

The microbial community of different environmental reservoirs was determined using 16S rRNA gene sequencing to evaluate shared microbial communities between stored water,

mother hands, child hands, and soil in Chapter 4. SourceTracker, a computational modeling tool, was used to estimate the relative contribution of child and maternal feces to the microbial community in these household reservoirs. I also investigate intra-household shared microbial communities using bootstrap iterations to quantify the number of bacteria that were identical in reservoirs in the same household, in comparison to other households.

The main conclusions of this dissertation are presented in Chapter 5. The findings are discussed in the context of the broader sanitation field and future research priorities are outlined.

Chapter 2

Effect of Sanitation Improvements on Pathogens and Microbial Source Tracking Markers

2.1 Introduction

Diarrheal illnesses cause nearly half a million deaths per year in children under five[1]. In South Asia, approximately 10% of all deaths in children under five are due to diarrhea[2]. These illnesses result from the transmission of enteric pathogens due to inadequate water, sanitation, and hygiene (WASH) practices. Pathogens can be transmitted from the feces of an infected host to a new host through the fecal- oral route, mediated through environmental reservoirs[8]. Specifically, transmission can occur through exposure to contaminated water, soil, hands, fomites, vectors (e.g., flies), and food. Drinking water treatment, sanitation, and hygiene interventions use different mechanisms to block these exposure pathways. Of these, sanitation has the potential to block the most upstream fecal- oral transmission pathways including feces to soil, feces to water, and feces to flies.

Both humans and animals can serve as hosts for enteric pathogens and can contribute to environmental fecal contamination. Recent studies highlight the importance of animal to human pathogen transmission pathways, which until recently have been largely ignored[10, 19, 40]. Globally, animals produce four times the amount of fecal waste as humans, which is concerning in low-and middle-income countries where animal husbandry in the household environment is common[41]. Livestock excrete pathogens such as *Giardia*, *Campylobacter*, and EHEC/STEC and the total number of pathogens excreted by infected animals can be much greater than that excreted by humans[12, 42]. Due to the potential hazard posed by animal feces, it has been suggested that animal fecal management (e.g. separating animals from living spaces, removal of animal feces from household environment with tools, and reducing the movement of animals) be integrated with human fecal management (e.g. latrines)

to broaden the scope of sanitation interventions[40].

Determining the levels of fecal contamination in environmental reservoirs, and the impact of sanitation, can provide more insight as to which exposure pathways are most important for pathogen transmission and provide a mechanistic understanding of how sanitation interventions are effective, or not, at reducing the spread of enteric pathogens. Existing studies have investigated the impact of sanitation on health outcomes but few have quantified the impact of sanitation on indicators of fecal contamination along exposure routes (soil, hands, water, food, fomites, flies)[20, 43–47]. Notably, fewer studies have measured the impact of sanitation on the presence of specific pathogens in these reservoirs[21].

In this study we quantify the impact of a sanitation intervention (a combined human and animal intervention which included dual pit latrines, sani-scoop, and child potties) on environmental reservoirs delivered as part of a randomized controlled trial (WASH Benefits) in rural Bangladesh. Specifically, we measured pathogen genes and microbial source tracking markers in soil, stored drinking water, and on caregiver and child hands in a subset of enrolled compounds to assess the impact of the intervention on: 1) prevalence of pathogen genes; 2) prevalence of human fecal markers; and 3) prevalence and concentration of animal fecal markers in the domestic environment. This chapter complements a previously published paper from the same field trial, which investigated the prevalence and concentration of indicator *E. coli* in all households enrolled in the control and sanitation arms[46]. Chapter 2 of this dissertation documents the relationship between indicator *E. coli*, microbial source tracking markers, and pathogen genes in the same subset of households sampled as this chapter. Future publications are planned that will report on the prevalence and concentration of indicator *E. coli* in a subset of sanitation and control households at multiple time points over two years to assess the long-term effect of sanitation.

2.2 Methods

2.2.1 Data Collection

Trained field staff from the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) visited households between March and October 2015. They collected environmental samples, interviewed the female caregiver of young children regarding household practices related to sanitation and handling of human and animal feces, and conducted environmental spot checks.

2.2.2 Sample Collection

We collected approximately 720 stored drinking water samples, 720 soil samples, 720 mother hand rinses, and 360 child hand rinses from 600 study households (some households were sampled twice, approximately four months apart to capture both the wet and dry seasons).

Hand rinse samples were collected by mothers placing their left hand into a sterile Whirlpak bag (Nasco, Modesto, CA) filled with 250 mL of distilled water. The hand was massaged from the outside of the bag for 15 seconds, followed by 15 seconds of shaking. The same procedure was repeated with the right hand in the same bag. To collect child hand rinse samples, respondents placed their child's hand into a separate Whirlpak bag a followed the same procedure. Soil samples were collected from a 30 x 30 cm^2 area as close to the house entrance as possible by scraping the top layer of soil within a stencil into a sterile Whirlpak bag using a sterile disposable plastic scoop. The sample area was scraped both vertically and horizontally. Stored water samples were collected by asking mothers to provide a glass of water as they would give it to their child under five. The provided water was poured into a sterile Whirlpak bag. All samples were transported to the icddr,b field laboratory on ice and processed within 12 hours of collection.

2.2.3 Detection of Pathogenic *E.coli* Genes

Pathogenic *E. coli* genes were identified using previously published methods[48, 49]. IDEXX Colilert-18 was used to determine samples that contained *E. coli*. All samples positive for *E. coli*, were archived for subsequent pathogenic *E. coli* gene analysis. Broth from up to 20 positive large wells was aseptically extracted from IDEXX trays, composited and centrifuged. Pellets were treated with 10x the pellet volume ($\approx$ 0.1 mL) of RNAlater (Qiagen, Germantown, MD), stored at -80 °C and transported to UC Berkeley at room temperature. Each lab technician analyzed one lab blank per week for the archiving process by archiving wells from IDEXX trays incubated with sterile distilled water. Pellets were stored at -80 °C upon arrival at UC Berkeley. DNA was extracted from bacteria pellets using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD). Multiplex reactions were used to detect seven *E. coli* virulence genes indicative of five possible pathotypes of *E. coli*: EAEC (*aggR*), EPEC (*eaeA*), EHEC (*stx1/ stx2/eaeA*), EIEC (*ipaH*), ETEC (*lt1/st1b*) (Table A.1). PCR products were run on 2% agarose gels at 110 V for 30 min. Additional information on DNA extraction and PCR protocols is available in Appendix A.

2.2.4 Filtration and Nucleic Acid Extraction for qPCR Targets

Laboratory technicians pre-processed samples in the field laboratory by filtering 50 mL of hand rinse samples and up to 500 mL of stored water (range:100-500 mL) through a 0.45 µm HA filter (Millipore, Burlington, MA) to capture bacteria and viruses. 0.5 mL of 2.5M $MgCl_2$ was added to every 50 mL of sample to increase retention of viruses[50]. Filters were treated with 0.5 mL of RNAlater, stored at -80 °C and transported to UC Berkeley at room temperature. Once per day, each lab technician processed a lab blank by filtering 5 mL of sterile distilled water with 0.5 mL of $MgCl_2$, followed by the addition of RNAlater. Five grams of soil were weighed and stored -80 °C until transport at room temperature to UC Berkeley in accordance with a USDA soil import permit (PPQ 525). Filter and soil samples were stored at -80 °C at UC Berkeley. DNA and RNA were extracted from filter and soil

samples using modified Mobio PowerWater and PowerViral (both now Qiagen, Germantown, MD) protocols, described in previous studies and in the supporting information[20, 37]. An extraction blank was included in each batch of 23 samples. Extraction efficiencies for DNA and RNA were determined in a subset of samples using *Pseudomonas syringae* pv. phaseolicola (pph6) and MS2 (Table A.2).

2.2.5 qPCR Assays

Samples were analyzed for norovirus GII, *Giardia lamblia*, *Cryptosporidium spp.* genes, and microbial source tracking markers using quantitative PCR (Table A.3). Specifically, to detect human fecal contamination we used the HumM2 assay which targets a gene for a hypothetical protein in human-associated *Bacteroidales*-like microorganisms[51, 52]. We obtained a research license (#864-15) from the U.S. Environmental Protection Agency to use the patented HumM2 assay. To detect non-human, animal-associated fecal contamination we used the BacCow assay which, although designed to target the 16S rRNA of *Bacteroidales* associated with ruminants, has been shown to detect *Bacteroidales* in cows, ducks, goats, and chickens but not humans in rural Bangladesh[20, 53]. The microbial source tracking markers were previously evaluated for sensitivity and specificity in rural Bangladesh[20]. BacCow was sensitive but not specific to ruminant feces and HumM2 performed the best out of all tested human-associated assays (HumM2, HF183 and BacHum), although it also amplified in the feces of chickens and goats[20].

Approximately 100 stored water, soil and mother and child hand rinse samples were processed for norovirus GII, *Giardia* and *Cryptosporidium* genes. Among this subset, <1 % of soil and stored water samples were positive for norovirus and*Giardia* genes. We therefore decided to analyze only hand rinse samples for norovirus and *Giardia* genes. Less than 1% of all environmental sample types were positive for the *Cryptosporidium* gene, consequently we did not continue to analyze for the *Cryptosporidium* gene in any sample type.

All samples were run in triplicate on a 96-well plate (Applied Biosystems, Foster City, CA) on a StepOnePlus thermal cycler (Applied Biosystems, Foster City, CA). Each run contained a standard curve and three no template controls. A subset of samples were tested for inhibition using the spike-and-dilute method[54]. We did not dilute any samples for any of the pathogen and indicator assays based on our results of inhibition testing (Tables A.4 and A.5). Standard curves for each assay were determined by pooling all results and using a linear mixed effects model (Table A.6)[55]. We used the curves from the linear mixed effects models to determine Ct values in samples for all assays. Assays and qPCR methods are described in more detail in Appendix A.

2.2.6 Data Analysis

The analysis plan for this research was pre-specified prior to data analysis and is publicly available on Open Science Framework (https://osf.io/xrbpz/) and was independently replicated by two researchers.

Samples were considered positive for qPCR targets if at least 1 of 3 replicates amplified, even if amplification was below the limit of quantification (BLOQ). The limit of quantification (LOQ) was defined as 10 gc/2 μL and corresponds to the lowest point on the pooled standard curve. The limit of detection (LOD) was based on the lowest gene copy that amplified in at least 1 of 3 replicates in each sample type. BacCow was the only quantitative assay and for quantitative analysis hand rinse samples below the limit of detection (BLOD) were assigned half the value of the LOD. Hand rinse sample BLOQ were assigned the midpoint between the LOD and LOQ.

To determine which quantities to assign to samples BLOD and BLOQ in stored water and soil samples, which were impacted by differences in processing volume and moisture content, we conducted a sensitivity analysis with four scenarios: 1.) Non-detects were assigned half the LOD accounting for the volume of water filtered and the moisture content for each sample. Samples BLOQ were assigned the midpoint between the LOD and LOQ. 2.) Non-detects were assigned the lowest detectable value. Samples BLOQ were assigned the lowest quantifiable value. Lowest detectable values and quantifiable values correspond to the maximum volume of water filtered (500 mL) and minimum soil moisture content (0%). 3.) Non-detects were assigned the LOD accounting for the volume of water filtered and the moisture content for each sample. Samples BLOQ were assigned the LOQ accounting for the volume of water filtered and the soil moisture content. 4.) Non-detects were assigned half the lowest detectable value. Samples BLOQ were assigned the midpoint between the lowest detectable value and lowest quantifiable value. The overall results for change in concentration of BacCow between arms were similar in all scenarios (Table A.7). Therefore, we conducted the primary analysis by assigning non-detects half the LOD accounting for the volume of water filtered and the moisture content for each sample. Samples BLOQ were assigned the midpoint between the LOD and LOQ (Scenario 1).

The LOD of BacCow was 285 gene copies (gc)/2 hands, 8.1-40.5 gc/100 mL stored water (depending on the volume of water filtered) and 130-244 gc/g dry soil weight (depending on the soil moisture content). The LOQ of BacCow was 2500 gene copies (gc)/2 hands, 100-500 gc/100 mL stored water and 2000-3760 gc/g dry soil weight. No samples were above the upper limit of quantification for BacCow (defined by the highest point on the standard curve).

Most samples did not amplify within the quantifiable range for norovirus GII, *G. lamblia* and HumM2. For binary analyses on pathogen/source tracker marker presence, all samples with positive lab, extraction, or archiving blanks were removed from the analysis according to

the date processed and lab technician. For BacCow, the blanks that amplified were BLOQ; samples that amplified in the same region (BLOQ) were treated as BLOD.

The impact of the sanitation intervention was assessed using generalized linear models with robust standard errors to account for the trial's clustered study design[39]. For binary outcomes, we determined the prevalence, prevalence differences (PD) and prevalence ratios (PR) for pathogen genes and microbial source tracking markers. For BacCow, the only continuous marker, we determined the mean concentration in each arm as well as the change in concentration between the arms. The input values for BacCow concentrations were $\log_{10}$ transformed.

Randomization in the parent trial resulted in balanced covariates between arms, thus we used the unadjusted models as our primary analysis[39]. We conducted three, pre-specified, subgroup analyses to estimate the parameters of interest (mentioned above) in a subset of the data. Datasets were separated by season (wet vs. dry), number of individuals living in a compound (<10 vs. ≥ 10), and number of animals in the compound (<20 vs. ≥ 20). In a deviation from the analysis plan, the wet season was defined as June-September to achieve better balance in sample numbers in the subgroups and reflect the actual rainfall in 2015. In a secondary analysis we adjusted the outcomes, excluding subgroup analysis. We investigated the same covariates as described in the main study's primary outcome plan with the addition of sampling condition covariates for each sample type. We corrected for multiple comparisons (i.e., the same marker evaluated in multiple sample types) using a Bonferroni correction[56].

2.3 Results

2.3.1 Prevalence of Pathogen Genes and Microbial Source Tracking Markers

The prevalence of each *E. coli* virulence gene was consistently higher on child hands than mother hands (Figure 2.1 and Table A.8). On child hands in both study arms, the most prevalent ECVG was *stx1/2* (Control (C): 13.4 (8.6-18.3) %, Sanitation (S): 16.9 (11.2-22.5) %). On mother hands, the most prevalent ECVGs were *eaeA*, *aggR*, and *stx1/2* genes (C and S: 6.1-9.3%). The prevalence of each *E. coli* virulence gene was consistently higher on child hands than mother hands. The prevalence of norovirus and *G. lamblia* genes was 1.7-5.1% in all hand samples in both arms. In other household reservoirs, ECVGs were present in 37.5 (33.0-42.1) % and 36.4 (31.3-41.6) % of stored water and 61.5 (56.4-66.6) % and 58.9 (53.9-64.0) % of soil in the control and sanitation arms, respectively (Figure 2.2 and Table A.8). The most commonly detected ECVG was *stx1/2* (C: 16.6 (12.6-20.6) %, S: 16.5 (12.6-20.4) %) in stored water and *eaeA* in soil (C: 38.6 (33.2-44.1) %, S: 36.7 (31.7-41.6) %). The least prevalent ECVG across all sample types was *ipaH* (<2 %).

Human and animal microbial source tracking markers were detected in all four reservoirs. The prevalence of BacCow was higher than any other target including both indicators and pathogens. 99.6 (98.4-100.0) % and 95.6 (92.4-98.9) % of child hands in the control and sanitation arms were positive for BacCow and a similar prevalence was also found on mother hands (C: 97.6 (96.0-99.1) %, S: 95.8 (93.9-97.7) %). BacCow was detected in 70 (65.1-74.8) % and 67 (61.6-72.3) % of stored water and in 91.2 (88.4-94.1) % and 90.2 (86.1-94.2) % of soil in the control and sanitation arms. The prevalence of the human fecal marker (HumM2) was substantially lower than the animal marker in all reservoirs. In the control and sanitation arms, the prevalence of HumM2 on child hands was 26.4 (19.7-33.0) and 17.7 (12.0-23.3) % and a similar level was found on mother hands (C: 18.5 (13.8-23.1) %, S: 17.7 (13.3-22.1) %). We detected HumM2 in <5% of stored water samples (C: 3.6 (1.5-5.6) %, S: 1.6 (0.2-2.9) %) and in approximately 20% of soil (S: 18.4 (14.3-22.5) %, C: 22.0 (17.3-26.6) %). Hand rinses, stored water, and soil samples have different units (per 2 hands, 100 mL stored water, and g of dry soil) therefore it is difficult to draw meaningful conclusions regarding the different prevalence between environmental sample types.

Considering that stored drinking water is a direct exposure route to ingestion of enteric pathogens, the high prevalence of BacCow (C: 70%, S: 67%) and *E. coli* virulence genes (C: 37.5%, S: 36.4%) in stored water in both study arms is a concern. As we did not measure pathogen genes or BacCow in tubewells (source water) we were unable to differentiate contamination in the tubewells from contamination introduced during storage. While it is possible that the fecal contamination in stored drinking water originated from the source water, as other studies have found indicators of fecal contamination and pathogen genes in these sources[57–60], we believe it is unlikely to be the main source of BacCow and *E. coli* virulence genes. In the WASH Benefits study, Ercumen et al. found that the prevalence of indicator *E. coli* was significantly higher in stored water (C:77% , S: 76%) than was found in tubewells (C: 25%, S: 23%)[46, 61]. One possible mechanism of water quality degradation during storage is through contact with hands,[22, 62, 63] which we have shown to be contaminated with multiple types of pathogenic *E.coli* virulence genes (Figure 2.1).

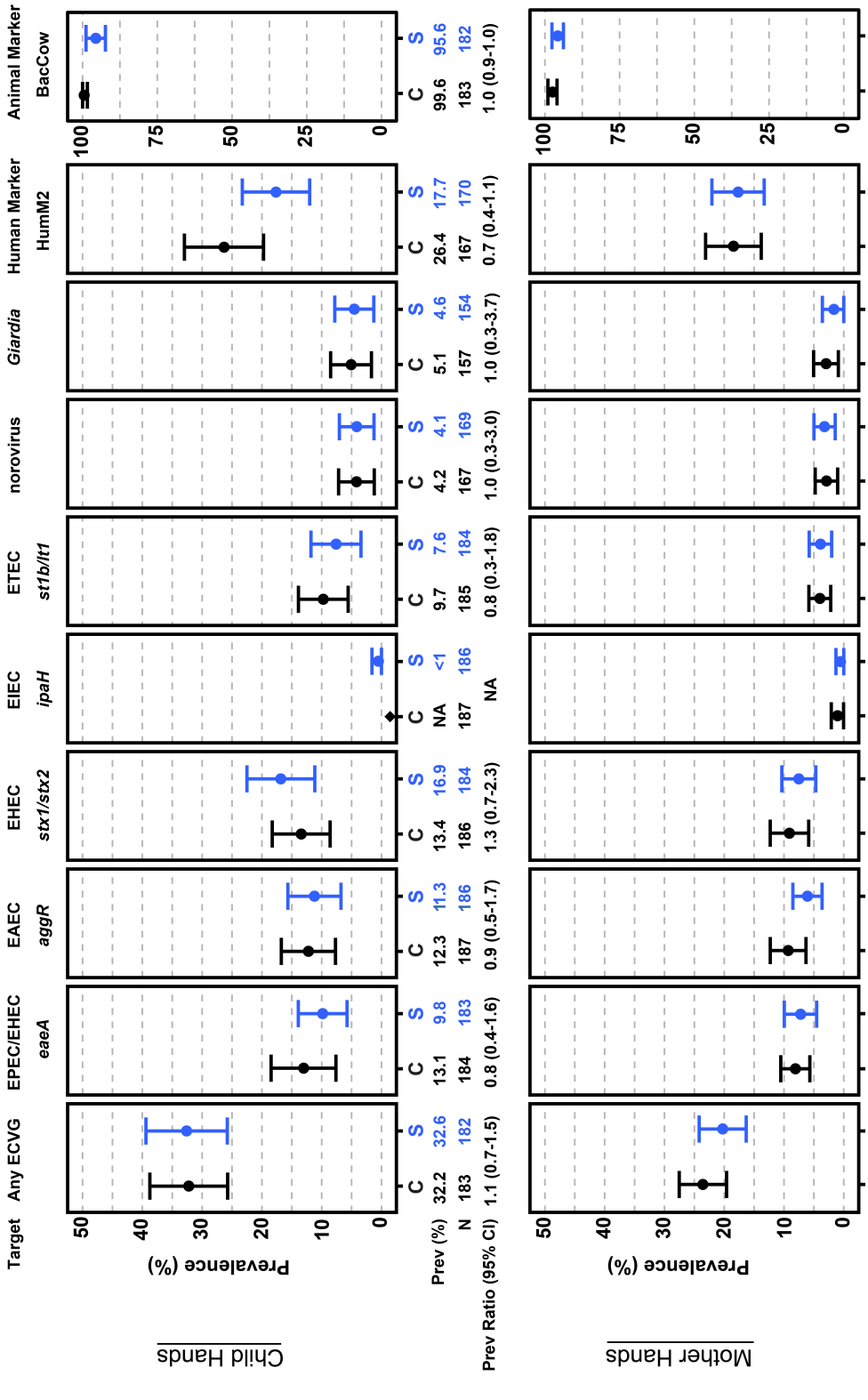


Figure 2.1: Prevalence (95% CI) and prevalence ratio (95% CI) of pathogenic *E. coli* virulence genes, norovirus GII gene, *Giardia lamblia* gene and microbial source tracking markers in mother and child hands in the sanitation (S) and control (C) arms.

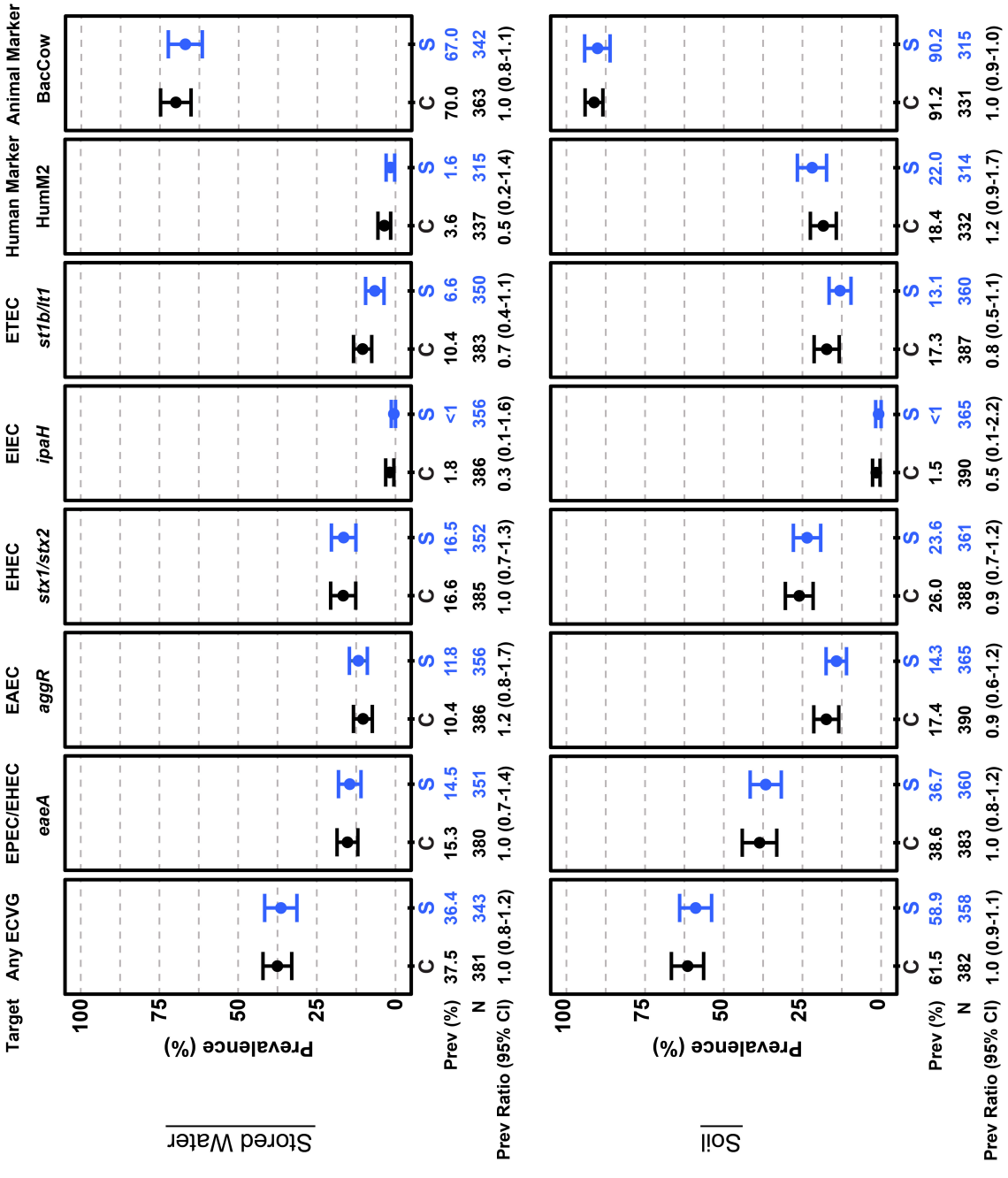


Figure 2.2: Prevalence (95% CI) and prevalence ratio (95% CI) of pathogenic *E. coli* virulence genes and microbial source tracking markers in stored water and soil in the sanitation (S) and control (C) arms.

We observed a slightly lower prevalence of any ECVG on hands (≈ 32% on child hands and ≈ 20% on mother hands), in stored water (≈ 36 %), and in soil (≈ 60 %) than a previous study on household level fecal contamination in Tanzania where 41% of caregiver hands, 59% of stored water, and 72% of soil were positive for any ECVG[37, 49]. The prevalence of EIEC was higher in Tanzania (20% of hands, 32% of stored water, and 51% of soil) than our study (<2% in all sample types). The observed prevalence of pathogenic *E. coli* virulence genes could reflect actual differences in the prevalence of these pathogens between locations or could be due to slight differences in detection methods or seasonal variation. We used IDEXX to enumerate *E. coli* and sampled in March through October, whereas the Tanzania study used membrane filtration for enumeration and sampled in March.

Previous studies have also investigated the presence of microbial source tracking markers in the household environment in rural India. We found a higher occurrence of the same animal fecal marker (BacCow) in hand rinses and stored water in rural Bangladesh than was observed in Odisha, India[21]. BacCow was present in 32.7% of stored drinking water and 71.5% of hands (combined mother and children) in India compared to approximately 70% of stored drinking water and >95% of hands in our study. However, the prevalence of the human fecal marker was higher in stored drinking water in India (19.4% BacHum compared to <5% HumM2 in rural Bangladesh)[21]. We found a similar level of human fecal markers on hands (18.0% BacHum in India and 17.7-26.2% HumM2 in our study). The reported 7-day diarrhea prevalence among children <5 years in the India study (C: 9.1% , S: 8.8%)[44] was slightly higher compared to our study (C: 5.7% , S: 3.5%)[39], which could account for the difference in the prevalence of human marker in stored water between the studies but the same trend is not evident on hands. Moreover, different human fecal markers were used in India and Bangladesh, making it difficult to compare across studies. The sensitivity and specificity of human makers varied geographically and the best performing marker was selected for each location[20, 64].

A comparable study of rural Bangladeshi households found lower prevalence of human fecal contamination using the same human marker (HumM2)[20]. This study was also conducted in a subset of WASH Benefits households in the sanitation and control arms but samples were collected at an earlier date (November 2013-March 2014). In the control arm, 9.0% of soil samples, 0% of stored drinking water, and 2.0% of child hands were positive for HumM2 compared to 18.4% soil, 3.6% stored water, and 26.4% of child hands in our study. Taken together, this indicates a high degree of temporal variability in human fecal contamination. These studies sampled in different years as well as months; the previous study was conducted during the dry season of Bangladesh while our sampling coincided with the both seasons. Therefore differences in the occurrence of HumM2 could be due to both seasonal and yearly variation. Previous studies have found higher levels of fecal contamination in various environmental compartments during the wet season in Bangladesh[10]. In both studies, the prevalence of viruses was low in child hands (<10%) although the studies investigated different viruses. In the control and sanitation arms, rotavirus was detected in 5.7% and 6.1%

of child hands in the earlier study and norovirus was prevalent in 4.2 (1.2-7.2) % and 4.1 (1.3-7.0) % of child hands in our study.

2.3.2 Intervention Impact

Almost all compounds in both the sanitation and control arms contained a latrine (C: 98.7%, S: 100%) (Table 2.1). Most sanitation arm compounds had a pour flush latrine (C: 62%, S: 100%) with a functional water seal (C: 39% S: 97%). Only 3.4% of intervention arm latrines drained into the environment whereas over 20% of control households had latrines draining into the environment or feces spilling out from the pit. While ownership does not guarantee use, most adults (94%) in the sanitation arm used the latrine during structured observations in the main field trial[65]. There were also more compounds with young children defecating in latrines or potties (C 13.0%, S: 55.6%) and using a scoop or hoe to handle animal feces (C: 47.0%, S: 94.1%) in the intervention arm. Only 0.3% of compounds in both arms treated their drinking water.

Table 2.1: Household drinking water, sanitation, and animal hygiene practices (% of households).

	Control	Sanitation
Drinking Water	n=300	n=297
Cover on stored water container[a]	23.2	25.4
Reported treating water[b]	0.4	0.4
Sanitation		
Latrine present in compound[c]	98.7	100
With slab	96.3	99.7
Pour flush	63.4	99.7
Functional water seal	39.7	97.0
Flow into environment[d]	21.7	3.4
Visible feces	8.4	1.7
Feces odor	64.9	23.6
Urine odor	19.3	2.0
Pit emptied since last visit	16.9	3.4
Reported using latrine always		
Children <3 yrs	1.4	13.8
Men	84.5	90.0
Women	93.2	92.6
Reported children <5 defecating in potty or latrine	13.0	55.6
Human feces visible in courtyard	0.3	1.0
Reported using scoop or hoe to handle child feces	36.0	36.7
Animal Feces		
Animal feces visible in courtyard		
Chicken/non-chicken poultry	89.7	90.9
Cow	35.7	26.9
Goat/sheep	24.3	22.2
Pig	0	0
Dog or cat	<1	<1
Cow patty	11.0	7.4
Non-poultry birds[e]	4.7	3.7
Reported using scoop or hoe to handle animal feces	47.0	94.1

[a]:Of households with stored water. Some households obtained water directly from tubewells, taps or piped connections.
[b]:Treated with household water filter
[c]:Characteristics for primary pit latrine used in compound
[d]:Latrine drains into environment or feces spilling out from pit
[e]:Sparrows, crows, pigeons

There was no significant difference, where p-value <0.01, in the prevalence of pathogen genes and microbial source tracking markers between the control and sanitation arms across all sample types using the unadjusted models and correcting for multiple comparisons (Figure 2.1, Figure 2.2, and Table A.9). On child hands, there was a 4% reduction in the prevalence of the BacCow marker in the sanitation arm, but this association was borderline significant after correcting for multiple comparisons (prevalence ratio [PR]: 0.96 (95% CI: 0.93-1.00), p=0.05) (Figure 2.11 and Table A.9). However, this association was statistically significant, correcting for multiple comparisons, in the adjusted model (PR: 0.95 (0.92-0.97), p <0.001) (Table A.9). There was a borderline significant 33% reduction in the human fecal marker on child hands in the adjusted model (0.66 (0.44-0.99), p=0.05). There was also a borderline significant reduction in *st1b/lt1* in stored water (PR: 0.64 (0.39-1.03), p=0.07) and soil (0.74 (0.54-1.03), p=0.07) in the adjusted model. The average $\log_{10}$ concentration of BacCow on mother hands was lower in the sanitation arm in the unadjusted (Δ : -0.16 (95% CI: -0.30 to -0.03, p=0.02) and adjusted models (0.37 (-0.66 to -0.08), p=0.01) but these associations were not significant after correcting for multiple comparisons (Figure 2.3 and Table A.10).

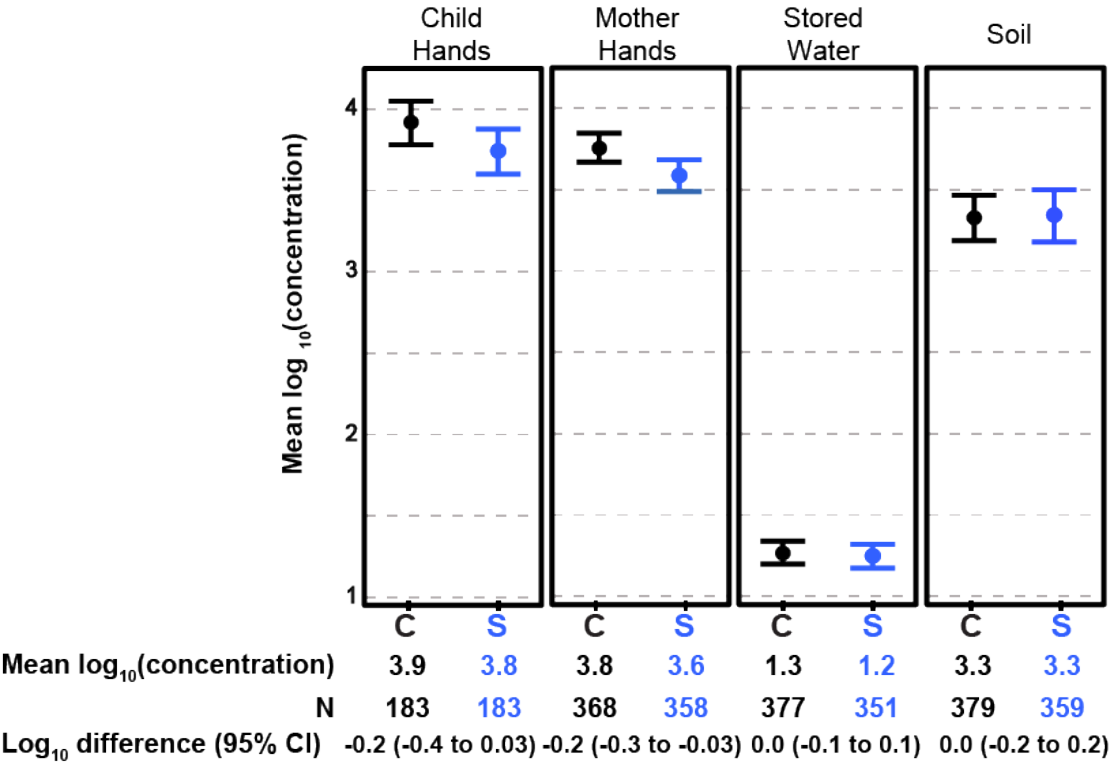


Figure 2.3: Mean $\log_{10}$ concentration of BacCow and change in concentration in hands, soil, and stored water in the sanitation (S) and control (C) arms. Concentrations are in units of gene copies per 2 hands, 100 mL, and dry gram.

In the subgroup analyses, there were no significant associations, correcting for multiple comparisons, separating the data by season or by the number of animals in compounds. (Table A.11 and A.12). In households with $\geq$ 20 animals in their compound, we observed borderline significant reductions in *st1b/lt1* on child hands (0.22 (0.04-1.06), p=0.04) and in soil (0.46 (0.23-0.92), p=0.05) (Table A.12). In households with $\geq$ 10 individuals in their compound, we observed an 89% reduction in the prevalence of *st1b/lt1* on child hands in the sanitation arm (0.11 (0.02-0.77), p <0.01) (Table A.13). We did not find any significant difference in $\log_{10}$ concentration of BacCow in any subgroup analyses, correcting for multiple comparisons, but there were some borderline significant effects on mother hands (Table A.14).

Overall, the improved pit latrines, child potties and sani-scoops in the WASH Benefits trial were not sufficient to reduce fecal contamination in the household environment, evident by the similar prevalence of pathogen genes and microbial source tracking markers in both study arms. This finding is consistent with the results of the indicator bacteria study which found no difference between indicator *E. coli* concentrations in tubewells, ponds, stored water, food, soil, and on child hands and fies in the sanitation arm compared to the control arm[46]. One possible explanation provided was that the observed concentrations of indicator *E. coli* were occluded by "naturalized" *E. coli* (*E. coli* naturally found in soil and water that is not derived from fecal sources), which masked the effect of the sanitation intervention. However, this explanation is unlikely given that we also did not find a difference in the prevalence of microbial source tracking markers and select pathogen genes between the sanitation and control arms. Rather, it is more likely that containment of feces was not significantly improved in the sanitation arm. Another suggested hypothesis from the indicator *E. coli* study was that human feces, but not animal feces, could have been impacted by the intervention. The authors reasoned that the indicator *E. coli* detected could have come predominantly from animal feces, which was still present in intervention households despite the provision of animal scoops. This is consistent with our finding that the animal fecal marker was substantially more prevalent than the human fecal marker among all sample types in our study. However, given that we saw no significant differences in the human fecal marker (HumM2) between the study arms, we demonstrate that human fecal contamination was not sufficiently impacted by the sanitation intervention. Similarly, there was no difference between the prevalence of the animal fecal marker (BacCow) between the two arms expect for a small difference in the prevalence on child hands, suggesting no intervention impact on animal fecal contamination.

The WASH Benefits study reported a lower prevalence of childhood diarrhea in households in the sanitation arm[39], while our study found no significant difference in the prevalence of pathogens or fecal indicators in environmental samples collected from a subset of these same households. We expected that because the prevalence of diarrhea was lower in the sanitation arm, containment of feces was better and therefore fewer pathogens would be present in the household environment. One possible explanation for why we did not measure a difference is that pathogen detection limits were too high, resulting in many false negatives,

and insufficient sensitivity to detect a difference between the two arms. In other words, our study may not have detected whether some of these samples (false negatives) contained pathogen concentrations sufficient to exceed the minimum infective dose. For example, WASH Benefits measured protozoan pathogens in the stool of study participants and found a 25% reduction in the prevalence of *Giardia* infections in the sanitation arm[66]. However, we did not observe a similar reduction in the prevalence of *Giardia* in environmental reservoirs. The concentration of *Giardia* in environmental samples is much lower than in the stool of infected individuals, which could have resulted in lower sensitivity to detect a difference in the environmental samples.

Another possibility that can explain the discrepancy between the prevalence of diarrhea and pathogens in environmental samples is that we did not measure all pathogens that could be etiological agents of diarrhea. Our study focused on pathogenic *E. coli*, *Giardia* and norovirus while rotavirus, *Shigella*, *Aeromonas*, and *Campylobacter jejuni* are other possible etiological agents of disease in Bangladesh[7]. Also, we did not measure pathogen genes and microbial source tracking markers on fomites or food which could be dominant reservoirs for enteric pathogen transmission[67]. It is important to note that 16-35 months, the time between latrine construction and sample collection, should have been sufficiently long for organisms and nucleic acids to degrade in environmental reservoirs, including soil[68–71]. Therefore, it is unlikely the genes detected in this study persisted from before the intervention. Moreover, the pathogenic *E. coli* genes detected were likely from viable organisms due to the enrichment step.

The detection methods used in this study have some limitations, and these could impact the measured impact of the sanitation intervention. The method we used for pathogenic *E. coli* was not quantitative, and a quantitative method such as qPCR could have determined if the concentrations of pathogenic *E. coli*, not just prevalence, were impacted in the trial. However, using qPCR does not guarantee that quantitative results will be produced, as occurred in this study. Despite using qPCR to detect norovirus, *Giardia*, and HumM2, most of the results were below the limit of quantification. Detection limits can be improved with the processing of larger sample volumes; however, available sample volumes are constrained in the household setting. For example, stored water samples are restricted to the volume that is being stored and that can be spared at the time of collection. More effort is needed to develop methods with lower detection limits compatible with complex environment samples. While there have been recent developments in aptamer based detection, nucleotides that bind to a target, with potential for lower detection limits, very few have been tested in environmental matrices[72].

Our finding that we did not observe consistent reductions in household fecal contamination is consistent with a previous study in rural India that found no reduction in pathogens and microbial source tracking markers in household reservoirs despite a 27% increase in latrine coverage[21]. One notable exception in our study was a slightly lower (5%) prevalence of

BacCow on child hands in the sanitation arm, although the prevalence was still high (100 versus 96% in control and sanitation arms, respectively). However, the lower prevalence of BacCow was not reflected in a difference in $\log_{10}$ concentration of BacCow on child hands. The lower prevalence of BacCow could be due to the use of sani-scoops for animal feces removal in 94% of the sanitation arm. We observed less cow feces in sanitation compound courtyards (C:36 %, S:27%), although poultry feces were still visible in 91% of courtyards (Table 2.1). We also observed a 89% lower prevalence of *st1b/lt1* genes, indicative of ETEC, in sanitation arm households with $\geq$ 10 individuals (Table A.13). It is possible that the effect of the intervention was greater in households where more human feces were produced and therefore more human feces were isolated from the environment with the introduction of improved latrines. However, we were unable to differentiate human-specific ETEC in our study and we did not observe the same trend in the prevalence of other human specific pathogen genes (e.g. *aggR*) or human fecal markers.

Many of the main reasons we likely did not see a difference in fecal contamination in household reservoirs between the control and sanitation arms are discussed in Ercumen et al., including poor child and animal feces management even in the intervention households, and the scale of the intervention being the compound level, rather than community level[46]. Additionally, the lack of improvement in household fecal contamination despite a high level of compliance and usage of latrines indicates the quality of the latrine may not have been sufficient to isolate fecal contamination from the household environment[65]. The latrine improvements in this study most closely resemble a scenario in which latrines are upgraded from unimproved to improved latrines as defined by the Millennium Development Goals[73]. Previous research suggests that improved sanitation alone does not adequately reduce fecal contamination in the environment[74, 75]. This shortcoming is reflected in the sustainable development goals where sanitation targets have shifted to include fecal sludge management[24, 76]. Fecal sludge management may reduce the potential for latrine contents to leak into the environment, which could reduce fecal contamination in groundwater, soil, and other environmental reservoirs.

In this study the access to latrines between both study arms was high (98.7 and 100% in control and sanitation arms, respectively). If there had been a greater difference in latrine access between the two arms, it is possible that we may have observed a difference in the prevalence of indicators and pathogen genes. To answer this question, Ercumen et al. recommended investigating the impact of varying levels of latrine coverage and compliance on fecal contamination[46]. We suggest a complementary approach in which modeling is used to estimate the impact of coverage and compliance on the concentration of pathogens in the environment and prevalence of diarrhea in children under five. Quantitative microbial risk assessment, including exposure models[67, 77], can be combined with pathogen measurements to predict the number of pathogens children are exposed to in their household environment and estimate the resulting prevalence of diarrhea. Our study, as well as others on child behavior and interaction with the environment in Bangladesh[78], provide a founda-

tion for exposure models to investigate fecal contamination pathways that result in pathogen ingestion by children in these households. The impact of compliance and coverage can be studied by altering the level of pathogen exposure according to different tiers of sanitation access and use. Previous work has modeled the reduction in infection risk due to different intervention compliance scenarios on household level water treatment[79–81], but we are unaware of studies that model the impact of varying levels of compliance with sanitation improvements. Modeling the impacts of sanitation coverage level and compliance could help determine the degree of intervention that is needed to significantly reduce fecal contamination levels in the household environment. Finally, this study was not designed to measure fecal pathogen levels in potential reservoirs outside of the household environment to which children may be exposed such as open drains, fecal sludge dumping sites, and open defecation sites. Additional research is needed to better understand exposure in public versus private realms[67].

Chapter 3

Predictors of Enteric Pathogens in the Domestic Environment from Human and Animal Sources

3.1 Introduction

In low-and middle-income countries diarrheal illness is a leading cause of morbidity and mortality[1]. In Bangladesh, 6% of the 129,000 deaths in children under five in 2013 were attributed to diarrheal diseases[82]. Diarrheal illness results from exposure to fecal pathogens which can be transmitted from feces to a new human host through a variety of environmental pathways, including fingers, fields, flies, fluids, and food, described in the F-diagram[8]. Recent additions to the F-diagram stress the importance of animal hosts by expanding fecal sources to include feces from livestock, free-roaming animals and synanthropic rodents[40]. While it is well known that enteric pathogens are transmitted through these pathways, few studies have measured pathogens to characterize exposure from different animal reservoirs.

There is a high potential for zoonotic enteric disease transmission in low-and middle-income countries where animal husbandry is a primary source of income[12]. For example, in Bangladesh raising livestock such as cows, goats and chickens results in animals roaming freely within the home environment[83]. Close proximity to domesticated animals can lead to human exposure to livestock feces. Many pathogens can be transmitted from animal feces to human hosts and result in diarrheal illnesses[12]. Of the four most common etiological agents of moderate to severe diarrhea in children 0-11 months in Bangladesh, two (*Cryptosporidium* and *Campylobacter*) are known to have important animal reservoirs[7, 11].

Measuring enteric pathogens in the environment can help identify reservoirs and potential exposure pathways, which can better inform design of engineered controls to reduce human exposure. However, there are many different fecal pathogens from humans and animals

capable of causing disease, making it infeasible to measure them all, especially given that most are difficult to measure in the environment due to their low concentrations in environmental matrices and costly and complex methods of detection[84]. Therefore, fecal indicators are commonly used to indicate the presence of fecal contamination, which may contain pathogens[84]. Thermotolerant coliforms and *Escherichia coli* are recommended indicators of water quality by the World Health Organization[28]. Other fecal indicators such as *Bacteroidales* are used for identifying sources of fecal contamination through microbial source tracking (MST)[36]. However, the degree to which indicators and pathogens co-occur in environmental samples varies depending on their concentrations in the original fecal source (pathogen concentrations depend on the infection status of the human or animal whereas indicator organism concentrations are expected to remain more stable) and the relative transport and die-off/growth rates of organisms once they are deposited in the environment. Thus, the relationship of indicator organisms and specific pathogens in or on different environmental matrices (e.g., water, soil, hands, fomites, food) may vary spatially and temporally depending on these factors[32]. Nonetheless, a better empirical understanding of the relationship between indicators and pathogens in specific environmental reservoirs and contexts in low-and middle-income countries may improve the ability to estimate human health risk and identify fecal sources and exposure pathways of greatest concern. Furthermore, pinpointing dominant sources and reservoirs will help improve the design of targeted interventions that can reduce exposure to fecal pathogens, mitigating the burden of diarrheal illnesses.

The aim of this study was to measure fecal indicators and select human pathogen genes in different reservoirs in the domestic environment to: 1) determine the association between indicator *E. coli* and pathogen genes; and 2) use microbial source tracking and spot-check observations of animal feces in compounds to investigate potential human and animal sources of detected pathogen genes.

3.2 Methods

3.2.1 Pathogen Selection

We selected multiple types of microorganisms that are leading causes of diarrheal illness in developing countries and have a range of host specificity between humans and animals. The selected targets were indicators (*E. coli*, human-associated *Bacteroidales*-like gene, and non-human animal-associated *Bacteroidales* gene) and pathogen genes (enteropathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC), norovirus GII, *Giardia lamblia*, and *Cryptosporidium spp.*).We measured these targets in different reservoirs in the domestic environment within rural Bangladeshi compounds - stored water, soil, and mother and child hands (Figure 3.1).

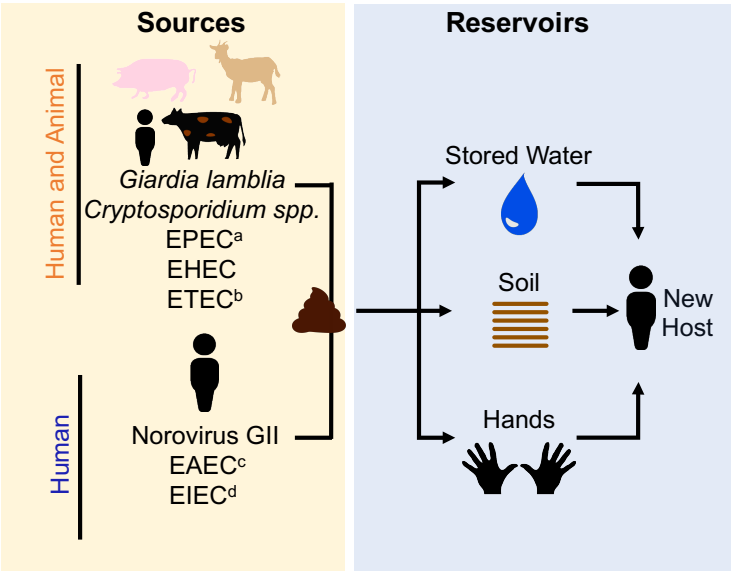


Figure 3.1: Pathogens associated with human and animal origin in different reservoirs investigated in this chapter.
[a]:Atypical enteropathogenic *E. coli* (EPEC) is transmitted by humans and animals, typical is transmitted by humans only
[b]:Enterotoxins found in enterotoxigenic *E. coli* (ETEC) can be from humans and animals but species-specific adhesion factors confer host specificity
[c]:*AggR* gene in enteroaggregative *E. coli* (EAEC) only found in strains isolated from humans
[d]:Enteroinvasive *E. coli* (EIEC) has been found in humans and primates, primates are not relevant to this study

We reviewed current literature to determine which of the selected pathogens had the potential to originate from both human and animal reservoirs or only human reservoirs. Pathogens associated with both human and animal feces include *Cryptosporidium*, *Giardia lamblia*, EPEC, EHEC and ETEC[42, 85, 86]. Humans serve as a reservoir of typical EPEC, while animals such as pigs and chickens can serve as reservoirs of atypical EPEC[86, 87]. Though typical and atypical EPEC can be differentiated by the presence of the gene *bfpA*, this study only detected *eaeA* and will therefore not distinguish between these strains[88]. ETEC is also found in both humans and animals but species-specific adhesion factors confer host specificity[11]. We were unable to differentiate ETEC with human adhesion factors in this study because the enterotoxins we detected, *st1b* and *lt1*, can be found in both humans and animals[89–92]. Previous studies have isolated *Giardia lamblia* and *Cryptosporidium spp.* in many fecal sources and identified closely related isolates from humans and animals[85, 93]. On the other hand, norovirus GII, EAEC and EIEC are associated with primarily human sources, largely based on genetic analysis of pathogens in human and animal fecal samples[86, 94, 95]. While EAEC is also found in animals, the gene we detected, *aggR*, has been identified in EAEC isolated from humans but not animals[95, 96]. EIEC has also been found in primates, which are unlikely to be important contributors to fecal pollution in the study communities[86]. It should be noted that the associations between pathogens and possible hosts in Figure 3.1 are based on current scientific understanding and evidence.

Zoonotic pathogen transmission is difficult to demonstrate and we are limited by the methods that have been employed and the pathogens, virulence genes, locations, and animal hosts that have been investigated.

3.2.2 Study Setting and Data Collection

The study setting of 600 households enrolled in the control and sanitation arms of WASH Benefits Trial is described in detail in Chapter 2. Field staff from the icddr,b collected environmental samples, interviewed the female caregiver of young children regarding household practices on sanitation and handling of animal feces, and conducted environmental spot-checks as outlined in Chapter 2. For this study, field surveyors also recorded the number and type of animals owned by the household as reported by the caregiver and observed the number and type of animal fecal piles in compound courtyards.

3.2.3 Molecular Methods

E. coli were enumerated (100 mL sample volumes) using IDEXX Colilert-18 (IDEXX Laboratories, Westbrook, Maine). Trays were incubated at 44.5 °C for 18 hours[97]. To prepare for IDEXX analysis, hand rinse samples were diluted 1:2 with sterile distilled water in Whirpak bags while stored water was analyzed undiluted. 20 g of soil was homogenized with 200 mL of distilled water and diluted $1{:}10^4$. Soil was dried at 110 °C for 24 hours to determine moisture content. Each lab technician processed one blank (sterile distilled water) for *E. coli* enumeration per day.

The method for identifying pathogenic *E. coli* genes, filtering and archiving samples, extracting nucleic acids, and qPCR are described in Chapter 2.

3.2.4 Data Analysis

Samples below the limit of detection of *E. coli* were set to 0.5 MPN and those above the limit of detection were set to 2420 MPN, which corresponds to a lower detection limit of 5 MPN/2 hands, 1 MPN/100 mL of stored water and 1000-1800 MPN/g dry soil (depending on the soil moisture content) (Table 3.1).

qPCR target limits of detection and limits of quantification are described in Table 3.1 and Chapter 2.

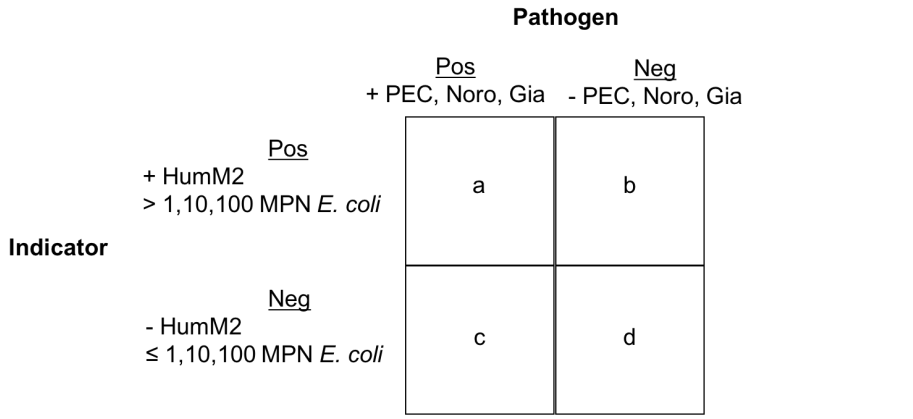
Data were analyzed using R (version 3.5.0). The association between indicators and pathogens was assessed using generalized linear models with robust standard errors to account for the trial's clustered study design[39]. The input values for *E. coli* and BacCow concentrations were $\log_{10}$- transformed. Adjusted models controlled for the effects of month of sample collection and study arm. Conditional probabilities (positive predictive value (PPV), negative

Table 3.1: Units, limits of detection, and mean concentrations in hand rinse, stored water and soil samples.

	Child Hands	**Mother Hands**	**Stored Water**	**Soil**
Sample matrix unit	2 hands	2 hands	100 mL	1 g dry weight
E. coli lower limit of detection[a]	5	5	1	1000-1880[d]
E. coli upper limit of detection[a]	12098	12098	2419	2.42×10^6 - 4.55×10^{6}[d]
n<lower limit of detection	100	183	93	39
n>upper limit of detection	7	17	26	97
Log_{10} mean concentration of *E. coli*[a]	**1.38**	**1.47**	**1.15**	**5.14**
BacCow lower limit of detection[b]	285	285	8.1-40.5[c]	130-244[d]
BacCow lower limit of quantification[b]	2500	2500	100-500[c]	2000-3760[d]
n<lower limit of detection	25	9	242	97
n<lower limit of quantification	218	79	414	281
Log_{10} mean concentration BacCow[b]	**3.84**	**3.67**	**1.25**	**3.34**

[a]:In units of MPN/sample matrix
[b]:In units of gene copies/sample matrix
[c]:Corresponds to volumes of 100 and 500 mL filtered
[d]:Corresponds to soil moisture contents ranging from 0 to 88%

predictive value (NPV), sensitivity and specificity) to assess the ability of indicators to correctly predict pathogen gene presence were calculated for binary data and for continuous data by binning continuous data at a specified cutoff point (Figure 3.2). We investigated threshold trends in stored water associated with the WHO water quality guidelines by evaluating PPV, NPV, sensitivity and specificity of pathogen presence and indicator *E. coli* concentrations above cutoff values corresponding to intermediate risk (MPN/100 mL > 1), high risk (MPN/100 mL > 10) and very high risk (MPN/100 mL > 100).



Sensitivity: Ability of an indicator to correctly identify the presence of a pathogen (a/(a+c))
Specificity: Ability of an indicator to correctly identify the absence of a pathogen (d/(d+b))
Positive Predictive Value: Likelihood there is a pathogen, given positive indicator results (a/(a+b))
Negative Predictive Value: Likelihood there is not a pathogen, given negative indicator results (d/(c+d))

Figure 3.2: Contingency table for conditional probabilities (sensitivity, specificity, positive predictive value, and negative predictive value) calculated in this chapter.

The association between the presence of animal fecal piles within the compound courtyard and presence of pathogens was also assessed using generalized linear models with robust standard errors. We assessed the association by binning animal feces into two categories for each animal fecal type. For cow feces, goat/sheep, cow patty, and non-poultry birds the prevalence ratio represents the prevalence of pathogen genes in households with any visible fecal piles (> 0) compared to those with no fecal piles. For chicken/non-chicken poultry feces the prevalence ratio represents the prevalence of pathogen genes in households with > 5 fecal piles compared to those with ≤ 5 fecal piles. Cow patties for cooking were differentiated from cow feces because they were formed into cakes and dried in the sun. Exposures (feces of different animals) were screened against outcomes (pathogen genes) in bivariate models and those with a p-value < 0.2 were included in the final multivariable model. Final models controlled for the month of sample collection. We corrected for multiple comparisons (i.e., the same pathogen evaluated in multiple sample types) using a Bonferroni correction[56].

3.3 Results and Discussion

3.3.1 Presence of Pathogen Genes and Indicators

Culturable *E. coli* were detected in 95% of soil, 88% of stored water, 75% of mother hands and 73% of child hands (Table 3.2). Culturable *E. coli* were the most commonly detected indicator in soil and stored water followed by BacCow and HumM2. On hands, BacCow was detected most frequently (97.5% of child hands and 96.5% of mother hands) followed by *E. coli* (see above) and HumM2 (21.9 % of child hands and 18.0% of mother hands).

Of the 360 child hand and 720 mother hand samples, 32% of child and 22% of mother hand samples were positive for at least one *E. coli* virulence gene (ECVG), the most common being *stx1/2* (15.1% of child and 8.3% of mother hand samples) (Table 3.2). *Stx1/2* were also the most abundant genes in stored water (16%) whereas *eaeA* was the most frequently detected ECVG in soil (37%). More than half (60%) of soil samples contained at least one ECVG. Across all sample types, few (≈ 1 mL %) were positive for *ipaH*. Norovirus and *Giardia* genes were found in < 5% of mother hand and child hand samples. As mentioned previously, we detected few positive samples for the *Cryptosporidium spp.* gene in our initial testing of 100 samples and thus did not assay the remainder of the samples. The *Cryptosporidium spp.* gene was inhibited in soil (see Appendix A), which could have impacted our ability to detect the gene in this sample type.

Our detection methods for indicators and pathogen genes have some limitations. First, the sampling unit was different for hand rinses, stored water and soil samples, limiting the ability to compare the relative percentage of pathogens in different environmental reservoirs. The binary data on pathogen presence and concentrations of BacCow in stored water and soil are influenced by different sampling volumes of stored water and varying moisture content in soil. Looking at the ECVGs detected, some such as *eaeA* are common to both EPEC

Table 3.2: Percentage of hand rinse, stored water, and soil samples positive for culturable *E. coli*, microbial source tracking markers, norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes.

	Percent Positive		Percent Positive
Child Hands (N=311-373)		**Mother Hands** (N=603-738)	
Culturable *E. coli*	73.2	Culturable *E. coli*	75.2
BacCow	97.5	BacCow	96.5
HumM2	21.9	HumM2	18
norovirus GII	4.2	norovirus GII	3.1
Giardia lamblia	4.8	*Giardia lamblia*	2.3
Any *E. coli* virulence gene	32.4	Any *E. coli* virulence gene	22.1
stx1/stx2 (EHEC)	15.1	*stx1/stx2* (EHEC)	8.3
eaeA (EPEC/EHEC)	11.4	*eaeA* (EPEC/EHEC)	7.7
aggR (EAEC)	11.8	*aggR* (EAEC)	7.7
st1b/lt1 (ETEC)	8.7	*st1b/lt1* (ETEC)	4
ipaH (EIEC)	0.3	*ipaH* (EIEC)	0.8
Stored Water (N=652-742)		**Soil** (N=644-755)	
Culturable *E. coli*	87.5	Culturable *E. coli*	94.8
BacCow	66.6	BacCow	86.8
HumM2	2.6	HumM2	20.2
Any *E. coli* virulence gene	37	Any *E. coli* virulence gene	60.3
stx1/stx2 (EHEC)	16.6	*stx1/stx2* (EHEC)	24.8
eaeA (EPEC/EHEC)	14.9	*eaeA* (EPEC/EHEC)	37.7
aggR (EAEC)	11	*aggR* (EAEC)	15.9
st1b/lt1 (ETEC)	8.6	*st1b/lt1* (ETEC)	15.3
ipaH(EIEC)	1.2	*ipaH*(EIEC)	1.2

and EHEC, while *ipaH* is common to both EIEC and *Shigella*[48, 98]. Although *ipaH* is present in both EIEC and *Shigella*, IDEXX Colilert-18 has been shown to be selective for *E. coli*[99, 100]. Lastly, the presence of PCR targeted genes for norovirus and *Giardia* does not necessarily indicate the presence of infectious organisms. *E. coli* virulence genes likely originated from viable *E. coli*, due to the IDEXX culturing step, similar to enrichment PCR. For *Giardia* and norovirus, our methods capture the presence of genetic material, but only viable pathogens are capable of causing illness and our results could therefore overestimate the potential for infection with *Giardia* and norovirus.

3.3.2 Indicator *E. coli* and Pathogen Genes

The concentration of indicator *E. coli* across all samples types was significantly associated with the prevalence of *eaeA*, *aggR*, *stx1/2*, and *st1b/lt1* (Fig 3.3 and Table B.1). The detection of ECVGs increased 86% (prevalence ratio [PR]: 1.86 (95% CI:1.65-2.11), p<0.001) on child hands, 93% (PR: 1.93 (1.75-2.12), p<0.001) on mother hands, 64% (PR: 1.64 (1.54-

1.74), p<0.001) in stored water, and 55% (PR: 1.55 (1.45-1.65), p<0.001) in soil with every $\log_{10}$ increase in *E. coli* concentration. A $\log_{10}$ increase in the concentration of *E. coli* was associated with a doubling of the prevalence of *eaeA* in all sample types: child hands (PR: 2.11 (1.68-2.64), p<0.001); mother hands (PR: 2.14 (1.79-2.56), p<0.001); stored water (PR: 2.02 (1.76-2.31), p<0.001); soil (PR: 1.93 (1.74-2.15), p<0.001). A similar magnitude of association was observed between *aggR*, *stx1/2*, and *st1b/lt1* and indicator *E. coli* concentration (Figure 3.3). However, no statistically significant association was observed between the prevalence of *ipaH* and *E. coli* concentration in any sample type. Similarly, there were no statistically significant associations between the prevalence of norovirus GII or *Giardia lamblia* genes and the concentration of indicator *E. coli*.

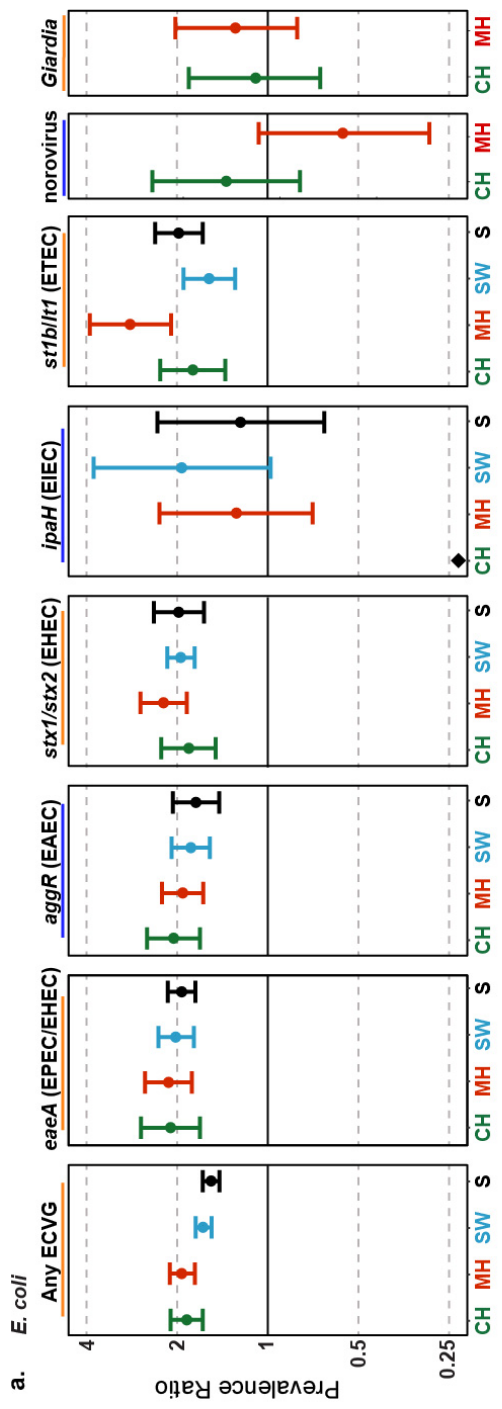


Figure 3.3: Prevalence ratios and 95% confidence intervals indicating the average increase in prevalence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes for a 1 $\log_{10}$ increase in a). MPN *E. coli*/sample matrix and b.) gene copies of BacCow/sample matrix. The diamond indicates too few samples were positive to model. Pathogens and virulence genes associated with human or human and animal sources are underlined in blue and orange, respectively.

Our results indicate that higher concentrations of *E. coli*, as detected by Colilert-18, is correlated with a higher prevalence of pathogenic *E. coli* genes. Our findings are consistent with previous studies that found higher concentrations of indicator *E. coli* were correlated with more ECVGs on mother hands, and in stored and source water in Tanzania[37, 101]. Unlike the Tanzania study, we did not find a significant association between concentration of *E. coli* and *ipaH* in stored water. However, the study in Tanzania reported a higher prevalence of *ipaH* (32.3%) in stored water compared to 1.2% in our study. There are several reasons why concentrations of indicator *E. coli* and pathogenic *E. coli* might not be correlated. Studies have identified "naturalized" *E. coli* in tropical environments; if these naturalized strains are non-pathogenic and are a significant contributor to the *E. coli* measured in a sample, then we would not expect a relationship between indicator and pathogenic *E. coli*[102–105]. Also, studies have elucidated physiological differences between strains of *E. coli* sourced from different environments, including pathogenic strains, that impact survival and transport in the environment[106–109]. Thus, it is meaningful that we found a significant relationship between indicator *E. coli* and ECVGs (PRs > 1.5 for any ECVG in all sample types). It should be noted that we did not measure genes from other bacterial pathogens to explicitly assess the ability of indicator *E. coli* to predict the presence of non *E. coli* bacterial pathogens; each bacterial pathogen may vary in its transport and survival in the environment.

The bacterial indicator *E. coli* was not a good indicator of norovirus or the protozoan *Giardia*. This could be due to differences in organism structure that dictate fate and survival in the environment[110]. Bacteria, viruses and protozoa respond differently to environmental conditions such as desiccation or sunlight exposure, and only bacteria have the capacity to grow in the environment and be naturalized[110, 111]. With regards to source tracking, the norovirus gene and *E. coli* were not correlated. This may be because norovirus only infects humans whereas indicator *E. coli* are present in both humans and animal feces[10, 94]. However, *Giardia* can infect both humans and animals, making this lack of association notable[93].

The positive predictive value for the presence of ECVGs increased from 45% (95% CI: 41-49%) using >1 MPN/100 mL as the cutoff for *E. coli* concentration to 65% (58-72%) using >100 MPN/100 mL as the cutoff (Table B.2). These findings are consistent with the results of the generalized linear model and suggest that higher concentrations of indicator *E. coli* are associated with an increase in the presence of pathogenic *E. coli* genes.

3.3.3 Human Fecal Marker and Pathogen Genes

We found no statistically significant relationships between the presence of the human fecal marker, HumM2, and any pathogen genetic marker on child hands, mother hands or in stored water (Figure B.1 and Table B.3).

The human marker was not a good predictor of enteric pathogen genes in hand rinses as indicated by the PPV (Table B.4). On mother and child hands there was only a 28% (95% CI: 18-39%) and 24% (17-32%) chance of observing any ECVG when HumM2 was present. The likelihood of detecting ECVGs when the human fecal marker was present was higher in stored water, 50% (26-75%) and soil, 61% (53-70%). Negative predictive values for many of the ECVGs were highest for mother hands. High NPVs indicate there was a low probability of ECVGs in the absence of HumM2. Therefore, HumM2 could be considered a somewhat conservative indicator of enteric pathogens on mother and child hands in the rural Bangladeshi environment.

We found no association between HumM2 and the norovirus gene. Norovirus infection is specific to human hosts; therefore, we would expect the human marker and norovirus to co-occur. On child hands we observed a prevalence ratio of 1.26 (0.36-4.41, p=0.72) and PPV of 5% (0-10%). However, the number of samples positive for norovirus (4.2%) was low, leading to an imprecise estimate of prevalence ratio and low PPV. We would need an even larger sample size to adequately test the relationship between HumM2 and norovirus, given the low prevalence. There have been mixed results in previous literature about the association between human-associated microbial source-tracking markers and human-associated viruses. In coastal waters both a positive association and no association have been reported[112, 113]. In our study, it is also possible we did not observe an association between norovirus and HumM2 due to the low specificity of the HumM2 assay (75% specific to human feces) in rural Bangladesh[20].

3.3.4 Animal Feces and Pathogen Genes

We observed a statistically significant increase in the prevalence of the *Giardia lamblia* gene (PR: 2.26 (1.50-3.40), p<0.001), any ECVG (PR: 1.37 (1.16-1.62), p<0.001), *eaeA* (PR: 1.63 (1.20-2.22), p<0.001), and *stx1/2* (PR: 1.79 (1.32-2.43), p<0.001) with every $\log_{10}$ increase in the concentration of the animal fecal marker (BacCow) on mother hands (Figure 3.3 and Table B.5).

We observed increases in the prevalence of many pathogen genes with an increase in the number of fecal piles from animals found in the courtyard of study households (Figure 3.4 and Table B.6), however only one association was statistically significant in the adjusted model and correcting for multiple comparisons. Specifically, there was over a 2-fold increase in the prevalence of *eaeA* on mother's hands in households with observed cow patties compared to those with no cow patties (PR: 2.82 (1.54-5.17), p<0.001).

On mothers' hands, both higher concentration of animal fecal marker (BacCow) and greater number of feces in household courtyards were positive predictors of zoonotic pathogen genes. We found the most significant association between BacCow and pathogenic genes *eaeA*, *stx1/2*, and the *Giardia* gene, which indicate the presence of EPEC or EHEC (*eaeA*), EHEC

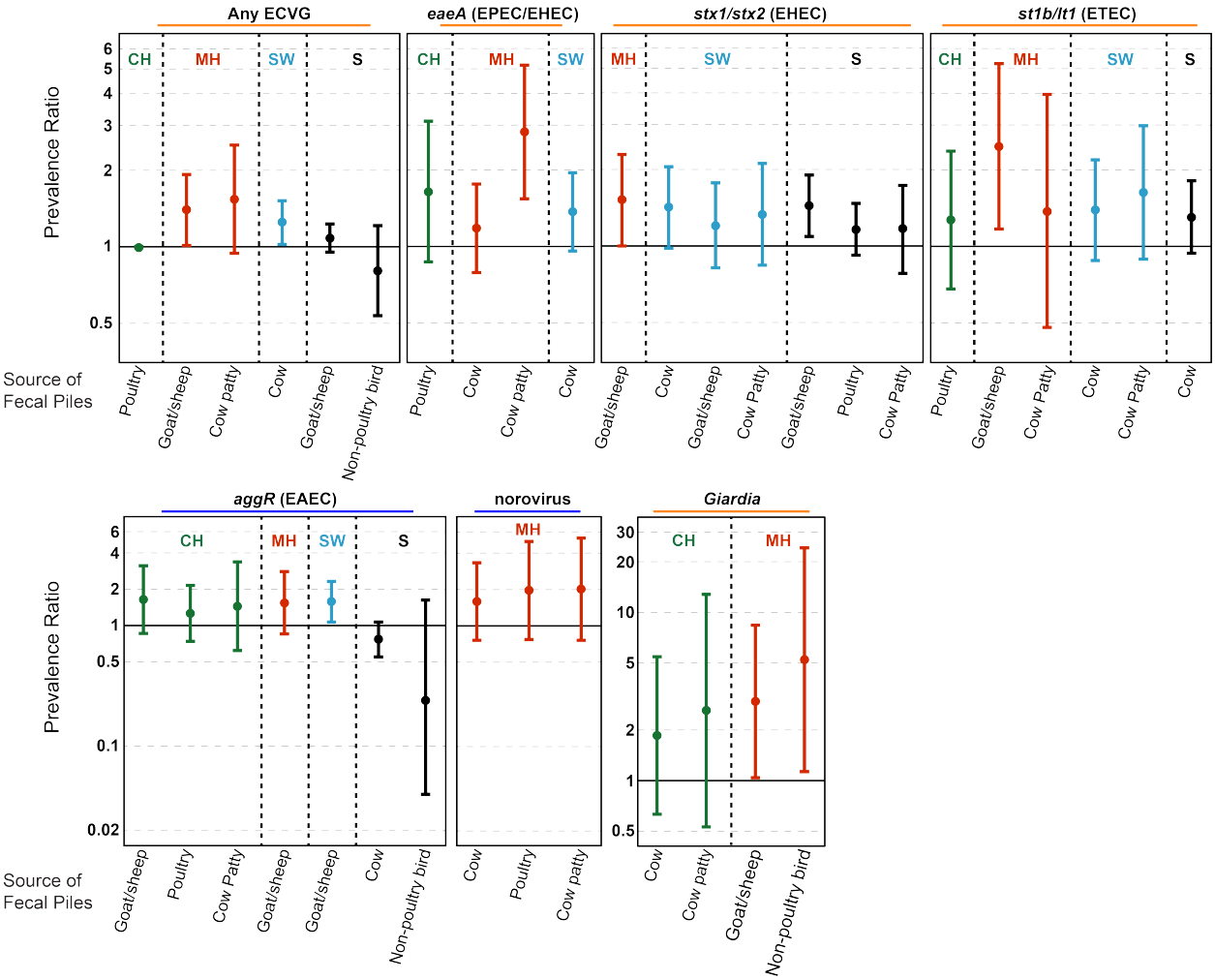


Figure 3.4: Adjusted prevalence ratios and 95% confidence intervals indicating the prevalence of norovirus GII, *Giardia lamblia* and *E. coli* virulence genes in households with fecal piles or cow patties above and below a threshold value. For cow, goat/sheep, non-poultry birds, the threshold value was zero. For chicken/non-chicken poultry the threshold value was 5 (>5 versus ≤ 5). Feces types with a p-value <0.2 in bivariate models between outcomes and exposures were included in the adjusted models. Pathogens and virulence genes associated with human or human and animal sources are underlined in blue and orange, respectively. Prevalence ratios for EIEC were omitted due to their large confidence intervals.

(*stx1/2*), and *Giardia lamblia*. EPEC, EHEC, and *Giardia lamblia* have been found in animal reservoirs and can be transmitted in the environment through animal feces[11]. The fecal pile analysis found that *eaeA* was most associated with cow patties and *Giardia*, EHEC/EPEC, and ETEC were associated with goat/sheep feces, although the latter were not significant when correcting for multiple comparisons (Table B.6). Together, the animal molecular marker and the fecal pile survey findings suggest specific sources for these pathogens in the

Table 3.3: Percentage of households with animals, human and animal feces, and reported fecal management practices. N=938[a]

	Animals	**Visible Feces**
Cow	70.4	33
Goats/Sheep	39.2/<1	22.9[b]
Chicken/Non-chicken poultry	94.1/16.2	88.8[b]
Pigs	<1	0
Dogs or cats	12.9	<1
Non-poultry birds[c]	NA	4.4
Cow Patty	NA	7.9
Human	NA	<1
		Animal Handling Practices
Reported using hands, cloth or scrap material to handle animal feces		34.4
Reported using scoop or hoe to handle animal feces		68.8

[a]:Each sampling event was counted individually as visible feces could change between visits to the same household
[b]:Feces from different animals not distinguishable
[c]:Sparrow, pigeon, crow

environment. Particularly, cows are likely candidates for the source of EPEC and EHEC, which can ultimately make their way to mothers' hands when they handle animal feces. While 68% of households reported using tools to clean up feces, 34% reported using their hands, cloth, or scraps for animal feces removal (Table 3.3). Additionally, cow patties are used for cooking fuel which results in caregivers frequently switching between food handling and handling dung patties as they cook[23]. Based on these household practices, it is possible mothers are both picking up animal feces and handling their children, which can provide an opportunity for transmission of zoonotic pathogens to children. Therefore, future efforts to reduce enteric pathogen transmission in rural environments, where livestock are common, should include animal fecal containment.

Chapter 4

Shared Bacterial Communities Between Soil, Stored Drinking Water, and Hands in Rural Bangladeshi Households

4.1 Introduction

Enteric pathogens are transmitted through a variety of pathways in the environment via the fecal-oral route. Pathogens can move through environmental reservoirs including flies, fomites, hands, soil, food, and water. Providing sufficient quantities of water, adequate drinking water treatment, sanitation, and hygiene (WaSH) are the main strategies to block pathogen transmission pathways and reduce the occurrence of diarrheal illnesses. Field trials, such as the WASH Benefits study, have investigated the impact of these interventions on diarrhea in children and found modest reductions in diarrhea and no additive benefit of multiple interventions[38, 44, 114, 115]. Improving our understanding of the role of environmental reservoirs in enteric pathogen transmission is important, especially given the limited efficacy of WaSH interventions in the field.

Previous studies investigating environmental pathways of pathogen transmission have largely focused on identifying fecal indicators and some pathogens in household reservoirs such as soil, hands, stored drinking water, and children's toys[10, 19–21, 37, 46, 49, 62, 101, 116–119]. This approach is useful because it can produce quantitative results and directly detect microorganisms of interest. However, identifying traditional fecal indicators relies on pre-specification of single targets (microorganisms) as an indicator of a whole microbial community (fecal microbial community). One key application is microbial source tracking, in which fecal contamination from humans and animals is identified by quantifying host-specific intestinal bacteria. However, host microbiomes vary between individuals and geographically,

impacting the sensitivity and specificity of assays when applied to populations outside of those used to develop these assays[120–123].

Here, we aimed to apply a non-target method of characterizing transmission pathways by applying high-throughput 16S rRNA gene sequencing to whole bacterial communities present in environmental samples. New and less expensive sequencing technology has given rise to the study of a wide variety of microbiomes to understand bacterial transmission in the built environment[124]. This research has shown microbiomes can be shared between environments, fomites, people and animals[124–126]. Shared microbiota of concern include pathogens but there is increasingly more evidence that non-pathogenic bacteria can also impact human health outcomes[127–130].

To better understand shared microbiomes, SourceTracker (a Bayesian modeling tool), has been used to explore potential sources of fecal contamination in environmental engineering studies as well as outside the environmental engineering field to study shared microbiomes in homes, kitchens, hospitals, and restrooms[131–134]. SourceTracker is unique in its ability to build models with hundreds of sequences and is particularly useful for sources in which there are no, or limited, single target markers that indicate the presence of the source[135]. Despite the rapid onset of high throughput sequencing and subsequent computational tools such as SourceTracker, few studies have investigated household level microbiomes in low-and middle-income settings[136–138].

The goal of this study was to apply bacterial community analysis to elucidate pathways of enteric pathogen transmission in 50 rural Bangladeshi households. We used SourceTracker to quantify the shared microbial community in stored water, soil, and hands. We also estimated the percentage of fecal-associated bacteria from child and maternal feces to household reservoirs, in an attempt to overcome the limitation of low host specificity exhibited by the single target markers in rural Bangladesh[20]. To observe intra-household bacterial community exchange, we then quantified the overlap in amplicon sequencing variants in and outside individual households. Finally, we demonstrate the limits to identifying potential pathogens, at the genus and species level, using 16S rRNA gene sequencing in this context.

4.2 Methods

4.2.1 Environmental Sample Collection and Processing

50 households out of 300 sampled in the control arm of WASH Benefits were selected for this analysis. Households were chosen based on availability of all environmental sample types (child hands, mother hands, stored water and soil) collected at the same visit. An additional 25 soil samples were selected from the 300 control households for the source tracking soil library. To attempt to decouple the contributions from soil and feces, soil samples in the library were selected based on the absence of the human (HumM2) and animal (BacCow)

fecal markers, as determined by qPCR. The methods for environmental sample collection, DNA extraction and qPCR are described in Chapter 2.

4.2.2 Fecal Sample Collection and Processing

25 maternal and 25 child fecal samples were obtained from households enrolled in the control and combined water, sanitation and hygiene arms of WASH Benefits. Maternal and child fecal samples were not from the same households as those used for environmental analysis. Field staff distributed sterile fecal collection containers to households. Caregivers were instructed to collect their feces and the feces of their child from the same evening or the following morning. Field staff returned to the household the following day to retrieve the fecal specimens, which were placed on ice after defecation, and then transported on dry ice to the laboratory. Specimens were stored at -80 °C until extraction. DNA was extracted from fecal samples using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Germantown, MD), with an additional bead beating step. A blank sample was included in each round of extraction. Nucleic acid extract was transported to Stanford University and then UC Berkeley on dry ice.

4.2.3 16S Sequencing

Library preparation and sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. The V4 region of the 16S rRNA gene was amplified using 515f and 806r primers. Samples were pooled and sequenced on two MiSeq lanes, yielding paired-end 250 bp reads. A mock community DNA standard (Zymo Research, Irvine, CA) was included on each MiSeq lane.

4.2.4 Data Analysis

Forward and reverse reads were processed using the DADA2 pipeline[139]. Reads were truncated to 180 nucleotides, after which the quality score dropped significantly. The error rate was determined from sample reads and samples were denoised using the learned error model. Paired-end reads were merged to yield 250 bp sequences and chimeras were removed. On average 79 and 76% of the input reads from runs one and two remained after chimera removal.

Taxonomy was assigned in DADA2 using a Naïve Bayes classifier that was trained on the Silva v132 database[140, 141]. Species level identification was based on 100% identity between the reference database and amplicon sequence variants (ASVs)[142]. To identify potential pathogens we filtered ASVs at the genus and species level to known enteric pathogens (*E. coli*, *Aeromonas hydrophila*, *Campylobacter jejuni*, *Vibrio cholera*, and *Salmonella enterica*). We included ASVs that matched multiple species if one of those was a pathogen. We omitted

ASVs that were classified to the above genera but were assigned species that do not contain enteric pathogens.

Data were then analyzed using phyloseq (version 1.24.2) in R (version 3.5.0)[143]. Some samples contained an ASV from *Pseudomonas syringae* pv. phaseolicola (pph6) that was spiked into the sample during extraction to estimate extraction efficiency. This ASV was removed from all samples that were spiked. We also removed all ASVs associated with the eukaryotic organelles chloroplasts and mitochondria (45900 ASVs before and 30900 after). ASVs were normalized using the inverse hyperbolic sine transformation[144]. Beta diversity was analyzed via PCoA using Bray-Curtis dissimilarity and variables contributing to differences between communities were identified with PERMANOVA using adonis in the vegan package for R. For all sample types we investigated the association between Bray-Curtis dissimilarity and sample type, location, season, number of animals in a compound (<20 vs. ≥ 20), and number of individuals living in a compound (<10 vs. ≥ 10). For stored water we also assessed the impact, if any, of duration of storage and for soil we included sun vs. shade in the model.

Overlapping ASVs (identical ASVs) inside and outside households were determined by identifying the number of identical ASVs present in each sample type within the same household and outside the household. In order to determine if the mean number of ASVs overlapping inside and outside households was significantly different, we used a bootstrap method[145]. The average number of ASVs matching between households in the real dataset was compared to the average number of ASVs matching in a randomly generated dataset (generated by randomizing household ID numbers) 10,000 times. For each iteration we calculated a test statistic which was the difference in the means between the number of ASVs matching in the same households in the real dataset and the number of ASVs matching between households in the randomly generated dataset. Statistical significance was determined if zero was not included in the distribution of the test statistic at specified alpha values, 0.05 and 0.0083. The latter alpha value was adjusted to correct for multiple comparisons using the Bonferroni correction[56].

To estimate the association between bacterial communities in different reservoirs and feces we used SourceTracker2 in python (version 3.7.0)[135]. ASVs with a mean relative abundance less than 0.001% were removed (30900 ASVs before and 5500 after filtering) and the remaining ASV counts were imported to SourceTracker2. SourceTracker2 was run with default parameters and each environmental sample type was designated as a sink with all other environmental sample types and feces designated as sources. We included all reservoirs in the source tracker analysis for all other reservoirs for consistency in the modeling although some pairs do not represent realistic scenarios (e.g. the percentage of bacteria in soil associated with hands).

The source tracking analysis was validated using two approaches: 1.)Spiking DNA extract

from source (mother feces, child feces, and soil) composite samples into child hands, mother hands, and stored water composite samples. Source composites were made by pooling DNA extract from five samples to an equal concentration in the composite. One composite was made for each source type (child feces, mother feces, soil). Similarly, for sink composite samples DNA extract from five samples was pooled to make one composite for each sample type (child hands, mother hands, soil). Source composites were spiked into the environmental sink composites to achieve DNA concentration ratios of 10% source/90% sink, 1% source/99% sink and 0.5% source/99.5% sink. 2.)Composite child feces, mother feces and soil source samples were combined in different DNA concentration ratios from 0 to 80%. All validation combinations are shown in Table 4.1. Source samples used in the validation composites were not included in source libraries.

4.3 Results

4.3.1 Quality Controls

The relative abundance of the ZymoBIOMICS mock DNA community was very similar between the two lanes (Table C.1). The eight most abundant taxa matched the reference sequences and the next most abundant ASV (1.4% of sample reads in both mock communities) had a one bp mismatch from the reference sequence of *Salmonella enterica*. Extraction blanks for environmental and fecal samples amplified poorly, or not at all, in PCR and were therefore not included in the pooled library. Note that the mother hand sample in household 48 was omitted due to poor PCR amplification.

4.3.2 Source Tracking Validation

There was good agreement between estimated source contributions and SourceTracker predicted contributions (Table 4.1). There was a strong linear relationship between the estimated and predicted source percentages (Pearson's ρ = 0.97, 0.95, 0.92 for child feces, mother feces and soil; $p<0.001$ for all). SourceTracker was sensitive to the lower percentages of source spike-in and was able to identify a qualitative difference between 1 and 0.5%. Estimated contribution from 1% was consistently higher than the estimated contribution from 0.5% in the spike-in samples. It should be noted that our estimates of the sources in the validation samples only account for the amount of DNA added to the samples, and not for any child feces, mother feces, and soil DNA present in the validation samples of mother hand, child hand and stored water composite samples. HumM2 and BacCow markers were present in some of the environmental samples that went into the composites. Unlike the soil samples used in the library and in the validation samples, we were unable to identify multiple hand rinse and stored water samples that tested negative for both fecal markers. This resulted in what appears to be false positives, but the estimated percentage of feces in Table 4.1 is likely above zero. In the source only composites, SourceTracker was able to differentiate the

relative contribution of different sources (mother feces, child feces, and soil). SourceTracker predicted low percentages from soil and mother feces when soil and mother feces were not part of the source composite sample (0-2.2%) (Table 4.1). In composites without child feces, SourceTracker predicted slightly greater contributions (5.4-9.3%) from child feces. Soil was also consistently underestimated in the observed contribution compared to the expected.

Table 4.1: Lab based estimated source percentages compared to SourceTracker estimates (standard deviation) of source contributions in validation samples made by: 1) Spiking DNA extract from source (mother feces, child feces, and soil) composite samples into child hands, mother hands, and stored water composite samples. Source composites were spiked into the environmental sink composites to achieve DNA concentration ratios of 10% source/90% sink, 1% source/99% sink and 0.5% source/99.5% sink. 2) Combining composite child feces, mother feces and soil source samples in different DNA concentration ratios from 0 to 80%.

Estimated								**Source Tracker Sources % (Std)**			
Source	%	Source	%	Source	%	Sink	%	CF	MF	Soil	Unknown
Spike-In											
CF	0	MF	10	Soil	0	SW C	90	2.2 (0.2)	14.3 (1.2)	1.0 (0.4)	82.5 (1.3)
CF	0	MF	1	Soil	0	SW C	99	0.4 (0.2)	2.3 (0.3)	1.2 (0.3)	96.1 (0.4)
CF	0	MF	0.5	Soil	0	SW C	99.5	0.4 (0.2)	0.4 (0.2)	1.0 (0.3)	98.3 (0.4)
CF	10	MF	0	Soil	0	SW C	90	11 (0.2)	0.8 (0.3)	1.6 (0.4)	86.7 (0.5)
CF	1	MF	0	Soil	0	SW C	99	1.2 (0.1)	0.1 (0.1)	1.4 (0.3)	97.4 (0.3)
CF	0.5	MF	0	Soil	0	SW C	99.5	0.5 (0.1)	0 (0)	1.1 (0.4)	98.4 (0.4)
CF	10	MF	0			S 357	90	16.3 (0.5)	0.8 (0.3)	52.8 (0.4)	30.2 (0.6)
CF	1	MF	0			S 357	99	2.8 (0.2)	0.4 (0.1)	63.3 (0.5)	33.5 (0.5)
CF	0.5	MF	0			S 357	99.5	1.7 (0.1)	0.6 (0.2)	62 (0.7)	35.7 (0.7)
CF	10	MF	0	Soil	0	CH C	90	17.3 (0.3)	0.5 (0.2)	3.5 (0.5)	78.8 (0.6)
CF	1	MF	0	Soil	0	CH C	99	2.7 (0.4)	0.2 (0.2)	4.0 (1.1)	93.1 (1.3)
CF	0.5	MF	0	Soil	0	CH C	99.5	2.0 (0.2)	0.3 (0.2)	3.1 (0.8)	94.6 (0.9)
CF	10	MF	0	Soil	0	MH C	90	19.4 (0.5)	1.1 (0.3)	12.8 (1.1)	66.8 (1.3)
CF	1	MF	0	Soil	0	MH C	99	18.1 (0.4)	0.9 (0.2)	12.3 (1.2)	68.8 (1.3)
CF	0.5	MF	0	Soil	0	MH C	99.5	0.8 (0.2)	1.4 (0.3)	11.0 (0.9)	86.8 (1.0)
CF	0	MF	0	Soil	10	SW C	90	0 (0)	0.3 (0.1)	4.9 (1.0)	94.9 (1.1)
CF	0	MF	0	Soil	1	SW C	99	0 (0)	0.2 (0.1)	1.8 (0.2)	98.0 (0.3)
CF	0	MF	0	Soil	0.5	SW C	99.5	0.1 (0.1)	0.1 (0.1)	1.3 (0.5)	98.5 (0.5)
CF	10	MF	10	Soil	10	SW C	70	14.1 (0.5)	15.3 (0.7)	4.0 (1.0)	66.6 (1.0)
CF	1	MF	1	Soil	1	SW C	97	2.4 (0.3)	2.5 (0.3)	1.6 (0.5)	93.6 (0.6)
Source Composites Only											
CF	10	MF	80	Soil	10			20.8 (1.0)	55.9 (0.9)	1.8 (0.2)	21.5 (1.0)
CF	80	MF	10	Soil	10			63 (0.9)	23.1 (1.0)	5.7 (0.1)	8.2 (0.4)
CF	10	MF	10	Soil	80			17.7 (1.0)	20.4 (0.6)	34.5 (0.6)	27.4 (1.0)
CF	0	MF	50	Soil	50			6.8 (1.1)	48.5 (1.0)	15.6 (0.6)	29.0 (1.1)
CF	50	MF	0	Soil	50			44.2 (0.6)	2.5 (0.6)	28.3 (0.4)	25.1 (0.8)
CF	50	MF	50	Soil	0			42 (1.3)	44.5 (1.1)	0 (0)	13.5 (0.4)
CF	0	MF	25	Soil	75			6.2 (1.4)	38.2 (1.5)	29.6 (0.5)	26.1 (0.8)
CF	0	MF	75	Soil	25			7.5 (1.2)	56.4 (1.1)	6.1 (0.3)	30.0 (0.7)
CF	25	MF	0	Soil	75			31.8 (0.5)	1.3 (0.4)	40.4 (0.5)	26.6 (0.7)
CF	75	MF	0	Soil	25			63.2 (0.9)	2.5 (0.6)	14.6 (0.3)	19.8 (0.8)
CF	75	MF	25	Soil	0			55.2 (0.9)	34.1 (0.9)	0 (0)	10.7 (0.3)
CF	25	MF	75	Soil	0			29.7 (0.7)	50.3 (1.1)	0 (0)	20 (1.3)

4.3.3 Beta Diversity

There were significant differences in the bacterial community between all sample types (Figure 4.1) (PERMANOVA R^2=0.27, p=0.001). In Figure 4.1 samples cluster by sample type,

including distinct clusters between child and mother feces. Among environmental samples, the union (local unit of governance in rural areas) where households were located also explained differences in the bacterial community (R^2=0.25, p=0.001). In fecal samples, there was a significant difference in the communities between mothers and children (R^2=0.22, p=0.001). All other variables (season (wet vs. dry), number of animals (<20 vs. $\geq$ 20), number of individuals (<10 vs. $\geq$ 10), drinking water storage time, and soil sampling condition (sun vs. shade)) explained a low percentage of the variance. PERMANOVA tests were also performed using the Aitchison distance as suggested by Gloor et al.[146]. R^2 values were similar and also significant in these models except for the association between stored water bacterial community and union.

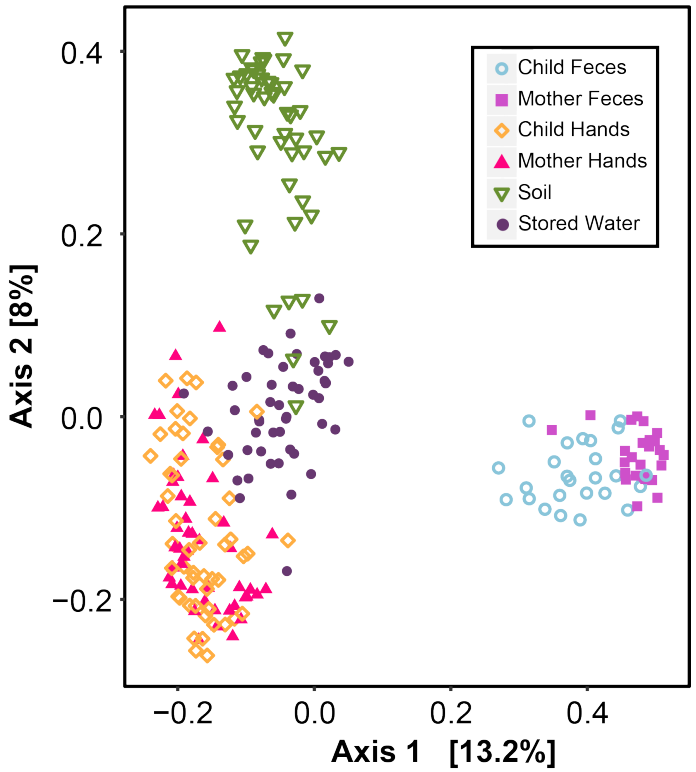


Figure 4.1: Principal Coordinates Analysis (PCoA) of bacterial community composition by sample type based on Bray-Curtis dissimilarity. Samples clustered by type which is indicated by color and shape.

4.3.4 SourceTracker Results

Mother and child hands had the highest estimated percentage of associated bacteria of all reservoirs (Figure 4.2). On child hands over half of the bacteria was attributed to mother hands(mean: 56.4 (95% CI: 52.3-60.5)%). Child feces (CF) (0.8 (0.5-1.2)%) and mother feces (MF) (1.1 (0.3-1.9)%) were minor contributors to the bacterial community. The estimated

percentage from soil (3.2 (2.0-4.3)%) was significantly greater than child and mother feces (t-test, $p<0.01$ for both). Similarly, on mother hands over 50% of the bacterial community was related to child hands (56.2 (51.5-60.9)%). Although the percentage of fecal- associated bacteria was low (<2%), there was a statistically significant difference in the percentage attributed to child feces and mother feces (CF: 0.4 (0.2-0.7)%; MF: 1.7 (1.0-2.5)%, $p=0.002$). There was also significantly more bacteria associated with soil (2.8 (1.8-3.8)%) than child feces ($p <0.001$).

In stored water, the estimated percentage of bacteria attributed to mother feces and child feces was less than 0.5%. A greater percentage of the bacterial community was associated with hands (mother hands (MH): 7.3 (4.6-10.2)%; child hands (CH): 1.9 (0.4-3.5)%), with a significantly greater percentage associated with mother hands ($p=0.002$). On average 2.3 (1.3-3.3)% of the microbial community was related to soil, which was significantly greater than feces ($p <0.001$ for both). However, the estimated percentage of bacteria attributed to mother hands was significantly greater than the estimated percentage attributed to soil ($p=0.002$). In soil, less than 0.5% of the bacteria were fecal-associated.

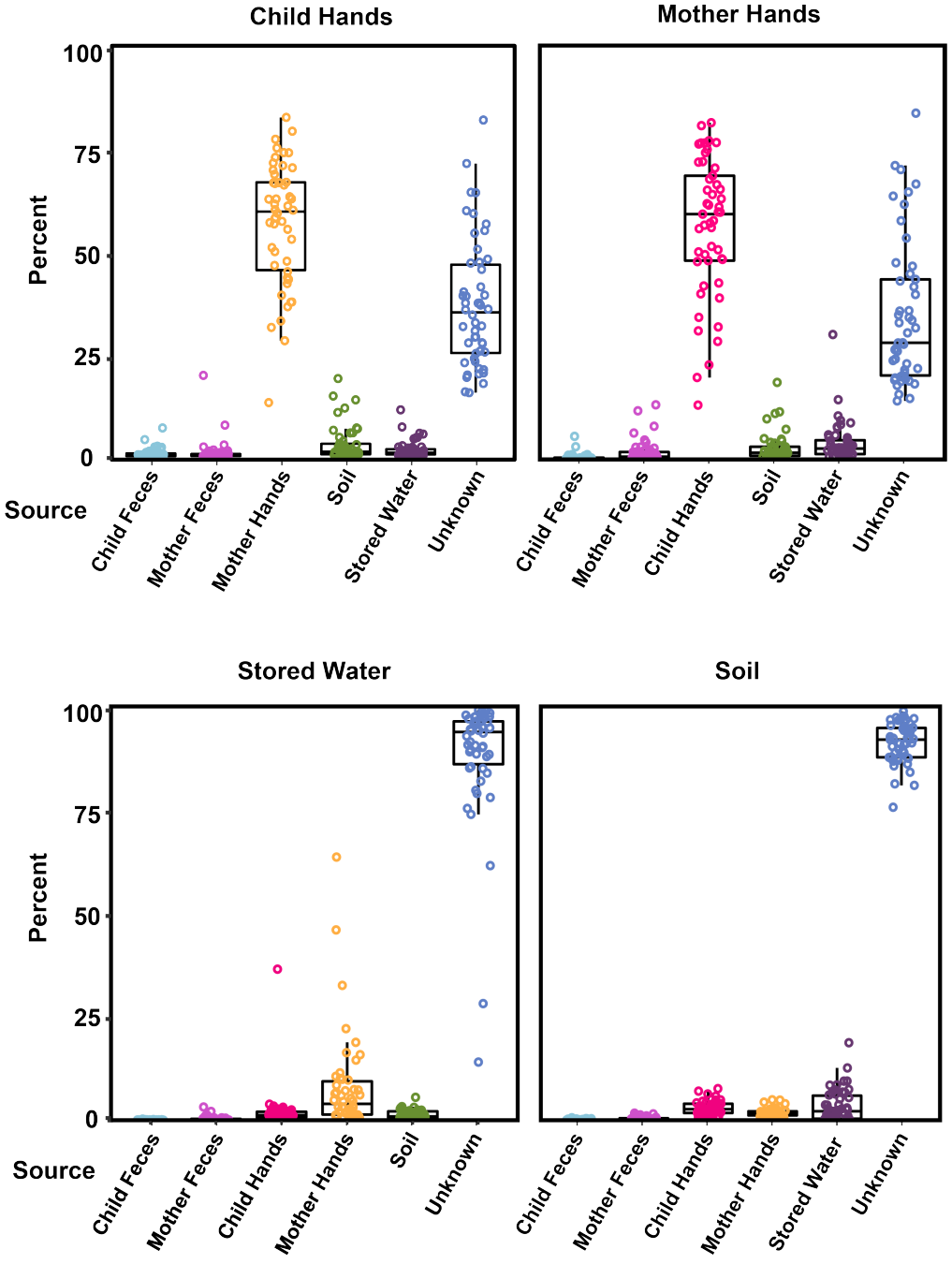


Figure 4.2: Percentage of bacteria in each sample that was associated with all other reservoir types, maternal feces, and child feces as determined by SourceTracker. Each reservoir was included in the model for all other reservoirs for model consistency despite illogical pairs. Box plots indicate the overall distribution (25th percentile, median, 75th percentile. Whiskers indicate at most 1.5 x interquartile range).

4.3.5 Intra-Household ASV Matching

The average (95% CI) number of ASVs on mother hands, child hands, and in stored water and soil was 381 (337-425), 401 (352-451), 412 (362-462), and 698 (637-759) respectively. We found a significant difference between the number of sequence variants that were identical between child hands and mother hands in the same household (μ =168) compared to other households (μ =98) (Figure 4.3). The lower bound of the distribution of the test statistic was well above zero (95% CI: 51.7 - 80.2) (Figure C.1 and Table C.2). The number of identical ASVs on mother and child hands and in soil was also significantly greater in the same household compared to other households. On average 84 ASVs overlapped between mother hands and soil, while approximately 59 ASVs matched outside the household. Similarly, the number of ASVs overlapping between child hands and soil inside (μ=96) households was greater than those that matched to outside (μ=65) households. The lower bound of the test statistic distribution was above zero for both comparisons (95% CI CH to S: 14.9 - 41.3; MH to S: 10.2 - 35.3)(Table C.2). There were no significant differences in the number of ASVs that overlapped between mother hands and stored water, child hands and stored water, and soil and stored water in and outside households, correcting for multiple comparisons.

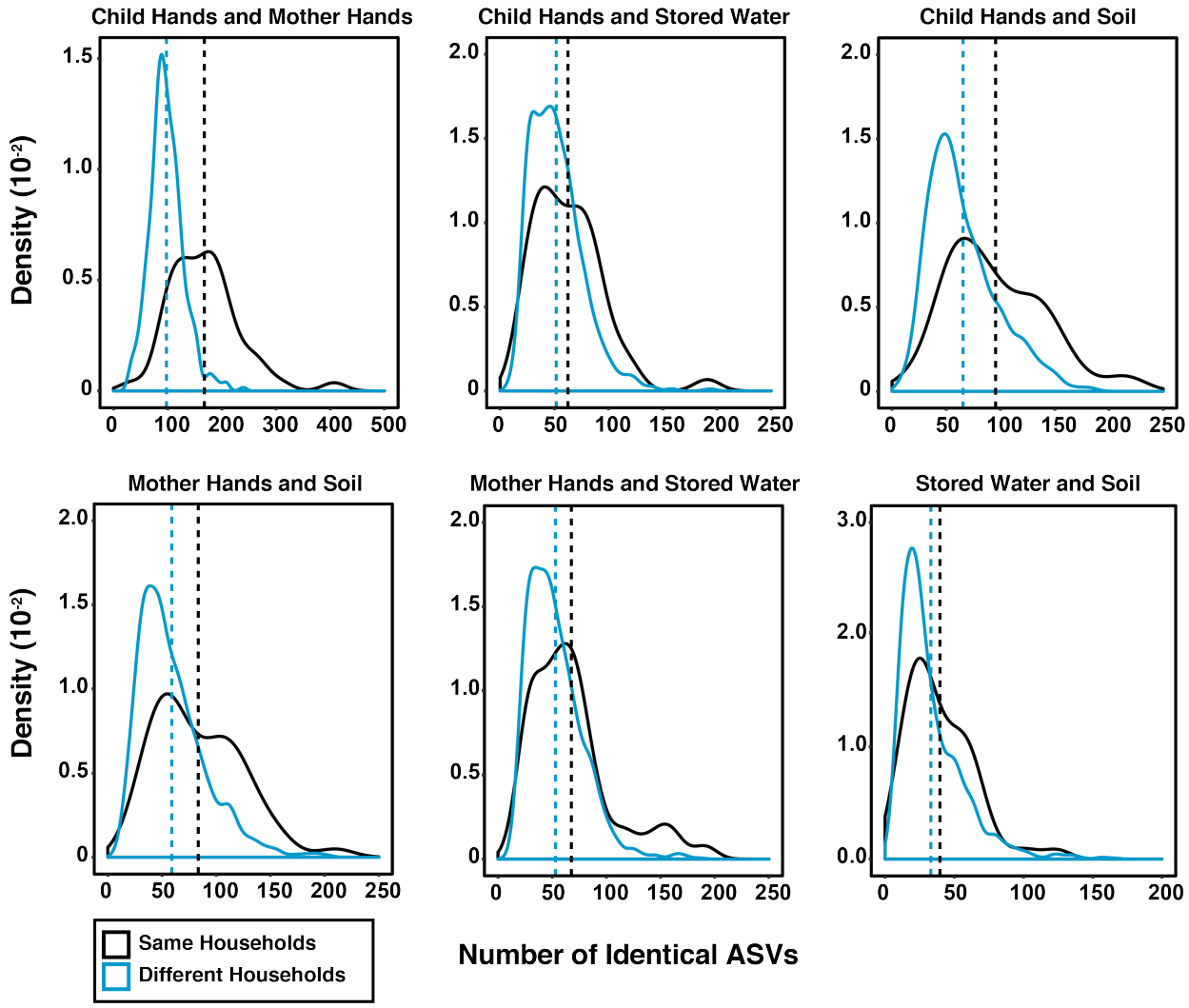


Figure 4.3: Density distribution of the number of identical ASVs in child hands and mother hands, child hands and stored water, child hands and soil, mother hands and soil, mother hands and stored water, and stored water and soil, for samples from the same households versus samples from different households. 50 households were sampled in total. Mean values for the number of ASVs matching inside and outside households are indicated with horizontal lines.

4.3.6 Potential Pathogens

We identified 22 ASVs from the genera *Campylobacter*, *Vibrio*, *Aeromonas*, *Salmonella*, and *Escherichia/Shigella* (Figures 4.4 and 4.5). Of those, 16 were identified to the genus level only, four were identified to the species level in which pathogenic species were one of multiple possible species, and two were identified to a single genus and species (*E. coli* and *V. cholera*). The relative abundance of these potential pathogens was low in all sample types (<1%). Of the ASVs classified to the genus *Campylobacter*, 10 were found on mother and

child hands, only two ASVs were found in stored water samples and none were present in soil. *V. cholera* was found on one child hand, on six mother hands, and in five stored water samples. In household 22, *V. cholera* was present in both mother and child hands and stored water. This ASV was also in the stored water and on the mother hands in households 1, 4, and 18. We also assigned taxonomy using the RDP database and taxonomic classifications of potential pathogen ASVs were nearly the same[147]. The only differences were apparent in ASVs classified to multiple species; the assigned non-pathogenic species varied slightly between databases (data not shown).

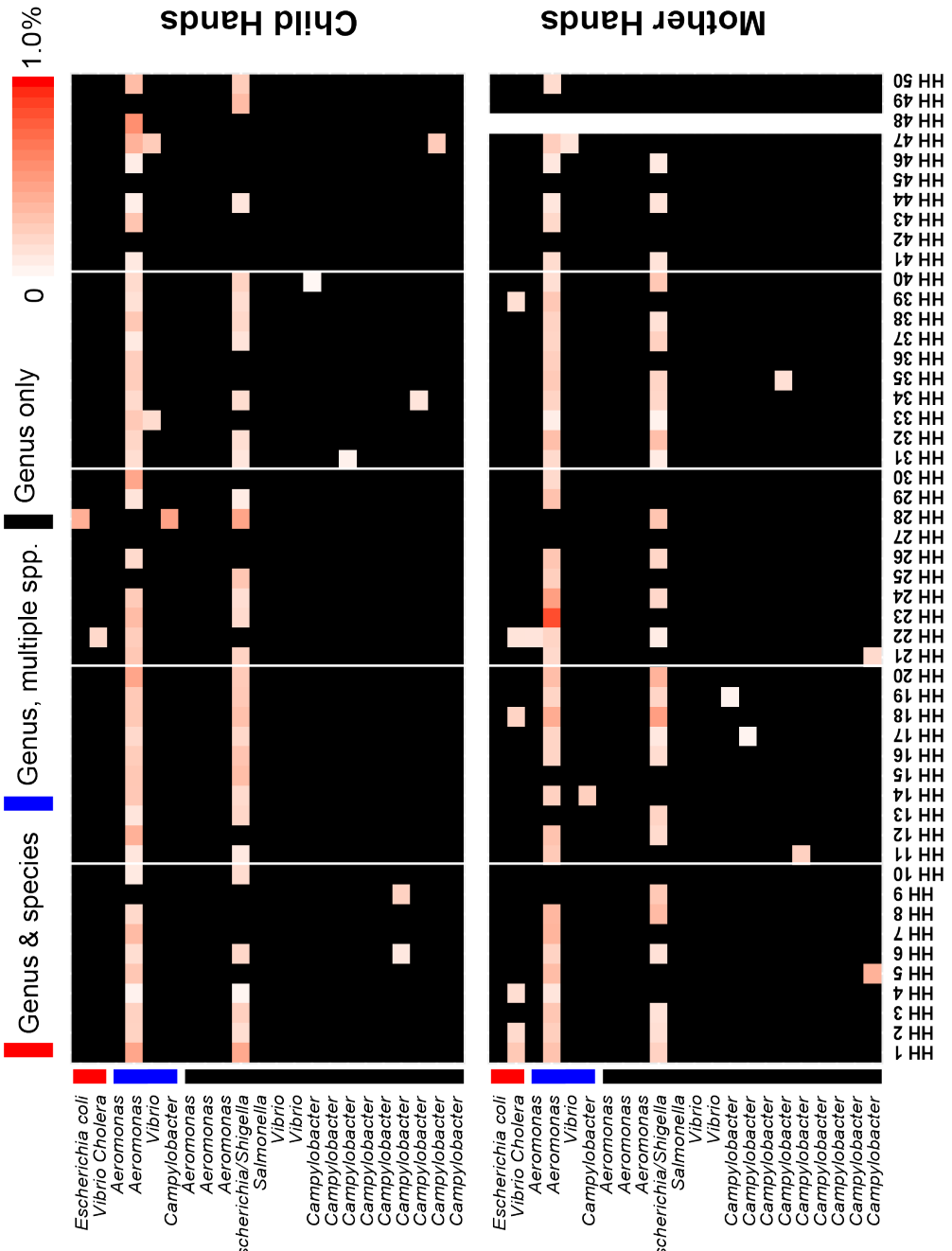


Figure 4.4: Relative abundance of ASVs that could be potential pathogens on mother and child hands in 50 households. Note: The mother hand sample in household 48 was omitted due to poor PCR amplification. Identification of pathogens at the genus level is indicated in black, multiple species (one of which could be a pathogen) in blue, and genus and species in red.

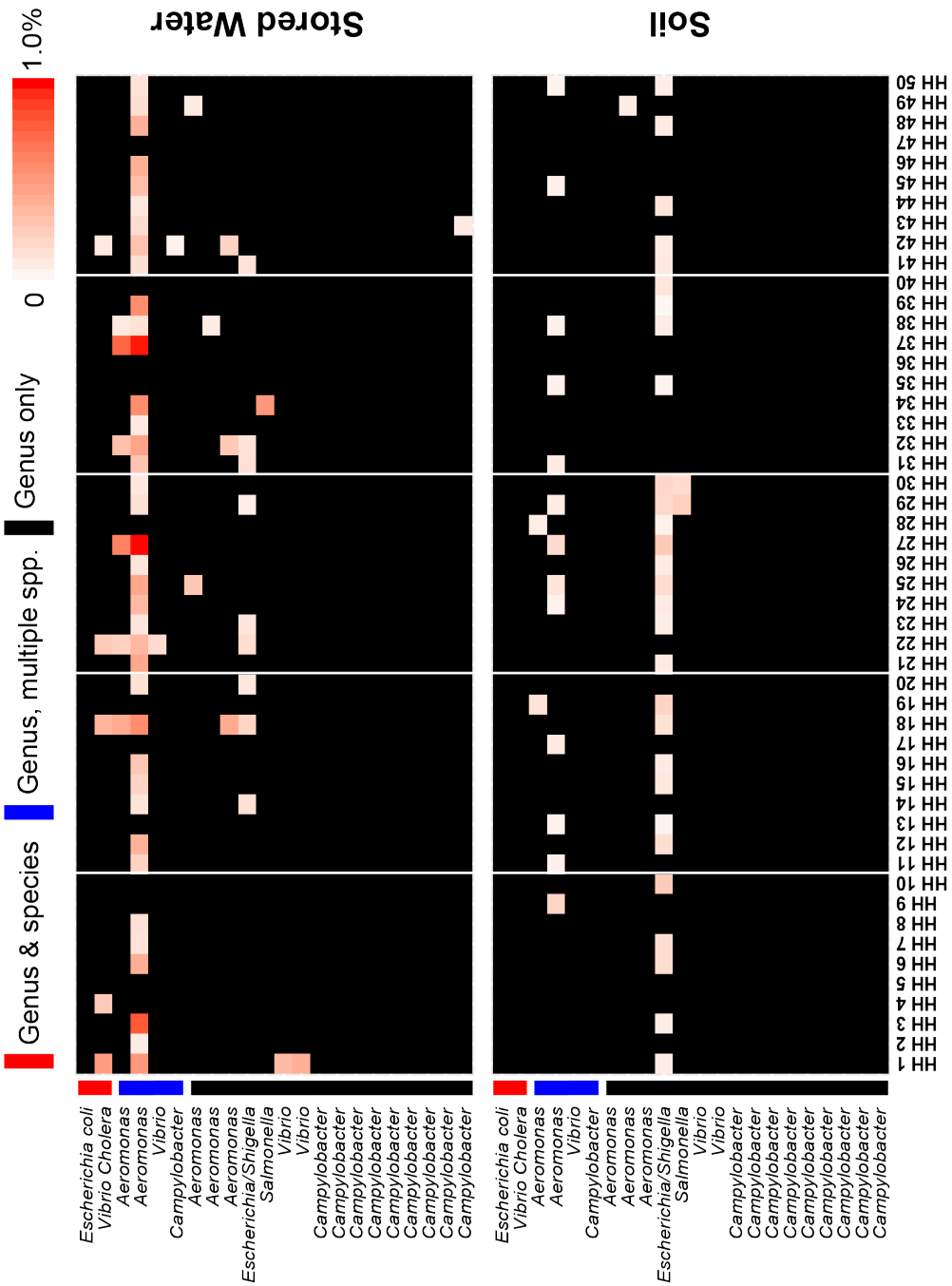


Figure 4.5: Relative abundance of ASVs that could be potential pathogens in soil and stored water in 50 households. Identification of pathogens at the genus level is indicated in black, multiple species (one of which could be a pathogen) in blue, and genus and species in red..

4.4 Discussion

The goal of this study was to explore the potential of bacterial community analysis to provide insight into household level transmission pathways of enteric pathogens. We first started with a broad investigation of the bacterial communities in different sample types in households using SourceTracker. We also aimed to estimate the relative importance of human feces to the bacterial community in household reservoirs. We then explored the overlap in communities inside and outside households by matching ASVs from different environmental reservoirs in order to understand which intra-household transmission pathways led to shared ASVs. Finally, we used the taxonomy assigned to ASVs to identify potential pathogens that could be shared in households at the genus and species level.

Using SourceTracker, we determined that the overall contribution of mother and child feces to the microbial communities was low in mother hands and child hands, and lowest in stored water and soil. In Chapters 2 and 3 we measured a single marker (HumM2) as an indicator of human fecal contamination in 600 households. HumM2 was present on $\approx$ 20 % of hands, in $\approx$ 20 % of soil, and in <4 % of stored water samples. Although it is difficult to directly compare these results because presence of HumM2 does not have a quantitative value and Sourcetracker attributed some fraction of each sample to feces, albeit very small, there are some noteworthy comparisons. In Figure 4.2, <0.5% of bacteria in stored water was sourced to feces which is consistent with the very low abundance of HumM2 in the single marker study. However, human fecal bacteria were found in <0.5 % of soil whereas 20% of soil samples were positive for HumM2. This could be due to low host specificity of HumM2 in the rural Bangladeshi environment, where HumM2 was also found in goat and chicken feces[20]. It is possible that the relative contribution from human feces to soil was low and feces from chickens and goats were detected by the single marker (HumM2) but not by SourceTracker.

Although the percentage of fecal bacteria was low on hands, our results are similar to those from a study of the bacterial community on hands of children, mothers and fathers in households in the United States[148]. This previous study, which also used SourceTracker, found that palms from children, mothers and fathers had a median of 1.7% (range 0-99%) of fecal bacteria. They found that 12% of hands had over 25% fecal bacteria, whereas none of the mother and child hands in our study had more than 13.5% of bacteria sourced to stool. It is notable that the percentage of fecal bacteria on hands in our study in a low-income area, with predominately pit latrines for sanitation, was not significantly greater than the percentage of fecal bacteria on hands in the high-income population.

While the percentage of bacteria in household reservoirs associated with mother and child feces was low in our study, the proportion of the microbial community on mother hands that was associated with mother feces was significantly greater than the percentage attributed to child feces. This is unexpected given that disposal of child feces in Bangladesh typically

involves caretaker handling of feces[149]. On child hands, which are important because children are more vulnerable to diarrheal illnesses, there was no significant difference between the percentage of the bacteria attributed to mother and child feces (CF: 0.8% and MF: 1.1%). About 3% of the bacteria on hands was soil-associated which was greater than the contribution from feces. In Bangladeshi households soil is a common flooring material and children, especially of crawling age, make frequent hand contact with soil[78]. The hand-soil pathway was also shown to be a significant intra-household transmission pathway in the overlapping ASV analysis. More bacteria on mother and child hands were identical to bacteria in the soil in their own household compared to soil from other households. Soil has also been shown to be a main environmental contributor to the bacterial community on caretaker hands in other LMICs, such as Tanzania[150]. In Tanzania, the most abundant families of bacteria on caretaker hands were soil-associated *Rhodobacteraceae* and *Nocardioidaceae*. The most abundant bacteria families in our study on mother hands were *Micrococcaceae* and *Staphylococcus*, common skin bacteria, followed by *Intrasporangiaceae* and *Moraxellaceae* which are soil-associated (Figure C.2).

The highest percentage of bacteria in the stored water microbial community was attributed to bacteria on mother hands, aside from the unknown category. In Bangladesh, water is typically collected from shallow tubewells and stored in a Kolshi (metal container) that can be covered or uncovered. Water is then dispensed by pouring directly into a vessel for drinking or into an intermediary container, such as a pitcher, or by reaching into the container with a cup. This can result in the re-contamination of source water, specifically by mother hands. However, there was only a borderline significant difference in the number of identical bacteria on hands and in stored water within the same household compared to outside households (Figure 4.3 and Table C.2). Nevertheless, the re-contamination of stored water was observed in another WASH Benefits study where the concentration of indicator *E. coli* was higher in stored water compared to source water[46]. Stored drinking water has also been studied using 16S amplicon sequencing in Cameroon, where they found higher alpha diversity in stored water compared to source water[138]. The study theorized that the higher alpha diversity in stored water was due to introduction of bacteria from soil, humans and air during storage. *Acintobacter*, a common water related bacteria, was one of the most abundant genera in Cameroon and in our study. Other abundant bacteria in stored water in Camaroon were from the genera *Pseudomonas* and *Enterobacteriacaea*, whereas the next most abundant genera in Bangladesh were *Brevundimonas* and *Jaribacter*. *Jaribacter* is from the family *Intrasporangiaceae* which was also one of the most abundant families on mother hands (Figures C.3 and C.2).

We demonstrated that the hand to hand transmission between mother and child is an important pathway for the exchange of bacteria. Looking at the overall microbial community composition, mother and child hands were very similar (Figure 4.2 and 4.1). This is unsurprising given that the two matrices can harbor similar skin-associated bacteria but there were significantly more identical bacteria on mother and child hands within the same house-

holds compared to other households. Shared skin microbiomes have also been observed in families in the United States[151, 152]. Although we identified very few ASVs that could be pathogens, we demonstrate the potential for pathogenic bacteria, if present, to be shared through mother and child hand interactions.

Of the ASVs that could be enteric bacterial pathogens, many were not present in reservoirs in the same household, with the exception of *V. cholera* where there was some overlap between mother hands and stored water. It should be noted that identification of potential pathogens using 16S rRNA gene sequencing has many limitations. The presence of the 16S rRNA gene does not indicate viable organisms. Also, most species have only a fraction of strains or subtypes that are pathogenic in comparison to all species. For example, there are numerous serotypes of *V. cholera*, but only two (serotypes 01 and 0129) are pathogenic. In addition, while DADA2 assigns species at 100% identity, the V3-V4 region of the rRNA gene is not enough to differentiate many strains. More genomic context is needed to identify pathogenic strains, which is possible with whole genome sequencing (metagenomics) or qPCR. Even using metagenomics, deep sequencing is needed to identify pathogens that are present in low relative abundance in environmental samples. Additionally, assigning species relies on the accuracy of existing databases. We found no difference in the taxonomy assigned to potential pathogen ASVs between Silva and RDP, but a previous study estimated annotation error rates in these databases to be 17% and 10%, respectively[153]. Therefore, results from 16S analysis should be coupled with specific and sensitive pathogen detection methods (PCR).

Our study has many important limitations. This study utilized samples originally collected for the purpose of other related work (Chapters 2 and 3) and a WASH Benefits stool study. As a result we were limited to the environmental sample types collected for the purpose of Chapters 2 and 3 and fecal samples that were not from the same households as the environmental samples. This study could be improved by investigating more reservoirs of bacteria in the household. For example, there was a high percentage of bacteria that was not sourced to human feces or other environmental sample types (unknown). There was likely a portion of the bacterial community in environmental samples sourced from animal fecal contamination, which was not included in our study. Animal feces from cows, goats, ducks and chickens is very common in rural Bangladesh and has been associated with increased concentrations of fecal indicator organisms as well as increased prevalence of enteric pathogens[10, 20, 154]. Sampling other microbiomes such as the forehead, mouth, and groundwater could have improved our ability to account for more of the bacteria found on hands and in stored water. Lastly, in our existing analysis of intra-household transmission, identifying the number of identical ASVs does not take into account relative abundance, which is included in Sourcetracker models. Using sourcetracker to investigate transmission pathways is also complicated by the inability to determine directionality. For example, mother hands and child hands had the most associated bacteria but it is unknown what proportion of the bacteria originated from mothers versus children.

We demonstrate the ability of bacterial community analysis to better understand potential pathways of enteric pathogen transmission from a non-targeted approach. We show relatively low levels of human fecal bacteria in the bacterial community in environmental reservoirs but even low levels of fecal contamination have the potential to transmit pathogens. We were able to identify potentially important pathways of bacteria transfer between mother and child hands, soil and child hands, and soil and mother hands at the household level. This non-targeted approach may be more appropriate to use in LMIC settings where sensitivity and specificity of source tracking markers is impacted by the potential sharing of microbial communities between humans and animals. Whole bacterial community analysis also does not require validation of a single marker in a region different from where it was designed. While 16S sequencing proved useful in determining the importance of different transmission pathways in households, 16S rRNA gene sequencing is limited in its ability to identify specific enteric pathogens. We identified sequences of potential pathogens but it is likely only a small subset of the classified species contain pathogens. Targeted gene amplification specific to pathogens (PCR, qPCR, ddPCR) is a more appropriate tool for identifying pathogens, however does require a priori selection of targets. In this study, over half of the ASVs that could be potential pathogens were assigned to the genus *Campylobacter* which was not a pathogen investigated using qPCR or PCR in Chapters 2 and 3. Future studies of enteric pathogen transmission in rural Bangladesh might consider targeted analysis for *Campylobacter jejuni* or *Aeromonas hydrophila* or conduct similar analyses using metagenomic pipelines.

Chapter 5

Conclusions

This research aimed to better understand household reservoirs of fecal contamination in low- and middle-income countries (LMICs), specifically Bangladesh. Investigating how enteric pathogens are transferred from these household reservoirs (hands, water, soil, fomites, food and flies) to new hosts through the fecal-oral route ultimately will help us better understand how to reduce pathogen infections at the household scale. To achieve these goals, I specifically: 1) quantified the impact of a household level sanitation intervention on concentration and prevalence of pathogens and indicators in soil, stored water and on hands; 2) investigated the relationship between indicators and pathogens in these reservoirs to both understand how well indicators perform in this context and identify specific sources of fecal contamination; 3) employed 16S rRNA gene sequencing to investigate interrelated pathways without relying on a singular indicator organism of fecal contamination.

5.1 Chapter 2

5.1.1 Summary

In Chapter 2, the impact of a sanitation intervention (dual pit latrines, sani-scoop, child potties) in rural Bangladeshi household compounds was measured by assessing prevalence ratios, differences, and changes in concentration of pathogen genes and host-specific fecal markers. There were few significant differences between the prevalence of pathogenic *E. coli*, norovirus,*Giardia* genes, and microbial source tracking markers in the sanitation and control arms. This work complimented many other studies from the WASH Benefits randomized controlled trial that investigated the impacts of WASH interventions on prevalence of diarrhea in children and child growth[39], and indicator *E. coli* in household reservoirs[46]. The main study on child health determined that the prevalence of diarrhea was lower in the sanitation arm than in the control arm although the overall prevalence was relatively low even in the control group (<6%)[39]. This difference in prevalence of diarrhea was not evident in the empirical measurements of indicators and pathogens reported in this dissertation and

in a complimentary study on indicator *E. coli*[46]. Inconsistencies between our findings and the health outcomes could be due to the difficulty of detecting pathogens in the environment (methods not sensitive enough), studying reservoirs that may not have been the main reservoirs of pathogens (food or fomites were not sampled), and investigating pathogens that may not have been the main etiologic agents of disease.

Looking at the findings in our study, one reason the prevalence of microbial source tracking markers and pathogen genes remained high in the sanitation arm in all sample types could be due to inadequate removal of child and animal feces from the household environment. The animal fecal marker, BacCow, was present in over 90% of soil and hand samples and 70% of stored water. Despite high reported use of the animal scooper in the sanitation arm, animal feces were visible in most household courtyards; 91% of households had visible chicken feces and 27% of household compounds had visible cow feces. Although the human marker was much less prevalent than the animal fecal marker, only 36% of both the sanitation and control arms reported using a scoop or hoe to handle child feces and just over half of the sanitation arm reported that children defecated in a latrine or the child potty. Another reason there was little impact of the sanitation intervention could be due to the scale of the intervention, which was limited to the household. Inter-household transmission can occur through a variety of pathways involving the transfer of pathogens through people, water bodies and surfaces. Adults and children can introduce fecal pathogens from outside the household environment from schools, neighboring compounds, agriculture fields and other places of work[67, 155].

5.1.2 Future Directions

This work investigated modest improvements in sanitation. Both sanitation and control arms had similar coverage of pit latrines (almost 100%), and the quality upgrade from unimproved to improved sanitation is a relatively small step. Sanitation services on more extreme ends of the spectrum of quality have a greater difference in containment of feces and therefore are more likely to result in detectable differences in household level fecal contamination. Future work should examine the impact of varying tiers of sanitation quality ranging from open defecation (no sanitation) to safely managed sanitation, as defined by the Sustainable Development Goals, with the inclusion of management of child and animal feces. For example, there may be a significant difference in levels of enteric pathogens between households with open defecation and improved pit latrines, open defecation and safely managed sanitation, or even improved pit latrines and safely managed sanitation.

Above I suggest that future sanitation efforts should include safe management of child and animal feces but effective strategies to handle these fecal sources are also areas that need more attention. Although scoops and child potties were included for containment of animal and child feces from the courtyard, the presence of the animal fecal marker was ubiquitous throughout the household and just over half of households used the child potty or a latrine.

There is still much work to be done to investigate the reduction in household level child and animal fecal contamination from other interventions such as household flooring, separation of animals from human living areas, reusable diapers, or child latrine seats[40, 156].

Finally, all potential household reservoirs of enteric pathogens were not investigated. In a study of the most important pathways of transmission of fecal indicator bacteria in Ghana, food was the largest contributor of indicator *E. coli* to total exposure to children under five[77]. Transmission through food is one of the most complex pathways as contamination can come from feces in the field (e.g., contaminated produce), water used to clean food, and surfaces and hands used to prepare the food. In the related WASH Benefits study on indicator *E. coli* there was no difference in the prevalence or concentration of indicator *E. coli* in food between the sanitation and control arms; however, the prevalence of pathogen genes was not measured[46]. Future research should investigate the many potential transmission routes that contribute to the contamination of food and identify how interventions can impact the presence of enteric pathogens along food related pathways.

5.2 Chapter 3

5.2.1 Summary

In Chapter 3, the relationships between indicator *E. coli*, microbial source tracking markers, select enteric pathogen genes and potential sources of enteric pathogens was evaluated in stored drinking water, soil and on mother and child hands. The concentration of indicator *E. coli* was positively associated with the prevalence of pathogenic *E. coli* genes in all sample types. Although this finding is intuitive, it was previously unknown how frequently *E. coli* virulence genes were present in samples positive for indicator *E. coli*. The presence of naturalized non-pathogenic *E. coli* in Bangladesh, a warm tropical environment, could have had a strong effect on this relationship, such that samples positive for indicator *E. coli* may not have been positive for pathogenic *E. coli*, and there would not have been a quantitative relationship between the two. Given the current need to rely on indicators to assess fecal contamination in the field, it is significant that in this study context indicator *E. coli* concentrations provided quantitative information on the relative levels of pathogenic *E. coli* in different household reservoirs.

As noted in Chapter 2, animal fecal management is extremely important in the rural Bangladeshi household environment. In this chapter, I provide the additional evidence that increased concentrations of the animal marker were associated with the presence of *Giardia lamblia*, any *E. coli* virulence gene, and the specific *E. coli* virulence genes *eaeA* on mothers' hands. This suggests that the more animal feces, the more likely we were to detect zoonotic pathogen genes. While it is known that pathogenic *E. coli* and *Giardia* can be transmitted by animals, the evidence for this is largely based on isolating pathogens from animal feces. The potential for zoonotic transmission is rarely evaluated in the context of environmental

samples and demonstrated in the field. This finding is also significant because it was only observed on mother hands and not in any other household reservoir. Mothers are the family member most likely to be performing housekeeping tasks such as cleaning animal feces from their home, resulting in close proximity to animal feces and increased risk of infection.

Another major finding of this chapter is the lack of association between the human fecal marker and human specific pathogens. This could be due to the poor performance of HumM2 as an indicator of human feces in this context (i.e. it was not 100% specific to humans) or the use of a bacteria (*Bacteroidales*) as an indicator of a virus (norovirus).

5.2.2 Future Directions

Future work could develop more sensitive and specific microbial source tracking markers for human feces in developing countries. Although most microbial source tracking markers are developed using feces from humans in high-income countries, they are arguably more useful in LMICs where the load of fecal contamination and resulting cases of diarrheal illnesses are a more pressing problem. More work is needed to identify appropriate bacterial markers in the intestinal flora of humans in rural Bangladesh that are both unique and abundant to improve specificity and sensitivity of assays. The poor specificity of existing markers can potentially also be overcome by utilizing whole bacterial communities as discussed in Chapter 4.

In this chapter, indicator *E. coli* was found to be a good indicator of pathogenic *E. coli* genes in rural Bangladesh which has implications for risk assessment. In these resource constrained areas, fecal indicators are commonly used to suggest the potential for pathogens and assess risk of infection. Higher concentrations of indicator bacteria were indicative of a higher risk of exposure to bacterial pathogens. This analysis was limited to pathogens and indicator from the same genus and it would be useful to explore the relationship between indicator *E. coli* and other bacterial pathogens such as *Salmonella enterica*, *Aeromonas hydrophila*, and *Campylobacter jejuni*.

5.3 Chapter 4

5.3.1 Summary

SourceTracker, a Bayesian computational tool, was used to quantify the overlap in bacterial community between related household reservoirs. There was a significant overlap in bacterial community between mother and child hands, which is not surprising given that both mother and child hands have skin-associated bacteria. In stored water, the highest percentage of the bacterial community was attributed to mother hands (aside from unknown). Recontamination of drinking water during storage is a concern in low-and middle-income countries without continuously supplied piped water on the premises. While other studies

have investigated contamination introduced during storage, I demonstrate the potential for bacterial contamination to occur specifically through contact with mother hands. In this chapter, I also estimated the relative contribution of bacteria from child and mother feces to soil, stored water, and hands. The percentage of fecal- associated bacteria in all sample types was very low (<2%) in all household reservoirs and was similar to percentages found on hands of adults and children in the United States. Although the percentage of fecal-associated bacteria was low, small amounts of fecal contamination can still contain enteric pathogens that lead to illness.

In this chapter I investigated intra-household transmission by quantifying the number of shared amplicon sequence variants (ASVs) between reservoirs inside and outside households. Through this analysis I show that there are multiple important potential pathways of pathogen transmission, specifically through contact between mother and child hands and contact between hands and soil. Interestingly, the number of sequence variants that were identical between mother hands and stored water from the same household was not significantly greater than the number of sequence variants shared between mother hands and stored water in other households. Although the ASV analysis found no significant overlap in ASVs on mother hands and in stored water within households, the community in stored water was most associated with mother hands in the Sourcetracker model. The discrepancy in the findings between the ASV analysis and Sourcetracker could be due to methodological differences (i.e. Sourcetracker accounts for relative abundance and overlapping ASVs does not). Lastly, I identified 22 amplicon sequence variants that could be derived from potential pathogens, half of which were from the genus *Campylobacter*. However, it is difficult to draw conclusions on the relative abundance of pathogens from 16S sequences because many sequence variants were identified at the genus level only, which is insufficient to determine pathogenicity.

5.3.2 Future Directions

The motivation for this chapter was determining how non-targeted bacterial analysis could be used to help identify important reservoirs of fecal contamination. One of the main limitations to this study was that it was not possible to obtain samples of animal feces for sequencing analysis thus animals were not included as potential sources of fecal contamination in the analysis. Given the high prevalence of BacCow in environmental samples, not including samples of animal feces is a major shortcomning of the Sourcetracker analysis. The single marker target (BacCow) used in Chapters 2 and 3 was present in the feces of all animals of interest (cows, ducks, goats, chickens). A strength of SourceTracker, compared to BacCow, is that it likely has the potential to differentiate the contributions of each of these animal feces types to household reservoirs. Identifying individual animal sources that contribute the most fecal-associated bacteria is important because different fecal types should be managed in different ways. For example, chicken feces were dispersed throughout courtyards and its small size makes it difficult to remove with scoops or hoes. Cow and goat feces were usually

limited to the areas where the animals were kept and these larger fecal piles are easier to remove with scooping tools.

Through this work I also demonstrated the limits to studying whole bacterial communities with 16S rRNA gene sequencing. While potentially pathogenic ASVs were identified, 16S amplicon sequencing is not sufficient to identify specific pathogens with certainty. PCR is still the most sensitive and reliable method to detect enteric pathogens. Other sequencing approaches such as whole genome sequencing (metagenomics) have yet to be applied to the household environment in low-and middle-income countries. Metagenomics has the potential to identify pathogens at the strain level, although is dependent on sufficient sequencing depth. Whole genome sequencing could also be used to explore the function of bacteria in microbial communities which could be relevant to understanding how pathogens survive in the environment.

5.4 Personal Reflection on Interdisciplinary Research

Through this dissertation research I had the unique opportunity to work at the intersection of environmental engineering and public health. Sanitation in low-and middle-income countries is inherently an interdisciplinary research area that involves the confluence of engineers, public health practitioners, and social scientists. Each discipline has a distinctly different outlook and scientific process to ultimately reach the same goal of reducing diarrheal illness in heavily burdened areas. My work focused on the first two fields, but the contribution from social scientists to the sanitation field of study is equally important.

From the perspective of engineering, the role of an engineer is to design appropriate technologies that will adequately contain and treat waste. Courses on wastewater have historically focused on conventional activated sludge treatment that is commonplace in centralized treatment plants in high-income countries. There has been very little attention given to the "engineering problem" of sanitation in rural areas which is constrained to decentralization. Traditionally, practitioners in the field have relied on pit latrines which, if designed well, adequately collect waste but are unsustainable in the long term because pits fill up. Pits take up a significant amount of land and can contaminate groundwater if not properly constructed and maintained. The added challenge in the environmental engineering field is how to detect pathogens in the environment. Measuring enteric pathogens is important for characterizing a system (like a rural household) in order to design effective sanitation technologies and for testing the efficacy of containment and treatment. Enteric pathogens are extremely difficult to detect due to their very low concentration in environmental matrices, yet these low concentrations are biologically relevant and can still cause illness. The most sensitive detection method is PCR but sensitivity is dependent on achieving a high degree of concentration of samples (e.g. filtration) which is difficult in the field, and still may be inadequate to detect pathogen levels that represent a risk to public health.

From the perspective of public health practitioners, the goal is to carefully design studies to investigate hypotheses of interest. The randomized controlled trial in this dissertation was implemented to understand the impacts of sanitation on human health and household level fecal contamination. Due to the low prevalence of pathogens in the environment, field trials require a very large sample size to be powered to detect a difference in control and intervention groups. Here is where the greatest challenges lie for combining the pathogen detection methods employed by engineers and field studies designed by public health researchers.

Field trials are expensive and require significant personnel to collect and process environmental samples. In this time and resource constrained setting it is difficult to collect adequate volumes of samples to concentrate them enough to reliably detect pathogens. For example, the volume of stored water processed in this study was limited to 100-500 mL. Stored water must be collected from water storage containers; therefore, the volume is limited to the amount stored at the time of collection and that can be transported to the lab. In the WASH Benefits trial, it was infeasible to transport more water or filter onsite as each field enumerator was visiting multiple households per day and samples/equipment had to be carried, in part, by hand. While the hand rinse samples and soil samples were not limited by sample collection, they were impacted by the time it takes to process samples in the lab. For example, we could have performed multiple extractions per soil sample to increase the weight processed, but that would have increased the amount of processing time.

I recommend that these sampling limitations be carefully considered when designing field trials that involve the collection of environmental samples. It is also important that engineers, public health practitioners, and social scientists work together at a very early stage of study design and spend time in the field to fully understand the context. Future collaborations of this nature could first model interventions through an exposure assessment. This has been difficult in the past as measurements of pathogens in household reservoirs and exposure assessments have only recently become more common[67]. Modeling interventions could be useful for determining effective sanitation strategies and generating hypotheses before designing a field trial. Modeling could also be used to estimate the concentration of pathogens in field samples needed to detect a difference in study arms and empirically demonstrate a reduction in risk of infection. Throughout my PhD, the field of public health also provided valuable insight into producing transparent and reproducible research by encouraging the posting of analysis plans prior to data analysis and suggesting replication of analyses by multiple researchers.

Appendix A

Supporting Information for Chapter 2

A.1 Molecular Methods for Processing of Soil, Stored Water, and Hand Rinse Samples

A.1.1 Pathogenic *E. coli* DNA Extraction

DNA was extracted from IDEXX-processed bacteria pellets using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD). DNA was extracted using a QIAcube (Qiagen, Germantown, MD) in batches of 12 (11 samples and 1 blank). The kit protocol for extraction of gram-negative bacteria was modified for use with the QIAcube according to the manufacturer's instructions. After the addition of Proteinase K, lysis proceeded at 56 °C for 5 minutes. Following the addition of Buffer AL, the samples were incubated and agitated at 70 °C for 12 minutes. Volumes of Buffers AW1 and AW2 were increased to 750 µl. 100 µl of buffer AE was added to the spin column, followed by incubation for 1 min and centrifugation for 1 minute at 7500 rpm. This procedure was repeated for a total eluent volume of 200 µl.

A.1.2 Pathogenic *E. coli* Gene Detection

Three sets of multiplexed PCR reactions were used to analyze samples for 5 pathotypes of *E. coli*[49]. PCR was performed using the Type-It Mutation Detection PCR Kit (Qiagen, Germantown, MD). Multiplex reactions were run according to Table A.1. PCR products (8 µl) were run on 2% agarose gels at 110 Volts for 30 minutes, stained with ethidium bromide, and imaged using a Gel Doc 2000 (Bio-Rad, Hercules, CA). In the cases where non-specific amplification was present in gels at a similar length as the product for *lt1*, PCR for the sample was re-run at an annealing temperature of 58 °C. PCR products from a subset of samples run at the higher temperature were positively confirmed by sanger sequencing.

A.1.3 Pathogenic *E. coli* Quality Control

We processed 1 extraction blank for every 2 QIAcube runs for pathogenic *E. coli* genes. All extraction blanks were negative for pathogenic *E. coli* genes. 2 of 72 lab blanks were positive for *eaeA*, 1 of 72 lab blanks was positive for *stx1* and *stx2*, and 2 of 72 lab blanks were positive for *lt1*. All samples corresponding to the date of the positive blanks were removed from analysis.

A.1.4 DNA/RNA Extraction from Filters

DNA and RNA were extracted according to the modified Mobio PowerWater RNA Extraction (Now Qiagen, Germantown, MD) protocol described in Mattioli et al. with a few added modifications[37]. 1 % (v/v) βME was added to PWR1 solution prior to extraction. Samples were vortexed on a horizontal vortex plate for 10 minutes at maximum speed. Typical centrifugation steps, from loading the lysed sample onto the spin columns to the addition of PWR4 and PWR5, were performed on a vacuum manifold. Following the addition of PWR4, samples were placed in a collection tube and centrifuged at 13,000 $\times$ g for 2 minutes. DNA digestion steps were skipped in order to co-extract RNA and DNA. Nucleic acids were eluted using 100 µl of pre-warmed (55 °C) DNase/RNase free water. A small aliquot of nucleic acids was digested with DNase I for norovirus GII analysis. 24 samples were extracted at one time (23 samples + 1 blank). Aliquots were stored at -80 °C prior to qPCR.

A.1.5 DNA/RNA Extraction from Soil

We used a modified Mobio PowerViral protocol described in Boehm et al.[20]. The protocol utilized the Mobio PowerViral Environmental DNA/RNA Isolation Kit (Now Qiagen, Germantown, MD) with the addition of GeneRite lysis buffer (GeneRite, North Brunswick, NJ) to extract DNA and RNA from soil in Bangladesh. The protocol was specifically tested on three different representative soil types from Bangladesh and maximized recovery efficiency of RNA and DNA in all representative soil samples. Our protocol included bead beating for 2 minutes at maximum speed on a Mini-Beadbeater (BioSpec, Bartlesville, OK). We also used a vacuum manifold for the centrifugation steps, as described above. Nucleic acids were eluted using 100 µl of pre-warmed (55 °C) DNase/RNase free water. A small aliquot of nucleic acids was digested with DNase I for norovirus GII analysis. 23 samples were extracted at a time (23 samples + 1 blank). Aliquots were stored at -80 °C prior to qPCR.

A.1.6 Quality Control

6 of 119 extraction blanks were positive for HumM2, 2 of 51 extraction blanks were positive for norovirus GII, 2 of 51 extraction blanks were positive for *G. lamblia*, 5 of 119 extraction blanks were positive for BacCow. 14 of 168 lab blanks were positive for HumM2, 4 of 168 lab blanks were positive for norovirus, 28 of 168 lab blanks were positive for *G. lambli*a, 1

of 168 lab blanks was positive for BacCow. For binary markers (HumM2, norovirus GII, *G. lamblia*), samples corresponding to positive blanks were removed from the analysis according to date and lab technician. For BacCow, all positive blanks amplified below the limit of quantification (BLOQ). Samples corresponding to positive blanks that amplified BLOQ were treated as below the limit of detection (LOD).

A.1.7 Recovery of Nucleic Acids from Bacteria and Viruses

A subset of filter and soil samples were spike with MS2 and *Pseudomonas syringae* pv. phaseolicola (pph6) to test for recovery of RNA and DNA. Recovery efficiencies for each sample type are shown in Table A.2. For DNA efficiency testing, $7.65\times$ 10^4 pph6 cells (determined by flow cytometry) were spiked in to each sample. QPCR was run on DNA extract to determine gene copies of pph6[157]. We assumed 1 gene copy per cell to estimate recovery efficiency. For RNA recovery efficiency, we were unable to measure the concentration of MS2 in units of virus particles or gene copies. Flow cytometry was unable to count viruses due to their small size, therefore we used a plaque assay to estimate the spike concentration. We spiked in 2.41×10^8 PFUs of MS2 to each sample. Recovery efficiency was estimated by normalizing the gene copies of MS2 in samples, determined by qPCR, to gene copies in recovery blanks[158]. Recovery blanks were empty tubes with the spike of MS2 and went through the same extraction process[159].

A.1.8 Recovery of *Cryptosporidium* DNA

Four test filters and four test soil samples were spiked with 10^3-10^4 *Cryptosporidium parvum* oocysts (Waterborne Inc. New Orleans, LA) to test filter and soil extraction protocols for recovery of protozoan DNA. The median (IQR) of recovery of *C. parvum* DNA in filter samples was 49.9 (46.2-58.8) % and 42.4 (27.8-60.7) % in soil samples. Spike concentrations were determined from the reported concentration of oocysts from the vendor, measured using the Neubauer RBC hemocytometer method. Gene copies were measured in the eluent using qPCR.

A.1.9 Quantitative PCR

A.1.9.1 Reaction Setup

QPCR assays, described in Table A.3 were run on a StepOnePlus (Applied Biosystems, Foster City, CA). All cycling conditions were run in accordance with the manufacturer's instructions for the corresponding mastermix. Samples were run in triplicate and each plate contained standards ranging from 10 to 10^5 gene copies/2 μL and 3 no template controls. 2 μL of template was added to each reaction.

A.1.9.2 Standards

Plasmid standards for HumM2, *G. lamblia*, and *Cryptosporidium spp.* assays were ordered from IDT (Coralville, IA) and transformed into *E. coli* using OneShot Top 10 Chemically Competent *E. coli* (Invitrogen, Carlsbad, CA). Plasmids for BacCow, pph6, and norovirus GII assays were generated from previous studies[157, 160]. All plasmids were extracted from *E. coli* using a QIAprep Spin MiniPrep Kit (Qiagen, Germantown, MD) followed by digestion with restriction endonuclease for 1hr at 37°C to linearize. Standards for HumM2, *Cryptosporidium spp.*, and pph6 assays were linearized using BamHI-HF (NEB, Ipswich, MA), BacCow was linearized using ScaI (NEB, Ipswich, MA),*G. lamblia* standard was linearized using Pvul (NEB, Ipswich, MA). Linearized plasmid was purified using a QIAquick PCR Purification Kit (Qiagen, Germantown, MD). Norovirus GII plasmid was then reverse transcribed using a MAXIscript T7 Transcription Kit (Invitrogen, Carlsbad, CA) following purification with a MEGAclear Kit (Invitrogen, Carlsbad, CA). Genomic RNA was extracted from MS2 using a QIAamp Viral RNA Mini Kit (Qiagen, Germantown, MD). Nucleic acids were quantified using a Nanophotometer P300 (Implen, Westlake Village, CA).

A.1.9.3 Inhibition Testing

To test for inhibition in all qPCR assays (BacCow, HumM2, *G. lamblia*, *Cryptosporidium spp.*, pph6, MS2 and norovirus GII), we used the spike and dilute method[54]. Undiluted nucleic acid extract was spiked with 3×10^3- 1×10^5 gene copies of standard. Inhibition was tested on a subset of samples from a total of 10 filters (collected for the purpose of preliminary testing) and 13 soil samples (9 collected for preliminary testing and 4 actual samples). Samples were subsequently diluted and assessed for inhibition based on a comparison of Ct values between diluted and undiluted samples. Inhibition was present in the undiluted sample if the Ct difference between the diluted sample and the undiluted sample was at least 1 cycle less than the expected difference for a specific dilution, accounting for the standard curve efficiency. The expected difference was calculated using the equation from Cao et al. (2012)[54]:

$$\Delta Ct = log_{\text{amp}} dil$$

Where Δ Ct = expected cycle threshold difference for a specific dilution, amp = standard curve amplification factor, dil = dilution factor

The amplification factor was calculated using the standard curve efficiency:

$$amp = 1 + (eff/100)$$

Filter samples were uninhibited for all DNA and RNA assays (Table A.4). In soil samples, *Cryptosporidium spp.* and MS2 assays were inhibited in some or all samples (Table A.5). HumM2, BacCow, *G. lamblia*, pph6, and Norovirus GII were uninhibited for all samples. A 1:10 dilution for extraction efficiency of RNA in soil was used for MS2. We decided not to dilute *Cryptosporidium* due to the resulting increase in the detection limit.

A.1.9.4 Standard Curves

The standard curves generated from each plate run were used to create a master standard curve. We used a linear model with batch effects to account for the plate to plate variation due to the high number of plates processed for many assays[55]. Standard curves determined by mixed models are shown in Table A.6 along with the standard curves that result from pooling the data for comparison.

A.1.9.5 Limits of Detection and Quantification

Most samples did not amplify within the quantifiable range for norovirus GII, *G. lamblia* and HumM2. Samples were considered positive if there was amplification in at least 1 of 3 replicates. Limits of detection (LOD) for norovirus GII and *Giardia* on hands were estimated to be 62.5 and 1020 target copies per 2 hands, respectively. For HumM2, the LOD was 173 target copies per 2 hands, 21-103 target copies per 100 ml of stored water, and 254-478 target copies per gram dry soil. The LOD was determined based on the lowest gene copy that amplified in at least 1 of 3 replicates in each sample type. The range of LODs for stored water and soil correspond to variation in volumes filtered (100 ml - 500 ml) and soil moisture content (0 - 88%). Many samples did amplify within the quantifiable range for BacCow. For BacCow, samples that amplified below the limit of quantification (LOQ) were assigned the midpoint between the LOQ and the LOD. Samples that did not amplify were treated as below the LOD and assigned a value of half the LOD.

A.2 Tables

Table A.1: Genes, primers, product size, and run conditions for three multiplex reactions used to detect pathogenic *E. coli*.

	Path *E. coli* Type	Gene	Primer Name	Sequence	Product Size	Anneal Time (sec)	Anneal Temp (°C)
Multiplex 1	EAEC	*aggR*	aggRKs1-F	GTATACACAAAAGAAGGAAGC	254	90	60
			aggRks2-R	ACAGAATCGTCAGCATCAGC			
	EPEC/EHEC	*eaeA*	eaeA-F	TCAATGCAGTTCCGTTATCAGTT	482		
			eaeA-R	GTAAAGTCCGTTACCCCAACCTG			
Multiplex 2	EHEC	stx1	stx1-F	AAATCGCCATTCGTTGACTACTTCT	370	90	58
			stx1-R	TGCCATTCTGGCAACTCGCGATGCA			
	EHEC	*stx2*	stx2-F	CAGTCGTCACTCACTGGTTTCATCA	283		
			stx2-R	GGATATTCTCCCCACTCTGACACC			
	EIEC	*ipaH*	ipaIII-F	GTTCCTTGACCGCCTTTCCGATACCGTC	619		
			ipaIV-R	GCCGGTCAGCCACCCTCTGAGAGTAC			
Multiplex 3	ETEC	*lt1*	lt1-F	TCTCTATGTGCATACGGAGC	322	180	51
			lt1-R	CCATACTGATTGCCGCAAT			
	ETEC	*st1b*	st1b-F	ATTTTTCTTTCTGTATTGTCTT	192		
			st1b-R	CACCCGGTACAAGCAGGATT			

Table A.2: Recovery efficiency of *Pseudomonas syringae* pv. phaseolicola (Pph6) and MS2 for all samples types and number of samples tested.

	Pph6 Recovery		**MS2 Recovery**	
	n	Recovery Efficiency (%)	n	Recovery Efficiency (%)
		median (IQR)		median (IQR)
Mother Hands	51	39.1 (30.1-44.1)	50	16.2 (5.7-38.4)
Child Hands	39	24.8 (10.7-38.2)	36	8.7 (4.8-15.5)
Stored Water	39	35.3 (29.1-36.5)	NA	NA
Soil	46	56.2 (36.5-97.0)	13	6.5 (6.1-9.0)

Table A.3: Primer sequences, probe sequences, reaction volume, bovine serum albumin (BSA) concentration, master mix, and qPCR standard for all qPCR assays used in this study.

Target	Sequence (5-3) (concentration in rxn)	Reaction Vol(ml)	BSA (mg/ml)	Master Mix	Standard
HumM2[51]	F-CGTCAGGTTTGTTTCGGTATTG (1.2 mM) R-TCATCACGTAACTTATTTATATGCATTAGC (1.2 mM) P-FAM /TATCGAAAATCTCACGGATTAACTCTTGTGTACGC/TAMRA (0.1mM)	25	0.2	TaqMan Env	Plasmid
BacCow[53]	F-CCAACYTTCCCGWTACTC (0.4 mM) R- GGACCGTGTCTCAGTTCCAGTG (0.4 mM) P-FAM/TAGGGGTTCTGAGAGGAAGGTCCCCC/TAMRA (0.08 mM)	25	0.05	TaqMan Env	Plasmid
Norovirus GII[113]	F-ATGTTCAGRTGGATGAGRTTCTCWGA (0.4 mM) R-TCGACGCCATCTTCATTCACA (0.4 mM) P-FAM/AGCACGTGGGAGGGCATCG/BHQ (0.2 mM)	20	0.4	TaqMan Fast Virus 1-Step	Plasmid
Giardia lamblia (β-Giardin P241)[161]	F -CATCCGCGAGGAGGTCAA (0.3 mM) R- GCAGCCATGGTGTCGATCT (0.3 mM) P- FAM/AAGTCCGCCGACAACATGTACCTAACGA/BHQ (0.2 mM)	25	0.2	TaqMan Universal	Plasmid
Cryptosporidium spp. (JVA)[162]	F-ATGACGGGTAACGGGGAAT (0.6 mM) R-CCAATTACAAAACCAAAAAGTCC (0.6 mM) P-FAM/CGCGCCCTGCCTGCCTTCCTTAGATG/BHQ (0.12 mM)	25	0.2	TaqMan Universal	Plasmid
Phi6[157]	F-GGTTCTGGCCCGATGATG (0.9 mM) R-CAGGCCCTAGCGTGGAAAC (0.9 mM) P-FAM/CACTGAAAAGGCTGTTCA/MGBNFQ (0.25 mM)	25	0.2	TaqMan Universal	Plasmid
MS2[158]	F-CGGCTGCTCGCCGATA (0.9 mM) R-ACTTCGCTTCTCGAGCGATAC (0.9 mM) P-FAM/CCGTACCTCGGGTTTCCGTCTTGCT/TAMARA (0.25 mM)	20	0.2	TaqMan Fast Virus 1-Step	Genomic RNA

Table A.4: Difference in Ct values between diluted and undiluted filter samples. Expected Ct values are embolden.

BacCow		1:2-und	1:5-und	1:10-und	
	MH 5	0.96	2.39	3.26	
	CH 10	0.97	2.27	3.26	
	CH 5	1.13	2.52	3.69	
	SW 9	0.85	2.03	2.85	
	Δ Ct	**1.03**	**2.39**	**3.42**	
HumM2		1:2-und	1:5-und		
	SW 3	1.14	2.58		
	CH 3	1.10	2.58		
	SW 9	1.02	2.32		
	MH 5	1.15	2.34		
	Δ Ct	**1.01**	**2.36**		
Giardia		1:2-und	1:5-und	1:20-und	
	SW 9	0.96	2.66	4.83	
	SW 4	1.02	2.64	4.84	
	MH 2	1.12	2.45	4.63	
	CH 3	1.16	2.50	4.61	
	Δ Ct	**1.03**	**2.38**	**4.44**	
Crypto		1:2-und	1:5-und	1:20-und	
	SW 9	1.15	3.06	5.08	
	SW 4	1.34	2.81	5.04	
	MH 2	1.17	2.65	4.63	
	CH 3	1.09	2.34	4.71	
	Δ Ct	**1.04**	**2.41**	**4.49**	
Pph6		1:2-und	1:5-und		
	SW 1	1.09	2.25		
	SW 5	1.02	2.44		
	CH 9	1.14	2.63		
	MH 5	1.18	2.51		
	Δ Ct	**1.05**	**2.43**		
MS2		1:2-und		1:10-und	1:60-und
	CH 5	0.86		3.44	11.20
	SW 9	0.89		3.04	8.23
	CH 10	0.94		3.51	5.46
	Δ Ct	**1.04**		**3.47**	**6.17**
Noro GII		1:2-und	1:5-und	1:10-und	
	MH 7	0.93	2.27	3.24	
	CH 6	0.84	2.09	3.09	
	SW 3	0.97	2.46	3.70	

SW 6	0.67	1.88	2.55
Δ Ct	**1.02**	**2.37**	**3.39**

Table A.5: Difference in Ct values between diluted and undiluted soil samples. Expected Ct values are embolden.

HumM2		1:2-und	1:5-und			
	S2	0.86	2.34			
	S4	1.10	2.35			
	S9	0.88	2.93			
	Δ Ct	**1.03**	**2.38**			
	43209	0.82	1.83			
	43247	1.18	2.15			
	43066	1.04	2.03			
	43165	0.90	1.99			
	Δ Ct	**1.04**	**2.42**			
Giardia		1:2-und	1:5-und	1:10-und	1:20-und	
	S1	0.81	2.24	3.29	4.44	
	S7	0.98	2.22	3.37	4.13	
	S10	0.78	2.12	3.21	4.09	
	Δ Ct	**1.00**	**2.32**	**3.32**	**4.32**	
	43209	0.59	2.25			
	43247	0.72	1.86			
	43066	0.48	1.63			
	43165	0.68	2.08			
	Δ Ct	**1.08**	**2.52**			
Crypto		1:2-und			1:20-und	1:100-und
	S1	-1.27			0.86	3.03
	S7	0.53			3.70	6.25
	S10	0.63			4.07	6.11
	43209	-0.08			3.06	5.27
	43247	-0.54			1.33	4.08
	43066	-0.94			0.71	3.11
	43165	-0.38			2.03	4.31
	Δ Ct	**1.09**			**4.69**	**7.21**
Pph6		1:2-und	1:5-und	1:10-und	1:20-und	
	S4	0.78	1.90	3.16	4.08	
	S5	1.17	2.46	3.59	4.76	
	S9	0.88	2.41	3.40	4.41	
	Δ Ct	**1.07**	**2.5**	**3.57**	**4.64**	
	43209	1.09	2.25			
	43247	1.02	2.44			
	43066	1.14	2.63			
	43165	1.18	2.51			
	Δ Ct	**1.05**	**2.43**			
		1:2-und		1:10-und	1:20-und	1:100-und
	43209	0.91		2.88	4.08	6.65

BacCow	43247	1.02		3.62	4.47		6.67	
	43066	0.84		2.89	3.99		5.99	
	43165	0.81		3.02	4.19		6.63	
	Δ Ct	**1.14**		**3.81**	**4.95**		**7.62**	
		1:2-und	1:5-und	1:10-und	1:20-und	1:50-und	1:100-und	1:10^3-und
	S2	-0.57	-0.48	-0.23	0.87	4.68	6.30	9.41
	S10	-0.87	-0.40	-0.34	1.45	2.22	3.99	8.80
	Δ Ct	**1.03**	**2.4**	**3.43**	**4.46**	**5.16**	**6.08**	**9.12**
MS2	43209		-3.71		-3.43		-2.19	1.16
	43247		-0.22		1.97		4.40	6.12
	43066		-2.26		-1.14		1.09	4.82
	43165		-3.86		-2.60		-0.99	2.27
	Δ Ct		**2.12**		**3.95**		**6.08**	**9.12**

Table A.6: Number of runs and threshold settings for all qPCR assays used in this study. Efficiency, R^2, and y-intercept for standard curves determined by mixed models and pooling data.

Assay	**Threshold**	**Runs**	**Mixed Model**			**Pooled**		
			Efficiency	R^2	Y_{int}	Efficiency	R^2	Y_{int}
HumM2	0.02	111	92.23	0.99	38.07	93.51	0.97	39.02
BacCow	0.03	115	90.26	0.99	38.54	90.25	0.99	38.56
Norovirus GII	0.03	49	91.87	0.99	39.03	92.28	0.96	39.42
G. lamblia	0.02	55	92.46	0.99	37.77	92.56	0.98	37.42
Crypto spp.	0.02	23	89.58	0.99	41.25	89.58	0.98	39.91
Pph6	0.02	11	92.78	0.99	39.80	93.95	0.91	39.40
MS2	0.10	7	100.92	0.99	44.70	99.46	0.92	42.47

Table A.7: Mean $\log_{10}$ concentrations of BacCow in soil (gene copies/g dry weight) and stored water (gene copies/100 ml) samples for each sensitivity analysis scenario.

	Mean $\log_{10}$ Concentration		Δ $\log_{10}$ Concentration	p-value
	C	S		
Scenario 1				
Stored Water	1.27 (1.20-1.34)	1.24 (1.17-1.32)	-0.02 (-0.13 to 0.09)	0.78
Soil	3.34 (3.20-3.47)	3.35 (3.19-3.51)	0.04 (-0.16 to 0.23)	0.72
Scenario 2				
Stored Water	1.70 (1.64-1.77)	1.67 (1.61-1.74)	-0.03 (-0.12 to 0.07)	0.58
Soil	3.66 (3.56-3.77)	3.67 (3.54-3.79)	0.02 (-0.13 to 0.17)	0.78
Scenario 3				
Stored Water	1.74 (1.67-1.80)	1.71 (1.64-1.78)	-0.02 (-0.12 to 0.07)	0.64
Soil	3.71 (3.61-3.81)	3.71 (3.59-3.83)	0.02 (-0.13 to 0.17)	0.81
Scenario 4				
Stored Water	1.24 (1.17-1.31)	1.21 (1.14-1.29)	-0.02 (-0.13 to 0.09)	0.71
Soil	3.29 (3.15-3.43)	3.31 (3.15-3.47)	0.04 (-0.16 to 0.24)	0.71

Table A.8: Prevalence and 95% confidence interval (95% CI) of pathogenic *E. coli* virulence genes, microbial source tracking markers, norovirus GII gene, and *Giardia* gene in the sanitation and control arms.

	Control		**Sanitation**	
	N	Prevalence (95% CI)	N	Prevalence (95% CI)
Child Hands				
Any ECVG	183	32.2 (25.7-38.7)	181	32.6 (25.8-39.4)
eaeA (EPEC/EHEC)	184	13.0 (7.6-18.5)	183	9.8 (5.8-13.9)
aggR (EAEC)	187	12.3 (7.8-16.8)	186	11.3 (6.9-15.7)
stx1/stx2 (EHEC)	186	13.4 (8.6-18.3)	184	16.8 (11.2-22.5)
ipaH (EIEC)	187	NA	186	0.5 (0.0-1.6)
st1b/lt1 (ETEC)	185	9.7 (5.6-13.9)	184	7.6 (3.4-11.8)
BacCow	183	99.5 (98.4-100.0)	182	95.6 (92.4-98.9)
HumM2	167	26.3 (19.7-33.0)	170	17.6 (12.0-23.3)
Norovirus	167	4.2 (1.2-7.2)	169	4.1 (1.3-7.0)
Giardia	157	5.1 (1.7-8.5)	154	4.5 (1.3-7.8)
Mother Hands				
Any ECVG	367	23.7 (19.7-27.7)	353	20.4 (16.5-24.3)
eaeA (EPEC/EHEC)	370	8.1 (5.7-10.5)	359	7.2 (4.5-9.9)
aggR (EAEC)	376	9.3 (6.3-12.3)	362	6.1 (3.6-8.5)
stx1/stx2 (EHEC)	374	9.1 (5.9-12.3)	359	7.5 (4.7-10.4)
ipaH (EIEC)	376	1.1 (0.0-2.1)	362	0.6 (0.0-1.3)
st1b/lt1 (ETEC)	374	4.0 (2.2-5.8)	358	3.9 (2.0-5.8)
BacCow	368	97.6 (96.0-99.1)	358	95.8 (93.9-97.7)
HumM2	325	18.5 (13.8-23.1)	328	17.7 (13.3-22.1)
Norovirus	344	2.9 (1.0-4.8)	342	3.2 (1.4-5.0)
Giardia	302	3.0 (0.9-5.0)	301	1.7 (0.0-3.6)
Stored Water				
Any ECVG	381	37.5 (33.0-42.1)	343	36.4 (31.3-41.6)
eaeA (EPEC/EHEC)	380	15.3 (12.0-18.6)	351	14.5 (11.0-18.1)
aggR (EAEC)	386	10.4 (7.4-13.4)	356	11.8 (9.0-14.6)
stx1/stx2 (EHEC)	385	16.6 (12.6-20.6)	352	16.5 (12.6-20.4)
ipaH (EIEC)	386	1.8 (0.5-3.1)	356	0.6 (0.0-1.3)
st1b/lt1 (ETEC)	383	10.4 (7.6-13.3)	350	6.6 (3.6-9.5)
BacCow	363	70.0 (65.1-74.8)	342	67.0 (61.6-72.4)
HumM2	337	3.6 (1.5-5.6)	315	1.6 (0.2-2.9)
Soil				
Any ECVG	382	61.5 (56.4-66.6)	358	58.9 (53.9-64.0)
eaeA (EPEC/EHEC)	383	38.6 (33.2-44.1)	360	36.7 (31.7-41.6)
aggR (EAEC)	390	17.4 (13.5-21.4)	365	14.2 (11.0-17.5)
stx1/stx2 (EHEC)	388	26.0 (21.6-30.4)	361	23.5 (19.2-27.9)
ipaH (EIEC)	390	1.5 (0.40-2.7)	365	0.8 (0.0-1.7)
st1b/lt1 (ETEC)	387	17.3 (13.3-21.3)	360	13.1 (9.6-16.5)

BacCow	331	91.2 (88.4-94.1)	315	90.2 (86.1-94.2)
HumM2	332	18.4 (14.3-22.5)	314	22.0 (17.3-26.6)

Table A.9: Unadjusted and adjusted prevalence ratios (PR) and 95% confidence intervals (95% CI) comparing the prevalence of pathogenic *E. coli* virulence genes, norovirus GII gene, *Giardia* gene, and microbial source tracking markers on child hands, mother hands, stored water and soil in the sanitation and control arms. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value <0.01, correcting for multiple comparisons.

	Unadjusted		Adjusted	
	PR (95% CI)	p-value	PR (95% CI)	p-value
Child Hands				
Any ECVG	1.06 (0.75-1.49)	0.75	1.00 (0.76-1.33)	0.99
eaeA (EPEC/EHEC)	0.82 (0.41-1.61)	0.56	0.73 (0.40-1.33)	0.30
aggR (EAEC)	0.91 (0.50-1.67)	0.76	0.90 (0.54-1.51)	0.69
stx1/stx2 (EHEC)	1.30 (0.73-2.29)	0.37	1.16 (0.72-1.85)	0.54
ipaH (EIEC)	NA		NA	
st1b/lt1 (ETEC)	0.77 (0.33-1.78)	0.54	0.80 (0.41-1.56)	0.52
BacCow	0.96 (0.93-1.00)	**0.05**	0.95 (0.92-0.97)	**<0.001**
HumM2	0.67 (0.41-1.11)	0.12	0.66 (0.44-0.99)	**0.05**
Norovirus	1.00 (0.33-3.03)	0.99	1.02 (0.39-2.64)	0.97
Giardia	1.00 (0.27-3.71)	1.00	1.00 (0.32-3.10)	1.00
Mother Hands				
Any ECVG	0.89 (0.68-1.15)	0.38	0.87 (0.68-1.11)	0.26
eaeA (EPEC/EHEC)	0.91 (0.51-1.62)	0.75	0.90 (0.51-1.58)	0.72
aggR (EAEC)	0.66 (0.38-1.14)	0.14	0.67 (0.40-1.13)	0.13
stx1/stx2 (EHEC)	0.84 (0.48-1.46)	0.54	0.82 (0.51-1.32)	0.42
ipaH (EIEC)	0.62 (0.10-3.72)	0.60	0.52 (0.09-2.88)	0.45
st1b/lt1 (ETEC)	0.98 (0.47-2.05)	0.96	0.97 (0.51-1.85)	0.93
BacCow	0.98 (0.96-1.01)	0.28	0.98 (0.96-1.01)	0.24
HumM2	1.01 (0.66-1.53)	0.97	0.95 (0.67-1.36)	0.79
Norovirus	1.16 (0.42-3.18)	0.78	1.07 (0.46-2.47)	0.88
Giardia	0.60 (0.15-2.42)	0.47	0.54 (0.19-1.50)	0.24
Stored Water				
Any ECVG	0.96 (0.79-1.17)	0.67	0.95 (0.79-1.15)	0.60
eaeA (EPEC/EHEC)	0.95 (0.67-1.36)	0.79	0.93 (0.67-1.30)	0.67
aggR (EAEC)	1.16 (0.79-1.70)	0.45	1.12 (0.78-1.60)	0.54
stx1/stx2 (EHEC)	0.97 (0.70-1.34)	0.84	0.98 (0.72-1.33)	0.91
ipaH (EIEC)	0.29 (0.05-1.58)	0.15	0.32 (0.07-1.43)	0.13
st1b/lt1 (ETEC)	0.66 (0.41-1.08)	0.10	0.64 (0.39-1.03)	0.07
BacCow	0.95 (0.85-1.07)	0.42	0.98 (0.88-1.08)	0.64
HumM2	0.46 (0.15-1.39)	0.17	0.45 (0.18-1.11)	0.08
Soil				
Any ECVG	0.96 (0.86-1.08)	0.52	0.96 (0.86-1.06)	0.42
eaeA (EPEC/EHEC)	0.96 (0.80-1.16)	0.68	0.97 (0.82-1.15)	0.71

aggR (EAEC)	0.87 (0.62-1.22)	0.42	0.82 (0.59-1.14)	0.23
stx1/stx2 (EHEC)	0.89 (0.68-1.15)	0.37	0.89 (0.69-1.14)	0.35
ipaH (EIEC)	0.55 (0.14-2.19)	0.40	0.43 (0.09-2.17)	0.31
st1b/lt1 (ETEC)	0.76 (0.53-1.07)	0.12	0.74 (0.54-1.03)	0.07
BacCow	0.99 (0.93-1.04)	0.60	1.00 (0.95-1.05)	0.99
HumM2	1.21 (0.88-1.65)	0.24	1.16 (0.86-1.56)	0.34

Table A.10: Unadjusted and adjusted difference and 95% confidence interval in the $\log_{10}$ concentration of BacCow on hands, in stored water, and soil. Units of concentration are gene copies of BacCow per 2 hands, 100 ml, and dry gram. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	Unadjusted	p-value	Adjusted	p-value
Child Hands	-0.19 (-0.42 to 0.03)	0.09	-0.10 (-0.47 to 0.28)	0.62
Mother Hands	-0.16 (-0.30 to -0.03)	**0.02**	-0.37 (-0.66 to -0.08)	**0.01**
Stored Water	-0.02 (-0.13 to 0.09)	0.78	0.01 (-0.11 to 0.13)	0.91
Soil	0.04 (-0.16 to 0.23)	0.72	0.07 (-0.27 to 0.40)	0.71

Table A.11: Prevalence ratios (PR) and 95% confidence intervals (95% CI) comparing the prevalence of pathogenic *E. coli* virulence genes, norovirus GII gene, *Giardia* gene, and microbial source tracking markers on child hands, mother hands, stored water and soil in the sanitation and control arms in the wet and dry season. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	Wet		Dry		
	N	PR (95% CI)	N	PR (95% CI)	Interaction p-value
Child Hands					
Any ECVG	176	1.08 (0.67-1.75)	188	1.06 (0.59-1.92)	0.96
eaeA (EPEC/EHEC)	180	0.94 (0.37-2.37)	187	0.69 (0.26-1.88)	0.66
aggR (EAEC)	180	0.52 (0.18-1.50)	193	1.46 (0.65-3.28)	0.15
stx1/stx2 (EHEC)	177	1.43 (0.70-2.95)	193	1.19 (0.45-3.17)	0.76
ipaH (EIEC)	180	NA	193	NA	NA
st1b/lt1 (ETEC)	178	0.55 (0.13-2.33)	191	0.98 (0.34-2.82)	0.53
BacCow	172	0.92 (0.86-1.00)	193	1.00 (0.97-1.04)	0.09
HumM2	162	0.54 (0.25-1.17)	175	0.86 (0.41-1.78)	0.40
Norovirus	146	2.49 (0.61-10.24)	190	0.49 (0.08-2.88)	0.16
Giardia	157	1.14 (0.13-9.90)	154	0.92 (0.21-3.95)	0.87
Mother Hands					
Any ECVG	351	0.76 (0.51-1.11)	369	1.06 (0.71-1.58)	0.25
eaeA (EPEC/EHEC)	357	0.57 (0.24-1.38)	372	1.45 (0.67-3.17)	0.12
aggR (EAEC)	357	0.59 (0.27-1.28)	381	0.77 (0.34-1.74)	0.64
stx1/stx2 (EHEC)	352	0.72 (0.33-1.56)	381	0.99 (0.47-2.08)	0.56
ipaH (EIEC)	357	1.46 (0.13-16.88)	381	NA	NA
st1b/lt1 (ETEC)	355	1.81 (0.59-5.59)	377	0.45 (0.14-1.42)	0.10
BacCow	345	0.96 (0.92-1.00)	381	1.01 (0.97-1.06)	0.08
HumM2	314	0.92 (0.51-1.66)	339	1.09 (0.62-1.91)	0.68
Norovirus	317	0.94 (0.22-4.05)	369	1.30 (0.32-5.30)	0.75
Giardia	304	0.59 (0.13-2.57)	299	0.59 (0.10-3.66)	1.00
Stored Water					
Any ECVG	386	0.95 (0.72-1.24)	338	0.97 (0.72-1.29)	0.92
eaeA (EPEC/EHEC)	393	0.96 (0.61-1.52)	338	0.92 (0.51-1.67)	0.91
aggR (EAEC)	393	1.29 (0.75-2.20)	349	1.03 (0.56-1.89)	0.60
stx1/stx2 (EHEC)	388	1.01 (0.64-1.58)	349	0.93 (0.57-1.49)	0.80
ipaH (EIEC)	393	0.50 (0.04-6.93)	349	0.24 (0.02-2.44)	0.69
st1b/lt1 (ETEC)	389	0.42 (0.19-0.95)	344	0.90 (0.45-1.77)	0.18
BacCow	355	0.96 (0.82-1.12)	350	0.95 (0.82-1.09)	0.89
HumM2	350	0.80 (0.15-4.13)	302	0.28 (0.05-1.67)	0.41
Soil					
Any ECVG	354	0.94 (0.80-1.10)	386	1.00 (0.82-1.21)	0.66
eaeA (EPEC/EHEC)	360	0.99 (0.76-1.28)	383	0.95 (0.72-1.25)	0.83
aggR (EAEC)	360	0.92 (0.59-1.44)	395	0.82 (0.45-1.48)	0.76
stx1/stx2 (EHEC)	354	0.99 (0.71-1.38)	395	0.77 (0.49-1.19)	0.38

ipaH (EIEC)	360	0.48 (0.07-3.19)	395	0.59 (0.05-6.42)	0.91
st1b/lt1 (ETEC)	357	0.73 (0.41-1.29)	390	0.79 (0.53-1.18)	0.81
BacCow	292	0.97 (0.90-1.05)	354	1.00 (0.92-1.08)	0.67
HumM2	284	0.81 (0.46-1.42)	362	1.52 (0.98-2.36)	0.10

Table A.12: Prevalence ratios (PR) and 95% confidence intervals (95% CI) comparing the prevalence of pathogenic *E. coli* virulence genes, norovirus GII gene, *Giardia* gene, and microbial source tracking markers on child hands, mother hands, stored water and soil in the sanitation and control arms in the households with < 20 and ≥20 animals. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	<20 animals		≥ 20 animals		
	N	PR (95% CI)	N	PR (95% CI)	Interaction p-value
Child Hands					
Any ECVG	186	1.54 (0.83-2.88)	178	0.71 (0.41-1.25)	0.12
eaeA (EPEC/EHEC)	189	1.33 (0.47-3.80)	178	0.49 (0.15-1.64)	0.28
aggR (EAEC)	194	1.22 (0.40-3.76)	179	0.67 (0.27-1.70)	0.48
stx1/stx2 (EHEC)	191	1.17 (0.46-2.98)	179	1.50(0.62-3.64)	0.74
ipaH (EIEC)	194	NA	179	NA	NA
st1b/lt1 (ETEC)	191	2.41 (0.65-8.94)	178	0.22 (0.04-1.06)	**0.04**
BacCow	189	0.96 (0.90-1.02)	176	0.97 (0.91-1.03)	0.80
HumM2	174	0.32 (0.12-0.82)	163	1.21 (0.62-2.35)	**0.03**
Norovirus	173	2.48 (0.22-27.4)	163	0.47 (0.08-2.87)	0.33
Giardia	163	0.33 (0.03-4.31)	148	4.20 (0.32-54.64)	0.17
Mother Hands					
Any ECVG	381	0.71 (0.47-1.06)	339	1.06 (0.70-1.61)	0.21
eaeA (EPEC/EHEC)	387	0.45 (0.19-1.05)	342	1.70 (0.71-4.08)	**0.03**
aggR (EAEC)	393	0.27 (0.09-0.83)	345	1.15 (0.54-2.48)	0.05
stx1/stx2 (EHEC)	389	0.79 (0.42-1.52)	344	0.92 (0.39-2.18)	0.78
ipaH (EIEC)	393	0.31 (0.03-3.04)	345	NA	NA
st1b/lt1 (ETEC)	390	1.34 (0.40-4.49)	342	0.75 (0.25-2.23)	0.52
BacCow	385	0.99 (0.94-1.04)	341	0.98 (0.95-1.01)	0.70
HumM2	343	0.74 (0.37-1.48)	310	1.32 (0.77-2.25)	0.22
Norovirus	359	0.64 (0.11-3.78)	327	1.66 (0.48-5.67)	0.40
Giardia	321	0.94 (0.17-5.07)	282	0.37 (0.03-5.16)	0.56
Stored Water					
Any ECVG	398	0.94 (0.70-1.27)	326	0.98 (0.72-1.32)	0.88
eaeA (EPEC/EHEC)	403	0.89 (0.50-1.60)	328	0.99 (0.56-1.72)	0.83
aggR (EAEC)	410	1.33 (0.74-2.41)	332	0.99 (0.48-2.06)	0.59
stx1/stx2 (EHEC)	406	0.86 (0.52-1.44)	331	1.07 (0.65-1.78)	0.58
ipaH (EIEC)	410	NA	332	0.30 (0.06-1.38)	NA
st1b/lt1 (ETEC)	405	0.68 (0.30-1.54)	328	0.65 (0.36-1.17)	0.92
BacCow	392	0.89 (0.74-1.07)	313	1.02 (0.88-1.18)	0.30
HumM2	364	0.72 (0.14-3.81)	288	0.19 (0.02-2.05)	0.42
Soil					
Any ECVG	390	1.01 (0.84-1.20)	350	0.92 (0.77-1.10)	0.52
eaeA (EPEC/EHEC)	392	0.84 (0.63-1.13)	351	1.08 (0.79-1.48)	0.30
aggR (EAEC)	400	0.70 (0.43-1.14)	355	1.17 (0.64-2.14)	0.25
stx1/stx2 (EHEC)	396	1.00 (0.66-1.50)	353	0.79 (0.58-1.09)	0.37
ipaH (EIEC)	400	0.84 (0.10-6.83)	355	0.32 (0.03-3.36)	0.57

st1b/lt1 (ETEC)	396	1.15 (0.71-1.86)	351	0.46 (0.23-0.92)	**0.05**
BacCow	341	1.01 (0.93-1.10)	305	0.96 (0.88-1.04)	0.39
HumM2	340	1.37 (0.81-2.32)	306	1.06 (0.69-1.65)	0.49

Table A.13: Prevalence ratios (PR) and 95% confidence intervals (95% CI) comparing the prevalence of pathogenic *E. coli* virulence genes, norovirus GII gene, *Giardia* gene, and microbial source tracking markers on child hands, mother hands, stored water and soil in the sanitation and control arms in the households with < 10 and ≥ 10 individuals. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	<10 individuals		10 individuals		
	N	PR (95% CI)	N	PR (95% CI)	Interaction p-value
Child Hands					
Any ECVG	181	1.36 (0.82-2.27)	183	0.80 (0.47-1.35)	0.18
eaeA (EPEC/EHEC)	181	1.83 (0.53-6.35)	186	0.40 (0.16-1.01)	0.07
aggR (EAEC)	184	0.85 (0.31-2.32)	189	0.94 (0.32-2.78)	0.90
stx1/stx2 (EHEC)	183	1.02 (0.53-1.97)	187	2.20 (0.70-6.90)	0.25
ipaH (EIEC)	184	NA	189	NA	NA
st1b/lt1 (ETEC)	183	5.08 (1.08-23.82)	186	0.11 (0.02-0.77)	**0.01***
BacCow	177	0.93 (0.85-1.01)	188	1.00 (0.97-1.03)	0.13
HumM2	170	0.85 (0.39-1.87)	167	0.54 (0.23-1.28)	0.49
Norovirus	162	1.21 (0.13-11.30)	174	0.81 (0.23-2.84)	0.76
Giardia	165	0.23 (0.01-4.60)	146	3.33 (0.37-30.28)	0.23
Mother Hands					
Any ECVG	356	0.96 (0.62-1.50)	364	0.83 (0.55-1.26)	0.69
eaeA (EPEC/EHEC)	360	0.89 (0.40-2.00)	369	0.93 (0.44-1.97)	0.94
aggR (EAEC)	361	0.62 (0.24-1.61)	377	0.64 (0.32-1.29)	0.96
stx1/stx2 (EHEC)	359	0.55 (0.24-1.25)	374	1.08 (0.50-2.33)	0.24
ipaH (EIEC)	361	0.74 (0.02-27.62)	377	0.45 (0.04-5.46)	0.83
st1b/lt1 (ETEC)	358	1.42 (0.42-4.81)	374	0.76 (0.27-2.19)	0.47
BacCow	354	0.96 (0.92-1.01)	372	1.00 (0.96-1.04)	0.20
HumM2	328	1.04 (0.54-1.99)	325	0.97 (0.55-1.72)	0.89
Norovirus	333	0.69 (0.13-3.76)	353	1.84 (0.42-7.96)	0.43
Giardia	320	0.63 (0.05-8.18)	283	0.54 (0.05-5.98)	0.94
Stored Water					
Any ECVG	364	1.07 (0.76-1.51)	360	0.86 (0.66-1.13)	0.36
eaeA (EPEC/EHEC)	366	0.99 (0.56-1.78)	365	0.92 (0.52-1.64)	0.87
aggR (EAEC)	369	2.05 (1.08-3.88)	373	0.70 (0.37-1.30)	**0.03**
stx1/stx2 (EHEC)	367	1.24 (0.73-2.09)	370	0.78 (0.48-1.28)	0.26
ipaH (EIEC)	369	0.33 (0.02-4.49)	373	0.26 (0.03-2.10)	0.89
st1b/lt1 (ETEC)	367	0.64 (0.28-1.45)	366	0.66 (0.34-1.26)	0.96
BacCow	356	1.00 (0.84-1.20)	349	0.91 (0.77-1.07)	0.42
HumM2	327	1.48 (0.21-10.54)	325	0.19 (0.03-1.29)	0.19
Soil					
Any ECVG	359	1.08 (0.88-1.31)	381	0.86 (0.72-1.04)	0.15
eaeA (EPEC/EHEC)	362	1.08 (0.82-1.43)	381	0.87 (0.65-1.16)	0.32
aggR (EAEC)	365	1.10 (0.64-1.87)	390	0.73 (0.44-1.19)	0.29
stx1/stx2 (EHEC)	363	0.90 (0.57-1.42)	386	0.87 (0.60-1.26)	0.90

ipaH (EIEC)	365	0.71 (0.06-8.88)	390	0.41 (0.04-4.17)	0.76
st1b/lt1 (ETEC)	362	0.72 (0.42-1.22)	385	0.78 (0.42-1.45)	0.87
BacCow	310	0.97 (0.89-1.06)	336	1.00 (0.94-1.08)	0.53
HumM2	306	1.69 (1.00-2.85)	340	0.90 (0.59-1.38)	0.08

Table A.14: Unadjusted difference and 95% confidence interval in the $\log_{10}$ concentration of BacCow on hands, in stored water, and soil by season, number of individuals in the household, and number of animals in the household. Units of concentration are gene copies of BacCow per 2 hands, 100 ml, and dry gram. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

Season	N	Dry	N	Wet	Interaction p-value
Child Hands	173	-0.13 (-0.46 to 0.21)	193	-0.26 (-0.55 to 0.02)	0.53
Mother Hands	345	-0.07 (-0.24 to 0.11)	381	-0.25 (-0.43 to -0.07)	0.13
Stored Water	377	0.02 (-0.14 to 0.17)	351	-0.05 (-0.20 to 0.09)	0.50
Soil	342	0.15 (-0.14 to 0.43)	396	-0.07 (-0.33 to 0.18)	0.25
Number of Individuals		<10 individuals		$\geq$ 10 individuals	
Child Hands	178	-0.27 (-0.6 to 0.06)	188	-0.13 (-0.46 to 0.21)	0.56
Mother Hands	354	-0.25 (-0.44 to -0.07)	372	-0.08 (-0.27 to 0.11)	0.19
Stored Water	365	0.04 (-0.10 to 0.17)	363	-0.07 (-0.25 to 0.10)	0.33
Soil	356	-0.02 (-0.30 to 0.26)	382	0.08 (-0.23 to 0.40)	0.65
Number of Animals		<20 animals		$\geq$ 20 animals	
Child Hands	189	-0.14 (-0.45 to 0.17)	177	-0.26 (-0.61 to 0.10)	0.64
Mother Hands	385	-0.24 (-0.44 to -0.05)	341	-0.10 (-0.27 to 0.08)	0.28
Stored Water	405	-0.07 (-0.20 to 0.06)	323	0.04 (-0.13 to 0.21)	0.31
Soil	391	0.14 (-0.08 to 0.37)	347	-0.10 (-0.45 to 0.25)	0.28

Appendix B

Supporting Information for Chapter 3

B.1 Tables

Table B.1: Unadjusted and adjusted prevalence ratios and 95% confidence intervals indicating the average increase in prevalence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes for a 1 $\log_{10}$ increase in $\log_{10}$ MPN *E. coli*/sample matrix. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	Prevalence ratio (95% CI) unadjusted	p-value	Prevalence ratio (95% CI) adjusted	p-value	N
Child Hands					
norovirus GII	1.37 (0.78-2.41)	0.27			334
Giardia lamblia	1.10 (0.67-1.83)	0.70			309
Any *E. coli* virulence gene	1.86 (1.65-2.11)	**<0.001***	1.78 (1.55-2.03)	**<0.001***	364
eaeA (EPEC/EHEC)	2.11 (1.68-2.64)	**<0.001***	2.04 (1.57-2.64)	**<0.001***	367
aggR (EAEC)	2.06 (1.68-2.52)	**<0.001***	1.94 (1.54-2.44)	**<0.001***	373
stx1/stx2 (EHEC)	1.83 (1.49-2.26)	**<0.001***	1.75 (1.37-2.23)	**<0.001***	370
ipaH (EIEC)	NA	NA			373
st1b/lt1 (ETEC)	1.77 (1.38-2.27)	**<0.001***	1.55 (1.20-2.01)	**<0.001***	369
Mother Hands					
norovirus GII	0.56 (0.29-1.07)	0.08	0.61 (0.34-1.12)	0.11	684
Giardia lamblia	1.28 (0.80-2.03)	0.30			601
Any *E. coli* virulence gene	1.93 (1.75-2.12)	**<0.001***	1.91 (1.74-2.10)	**<0.001***	720
eaeA (EPEC/EHEC)	2.14 (1.79-2.56)	**<0.001***			729
aggR (EAEC)	1.94 (1.64-2.25)	**<0.001***	1.91 (1.63-2.24)	**<0.001***	738
stx1/stx2 (EHEC)	2.22 (1.86-2.65)	**<0.001***			733
ipaH (EIEC)	1.27 (0.71-2.29)	0.42	1.17 (0.65-2.11)	0.59	738
st1b/lt1 (ETEC)	2.85 (2.09-3.89)	**<0.001***	2.76 (2.03-3.74)	**<0.001***	732
Stored Water					
Any *E. coli* virulence gene	1.64 (1.54-1.74)	**<0.001***			724
eaeA (EPEC/EHEC)	2.02 (1.76-2.31)	**<0.001***			731
aggR (EAEC)	1.80 (1.56-2.09)	**<0.001***			742
stx1/stx2 (EHEC)	1.94 (1.75-2.16)	**<0.001***			737
ipaH (EIEC)	1.93 (0.98-3.79)	0.06	1.82 (0.89-3.72)	0.10	742
st1b/lt1 (ETEC)	1.56 (1.28-1.90)	**<0.001***	1.54 (1.27-1.87)	**<0.001***	733
Soil					
Any *E. coli* virulence gene	1.55 (1.45-1.65)	**<0.001***	1.53 (1.43-1.63)	**<0.001***	740
eaeA (EPEC/EHEC)	1.93 (1.74-2.15)	**<0.001***	1.91 (1.69-2.15)	**<0.001***	743
aggR (EAEC)	1.73 (1.45-2.07)	**<0.001***	1.71 (1.40-2.11)	**<0.001***	755
stx1/stx2 (EHEC)	1.97 (1.63-2.39)	**<0.001***	2.00 (1.64-2.45)	**<0.001***	749
ipaH (EIEC)	1.23 (0.65-2.33)	0.52	1.05 (0.40-2.73)	0.92	755
st1b/lt1 (ETEC)	1.97 (1.64-2.36)	**<0.001***	1.71 (1.39-2.11)	**<0.001***	747

Table B.2: Positive predictive value, negative predictive value, sensitivity and specificity in stored water samples determined at 3 cutoff points of *E. coli* concentration associated with intermediate risk (MPN>1), high risk (MPN>10) and very high risk (MPN>100).

	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Stored Water >1 MPN				
Any *E. coli* virulence gene	45 (41-49)	93 (90-97)	96 (94-99)	31 (27-36)
eaeA (EPEC/EHEC)	18 (15-21)	96 (93-99)	94 (90-99)	24 (20-27)
aggR (EAEC)	14 (11-17)	100 (100-100)	100 (100-100)	24 (20-27)
stx1/stx2 (EHEC)	21 (17-24)	99 (97-100)	98 (96-100)	25 (22-28)
ipaH (EIEC)	2 (1-3)	100 (100-100)	100 (100-100)	21 (18-24)
st1b/lt1 (ETEC)	11 (8-13)	99 (97-100)	97 (92-100)	23 (20-26)
Stored Water >10 MPN				
Any *E. coli* virulence gene	55 (50-60)	81 (77-85)	75 (69-80)	64 (60-69)
eaeA (EPEC/EHEC)	25 (21-29)	95 (93-97)	83 (77-90)	56 (52-60)
aggR (EAEC)	17 (13-21)	95 (93-97)	78 (69-87)	54 (50-57)
stx1/stx2 (EHEC)	27 (23-32)	94 (92-97)	83 (76-89)	57 (53-61)
ipaH (EIEC)	2 (0-3)	99 (98-100)	67 (36-97)	50 (47-54)
st1b/lt1 (ETEC)	12 (8-15)	94 (92-97)	67 (55-78)	52 (48-56)
Stored Water >100 MPN				
Any *E. coli* virulence gene	65 (58-72)	71 (67-75)	40 (34-45)	88 (84-91)
eaeA (EPEC/EHEC)	33 (26-40)	90 (88-93)	50 (40-59)	82 (79-85)
aggR (EAEC)	21 (15-27)	92 (90-94)	43 (32-53)	80 (77-83)
stx1/stx2 (EHEC)	36 (28-43)	89 (86-92)	48 (39-57)	83 (80-86)
ipaH (EIEC)	2 (0-5)	99 (98-100)	44 (12-77)	78 (75-81)
st1b/lt1 (ETEC)	14 (9-19)	93 (91-95)	37 (25-48)	79 (76-82)

Table B.3: Unadjusted and adjusted prevalence ratios determined by a generalized linear model of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes as a function of HumM2 occurrence. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	Prevalence ratio (95% CI) unadjusted	p-value	Prevalence ratio (95% CI) adjusted	p-value	N
Child Hands					
norovirus GII	1.26 (0.36-4.41)	0.72			306
Giardia lamblia	1.36 (0.42-4.40)	0.61			299
Any *E. coli* virulence gene	0.84 (0.55-1.28)	0.42	0.86 (0.57-1.30)	0.48	325
eaeA (EPEC/EHEC)	0.91 (0.40-2.05)	0.81	0.93 (0.42-2.08)	0.86	328
aggR (EAEC)	1.40 (0.72-2.74)	0.32	1.45 (0.75-2.81)	0.27	334
stx1/stx2 (EHEC)	0.54 (0.25-1.18)	0.12	0.55 (0.26-1.19)	0.13	331
ipaH (EIEC)	NA	NA			334
st1b/lt1 (ETEC)	0.93 (0.36-2.43)	0.88	0.97 (0.38-2.46)	0.94	330
Mother Hands					
norovirus GII	1.19 (0.40-3.58)	0.75	1.06 (0.35-3.21)	0.92	635
Giardia lamblia	1.75 (0.50-6.15)	0.38			587
Any *E. coli* virulence gene	1.17 (0.82-1.65)	0.39	1.21 (0.85-1.71)	0.29	632
eaeA (EPEC/EHEC)	0.50 (0.22-1.13)	0.10			641
aggR (EAEC)	1.80 (1.02-3.19)	**0.04**			649
stx1/stx2 (EHEC)	0.92 (0.49-1.74)	0.80	0.97 (0.52-1.81)	0.92	644
ipaH (EIEC)	NA	NA			649
st1b/lt1 (ETEC)	2.01 (0.85-4.74)	0.11	2.38 (1.03-5.48)	**0.04**	643
Stored Water					
Any *E. coli* virulence gene	1.38 (0.82-2.32)	0.22	1.38 (0.82-2.32)	0.22	629
eaeA (EPEC/EHEC)	0.40 (0.06-2.73)	0.35			636
aggR (EAEC)	2.18 (0.88-5.43)	0.09			647
stx1/stx2 (EHEC)	0.75 (0.23-2.49)	0.64	0.79 (0.24-2.63)	0.70	642
ipaH (EIEC)	5.29 (0.66-42.25)	0.12	7.78 (0.91-66.92)	0.06	647
st1b/lt1 (ETEC)	0.75 (0.10-5.35)	0.77	0.77 (0.11-5.57)	0.79	638
Soil					
Any *E. coli* virulence gene	0.99 (0.83-1.18)	0.91	1.07 (0.91-1.26)	0.41	630
eaeA (EPEC/EHEC)	0.99 (0.75-1.31)	0.97	1.11 (0.85-1.45)	0.44	632
aggR (EAEC)	1.23 (0.83-1.83)	0.30	1.33 (0.89-1.98)	0.16	642
stx1/stx2 (EHEC)	1.16 (0.85-1.60)	0.35	1.29 (0.93-1.78)	0.13	638
ipaH (EIEC)	1.58 (0.31-7.96)	0.58			642
st1b/lt1 (ETEC)	0.58 (0.35-0.95)	**0.03**	0.69 (0.42-1.12)	0.13	634

Table B.4: Conditional probabilities for the occurrence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes when HumM2 was present or absent.

	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Child Hands				
norovirus GII	5 (0-10)	96 (94-99)	25 (1-50)	79 (75-84)
Giardia lamblia	6 (0-11)	96 (93-98)	29 (5-52)	78 (73-82)
Any *E. coli* virulence gene	28 (18-39)	67 (61-72)	19 (12-27)	77 (71-82)
eaeA (EPEC/EHEC)	10 (3-16)	89 (86-93)	21 (7-34)	78 (73-82)
aggR (EAEC)	15 (7-23)	89 (86-93)	28 (14-42)	79 (74-84)
stx1/stx2 (EHEC)	10 (3-16)	82 (77-87)	13 (4-22)	76 (71-81)
ipaH (EIEC)	0 (0-0)	100 (99-100)	0 (0-0)	78 (74-83)
st1b/lt1 (ETEC)	7 (1-13)	92 (89-96)	20 (4-36)	79 (74-83)
Mother Hands				
norovirus GII	3 (0-7)	97 (96-99)	21 (3-39)	82 (79-85)
Giardia lamblia	4 (0-7)	98 (97-99)	29 (5-52)	82 (79-85)
Any *E. coli* virulence gene	24 (17-32)	79 (76-83)	21 (14-27)	82 (79-86)
eaeA (EPEC/EHEC)	4 (1-8)	92 (89-94)	10 (2-19)	81 (78-84)
aggR (EAEC)	12 (6-18)	93 (91-96)	29 (16-41)	83 (80-86)
stx1/stx2 (EHEC)	8 (3-13)	92 (89-94)	17 (7-27)	82 (79-85)
ipaH (EIEC)	0 (0-0)	99 (99-100)	0 (0-0)	82 (79-85)
st1b/lt1 (ETEC)	6 (2-10)	97 (96-98)	30 (12-49)	83 (80-86)
Stored Water				
Any *E. coli* virulence gene	50 (26-75)	64 (60-68)	3 (1-6)	98 (97-99)
eaeA (EPEC/EHEC)	6 (0-17)	85 (82-88)	1 (0-3)	97 (96-98)
aggR (EAEC)	24 (3-44)	89 (87-92)	6 (0-11)	98 (97-99)
stx1/stx2 (EHEC)	13 (0-29)	83 (80-86)	2 (0-4)	97 (96-99)
ipaH (EIEC)	6 (0-17)	99 (98-100)	13 (0-35)	97 (96-99)
st1b/lt1 (ETEC)	6 (0-17)	92 (90-94)	2 (0-6)	97 (96-99)
Soil				
Any *E. coli* virulence gene	61 (53-70)	38 (34-42)	20 (16-24)	80 (74-85)
eaeA (EPEC/EHEC)	38 (30-47)	62 (57-66)	20 (15-25)	80 (76-84)
aggR (EAEC)	20 (13-27)	84 (81-87)	24 (16-32)	80 (77-84)
stx1/stx2 (EHEC)	28 (20-35)	76 (72-80)	23 (16-30)	80 (77-84)
ipaH (EIEC)	2 (0-4)	99 (98-100)	29 (0-62)	80 (77-83)
st1b/lt1 (ETEC)	10 (5-16)	82 (79-86)	13 (6-19)	79 (75-82)

Table B.5: Unadjusted and adjusted prevalence ratios and 95% confidence intervals indicating the average increase in prevalence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes for a 1 $\log_{10}$ increase in the $\log_{10}$ gene copies BacCow/sample matrix. Embolden values are significant at a p-value < 0.05. * denotes significance at a p-value < 0.01, correcting for multiple comparisons.

	Prevalence ratio (95% CI) unadjusted	p-value	Prevalence ratio (95% CI) adjusted	p-value	N
Child Hands					
norovirus GII	0.93 (0.64-1.36)	0.72			336
Giardia lamblia	1.13 (0.75-1.70)	0.56			311
Any E. coli virulence gene	1.22 (1.03-1.43)	**0.02**	1.15 (0.98-1.34)	0.09	355
eaeA (EPEC/EHEC)	1.41 (0.96-2.07)	0.08	1.31 (0.91-1.88)	0.15	358
aggR (EAEC)	1.20 (0.94-1.53)	0.14	1.11 (0.88-1.40)	0.38	364
stx1/stx2 (EHEC)	1.31 (0.97-1.77)	0.08	1.23 (0.92-1.64)	0.16	361
ipaH (EIEC)	NA	NA			364
st1b/lt1 (ETEC)	1.18 (0.8-1.75)	0.39	1.04 (0.74-1.46)	0.84	360
Mother Hands					
norovirus GII	0.88 (0.57-1.38)	0.59	0.90 (0.58-1.38)	0.61	685
Giardia lamblia	2.26 (1.50-3.40)	**<0.001***			602
Any *E. coli* virulence gene	1.37 (1.16-1.62)	**<0.001***	1.37 (1.15-1.62)	**<0.001***	706
eaeA (EPEC/EHEC)	1.63 (1.20-2.22)	**<0.001***	1.63 (1.19-2.21)	**0.002***	715
aggR (EAEC)	1.10 (0.84-1.45)	0.47	1.08 (0.82-1.42)	0.57	724
stx1/stx2 (EHEC)	1.79 (1.32-2.43)	**<0.001***	1.79 (1.31-2.44)	**<0.001***	719
ipaH (EIEC)	0.69 (0.20-2.37)	0.56	0.65 (0.18-2.39)	0.51	724
st1b/lt1 (ETEC)	0.98 (0.61-1.56)	0.92	0.95 (0.59-1.53)	0.84	718
Stored Water					
Any *E. coli* virulence gene	1.13 (0.98-1.30)	0.09			705
eaeA (EPEC/EHEC)	1.32 (1.01-1.73)	**0.05**			712
aggR (EAEC)	1.10 (0.83-1.46)	0.49			723
stx1/stx2 (EHEC)	1.26 (0.98-1.61)	0.07	1.27 (1.00-1.63)	0.05	718
ipaH (EIEC)	1.09 (0.48-2.46)	0.83	1.16 (0.52-2.59)	0.71	723
st1b/lt1 (ETEC)	1.11 (0.77-1.60)	0.59	1.14 (0.78-1.66)	0.50	714
Soil					
Any *E. coli* virulence gene	0.99 (0.94-1.05)	0.84	1.01 (0.96-1.06)	0.82	722
eaeA (EPEC/EHEC)	1.03 (0.94-1.12)	0.54	1.04 (0.96-1.13)	0.29	724
aggR (EAEC)	1.02 (0.89-1.16)	0.81	1.03 (0.90-1.18)	0.67	736
stx1/stx2 (EHEC)	1.12 (1.01-1.25)	**0.04**	1.14 (1.02-1.28)	**0.02**	731
ipaH (EIEC)	1.50 (1.04-2.16)	**0.03**			736
st1b/lt1 (ETEC)	1.08 (0.95-1.23)	0.23	1.11 (0.99-1.26)	0.08	728

Table B.6: Unadjusted and adjusted prevalence ratios and 95% confidence intervals indicating the prevalence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes in households with fecal piles or cow patties above and below threshold values. For cow, goat/sheep, non-poultry birds, the threshold value was zero. For chicken/non-chicken poultry the threshold value was 5 (>5 vs ≤ 5). Only feces types with a p-value <0.2 in bivariate models between outcomes and exposures were included in the adjusted models. Embolden values are significant at a p-value <0.05. * denotes significance at a p-value <0.01, correcting for multiple comparisons.

Pathogen	Sample Type	Animal	PR (95% CI) unadjusted	p-value	PR (95% CI) adjusted	p-value
Giardia						
	CH	Cow	2.05 (0.75-5.59)	0.16	1.85 (0.63-5.45)	0.27
		Goat/sheep	1.54 (0.54-4.35)	0.42		
		Poultry	0.71 (0.26-1.94)	0.51		
		Cow patty	3.25 (0.77-13.69)	0.11	2.61 (0.53-12.79)	0.24
		Non-poultry birds	NA			
	MH	Cow	1.53 (0.58-4.05)	0.40		
		Goat/sheep	2.49 (0.92-6.75)	0.07	2.95 (1.04-8.39)	**0.04**
		Poultry	0.48 (0.16-1.48)	0.20		
		Cow patty	1.52 (0.24-9.54)	0.66		
		Non-poultry birds	3.84 (0.90-16.42)	0.07	5.23 (1.13-24.19)	**0.03**
Norovirus						
	CH	Cow	1.72 (0.65-4.55)	0.27		
		Goat/sheep	1.23 (0.33-4.63)	0.76		
		Poultry	1.69 (0.62-4.60)	0.31		
		Cow patty	3.38 (1.02-11.22)	**0.05**		
		Non-poultry birds	NA			
	MH	Cow	1.76 (0.82-3.77)	0.15	1.59 (0.76-3.33)	0.22
		Goat/sheep	1.49 (0.60-3.69)	0.39		
		Poultry	1.91 (0.79-4.61)	0.15	1.97 (0.77-5.00)	0.16
		Cow patty	2.37 (0.79-7.11)	0.12	2.01 (0.76-5.32)	0.16
		Non-poultry birds	1.37 (0.20-9.50)	0.75		
Any ECVG						
	CH	Cow	0.94 (0.70-1.26)	0.67		
		Goat/sheep	1.17 (0.83-1.65)	0.36		
		Poultry	1.28 (0.93-1.77)	0.13	1.00 (1.00-1.01)	0.46
		Cow patty	1.29 (0.76-2.18)	0.34		
		Non-poultry birds	0.66 (0.25-1.70)	0.39		
	MH	Cow	1.06 (0.79-1.41)	0.72		
		Goat/sheep	1.38 (1.08-1.76)	**0.01***	1.41 (1.02-1.94)	**0.04**
		Poultry	1.07 (0.79-1.44)	0.67		
		Cow patty	1.33 (0.86-2.07)	0.20	1.55 (0.95-2.53)	0.08
		Non-poultry birds	1.05 (0.58-1.90)	0.87		
	S	Cow	1.00 (0.87-1.14)	0.99		
		Goat/sheep	1.10 (0.97-1.25)	0.13	1.09 (0.96-1.24)	0.18
		Poultry	1.04 (0.93-1.16)	0.47		
		Cow patty	1.05 (0.85-1.28)	0.67		
		Non-poultry birds	0.76 (0.50-1.16)	0.20	0.81 (0.54-1.22)	0.32
	SW	Cow	1.26 (1.03-1.53)	0.02	1.26 (1.03-1.53)	**0.02**
		Goat/sheep	1.09 (0.90-1.32)	0.39		
		Poultry	0.98 (0.81-1.20)	0.88		
		Cow patty	1.14 (0.86-1.53)	0.36		
		Non-poultry birds	0.86 (0.53-1.41)	0.56		
***eaeA* (EPEC/EHEC)**						
	CH	Cow	1.39 (0.79-2.46)	0.25		
		Goat/sheep	1.23 (0.66-2.30)	0.51		
		Poultry	1.81 (0.97-3.39)	0.06	1.64 (0.87-3.11)	0.13
		Cow patty	0.99 (0.32-3.08)	0.99		
		Non-poultry birds	1.29 (0.37-4.53)	0.69		
	MH	Cow	1.31 (0.79-2.16)	0.30	1.18 (0.79-1.76)	0.41

		Goat/sheep	1.23 (0.73-2.07)	0.43		
		Poultry	0.83 (0.50-1.39)	0.48		
		Cow patty	2.82 (1.54-5.17)	**<0.001***	2.82 (1.54-5.17)	**<0.001***
		Non-poultry birds	1.00 (0.28-3.53)	1.00		
	S	Cow	1.11 (0.89-1.38)	0.37		
		Goat/sheep	1.04 (0.84-1.30)	0.70		
		Poultry	1.01 (0.84-1.22)	0.89		
		Cow patty	0.98 (0.66-1.45)	0.91		
		Non-poultry birds	0.91 (0.51-1.61)	0.75		
	SW	Cow	1.37 (0.96-1.95)	0.08	1.37 (0.96-1.95)	0.08
		Goat/sheep	0.77 (0.50-1.19)	0.24		
		Poultry	1.16 (0.82-1.64)	0.41		
		Cow patty	0.74 (0.38-1.42)	0.37		
		Non-poultry birds	1.05 (0.50-2.21)	0.90		
***aggR* (EAEC)**						
	CH	Cow	1.11 (0.60-2.07)	0.75		
		Goat/sheep	1.67 (0.88-3.14)	0.11	1.64 (0.86-3.13)	0.13
		Poultry	1.54 (0.89-2.68)	0.13	1.26 (0.74-2.15)	0.39
		Cow patty	2.02 (0.92-4.42)	0.08	1.44 (0.62-3.37)	0.40
		Non-poultry birds	0.59 (0.08-4.20)	0.60		
	MH	Cow	1.01 (0.59-1.72)	0.98		
		Goat/sheep	1.54 (0.85-2.80)	0.15	1.54 (0.85-2.80)	0.15
		Poultry	1.44 (0.82-2.52)	0.21		
		Cow patty	0.73 (0.24-2.23)	0.58		
		Non-poultry birds	0.99 (0.28-3.56)	0.99		
	S	Cow	0.79 (0.57-1.10)	0.16	0.77 (0.55-1.07)	0.12
		Goat/sheep	0.93 (0.63-1.37)	0.70		
		Poultry	0.94 (0.68-1.31)	0.72		
		Cow patty	0.69 (0.31-1.53)	0.36		
		Non-poultry birds	0.24 (0.03-1.63)	0.14	0.24 (0.04-1.62)	0.14
	SW	Cow	1.10 (0.75-1.61)	0.63		
		Goat/sheep	1.58 (1.07-2.32)	**0.02**	1.58 (1.07-2.32)	**0.02**
		Poultry	0.87 (0.56-1.33)	0.51		
		Cow patty	1.18 (0.61-2.31)	0.62		
		Non-poultry birds	0.55 (0.15-2.05)	0.38		
***stx1/stx2* (EHEC)**						
	CH	Cow	0.81 (0.48-1.38)	0.44		
		Goat/sheep	0.80 (0.46-1.38)	0.42		
		Poultry	0.99 (0.58-1.68)	0.97		
		Cow patty	0.97 (0.37-2.54)	0.96		
		Non-poultry birds	0.94 (0.27-3.26)	0.92		
	MH	Cow	1.05 (0.62-1.77)	0.86		
		Goat/sheep	1.52 (1.00-2.29)	**0.05**	1.52 (1.00-2.29)	**0.05**
		Poultry	1.02 (0.60-1.72)	0.94		
		Cow patty	1.42 (0.70-2.91)	0.33		
		Non-poultry birds	1.40 (0.52-3.77)	0.50		
	S	Cow	1.13 (0.88-1.45)	0.32		
		Goat/sheep	1.47 (1.12-1.93)	**0.01***	1.44 (1.09-1.90)	**0.01**
		Poultry	1.27 (1.02-1.59)	0.03	1.16 (0.92-1.47)	0.22
		Cow patty	1.38 (0.93-2.04)	0.11	1.17 (0.78-1.73)	0.45
		Non-poultry birds	1.09 (0.57-2.05)	0.80		
	SW	Cow	1.49 (1.05-2.13)	**0.03**	1.42 (0.98-2.05)	0.06
		Goat/sheep	1.28 (0.88-1.86)	0.20	1.20 (0.82-1.77)	0.35
		Poultry	1.00 (0.76-1.33)	0.99		
		Cow patty	1.52 (0.99-2.34)	0.06	1.33 (0.84-2.11)	0.22
		Non-poultry birds	0.94 (0.39-2.27)	0.89		
***ipaH* (EIEC)**						
	CH	Cow	NA			
		Goat/sheep	NA			
		Poultry	NA			
		Cow patty	NA			
		Non-poultry birds	NA			

	MH	Cow	0.37 (0.04-3.26)	0.37		
		Goat/sheep	0.62 (0.07-5.22)	0.66		
		Poultry	0.60 (0.11-3.38)	0.56		
		Cow patty	NA			
		Non-poultry birds	5.46 (0.64-46.64)	0.12	5.71 (0.64-51.02)	0.12
	S	Cow	2.44 (0.55-10.89)	0.24		
		Goat/sheep	4.07 (0.92-17.94)	0.06	5.17 (0.81-32.84)	0.08
		Poultry	0.50 (0.10-2.51)	0.40		
		Cow patty	2.11 (0.26-17.12)	0.48		
		Non-poultry birds	4.56 (0.59-35.28)	0.15	8.10 (0.74-88.42)	0.09
	SW	Cow	0.57 (0.12-2.71)	0.48		
		Goat/sheep	0.45 (0.05-3.68)	0.46		
		Poultry	0.98 (0.26-3.61)	0.97		
		Cow patty	3.12 (0.66-14.79)	0.15	2.73 (0.53-14.17)	0.23
		Non-poultry birds	2.77 (0.37-20.48)	0.32		
***st1b/lt1* (ETEC)**						
	CH	Cow	1.08 (0.55-2.12)	0.83		
		Goat/sheep	0.97 (0.46-2.06)	0.94		
		Poultry	1.56 (0.81-3.01)	0.18	1.27 (0.68-2.37)	0.45
		Cow patty	1.80 (0.68-4.77)	0.23		
		Non-poultry birds	0.82 (0.12-5.72)	0.84		
	MH	Cow	1.12 (0.54-2.34)	0.75		
		Goat/sheep	2.52 (1.20-5.30)	**0.01**	2.47 (1.17-5.23)	**0.02**
		Poultry	0.85 (0.42-1.72)	0.65		
		Cow patty	2.09 (0.74-5.90)	0.17	1.37 (0.48-3.96)	0.56
		Non-poultry birds	0.97 (0.13-7.09)	0.97		
	S	Cow	1.33 (0.95-1.85)	0.09	1.30 (0.94-1.81)	0.11
		Goat/sheep	0.91 (0.60-1.40)	0.68		
		Poultry	1.06 (0.73-1.54)	0.75		
		Cow patty	0.95 (0.52-1.73)	0.88		
		Non-poultry birds	0.49 (0.14-1.76)	0.27		
	SW	Cow	1.48 (0.95-2.30)	0.08	1.39 (0.88-2.19)	0.16
		Goat/sheep	1.12 (0.65-1.94)	0.67		
		Poultry	1.25 (0.77-2.05)	0.37		
		Cow patty	1.80 (1.00-3.23)	**0.05**	1.63 (0.89-2.98)	0.11
		Non-poultry birds	0.74 (0.19-2.96)	0.67		

B.2 Figures

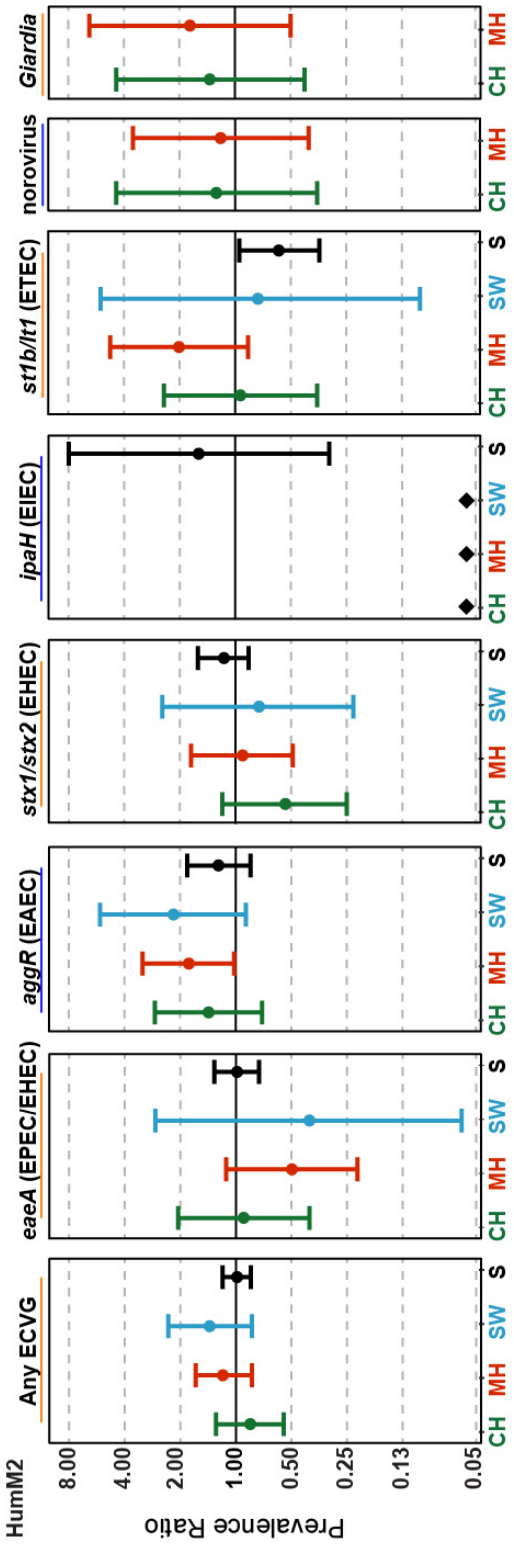


Figure B.1: Prevalence ratios and 95% confidence intervals showing the prevalence of norovirus GII, *Giardia lamblia*, and *E. coli* virulence genes in samples positive for HumM2 compared to those negative for HumM2. The diamond indicates too few samples were positive to model. Pathogens and virulence genes associated with human or human and animal sources are underlined in blue and orange, respectively.

Appendix C

Supporting Information for Chapter 4

C.1 Mock Community

We found 20 ASVs in the mock community run on lane one and 11 ASVs in the mock community run on lane two. Eight of the ASVs from both lanes were identical to the reference sequences for bacteria in the ZymoBIOMICS DNA standard C.1. Two ASVs in the mock community on lane one matched the reference sequences for the eukaryotes in the mock community, *Cryptococcus neoformans* and *Saccharomyces cerevisiae*, but accounted for less than <0.001% of total reads. The next most abundant ASV (after the exact matches) accounted for 1.4% of the total sample reads in both mock communities and had a single bp mismatch from the reference sequence of *Salmonella enterica.* A previous study from the Nelson research group observed the same ASV with a single bp mismatch in 3% of their total mock community sample reads[163]. The remained ASVs were <0.17% of the total read counts. The relative abundance of mock community members was similar to the theoretical values and very similar between lanes. However, the relative abundance of *Pseudomonas aeruginosa* (9.4-9.5%) was approximately twice what was expected (4.2%). Differences in the relative abundance between actual and theoretical could be due to primer bias[164]. The similarity in the relative abundance of mock community members between the lanes indicated there were minimal batch effects.

C.2 Tables

Table C.1: Relative abundance of mock community members for each sequencing run in comparison to the theoretical relative abundances from ZymoBIOMICS

Species	**Mock Lane 1**	**Mock Lane 2**	**Theoretical**
Pseudomonas aeruginosa	9.5	9.4	4.2
Escherichia coli	10.8	10.1	10.1
Salmonella enterica	8.2	8.1	10.4
Lactobacillus fermentum	17.0	17.9	18.4
Enterococcus feacalis	6.9	6.8	9.9
Staphylococcus aureus	16.0	16.6	15.5
Listeria monocytogenes	13.7	13.8	14.1
Bacillus subtilis	15.9	15.5	17.4

Table C.2: 95% CI and 99.15% confidence interval of the test-statistic (difference in mean between the number of identical intra-household ASVs in bootstrapped iterations of randomly assigned households and actual data). Distributions that do not include zero are statistically significant at alpha=0.05 and 0.0083. Alpha=0.0083 is corrected for multiple comparisons.

	Test Statistic Distribution 2.5-97.5% CI [alpha =0.05]	Test Statistic Distribution 0.42-99.58%CI [alpha=0.0083]
Child hands to Mother hands	51.7 - 80.2	45.1 - 84.6
Child hands to Stored water	0.3 - 19.5	-4.4 - 22.7
Child hands to Soil	14.9 - 41.3	10.0 - 45.7
Mother hands to Stored water	3.1 - 24.0	-1.1 - 27.3
Mother hands to Soil	10.2 - 35.3	4.2 - 39.8
Stored water to Soil	-3.6 -15.0	-8.2 - 17.4

C.3 Figures

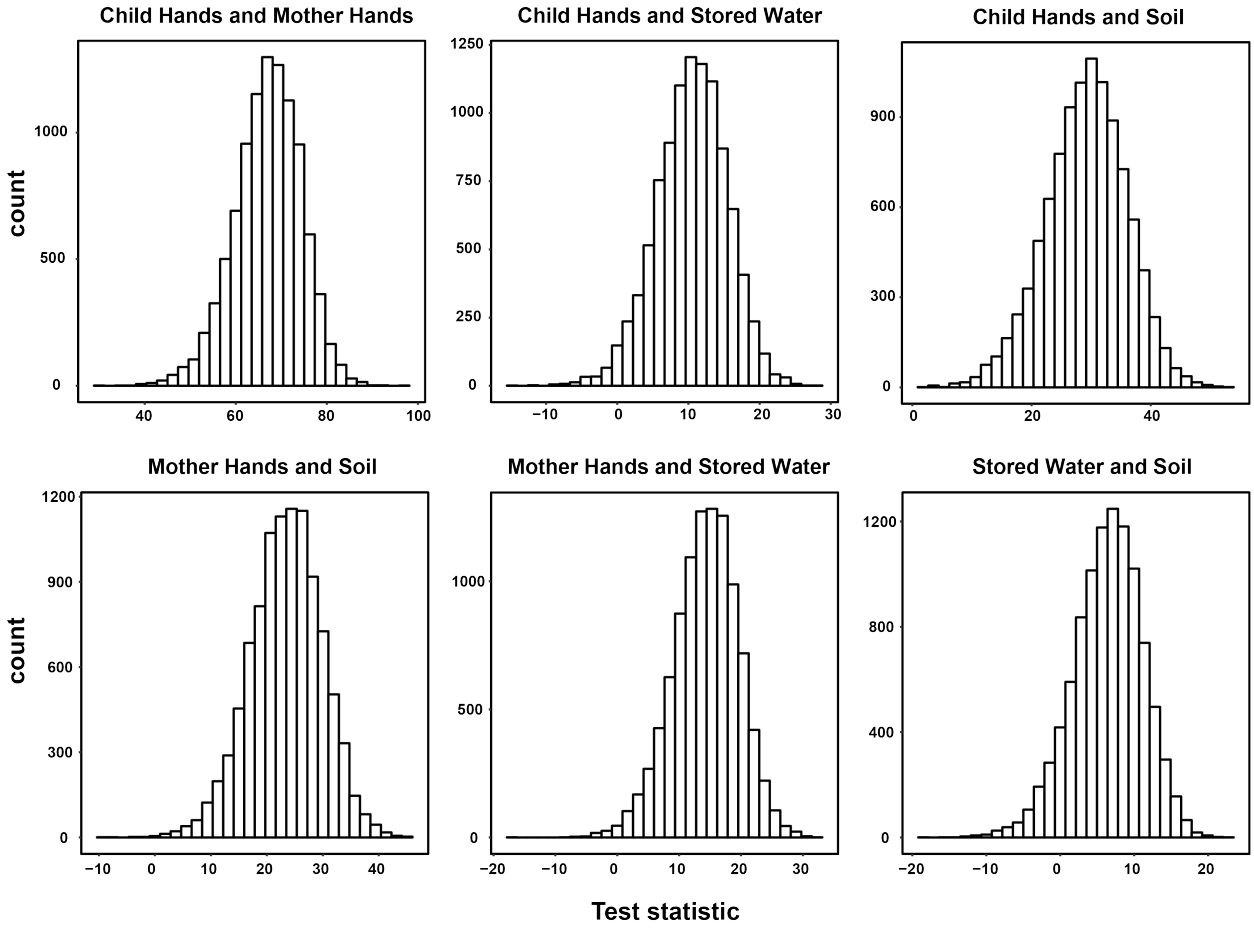


Figure C.1: Distribution of the test-statistic (difference in mean between the number of intra-household ASVs in bootstrapped iterations of randomly assigned households and actual data). Distributions that do not include zero are statistically significant at alpha=0.05 and 0.0083. Alpha=0.0083 is corrected for multiple comparisons.

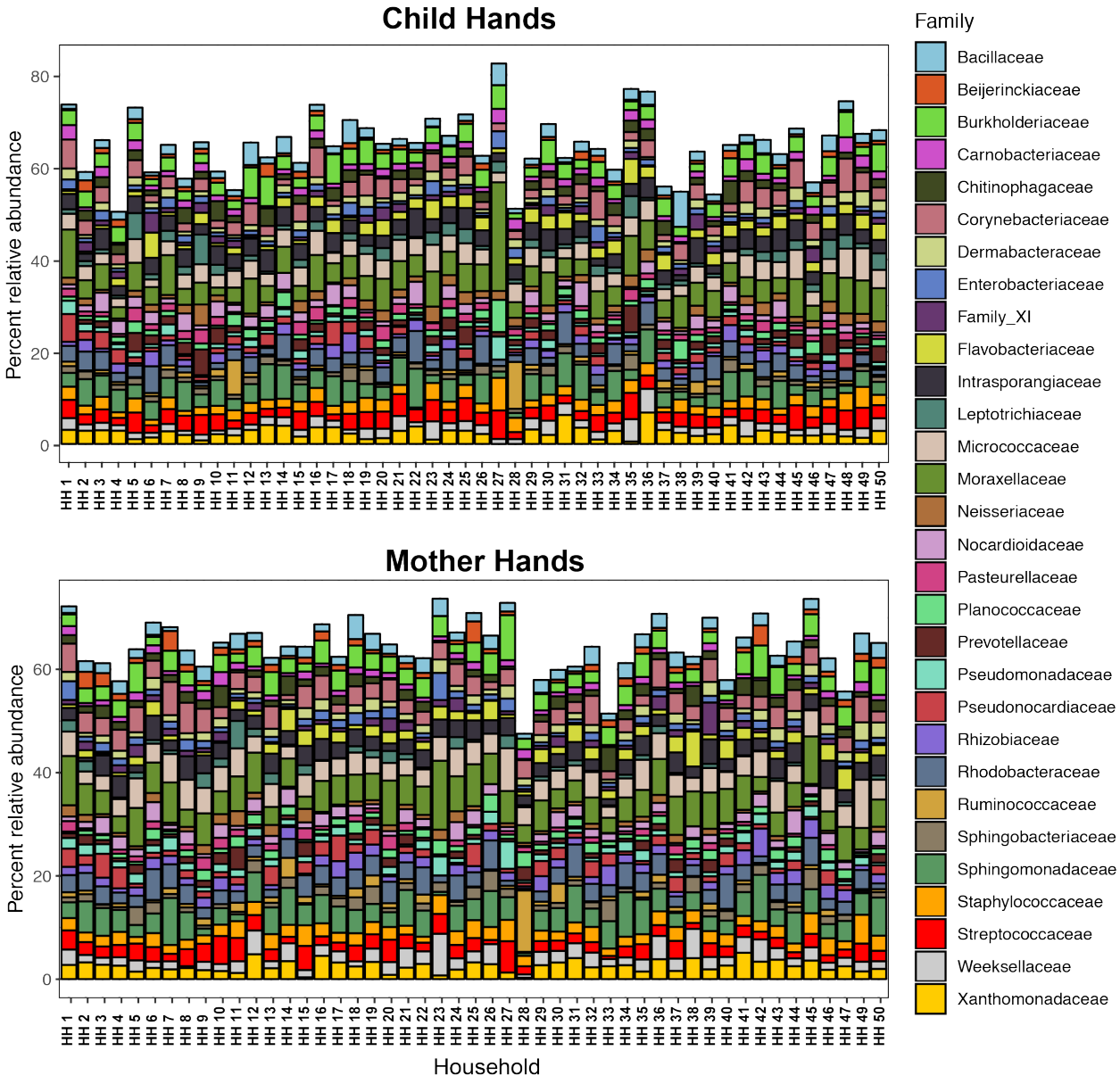


Figure C.2: Relative abundance (percentage of total sample reads) of the 30 most abundant families on hands.

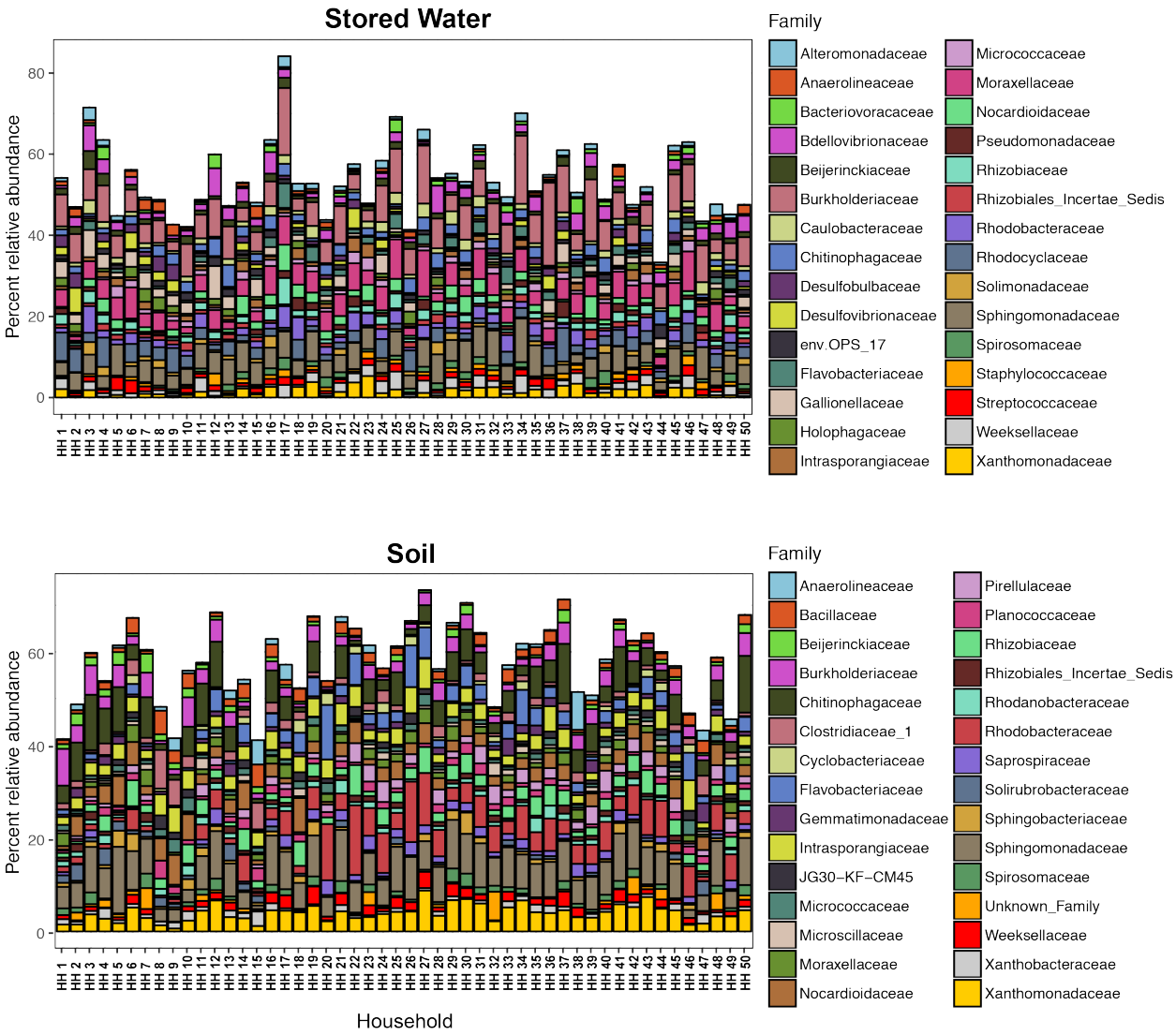


Figure C.3: Relative abundance (percentage of total sample reads) of the 30 most abundant families in stored water and soil, determined separately.

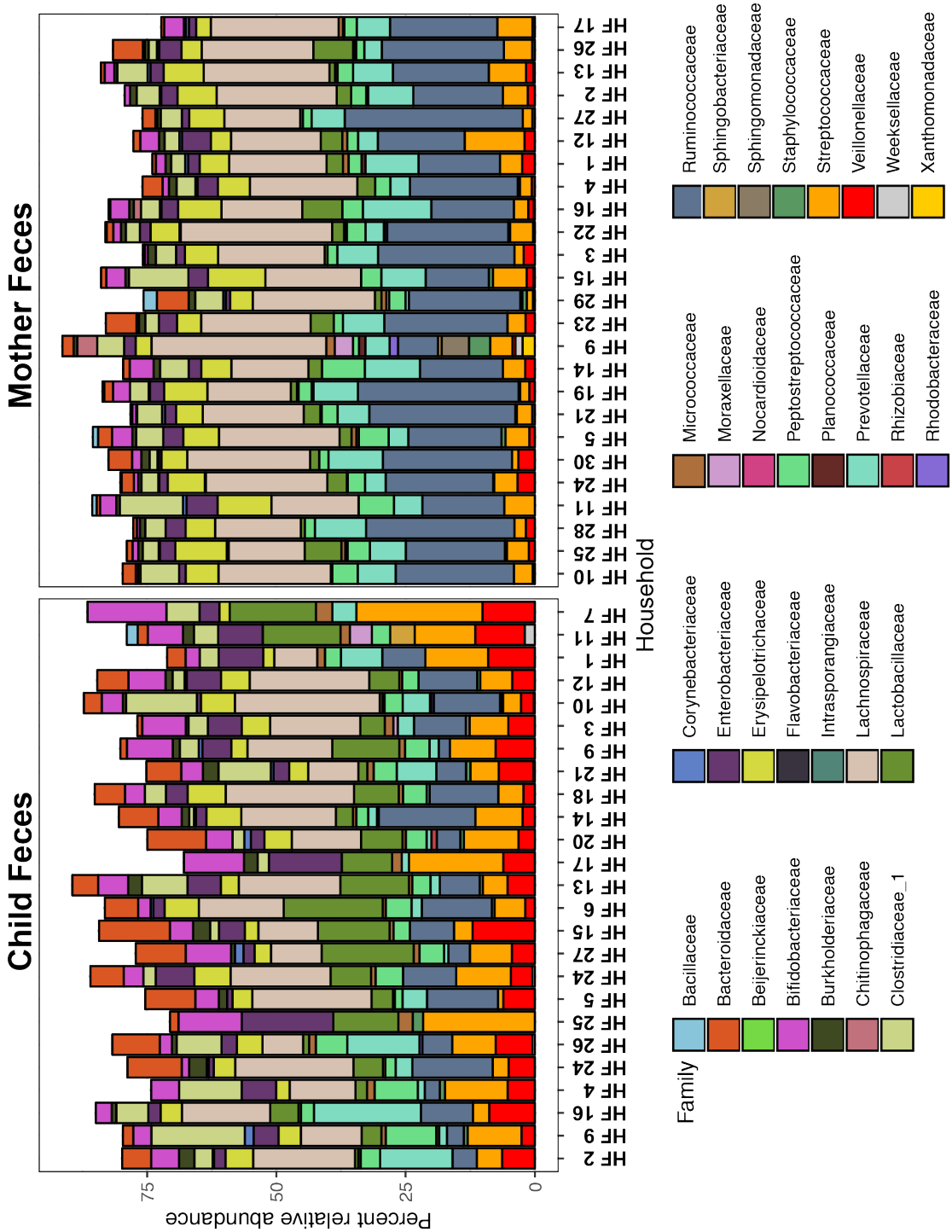


Figure C.4: Relative abundance (percentage of total sample reads) of the 30 most abundant families in feces.

Bibliography

(1) GBD 2015 Mortality and Causes of Death Collaborators Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **Oct. 2016**, *388*, 1459–1544.

(2) UNICEF Diarrhoeal disease - UNICEF DATA., 2018.

(3) Black, R. E.; Victora, C. G.; Walker, S. P.; Bhutta, Z. A.; Christian, P.; de Onis, M.; Ezzati, M.; Grantham-McGregor, S.; Katz, J.; Martorell, R.; Uauy, R. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet* **Aug. 2013**, *382*, 427–451.

(4) Pinkerton, R.; Oria, R. B.; Lima, A. A. M.; Rogawski, E. T.; Oria, M. O. B.; Patrick, P. D.; Moore, S. R.; Wiseman, B. L.; Niehaus, M. D.; Guerrant, R. L. Early childhood diarrhea predicts cognitive delays in later childhood independently of malnutrition. *Am. J. Trop. Med. Hyg* **2016**, *95*, 1004–1010.

(5) Lorntz, B.; Soares, A. M.; Moore, S. R.; Pinkerton, R.; Gansneder, B.; Bovbjerg, V. E.; Guyatt, H.; Lima, A. M.; Guerrant, R. L. Early Childhood Diarrhea Predicts Impaired School Performance. *The Pediatric Infectious Disease Journal* **June 2006**, *25*, 513–520.

(6) Hutton, G.; Haller, L. *Evaluation of the Costs and Benefits of Water and Sanitation Improvements at the Global Level*; tech. rep.; Geneva: World Health Organization, 2004.

(7) Kotloff, K. L. et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. *The Lancet* **July 2013**, *382*, 209–222.

(8) Wagner, E. G.; Lanoix, J. N. *Excreta disposal for rural areas and small communities*; tech. rep.; 1958, pp 1–182.

(9) Bauza, V.; Guest, J. S. The effect of young children's faeces disposal practices on child growth: evidence from 34 countries. *Tropical Medicine & International Health* **Oct. 2017**, *22*, 1233–1248.

(10) Ercumen, A.; Pickering, A. J.; Kwong, L. H.; Arnold, B. F.; Parvez, S. M.; Alam, M.; Sen, D.; Islam, S.; Kullmann, C.; Chase, C.; Ahmed, R.; Unicomb, L.; Luby, S. P.; Colford, J. M. Animal feces contribute to domestic fecal contamination: Evidence from E. coli measured in water, hands, food, flies, and soil in Bangladesh. *Environmental Science & Technology* **July 2017**, *51*, 8725–8734.

(11) Delahoy, M. J.; Wodnik, B.; McAliley, L.; Penakalapati, G.; Swarthout, J.; Freeman, M. C.; Levy, K. Pathogens transmitted in animal feces in low- and middle-income countries. *International Journal of Hygiene and Environmental Health* **Mar. 2018**, *221*, 661–676.

(12) Zambrano, L. D.; Levy, K.; Menezes, N. P.; Freeman, M. C. Human diarrhea infections associated with domestic animal husbandry: A systematic review and meta-analysis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **June 2014**, *108*, 313–25.

(13) Fischer Walker, C. L.; Perin, J.; Aryee, M. J.; Boschi-Pinto, C.; Black, R. E. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC public health* **Jan. 2012**, *12*, 220.

(14) Water and Sanitation Program *Management of Child Feces: Current Disposal Practices*; tech. rep.; World Bank Group, 2015.

(15) Gil, A.; Lanata, C.; Kleinau, E.; Penny, M. *Children's Feces Disposal Practices in Developing Countries and Interventions to Prevent Diarrheal Diseases*; tech. rep.; Washington, DC: US Agency for International Development, 2004.

(16) Hussain, L.; Saharia, K. K.; Sheikh, I. U.; Borgohain, A.; Kalita, S.; Ray, M. N.; Bulbul, K. H. Common animal husbandry practices by the farmers of Indo-Bangladesh international border areas of Assam, India. *Int.J.Curr.Microbiol.App.Sci* **2018**, *7*, 3570–3578.

(17) Talukder, M. A. I.; Shahjahan, M.; Haque, M. N.; Islam, M. S.; Rahman, S. M.; Yeasmin, M. N.; Munira, S.; Habib, M. R.; Salahuddin, M.; Billah, M. M.; Miah, M. A. H. Socio-economic condition, cattle production and management system in a community at Pabna district of Bangladesh. *Research in Agriculture Livestock and Fisheries* **Dec. 2017**, *4*, 181–185.

(18) Harvey, S. A.; Winch, P. J.; Leontsini, E.; Torres Gayoso, C.; Lo Pez Romero, S.; Gilman, R. H.; Oberhelman, R. A. Domestic poultry-raising practices in a Peruvian shantytown: implications for control of Campylobacter jejuni-associated diarrhea. *Acta Tropica* **2003**, *86*, 41–54.

(19) Harris, A. R.; Pickering, A. J.; Harris, M.; Doza, S.; Islam, M. S.; Unicomb, L.; Luby, S.; Davis, J.; Boehm, A. B. Ruminants contribute fecal contamination to the urban household environment in Dhaka, Bangladesh. *Environmental Science & Technology* **May 2016**, *50*, 4642–4649.

(20) Boehm, A. B.; Wang, D.; Ercumen, A.; Shea, M.; Harris, A. R.; Shanks, O. C.; Kelty, C.; Ahmed, A.; Mahmud, Z. H.; Arnold, B. F.; Chase, C.; Kullmann, C.; Colford, J. M.; Luby, S. P.; Pickering, A. J. Occurrence of host-associated fecal markers on child hands, household soil, and drinking water in rural Bangladeshi households. *Environmental Science & Technology Letters* **Oct. 2016**, *3*, 393–398.

(21) Odagiri, M.; Schriewer, A.; Daniels, M. E.; Wuertz, S.; Smith, W. A.; Clasen, T.; Schmidt, W. P.; Jin, Y.; Torondel, B.; Misra, P. R.; Panigrahi, P.; Jenkins, M. W. Human fecal and pathogen exposure pathways in rural Indian villages and the effect of increased latrine coverage. *Water Research* **2016**, *100*, 232–244.

(22) Schriewer, A.; Odagiri, M.; Wuertz, S.; Misra, P. R.; Panigrahi, P.; Clasen, T.; Jenkins, M. W. Human and animal fecal contamination of community water sources, stored drinking water and hands in rural India measured with validated microbial source tracking assays. *The American journal of tropical medicine and hygiene* **Sept. 2015**, *93*, 509–16.

(23) Nizame, F. A.; Leontsini, E.; Luby, S. P.; Nuruzzaman, M.; Parveen, S.; Winch, P. J.; Ram, P. K.; Unicomb, L. Hygiene practices during food preparation in rural Bangladesh: Opportunities to improve the impact of handwashing interventions. *The American journal of tropical medicine and hygiene* **2016**, *95*, 288–97.

(24) WHO; UNICEF *Progress on Drinking Water, Sanitation and Hygiene Update and SDG Baselines*; tech. rep.; 2017.

(25) Julian, T. R.; Schwab, K. J. Challenges in environmental detection of human viral pathogens. *Current Opinion in Virology* **Feb. 2012**, *2*, 78–83.

(26) Yazmin Ramirez-Castillo, F.; Loera-Muro, A.; Jacques, M.; Garneau, P.; Avelar-Gonzalez, F. J.; Harel, J.; Lilian Guerrero-Barrera, A. Waterborne pathogens: Detection methods and challenges. **2015**, *4*, 307–334.

(27) *Environmental Microbiology*, Second; Pepper, I. L., Gerba, C. P., Gentry, T. J., Maier, R., Eds.; Elsevier Inc.: Burlington, MA, 2009, pp 485–496.

(28) World Health Organization, *Guidelines for Drinking-water Quality*, 3rd; WHO Press: Geneva, 2008.

(29) Office of Water; Office of Science and Technology *2017 Five-Year Review of the 2012 Recreational Water Quality Criteria*; tech. rep.; Washington, D.C.: U.S. Environmental Protection Agency, 2018.

(30) Toranzos, G. A. Current and possible alternate indicators of fecal contamination in tropical waters: A short review. *Environmental Toxicology & Water Quality* **May 1991**, *6*, 121–130.

(31) Levy, K.; Nelson, K. L.; Hubbard, A.; Eisenberg, J. N. S. Rethinking Indicators of Microbial Drinking Water Quality for Health Studies in Tropical Developing Countries: Case Study in Northern Coastal Ecuador. *Am. J. Trop. Med. Hyg* **2012**, *86*, 499–507.

(32) Wu, J.; van Geen, A.; Ahmed, K. M.; Alam, Y. A. J.; Culligan, P. J.; Escamilla, V.; Feighery, J.; Ferguson, A. S.; Knappett, P.; Mailloux, B. J.; McKay, L. D.; Serre, M. L.; Streatfield, P. K.; Yunus, M.; Emch, M. Increase in diarrheal disease associated with arsenic mitigation in Bangladesh. *PloS one* **Jan. 2011**, *6*, DOI: 10.1371/journal.pone.0029593.

(33) Hodge, J.; Chang, H. H.; Boisson, S.; Collin, S. M.; Peletz, R.; Classes, T. Assessing the association between thermotolerant coliforms in drinking water and diarrhea: An analysis of individual level data from multiple studies. *Environmental Health Perspectives* **2016**, *124*, 1560–1567.

(34) Ravaliya, K.; Gentry-Shields, J.; Garcia, S.; Heredia, N.; Fabiszewski De Aceituno, A.; Bartz, F. E.; Leon, J. S.; Jaykus, L.-A. Use of Bacteroidales Microbial Source Tracking To Monitor Fecal Contamination in Fresh Produce Production., DOI: 10.1128/AEM.02891-13.

(35) Layton, B. A. et al. Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Research* **Nov. 2013**, *47*, 6897–6908.

(36) Harwood, V. J.; Staley, C.; Badgley, B. D.; Borges, K.; Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiology Reviews* **Jan. 2014**, *38*, 1–40.

(37) Mattioli, M. C.; Pickering, A. J.; Gilsdorf, R. J.; Davis, J.; Boehm, A. B. Hands and water as vectors of diarrheal pathogens in Bagamoyo, Tanzania. *Environmental Science & Technology* **Jan. 2013**, *47*, 355–363.

(38) Arnold, B. F. et al. Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: The WASH Benefits study design and rationale. *BMJ open* **Jan. 2013**, *3*, e003476.

(39) Luby, S. P. et al. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: A cluster randomised controlled trial. *The Lancet. Global health* **Mar. 2018**, *6*, e302–e315.

(40) Penakalapati, G.; Swarthout, J.; Delahoy, M. J.; McAliley, L.; Wodnik, B.; Levy, K.; Freeman, M. C. Exposure to animal feces and human health: A systematic review and proposed research priorities. *Environmental Science & Technology* **Oct. 2017**, *51*, 11537–11552.

(41) Berendes, D. M.; Yang, P. J.; Lai, A.; Hu, D.; Brown, J. Estimation of global recoverable human and animal faecal biomass. *Nature Sustainability* **Nov. 2018**, *1*, 679–685.

(42) Daniels, M. E.; Shrivastava, A.; Smith, W. A.; Sahu, P.; Odagiri, M.; Misra, P. R.; Panigrahi, P.; Suar, M.; Clasen, T.; Jenkins, M. W. Cryptosporidium and Giardia in humans, domestic animals, and village water sources in rural India. *The American journal of tropical medicine and hygiene* **Sept. 2015**, *93*, 596–600.

(43) Patil, S. R.; Arnold, B. F.; Salvatore, A. L.; Briceno, B.; Ganguly, S.; Colford, J. M.; Gertler, P. J. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLoS medicine* **Aug. 2014**, *11*, e1001709.

(44) Clasen, T.; Boisson, S.; Routray, P.; Torondel, B.; Bell, M.; Cumming, O.; Ensink, J.; Freeman, M.; Jenkins, M.; Odagiri, M.; Ray, S.; Sinha, A.; Suar, M.; Schmidt, W.-P. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *The Lancet. Global health* **Nov. 2014**, *2*, 645–53.

(45) Pickering, A. J.; Djebbari, H.; Lopez, C.; Coulibaly, M.; Alzua, M. L. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial., English *The Lancet. Global health* **Nov. 2015**, *3*, 701–11.

(46) Ercumen, A. et al. Do sanitation improvements reduce fecal contamination of water, hands, food, soil, and flies? Evidence from a cluster-randomized controlled trial in rural Bangladesh. *Environmental Science & Technology* **2018**, *52*, 12089–12097.

(47) Sclar, G. D.; Penakalapati, G.; Amato, H. K.; Garn, J. V.; Alexander, K.; Freeman, M. C.; Boisson, S.; Medlicott, K. O.; Clasen, T. Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *International Journal of Hygiene and Environmental Health* **Nov. 2016**, *219*, 709–723.

(48) Omar, K.; Potgieter, N.; Barnard, T. G. Development of a rapid screening method for the detection of pathogenic Escherichia coli using a combination of Colilert Quanti-Trays/2000 and PCR., en *Water Science and Technology* **Mar. 2010**, *10*, 7–13.

(49) Pickering, A. J.; Julian, T. R.; Marks, S. J.; Mattioli, M. C.; Boehm, A. B.; Schwab, K. J.; Davis, J. Fecal contamination and diarrheal pathogens on surfaces and in soils among Tanzanian households with and without improved sanitation. *Environmental science & technology* **June 2012**, *46*, 5736–43.

(50) Victoria, M.; Guimaraes, F.; Fumian, T.; Ferreira, F.; Vieira, C.; Leite, J. P.; Miagostovich, M. Evaluation of an adsorption-elution method for detection of astrovirus and norovirus in environmental waters. *Journal of Virological Methods* **Mar. 2009**, *156*, 73–76.

(51) Shanks, O. C.; Kelty, C. A.; Sivaganesan, M.; Varma, M.; Haugland, R. A. Quantitative PCR for genetic markers of human fecal pollution. *Applied and environmental microbiology* **Sept. 2009**, *75*, 5507–13.

(52) Shanks, O. C.; Domingo, J. W. S.; Lu, J.; Kelty, C. A.; Graham, J. E. Identification of bacterial DNA markers for the detection of human fecal pollution in water. *Applied and environmental microbiology* **Apr. 2007**, *73*, 2416–22.

(53) Kildare, B. J.; Leutenegger, C. M.; McSwain, B. S.; Bambic, D. G.; Rajal, V. B.; Wuertz, S. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water Research* **Aug. 2007**, *41*, 3701–3715.

(54) Cao, Y.; Griffith, J.; Dorevitch, S.; Weisberg, S. Effectiveness of qPCR permutations, internal controls and dilution as means for minimizing the impact of inhibition while measuring Enterococcus in environmental waters. *Journal of Applied Microbiology* **July 2012**, *113*, 66–75.

(55) Sivaganesan, M.; Haugland, R. A.; Chern, E. C.; Shanks, O. C. Improved strategies and optimization of calibration models for real-time PCR absolute quantification. *Water Research* **2010**, *44*, 4726–4735.

(56) Abdi, H. In *Encyclopedia of Measurment and Statistics*, Salkind, N. J., Ed.; Sage Publications: Thousand Oaks, CA, 2007.

(57) Luby, S. P.; Gupta, S. K.; Sheikh, M. a.; Johnston, R. B.; Ram, P. K.; Islam, M. S. Tubewell water quality and predictors of contamination in three flood-prone areas in Bangladesh. *Journal of applied microbiology* **Oct. 2008**, *105*, 1002–1008.

(58) Wu, J.; Yunus, M.; Islam, M. S.; Emch, M. Influence of climate extremes and land use on fecal contamination of shallow tubewells in Bangladesh. *Environmental Science & Technology* **Mar. 2016**, *50*, 2669–2676.

(59) Islam, M. S.; Siddika, A.; Khan, M. N.; Goldar, M. M.; Sadique, M. A.; Kabir, A. N.; Huq, A.; Colwell, R. R. Microbiological analysis of tube-well water in a rural area of Bangladesh. *Applied and environmental microbiology* **July 2001**, *67*, 3328–30.

(60) Ferguson, A. S. et al. Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. *Science of the Total Environment* **Aug. 2012**, *431*, 314–22.

(61) Ercumen, A.; Naser, A. M.; Unicomb, L.; Arnold, B. F.; Colford Jr., J. M.; Luby, S. P. Effects of source- versus household contamination of tubewell water on child diarrhea in rural Bangladesh: A randomized controlled trial. *PLOS ONE* **Mar. 2015**, *10*, ed. by Ali, M., e0121907.

(62) Pickering, A. J.; Davis, J.; Walters, S. P.; Horak, H. M.; Keymer, D. P.; Mushi, D.; Strickfaden, R.; Chynoweth, J. S.; Liu, J.; Blum, A.; Rogers, K.; Boehm, A. B. Hands, water, and health: Fecal contamination in Tanzanian communities with improved, non-networked water supplies. *Environmental Science & Technology* **May 2010**, *44*, 3267–3272.

(63) Trevett, A. F.; Carter, R. C.; Tyrrel, S. F. Water quality deterioration: A study of household drinking water quality in rural Honduras. *International Journal of Environmental Health Research* **Aug. 2004**, *14*, 273–283.

(64) Odagiri, M.; Schriewer, A.; Hanley, K.; Wuertz, S.; Misra, P. R.; Panigrahi, P.; Jenkins, M. W. Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *The Science of the total environment* **Jan. 2015**, *502*, 462–470.

(65) Parvez, S. M.; Azad, R.; Rahman, M.; Unicomb, L.; Ram, P. K.; Naser, A. M.; Stewart, C. P.; Jannat, K.; Rahman, M. J.; Leontsini, E.; Winch, P. J.; Luby, S. P. Achieving optimal technology and behavioral uptake of single and combined interventions of water, sanitation hygiene and nutrition, in an efficacy trial (WASH benefits) in rural Bangladesh. *Trials* **Dec. 2018**, *19*, DOI: 10.1186/s13063-018-2710-8.

(66) Lin, A.; Ercumen, A.; Benjamin-Chung, J.; Arnold, B. F.; Das, S.; Haque, R.; Ashraf, S.; Parvez, S. M.; Unicomb, L.; Rahman, M.; Hubbard, A. E.; Stewart, C. P.; Colford, J. M.; Luby, S. P. Effects of water, sanitation, handwashing, and nutritional interventions on child enteric protozoan infections in rural Bangladesh: A cluster-randomized controlled trial. *Clinical Infectious Diseases* **Apr. 2018**, *67*, 1515–1522.

(67) Robb, K.; Null, C.; Teunis, P.; Yakubu, H.; Armah, G.; Moe, C. L. Assessment of fecal exposure pathways in low-income urban neighborhoods in Accra, Ghana: Rationale, design, methods, and key findings of the SaniPath Study. *Am. J. Trop. Med. Hyg* **2017**, *97*, 1020–1032.

(68) Levy-Booth, D. J.; Campbell, R. G.; Gulden, R. H.; Hart, M. M.; Powell, J. R.; Klironomos, J. N.; Pauls, K. P.; Swanton, C. J.; Trevors, J. T.; Dunfield, K. E. Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry* **Dec. 2007**, *39*, 2977–2991.

(69) England, L. S.; Holmes, S. B.; Trevors, J. T. Review: Persistence of viruses and DNA in soil. *World Journal of Microbiology & Biotechnology* **1998**, *14*, 163–169.

(70) Olson, M. E.; Goh, J.; Phillips, M.; Guselle, N.; McAllister, T. A. Giardia cyst and Cryptosporidium oocyst survival in water, soil, and cattle feces. *Journal of Environment Quality* **1999**, *28*, 1991.

(71) Ogorzaly, L.; Bertrand, I.; Paris, M.; Maul, A.; Gantzer, C. Occurrence, survival, and persistence of human adenoviruses and F-specific RNA phages in raw groundwater. *Applied and environmental microbiology* **2010**, *76*, 8019–8025.

(72) Majdinasab, M.; Hayat, A.; Marty, J. L. Aptamer-based assays and aptasensors for detection of pathogenic bacteria in food samples. *TrAC Trends in Analytical Chemistry* **Oct. 2018**, *107*, 60–77.

(73) WHO; UNICEF *Progress on Drinking Water and Sanitation- 2012 Update*; tech. rep.; Geneva: World Health Organization, 2012, pp 1–61.

(74) Fuhrmeister, E. R.; Schwab, K. J.; Julian, T. R. Estimates of nitrogen, phosphorus, biochemical oxygen demand, and fecal coliforms entering the environment due to inadequate sanitation treatment technologies in 108 low and middle income countries., EN *Environmental science & technology* **Oct. 2015**, *49*, 11604–11.

(75) Tilley, E.; Strande, L.; Luthi, C.; Mosler, H. J.; Udert, K. M.; Gebauer, H.; Hering, J. G. Looking beyond technology: an integrated approach to water, sanitation and hygiene in low income countries. *Environmental science & technology* **Sept. 2014**, *48*, 9965–9970.

(76) Berendes, D. M.; Sumner, T. A.; Brown, J. M. Safely managed sanitation for all means fecal sludge management for at least 1.8 billion people in low and middle income countries. *Environmental Science & Technology* **Mar. 2017**, *51*, 3074–3083.

(77) Wang, Y.; Moe, C. L.; Null, C.; Raj, S. J.; Baker, K. K.; Robb, K. A.; Yakubu, H.; Ampofo, J. A.; Wellington, N.; Freeman, M. C.; Armah, G.; Reese, H. E.; Peprah, D.; Teunis, P. F. M. Multipathway quantitative assessment of exposure to fecal contamination for young children in low-income urban environments in Accra, Ghana: The SaniPath analytical approach. *Am. J. Trop. Med. Hyg* **2017**, *97*, 1009–1019.

(78) Kwong, L.; Ercumen, A.; Pickering, A.; Unicomb, L.; Davis, J.; Luby, S. Hand- and object-mouthing of rural Bangladeshi children 3-18 months old. *International Journal of Environmental Research and Public Health* **June 2016**, *13*, 563.

(79) Enger, K. S.; Nelson, K. L.; Rose, J. B.; Eisenberg, J. N. The joint effects of efficacy and compliance: A study of household water treatment effectiveness against childhood diarrhea. *Water Research* **Mar. 2013**, *47*, 1181–1190.

(80) Enger, K. S.; Nelson, K. L.; Clasen, T.; Rose, J. B.; Eisenberg, J. N. S. Linking quantitative microbial risk assessment and epidemiological data: informing safe drinking water trials in developing countries. *Environmental Science & Technology* **2012**, *46*, 5160–5167.

(81) Mellor, J.; Abebe, L.; Ehdaie, B.; Dillingham, R.; Smith, J. Modeling the sustainability of a ceramic water filter intervention. *Water Research* **Feb. 2014**, *49*, 286–299.

(82) Unicef *Committing to Child Survival: A Promise Renewed. Progress Report 2014*; tech. rep.; 2014, pp 1–104.

(83) Roess, A. A.; Winch, P. J.; Ali, N. A.; Akhter, A.; Afroz, D.; El Arifeen, S.; Darmstadt, G. L.; Baqui, A. H. Animal husbandry practices in rural Bangladesh: Potential risk factors for antimicrobial drug resistance and emerging diseases. *The American Journal of Tropical Medicine and Hygiene* **Nov. 2013**, *89*, 965–970.

(84) Leclerc, H.; Mossel, D. A. A.; Edberg, S. C.; Struijk, C. B. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annual Review of Microbiology* **2001**, *55*, 201–234.

(85) Ryan, U.; Fayer, R.; Xiao, L. Cryptosporidium species in humans and animals: current understanding and research needs. *Parasitology* **2014**, *141*, 1667–1685.

(86) Croxen, M. A.; Law, R. J.; Scholz, R.; Keeney, K. M.; Wlodarska, M.; Finlay, B. B. Recent advances in understanding enteric pathogenic Escherichia coli. *Clinical Microbiology Reviews* **Oct. 2013**, *26*, 822–880.

(87) Vasco, K.; Graham, J. P.; Trueba, G. Detection of zoonotic enteropathogens in children and domestic animals in a semirural community in Ecuador. *Applied and environmental microbiology* **2016**, *82*, 4218–4224.

(88) Blanco, M.; Blanco, J. E.; Dahbi, G.; Alonso, M. P.; Mora, A.; Coira, M. A.; Madrid, C.; Juarez, A.; Bernardez, M. I.; Gonzalez, E. A.; Blanco, J. Identification of two new intimin types in atypical enteropathogenic Escherichia coli. *International microbiology : the official journal of the Spanish Society for Microbiology* **June 2006**, *9*, 103–10.

(89) Vidal, R. M.; Chamorro, N. L.; Giron, J. A. In *Escherichia coli in the Americas*, Torres, A., Ed.; Springer International Publishing: Cham, 2016, pp 1–26.

(90) Etcheverria, A. I.; Lucchesi, P. M. A.; Kruger, A.; Bentancor, A. B.; Padola, N. L. In *Escherichia coli in the Americas*, Torres, A., Ed.; Springer International Publishing: Cham, 2016, pp 149–172.

(91) *Pathogenesis of Bacterial Infections in Animals*, 3rd; Gyles, C. L., Prescott, J. F., Songer, G. J., Theon, C. O., Eds.; Blackwell Publishing: 2010.

(92) Singh, G.; Vajpayee, P.; Ram, S.; Shanker, R. Environmental reservoirs for enterotoxigenic Escherichia coli in South Asian Gangetic riverine system. *Environmental Science & Technology* **Aug. 2010**, *44*, 6475–6480.

(93) Feng, Y.; Xiao, L. Zoonotic potential and molecular epidemiology of Giardia species and giardiasis. *Clinical microbiology reviews* **Jan. 2011**, *24*, 110–40.

(94) De Graaf, M.; van Beek, J.; Koopmans, M. P. G. Human norovirus transmission and evolution in a changing world. *Nature Reviews Microbiology* **July 2016**, *14*, 421–433.

(95) Uber, A. P.; Trabulsi, L. R.; Irino, K.; Beutin, L.; Ghilardi, A. C. R.; Gomes, T. A. T.; Liberatore, A. M. A.; Castro, A. F. P.; Elias, W. P. Enteroaggregative Escherichia coli from humans and animals differ in major phenotypical traits and virulence genes. *FEMS Microbiology Letters* **Mar. 2006**, *256*, 251–257.

(96) Okhuysen, P. C.; Dupont, H. L. Enteroaggregative Escherichia coli (EAEC): A cause of acute and persistent diarrhea of worldwide importance. *The Journal of Infectious Diseases* **2010**, *202*, 503–505.

(97) Yakub, G. P.; Castric, D. A.; Stadterman-Knauer, K. L.; Tobin, M. J.; Blazina, M.; Heineman, T. N.; Yee, G. Y.; Frazier, L. Evaluation of Colilert and Enterolert defined substrate methodology for wastewater applications. *Water Environment Research* **2002**, *74*, 131–135.

(98) Beld, M. J. C.; Reubsaet, F. A. G. Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli. *European Journal of Clinical Microbiology & Infectious Diseases* **June 2012**, *31*, 899–904.

(99) Julian, T. R.; Islam, M. A.; Pickering, A. J.; Roy, S.; Fuhrmeister, E. R.; Ercumen, A.; Harris, A.; Bishai, J.; Schwab, K. J. Genotypic and phenotypic characterization of Escherichia coli isolated from feces, hands, and soils in rural Bangladesh using the Colilert Quanti-Tray System (IDEXX). *Applied and environmental microbiology* **Dec. 2014**, 03214–14.

(100) Maheux, A. F.; Huppe, V.; Boissinot, M.; Picard, F. J.; Bissonnette, L.; Bernier, J. L. T.; Bergeron, M. G. Analytical limits of four β-glucuronidase and β-galactosidase-based commercial culture methods used to detect Escherichia coli and total coliforms. *Journal of Microbiological Methods* **Dec. 2008**, *75*, 506–514.

(101) Pickering, A. J.; Julian, T. R.; Mamuya, S.; Boehm, A. B.; Davis, J. Bacterial hand contamination among Tanzanian mothers varies temporally and following household activities. *Tropical Medicine and International Health* **Feb. 2011**, *16*, 233–239.

(102) Rivera, S. C.; Hazen, T. C.; Toranzos, G. A. Isolation of fecal coliforms from pristine sites in a tropical rain forest. *Applied and environmental microbiology* **Feb. 1988**, *54*, 513–7.

(103) Fujioka, R.; Sian-Denton, C.; Borja, M.; Castro, J.; Morphew, K. Soil: The environmental source of Escherichia coli and Enterococci in Guam's streams. *Journal of Applied Microbiology* **Dec. 1998**, *85*, 83S–89S.

(104) Oh, S.; Buddenborg, S.; Yoder-Himes, D. R.; Tiedje, J. M.; Konstantinidis, K. T. Genomic diversity of Escherichia isolates from diverse habitats. *PLoS ONE* **Oct. 2012**, *7*, ed. by Horn, M., e47005.

(105) Goto, D. K.; Yan, T. Genotypic diversity of Escherichia coli in the water and soil of tropical watersheds in Hawaii. *Applied and environmental microbiology* **June 2011**, *77*, 3988–97.

(106) Marcus, I. M.; Bolster, C. H.; Cook, K. L.; Opot, S. R.; Walker, S. L. Impact of growth conditions on transport behavior of E. coli. *Journal of Environmental Monitoring* **Feb. 2012**, *14*, 984.

(107) Bolster, C.; Haznedaroglu, B.; Walker, S. Diversity in cell properties and transport behavior among 12 different Escherichia coli Iisolates. *Journal of Environmental Quality* **2009**, *38*, 465–472.

(108) Durso, L. M.; Smith, D.; Hutkins, R. W. Measurements of fitness and competition in commensal Escherichia coli and E. coli O157:H7 strains. *Applied and environmental microbiology* **Nov. 2004**, *70*, 6466–72.

(109) Montealegre, M. C.; Roy, S.; Boni, F.; Hossain, M. I.; Navab-Daneshmand, T.; Caduff, L.; Faruque, A. S. G.; Islam, M. A.; Julian, T. R. Risk factors for detection, survival, and growth of antibiotic-resistant and pathogenic Escherichia coli in household soils in rural Bangladesh. *Applied and environmental microbiology* **Oct. 2018**, 01978–18.

(110) Sinclair, R. G.; Rose, J. B.; Hashsham, S. A.; Gerba, C. P.; Haas, C. N. A criteria for selection of surrogates used to study the fate and control of pathogens in the environment. *Appl. Environ. Microbiol* **2012**, *78*, 1969–1977.

(111) Ferguson, C.; Husman, A. M. d. R.; Altavilla, N.; Deere, D.; Ashbolt, N. Fate and Transport of Surface Water Pathogens in Watersheds. *Critical Reviews in Environmental Science and Technology* **July 2003**, *33*, 299–361.

(112) Boehm, A. B.; Yamahara, K. M.; Love, D. C.; Peterson, B. M.; McNeill, K.; Nelson, K. L. Covariation and photoinactivation of traditional and novel indicator organisms and human viruses at a sewage-impacted marine beach. *Environmental Science & Technology* **Nov. 2009**, *43*, 8046–8052.

(113) Viau, E. J.; Lee, D.; Boehm, A. B. Swimmer risk of gastrointestinal illness from exposure to tropical coastal waters impacted by terrestrial dry-weather runoff. *Environmental Science & Technology* **Sept. 2011**, *45*, 7158–7165.

(114) Null, C. et al. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. *The Lancet. Global health* **Mar. 2018**, *6*, e316–e329.

(115) Luby, S. P.; Agboatwalla, M.; Painter, J.; Altaf, A.; Billhimer, W.; Keswick, B.; Hoekstra, R. M. Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Tropical Medicine & International Health* **Apr. 2006**, *11*, 479–489.

(116) Ercumen, A. et al. Effects of single and combined water, sanitation and handwashing interventions on fecal contamination in the domestic environment: A cluster-randomized controlled trial in rural Bangladesh. *Environmental Science & Technology* **2018**, *52*, 12078–12088.

(117) Vujcic, J.; Ram, P. K.; Hussain, F.; Unicomb, L.; Gope, P. S.; Abedin, J.; Mahmud, Z. H.; Sirajul Islam, M.; Luby, S. P. Toys and toilets: cross-sectional study using children's toys to evaluate environmental faecal contamination in rural Bangladeshi households with different sanitation facilities and practices. *Tropical Medicine & International Health* **May 2014**, *19*, 528–536.

(118) Schriewer, A. et al. Performance evaluation of canine-associated Bacteroidales assays in a multi-laboratory comparison study. *Water Research* **Nov. 2013**, *47*, 6909–6920.

(119) Boehm, A. B.; Mrisho, M.; Davis, J.; Mattioli, M. C. M. Quantification of Human Norovirus GII on Hands of Mothers with Children Under the Age of Five Years in Bagamoyo, Tanzania. *The American Journal of Tropical Medicine and Hygiene* **Sept. 2015**, *93*, 478–484.

(120) Jenkins, M. W.; Tiwari, S.; Lorente, M.; Gichaba, C. M.; Wuertz, S. Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Research* **Nov. 2009**, *43*, 4956–4966.

(121) Gawler, A. H.; Beecher, J. E.; Brandao, J.; Carroll, N. M.; Falcao, L.; Gourmelon, M.; Masterson, B.; Nunes, B.; Porter, J.; Rince, A.; Rodrigues, R.; Thorp, M.; Martin Walters, J.; Meijer, W. G. Validation of host-specific Bacteriodales 16S rRNA genes as markers to determine the origin of faecal pollution in Atlantic Rim countries of the European Union. *Water Research* **Aug. 2007**, *41*, 3780–3784.

(122) Balleste, E.; Bonjoch, X.; Belanche, L. A.; Blanch, A. R. Molecular indicators used in the development of predictive models for microbial source tracking. *Applied and environmental microbiology* **Mar. 2010**, *76*, 1789–95.

(123) Reischer, G. H. et al. Performance characteristics of qPCR assays targeting human- and ruminant-associated Bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environmental Science & Technology* **Aug. 2013**, *47*, 8548–8556.

(124) Trinh, P.; Zaneveld, J. R.; Safranek, S.; Rabinowitz, P. M. One Health Relationships Between Human, Animal, and Environmental Microbiomes: A Mini-Review. *Frontiers in Public Health* **Aug. 2018**, *6*, 235.

(125) Pehrsson, E. C.; Tsukayama, P.; Patel, S.; Mejia-Bautista, M.; Sosa-Soto, G.; Navarrete, K. M.; Calderon, M.; Cabrera, L.; Hoyos-Arango, W.; Bertoli, M. T.; Berg, D. E.; Gilman, R. H.; Dantas, G. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **May 2016**, *533*, 212–216.

(126) Song, S. J.; Lauber, C.; Costello, E. K.; Lozupone, C. A.; Humphrey, G.; Berg-Lyons, D.; Caporaso, J. G.; Knights, D.; Clemente, J. C.; Nakielny, S.; Gordon, J. I.; Fierer, N.; Knight, R. Cohabiting family members share microbiota with one another and with their dogs. *eLife* **Apr. 2013**, *2*, e00458.

(127) Baothman, O. A.; Zamzami, M. A.; Taher, I.; Abubaker, J.; Abu-Farha, M. The role of Gut Microbiota in the development of obesity and Diabetes. *Lipids in health and disease* **June 2016**, *15*, 108.

(128) Clemente, J. C.; Ursell, L. K.; Parfrey, L. W.; Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **Mar. 2012**, *148*, 1258–70.

(129) Du Toit, A. The gut microbiome and mental health. *Nature Reviews Microbiology* **Apr. 2019**, *17*, 196–196.

(130) Jatzlauk, G.; Bartel, S.; Heine, H.; Schloter, M.; Krauss-Etschmann, S. Influences of environmental bacteria and their metabolites on allergies, asthma, and host microbiota. *Allergy* **Dec. 2017**, *72*, 1859–1867.

(131) Flores, G. E.; Bates, S. T.; Knights, D.; Lauber, C. L.; Stombaugh, J.; Knight, R.; Fierer, N. Microbial Biogeography of Public Restroom Surfaces. *PLoS ONE* **Nov. 2011**, *6*, ed. by Liles, M. R., e28132.

(132) Flores, G. E.; Bates, S. T.; Caporaso, J. G.; Lauber, C. L.; Leff, J. W.; Knight, R.; Fierer, N. Diversity, distribution and sources of bacteria in residential kitchens. *Environmental Microbiology* **Feb. 2013**, *15*, 588–596.

(133) Lax, S. et al. Bacterial colonization and succession in a newly opened hospital. *Science translational medicine* **May 2017**, *9*, eaah6500.

(134) Ruiz-Calderon, J. F.; Cavallin, H.; Song, S. J.; Novoselac, A.; Pericchi, L. R.; Hernandez, J. N.; Rios, R.; Branch, O. H.; Pereira, H.; Paulino, L. C.; Blaser, M. J.; Knight, R.; Dominguez-Bello, M. G. Walls talk: Microbial biogeography of homes spanning urbanization. *Science Advances* **Feb. 2016**, *2*, e1501061.

(135) Knights, D.; Kuczynski, J.; Charlson, E. S.; Zaneveld, J.; Mozer, M. C.; Collman, R. G.; Bushman, F. D.; Knight, R.; Kelley, S. T. Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **July 2011**, *8*, 761–763.

(136) Hospodsky, D.; Pickering, A. J.; Julian, T. R.; Miller, D.; Gorthala, S.; Boehm, A. B.; Peccia, J. Hand bacterial communities vary across two different human populations. **2017**, DOI: 10.1099/mic.0.075390-0.

(137) Mosites, E. et al. Microbiome sharing between children, livestock and household surfaces in western Kenya. **2017**, DOI: 10.1371/journal.pone.0171017.

(138) Bae, S.; Lyons, C.; Onstad, N. A culture-dependent and metagenomic approach of household drinking water from the source to point of use in a developing country. *Water Research X* **Feb. 2019**, *2*, 100026.

(139) Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* **2016**, *13*, 581–3.

(140) Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* **Aug. 2007**, *73*, 5261–7.

(141) Yilmaz, P.; Parfrey, L. W.; Yarza, P.; Gerken, J.; Pruesse, E.; Quast, C.; Schweer, T.; Rg Peplies, J.; Ludwig, W.; Glo Ckner, F. O. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks., DOI: 10.1093/nar/gkt1209.

(142) Edgar, R. C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **July 2018**, *34*, ed. by Valencia, A., 2371–2375.

(143) McMurdie, P. J.; Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **Apr. 2013**, *8*, ed. by Watson, M., e61217.

(144) Holmes, S.; Wolfgang, H., *Modern Statistics for Modern Biology*, 2014.

(145) Efron, B.; Tibshirani, R. J., *An Introduction to the Bootstrap*; Chapman and Hall: New York, 1993.

(146) Gloor, G. B.; Macklaim, J. M.; Pawlowsky-Glahn, V.; Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* **Nov. 2017**, *8*, 2224.

(147) Cole, J. R.; Wang, Q.; Fish, J. A.; Chai, B.; McGarrell, D. M.; Sun, Y.; Brown, C. T.; Porras-Alfaro, A.; Kuske, C. R.; Tiedje, J. M. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* **Jan. 2014**, *42*, D633–D642.

(148) Shaffer, M.; Lozupone, C.; Wilmes, P. Prevalence and Source of Fecal and Oral Bacteria on Infant, Child, and Adult Hands. *mSystems* **2018**, *3*, 192–209.

(149) Islam, M.; Ercumen, A.; Ashraf, S.; Rahman, M.; Shoab, A. K.; Luby, S. P.; Unicomb, L. Unsafe disposal of feces of children ¡3 years among households with latrine access in rural Bangladesh: Association with household characteristics, fly presence and child diarrhea. *PLOS ONE* **Apr. 2018**, *13*, ed. by Dearden, K. A., e0195218.

(150) Hospodsky, D.; Pickering, A. J.; Julian, T. R.; Miller, D.; Gorthala, S.; Boehm, A. B.; Peccia, J. Hand bacterial communities vary across two different human populations. *Microbiology* **June 2014**, *160*, 1144–1152.

(151) Song, S. J.; Lauber, C.; Costello, E. K.; Lozupone, C. A.; Humphrey, G.; Berg-Lyons, D.; Caporaso, J. G.; Knights, D.; Clemente, J. C.; Nakielny, S.; Gordon, J. I.; Fierer, N.; Knight, R. Cohabiting family members share microbiota with one another and with their dogs. *eLife* **Apr. 2013**, *2*, e00458.

(152) Ross, A. A.; Doxey, A. C.; Neufeld, J. D. The Skin Microbiome of Cohabiting Couples. *mSystems* **Aug. 2017**, *2*, 00043–17.

(153) Edgar, R. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* **2018**, *6*, e5030.

(154) Fuhrmeister, E. R. et al. Effect of sanitation improvements on pathogens and microbial source tracking markers in the rural Bangladeshi household environment. **2019**, In prep.

(155) Medgyesi, D.; Brogan, J.; Sewell, D.; Creve-Coeur, J.; Kwong, L.; Baker, K.; Medgyesi, D. N.; Brogan, J. M.; Sewell, D. K.; Creve-Coeur, J. P.; Kwong, L. H.; Baker, K. K. Where children play: Young child exposure to environmental hazards during play in public areas in a transitioning Internally displaced persons community in Haiti. *International Journal of Environmental Research and Public Health* **Aug. 2018**, *15*, 1646.

(156) Miller-Petrie, M. K.; Voigt, L.; Mclennan, L.; Cairncross, S.; Jenkins, M. W. Infant and Young Child Feces Management and Enabling Products for Their Hygienic Collection, Transport, and Disposal in Cambodia. *Am. J. Trop. Med. Hyg* **2016**, *94*, 456–465.

(157) Silkie, S. S.; Nelson, K. L. Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. *Water Research* **2009**, *43*, 4860–4871.

(158) Williams, G. The effect of loading rate on tertiary wastewater filtration., Ph.D. Thesis, University of California Berkeley, 2009.

(159) Silverman, A. I.; Akrong, M. O.; Amoah, P.; Nelson, K. L.; Drechsel, P. Quantification of human norovirus GII, human adenovirus, and fecal indicator organisms in wastewater used for irrigation in Accra, Ghana. *Journal of Water and Health* **2013**, *11*, 473–488.

(160) Chaudhry, R. M.; Nelson, K. L.; Drewes, J. E. Mechanisms of pathogenic virus removal in a full-scale membrane bioreactor. *Environmental Science & Technology* **Mar. 2015**, *49*, 2815–2822.

(161) Guy, R. A.; Payment, P.; Krull, U. J.; Horgen, P. A. Real-time PCR for quantification of Giardia and Cryptosporidium in environmental water samples and sewage. *Applied and environmental microbiology* **Sept. 2003**, *69*, 5178–85.

(162) Staggs, S. E.; Beckman, E. M.; Keely, S. P.; Mackwan, R.; Ware, M. W.; Moyer, A. P.; Ferretti, J. A.; Sayed, A.; Xiao, L.; Villegas, E. N. The applicability of TaqMan-based quantitative real-time PCR assays for detecting and enumerating Cryptosporidium spp. oocysts in the environment. *PLoS ONE* **June 2013**, *8*, ed. by Schallig, H. D. F. H., e66562.

(163) Kantor, R. S.; Miller, S. E.; Nelson, K. L. The water microbiome through an advanced treatment facility for direct potable reuse. *Frontiers in Microbiology* **2019**.

(164) Aird, D.; Ross, M. G.; Chen, W.-S.; Danielsson, M.; Fennell, T.; Russ, C.; Jaffe, D. B.; Nusbaum, C.; Gnirke, A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* **Feb. 2011**, *12*, R18.