Supporting Information

Fecal Fingerprints of Enteric Pathogen Contamination in Public Environments of Kisumu, Kenya associated with Human Sanitation Conditions and Domestic Animals

Kelly K. Baker*, Reid Senesac, Daniel Sewell, Ananya Sen Gupta, Oliver Cumming, Jane Mumma

* Corresponding author: Kelly K. Baker, University of Iowa College of Public Health, 145 S. Riverside Dr., S316, Iowa City, IA 52246; (001) 319-384-4008; kelly-k-baker@uiowa.edu.

Number of Pages: 12

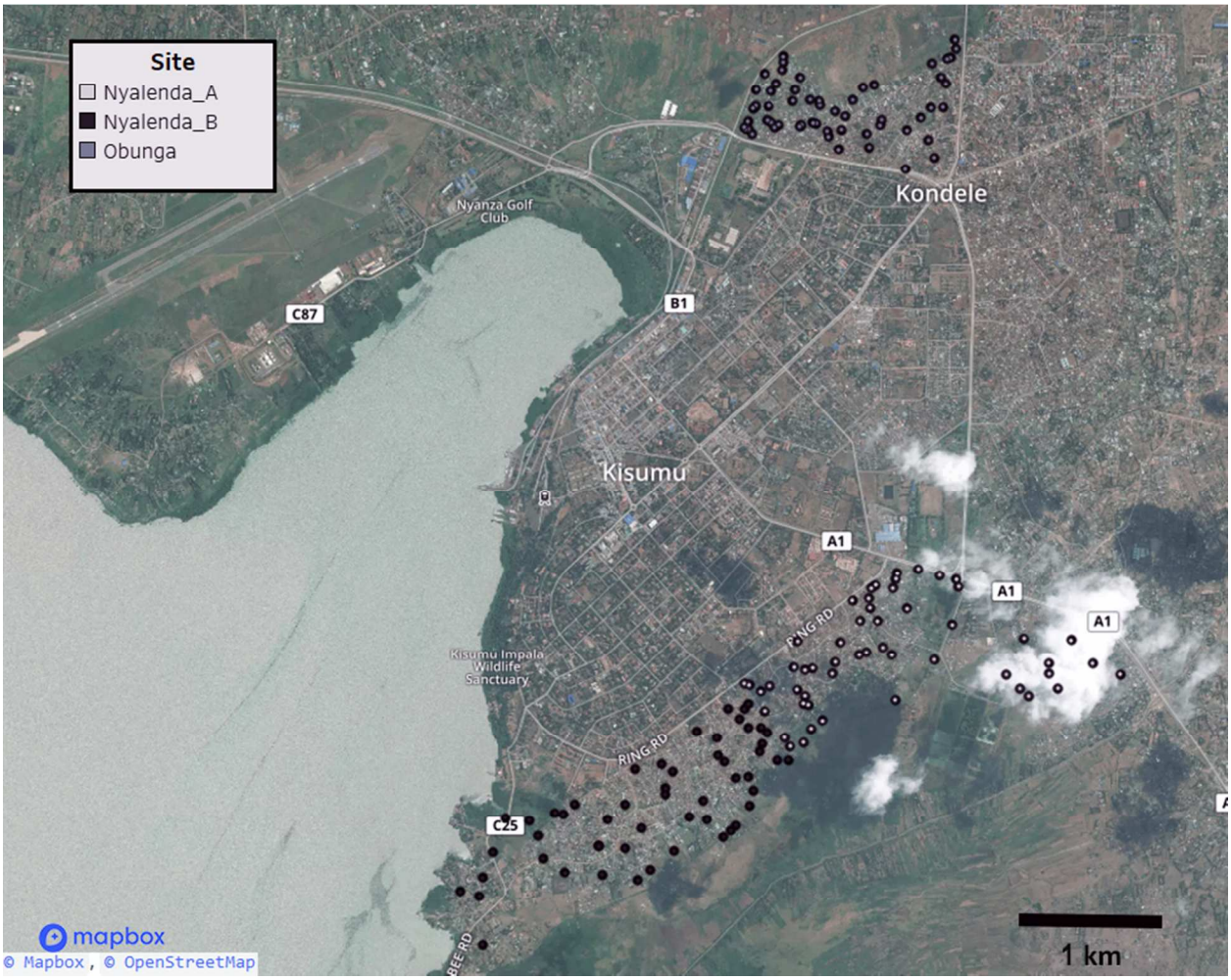
Number of Figures: 2

Number of Tables: 3

Table of Contents

I. Table S3. Concentration of enterococcus indicator bacteria and enteric pathogen DNA and RNA gene copy in soil and surface water from Kisumu neighborhood public domains.

A. Figure S1. Location of public domain sampling sites in three neighborhoods of Kisumu, Kenya.



B. Assessment of RT-PCR inhibition

Environmental samples commonly contain certain compounds (phenolics, humic acid, fulvic acid, humin), enzymes, and ions in environmental sample nucleic acid extracts that can inhibit the proper functioning of reverse transcription and polymerase enzymes during molecular testing. A subset of all field samples were initially tested for evidence of inhibition using the QuantiFast Pathogen + IC RNA extraction kit (Qiagen, Hilden, Germany). Any samples where the average (for two duplicates) cycle threshold (Ct) for the Qiagen Internal Control was >= 4 Ct from the Ct value of the water only control, then the sample was deemed inhibited per manufacturer's recommendations, and a 1:10 dilution was performed and retested. PCR amplification was inhibited in nearly all undiluted soil samples, but was mitigated with 1:10 dilution (Table S1). So all soil samples were performed at 1:10 dilution for consistency. Inhibition was rare in surface water samples, probably due to the use of a membrane absorb-elute method for concentrating organisms (Ahmeda et al. Comparison of Concentration Methods for Quantitative Detection of Sewage-Associated Viral Markers in Environmental Waters).

C. Table S1. Number of soil and surface water samples from Kenya with evidence of inhibition of the QuantiFast Internal Control.

Sample type	Undiluted, n (%)	1:10 Dilution, n (%)
Soil, N=62	54 (87%)	2/64 (3%)
Surface Water, n=51	10 (20%)	2/51 (4%)

D. RT-PCR Standard curves.

The bacteria standards were verified by growing the bacteria from purified glycerol stocks and then extracting DNA from volumes with known cfu/mL. The concentrations for viruses, protozoans, and helminths were provided by ATCC, BEI Resources, or IDT Technologies sources. In the case of ATCC and BEI Resources, concentrations were assessed using a range of pathogen-appropriate methods (ELISA, microscopy, flow cytometry), as reported on their data sheets. IDT Technologies gBlocks are generated to be a specific concentration. We examined DNA or RNA concentration for each of these by examining nanodrop and LLOD data from the serial dilutions to assess whether observed values were consistent with expected concentrations. Acceptable PCR conditions for standard curves were: agreement within 2 Ct of at least two repeats for each dilution of a specific pathogen, where the internal control amplified within the expected range (Ct 30-34), plus Ct values between 3 and 5 Ct in difference between consecutive serial dilutions of a specific pathogen, plus R^2 values for the slopes of all consecutive serial dilutions of a pathogen positive control > 0.95, including the outliers that were > 2 Ct difference from repeats of the same dilution. All R^2 values were greater than 0.95, with most greater than 0.99, and when R^2 was < 0.99, it was usually due to outliers (Table S3).

E. Table S2. Source of reference DNA or RNA use for standard curves in this study.

Pathogen Targeted by TAC	Source of DNA or RNA
All pathogenic bacteria, except *C. difficile*	Strains isolated from purified clinical isolates available through BEI Resources (Manassas, VA). DNA purified from bacterial culture by FastDNA kit.
C. difficile DNA	Strains isolated from purified clinical isolates from infected humans by Craig Ellermeier lab (University of Iowa)
adenovirus 40/41 DNA sapovirus RNA astrovirus RNA norovirus GII RNA	American Type Culture Collection (ATCC, Manassas, VA)
Rotavirus RNA	Stools provided by Cincinnati Children's Hospital
Cryptosporidium spp. DNA *Giardia lamblia* DNA *Entamoeba histolytica* DNA	BEI Resources (Manassas, VA)
Ascaris *Trichuris*	IDT Technologies gBlocks gene fragments (https://www.idtdna.com/pages/products/genes-and-gene-fragments/gblocks-gene-fragments)

F. Back calculations of final concentrations.

Final Surface Water Concentration =

$$1 \text{ dil.} \times \frac{\{\text{Conc. TAC rxn}\} \times \{100\ \mu\text{L total TAC volume}\} \times \{100\ \mu\text{L total extraction}/20\ \mu\text{L DNA or RNA}\}}{10\ \text{mL concentrate filtered and extracted}}$$

Final Soil Concentration =

$$10 \text{ dil.} \times \frac{\{\text{Conc. TAC rxn}\} \times \{100\ \mu\text{L total TAC volume}\} \times \{100\ \mu\text{L total extraction}/20\ \mu\text{L DNA or RNA}\}}{0.5\ \text{gram concentrate extracted}}$$

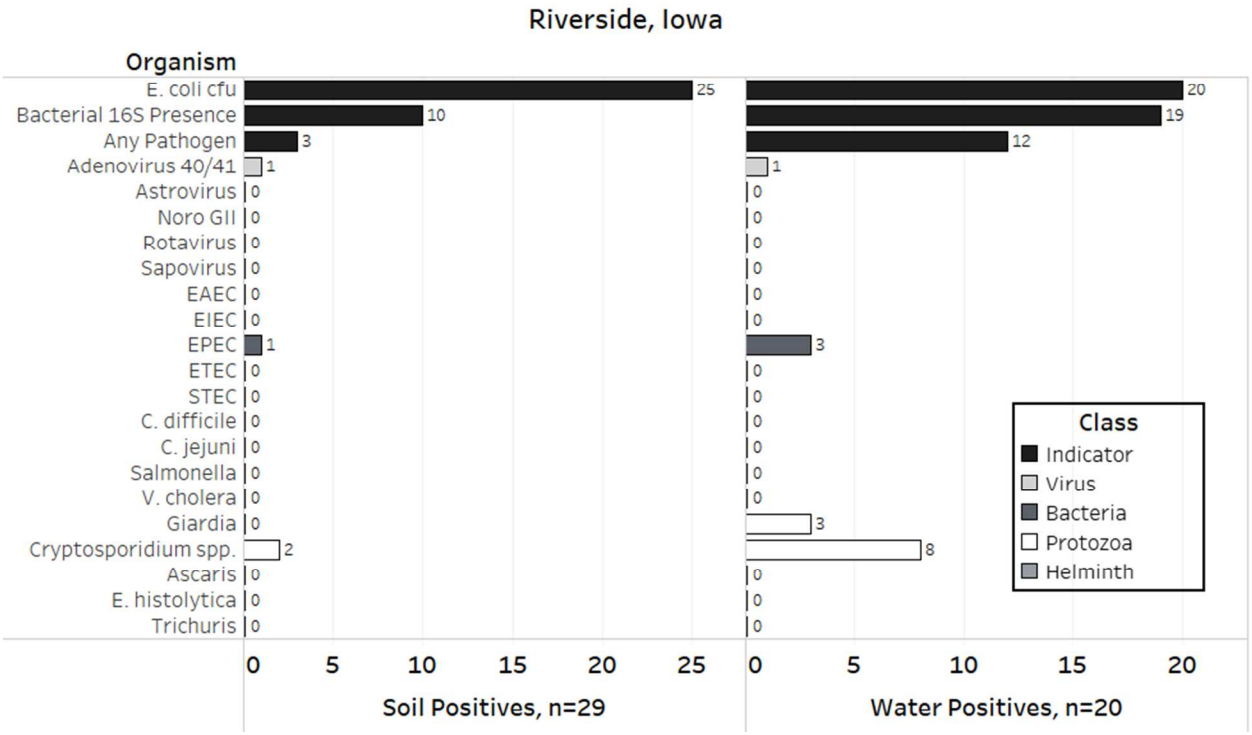
Where dil. Is dilution factor of purified nucleic acid used for PCR amplification, Conc. is concentration of pathogen in a 1µL TAC well, as estimated by comparison to standard curve, rxn is reaction, TAC is Taqman Array card, µL is microliter, mL is milliliter, and concentrate refers to the initial concentrated volume of water or soil used for DNA or RNA extraction. The {100 µL total extraction/20 µL DNA or RNA} equation is based upon the fact that DNA and RNA were separately purified to isolate DNA-based microbes, as well as RNA-based viruses, and then 20 µL of each was combined (total 40 µL) in the mastermix. The methodological lower limit of detection (LLOD) for each pathogen is reported as the lowest concentration of standard consistently detected by TAC, adjusted for processing.

G. **Microbial Detection and Diversity in Iowa watershed.**

Iowa is a mostly rural, agricultural state in the United States with approximately 92% of its land developed for farming or animal food production. Human open defecation is extremely rare. However, high rates of non-point source pollution from fertilizer runoff on crop land and waste generated by large animal containment buildings has led to significant deterioration of water quality in recent decades [23, 24] with 69% of surfaces waters considered impaired. Twenty-nine study sites in Iowa were randomly selected along a public watershed directly impacted by farms. Sites were visited during June 2015 (average temperature 21.8 C, relative humidity 89), and one soil sample and one surface water sample (if relevant) were collected per site. Nine out of 29 Iowa sites (31%) lacked a water source within 15 meters of the central site coordinates. Samples were analyzed using the same microbial protocols as described in the manuscript.

Pathogen detection was far less frequent in soil and surface water in Iowa, although enterococcus bacteria were detected at similarly high frequencies (Figure S2). In Iowa, 100% of surface water (n=20) and 86% of soil (n=29) was contaminated with enterococcus, but only 60% of surface waters and 10% of soils contained an enteric pathogen. Only four types of pathogens were detected in soil and in water, with *Cryptosporidium spp*. representing the majority of pathogen contamination (7% of 29 of soils, 40% of 20 waters). No soil or water samples were positive for multiple pathogens.

H. Figure S2. Detection of *E. coli* bacteria, bacterial 16S DNA, and enteric virus, bacteria, protozoan, or helminth pathogens in soil (A) and water (B) from an agricultural Iowa watershed.



I. Table S3. Concentration of enterococcus indicator bacteria and enteric pathogen DNA and RNA gene copy in soil and surface water from Kisumu neighborhood public domains.

		SOIL, N=185			WATER, N=51		
Organism	**Gene Target**	***N* pos.**	**LLOD/ Gram**	**log_{10} Mean (SD)/gram**	***N* pos.**	**LLOD/ mL**	**log_{10} Mean (SD)/mL**
Enterococcus bacteria	NA	138	0.3	1.9 (1.8)	51	0	4.8 (0.8)
Adenovirus 40/41	Hexon	13	4.4	7.2 (7.2)	19	3.1	6.6 (7.0)
Astrovirus	Capsid	1	2.7	5.9 (-)	3	$2.4x10^{1}$	2.0 (1.9)
Norovirus GII	ORF1-2	0	2.7	- (-)	2	1.4	2.1 (2.1)
Rotavirus	NSP3	0	1.3	- (-)	1	2.0	4.3 (-)
Sapovirus	RdRp	1	2.7	4.8 (-)	4	1.4	4.9 (5.0)
Enteroaggregative *E. coli* (EAEC)							
	aaiC	12	$^{3.5}$	6.9 7.3	20	2.2	5.6 6.2
	aatA	13	3.5	5.8 6.0	26	2.2	5.3 5.8
EIEC/Shigella	*ipaH*	1	3.6	6.1 -	7	2.3	4.0 4.2
Enteropathogenic *E. coli* (EPEC)							
	bfpA	5	3.6	5.2 (5.1)	1	2.3	3.3 (-)
	eae	11	3.6	6.8 (6.9)	22	2.3	5.5 (5.9)
Enterotoxigenic *E. coli* (ETEC)							
	eltB	18	3.2	7.2	24	1.9	5.6

				7.4			6.1
	estA	14	3.2	5.4 5.2	17	1.9	4.1 4.2
Shiga Toxin Expressing *E. coli* (STEC)							
	stx1	1	4.0	6.8 (-)	6	2.7	5.6 (5.7)
	stx2	1	4.0	7.0 (-)	1	2.7	5.6 (-)
Clostridium difficile	*tcdB*	1	4.7	4.2 (-)	0		- (-)
Campylobacter jejuni	*cadF*	5	3.8	4.8 (4.9)	5	3.[5]	2.4 (2.5)
Salmonella	*invA*	3	2.9	7.1 (7.3)	0	1.6	- (-)
Vibrio cholera	*toxR*	5	2.4	5.2 (5.5)	13	1	2.9 (3.0)
Cryptosporidium spp.	18S	125	3.3	6.0 (6.1)	36	2.0	4.8 (5.1)
Giardia lamblia	18S	33	3.0	6.6 (7.2)	17	1.7	4.1 (4.2)

[a] Five surface water samples excluded due to evidence of inhibited amplification. Standard deviations are reported if $n > 1$ positive sample. Lower limit of Detection (LLOD) for method from three replicates of each pathogen target gene; Standard Deviation (SD); *Escherichia coli* (*E. coli*).