

# Application of high-throughput 16S rRNA sequencing to identify fecal contamination sources and to complement the detection of fecal indicator bacteria in rural groundwater

Paul Naphtali, Mahi M. Mohiuddin, Athanasios Paschos and Herb E. Schellhorn

## ABSTRACT

Residents in rural communities across Canada collect potable water from aquifers. Fecal contaminants from sewage and agricultural runoffs can penetrate aquifers, posing a public health risk. Standard methods for detecting fecal contamination test for fecal indicator bacteria (FIB), but the presence of these do not identify sources of contamination. In contrast, DNA-based diagnostic tools can achieve this important objective. We employed quantitative polymerase chain reaction (qPCR) and high-throughput DNA sequencing to trace fecal contamination sources in Wainfleet, a rural Ontario township that has been under the longest active boil water advisory in Canada due to FIB contamination in groundwater wells. Using traditional methods, we identified FIBs indicating persistent fecal pollution in well waters. We used 16S rRNA sequencing to profile groundwater microbial communities and identified *Campylobacteraceae* as a fecal contamination DNA marker in septic tank effluents (STEs). We also identified *Turicibacter* and *Gallicola* as a potential cow and chicken fecal contamination marker, respectively. Using human specific *Bacteroidales* markers, we identified leaking septic tanks as the likely primary fecal contamination source in some of Wainfleet's groundwater. Overall, the results support the use of sequencing-based methods to augment traditional water quality testing methods and help end-users assess fecal contamination levels and identify point and non-point pollution sources.

**Key words** | amplicon sequencing, fecal marker, fecal pollution, groundwater, qPCR

Paul Naphtali  
Mahi M. Mohiuddin  
Athanasios Paschos  
Herb E. Schellhorn (corresponding author)  
Department of Biology,  
McMaster University,  
Hamilton, ON,  
Canada  
E-mail: schell@mcmaster.ca

This article has been made Open Access thanks to the generous support of a global network of libraries as part of the Knowledge Unlatched Select initiative.

## INTRODUCTION

Fecal pollutants from sewage and agricultural runoff can penetrate decaying groundwater wells and render the well water unsafe to drink (USEPA 1993). Fecal contaminants may contain waterborne pathogens that transfer into aquatic environments and cause infectious disease (Harwood *et al.* 2014). Waterborne pathogen detection remains a challenge

because of the diversity and low abundance of pathogens in water. *Escherichia coli* and *Enterococcus*, the standard fecal indicator bacteria (FIB), are present in high densities within the intestine of warm-blooded animals. FIB detection act as proxies for high fecal contamination levels (Field & Samadpour 2007). Public officials in rural communities enact boil water advisories upon FIB detection to lower the risk of waterborne disease outbreaks.

While FIBs indicate fecal contamination, the source of contamination and the true abundance of pathogens

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/wh.2019.295

cannot be identified using FIBs alone. Current practices for monitoring water quality include the use of microbial source tracking (MST) methods that target genetic markers such as the 16S rRNA gene to quantify and source fecal contamination (Field & Samadpour 2007). The HF183 16S rRNA sequence, belonging to a human *Bacteroidales* 16S rRNA gene fragment, was the first genetic DNA marker used to detect human fecal contamination in drinking water (Bernhard & Field 2000). Quantitative polymerase chain reaction (qPCR) assays are now commonly used to quantify the HF183 marker as an indicator of human fecal pollution (Seurinck *et al.* 2005). Human and animal-associated DNA markers can also be used to measure point and non-point source contamination inputs in freshwater environments (Staley *et al.* 2013).

In addition to MST-based approaches, next-generation DNA sequencing is now being used to identify fecal contamination sources and to examine the co-occurrence of fecal and source water bacteria in respective environments (Unno *et al.* 2010). Next-generation sequencing based methods are also used to identify FIBs and pathogens in freshwater reservoirs (Mohiuddin *et al.* 2017, 2018). Aquatic microbiota containing taxa belonging to fecal bacteria are more likely to be contaminated through fecal sources (Cao *et al.* 2013). A 16S rRNA gene-based sequencing approach can also identify additional human fecal markers, such as *Lachnospiraceae*, for further characterization with other genetic methods such as oligotyping (McLellan *et al.* 2013). Despite the decreasing cost and increasing usefulness of next-generation sequencing methods to identify fecal contamination sources, MST protocols have not yet, however, integrated next-generation DNA sequencing as a standard for monitoring Canadian drinking water sources.

Wainfleet, a rural Ontario Township by Lake Erie, is under the longest active boil water advisory in Canada. A previous monitoring study on fecal contamination in 280 private residential groundwater wells determined that  $\geq 50\%$  contained detectable FIBs (Niagara Region Report 2007). The town's boil water advisory provides an opportunity to test next-generation sequencing DNA methods in MST monitoring. Combining next-generation DNA sequencing approaches with traditional MST-based methods can augment current water quality monitoring approaches by incorporating new fecal-specific DNA markers to quantify host-specific contamination.

In this proof-of-principle study, we tested whether a next-generation DNA sequencing approach can be used to trace fecal contamination sources in Wainfleet's private well waters. Using both traditional methods of FIB detection and 16S rRNA amplicon sequencing, we quantified fecal contamination levels and identified likely fecal pollution sources. We also measured the concentration of human *Bacteroidales* gene markers as a proxy of sewage-based contamination. Information obtained from our analyses can augment traditional methods of water quality monitoring by providing additional information on fecal pollution markers in potable waters.

## METHODS

### Study site description

Wainfleet is situated in the southwest portion of the Niagara Region (42.92°N, 79.38°W). Residents obtain potable water using on-site groundwater wells. Many residences are built in low-lying areas close to Lake Erie. Of the 107 residences surveyed in March 2005 that use on-site groundwater wells, 44 had septic tanks that are  $\geq 20$  years old, and 49% of the residences do not comply with current provincial building codes (Niagara Region Report 2007). Most homeowners install septic tanks to discharge septic tank effluents (STEs) into the underlying soils through tile beds. Many of the plots have an area too small to install functioning septic tanks to current building standards. Besides the potential for leaking, concentrated raw sewage seeps through the underlying bed and into aquifers that supply wells (Niagara Region Report 2007).

### Groundwater tap sample collection

We identified nine test sampling sites (Sites A–K, Figure 1) and 21 wells based on FIB detection in the previous independent study (Niagara Region Report 2007). We received written consent from the township and identified volunteers based on the sampling site selection process. We collected tap water samples every month from April to November 2015 and grouped the samples based on season (spring, summer, and fall). A total of 48 samples were collected



**Maps Generated With Google Earth and Simplemapp**

**Figure 1** | Location of sampling sites.

from nine test sites. For each sampling event, town technical staff collected groundwater tap samples from homeowners by filling 500 mL autoclaved plastic bottles (Nalgene). The water samples were then kept on ice, transported to the laboratory, and processed within 6 h of collection.

### Septic tank effluent and manure sample collection

STEs were collected from two septic tanks owned by two homeowners that participated in this study (fall 2015). Three biological replicates for each sewage sample were collected on three consecutive days. Manure samples were collected from manure mounds stored by a concentrated animal feeding operation for chickens, a cow farm, a horse hobby farm, and a pig farm. Similar to septic samples, three biological replicates for each manure type were collected on three consecutive days. All samples were transferred into 500 mL autoclaved bottles, kept in ice and transported to the laboratory. The samples were then stored at  $-80^{\circ}\text{C}$  until further analysis.

### FIB detection assay

FIBs in water samples were detected using standard methods (APHA 2012). To enumerate *E. coli* and *Enterococcus* spp., 100-fold serial dilutions were prepared by transferring 1 mL

of well water into 100 mL of  $1\times$  PBS solution (pH 7.0), to a  $10^4$ -fold dilution. One hundred mL of the well water sample and PBS diluted samples were passed through  $0.45\ \mu\text{m}$  pore-size 47-mm-diameter sterile mixed cellulose ester membrane filters (Thermo Fisher Scientific, Burlington, ON, Canada). Each membrane filter was placed on differential coliform (DC, Oxoid) and mEI agar (BD Difco) plate and incubated at  $42^{\circ}\text{C}$  for 24 h. After incubation, *E. coli* and *Enterococcus* spp. colony forming units (CFUs) for filters containing stock and diluted well water samples were enumerated. To determine the concentration of FIBs (CFU/mL) in water samples, we multiplied the CFU count by its associated dilution factor. Detection of one (or more) CFU per 100 mL of drinking water samples were considered positive for FIBs (deemed unsafe for drinking (Health-Canada 2013)).

### DNA extraction and library generation for sequencing

DNA was extracted from water samples as described earlier (Mohiuddin *et al.* 2018). Briefly, 500 mL of each groundwater sample was passed through  $0.45\text{-}\mu\text{m}$  pore-size 47-mm-diameter sterile mixed cellulose ester membrane filters. The filters were then cut into fragments ( $1\ \text{cm}^2$  size) with sterile scissors and the cut fragments were aseptically transferred with sterile forceps into 1.7 mL microfuge tubes for

DNA extraction. DNA was extracted from the filters using a Soil DNA Isolation Kit (Norgen Biotek, Thorold, ON, Canada) according to manufacturers' instruction and isolated DNA was stored at  $-20^{\circ}\text{C}$  for further analysis. DNA quantity and purity were measured using the Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Burlington, ON, Canada). Manure samples stored at  $-80^{\circ}\text{C}$  were taken out overnight to thaw the samples and 200 mg of thawed samples were used for DNA extraction using a Stool Nucleic Acid Isolation Kit (Norgen Biotek, Thorold, ON, Canada). Similar to manure samples, septic tank samples were taken out overnight to thaw the samples and 10 mL of the thawed samples were centrifuged at  $10,000\times g$  for 10 min at  $4^{\circ}\text{C}$ . After centrifugation, the pellet was resuspended with resuspension buffer provided with Stool Nucleic Acid Isolation Kit (Norgen Biotek, Thorold, ON, Canada) and DNA was extracted according to manufacturers' instructions.

After DNA extraction, PCR assays were conducted to amplify the V3–V4 region of the 16S rRNA gene using an optimized primer pair (S-D-Bact-0341-b-102 S-17/S-D-Bact-0785-a-A-21) evaluated elsewhere (Klindworth *et al.* 2013). Reaction mixes of 25.0  $\mu\text{L}$  were prepared as follows: 2.5  $\mu\text{L}$  10 $\times$  PCR buffer minus  $\text{Mg}^{2+}$  (Invitrogen, Burlington, ON, Canada), 0.5  $\mu\text{L}$  of 100 mM dNTP solution, 1.0  $\mu\text{L}$  each of the forward and reverse primer (1.0  $\mu\text{M}$  each) with unique Illumina adapter sequences attached to them, 1.0  $\mu\text{L}$  10 mg/mL UV-treated BSA solution, 0.75  $\mu\text{L}$  25 mM  $\text{MgCl}_2$  solution (Invitrogen, Burlington, ON, Canada), 0.25  $\mu\text{L}$  Taq DNA Polymerase (Invitrogen), 2.0  $\mu\text{L}$  of DNA template, and 16.0  $\mu\text{L}$  ddH<sub>2</sub>O. The target DNA was amplified using the T100 Thermal Cycler (Bio-Rad, Mississauga, ON, Canada) and conducted as follows: Initial denaturation at  $95^{\circ}\text{C}$  for 3 min, 40 cycles of denaturation at  $94^{\circ}\text{C}$  for 30 s, annealing at  $50^{\circ}\text{C}$  for 30 s, and elongation at  $72^{\circ}\text{C}$  for 1 min, and a final extension step at  $72^{\circ}\text{C}$  for 10 min.

All PCR products were examined in 0.9% agarose gels, the desired band was cut from the gel and DNA was extracted from the cut band with the Nucleospin PCR and Gel Cleanup Kit (Macherey-Nagel, Bethelhem, PA, USA). All amplicon extracts were pooled in equal mass amounts and sent to the Farncombe Metagenomics Facility at McMaster University for further processing and sequencing. Product size and the presence of primer dimers in each

library were checked using BioAnalyzer and High Sensitivity DNA Kit (Agilent, Mississauga, ON, Canada). Each library was then quantified by qPCR and sequencing was performed on the Illumina MiSeq platform which generated  $2\times 300$  bp paired-end sequences.

### Processing and analysis of 16S rRNA gene sequences

Paired-end sequences generated by Illumina MiSeq were demultiplexed and quality-filtered using methods described earlier (Bokulich *et al.* 2013; Mohiuddin *et al.* 2018). Only full amplicons with 464 bp sequences were considered for further analysis. Before downstream analysis, all sequences were trimmed using a quality score threshold (Q score) of 25 over at least 75% of the sequence read. Sequences that contained ambiguous bases, errors in barcode sequences, >10 consecutive low-quality base pairs and >2 nt mismatches from the primer sequences were also removed (Bokulich *et al.* 2013). Processed sequence reads were then analyzed using software package QIIME v.1.9.0 (Caporaso *et al.* 2010). Sequences were clustered into Operational Taxonomic Units (OTUs) sharing 97% nucleotide sequence identity and a minimum query alignment length of 50% to the Greengenes 2013 reference database (McDonald *et al.* 2012) with a sequence similarity threshold of 97% and a minimum query alignment length of 50% using UCLUST (Edgar 2010). The resulting OTU table was then rarified to 5,000 OTU counts/sample before statistical analyses were performed. A potential DNA marker of a human or animal-specific fecal or manure contaminant was identified if it was detected in  $\geq 10\%$  abundance in a fecal sample but  $\leq 1\%$  abundance in all other fecal samples. The relative abundance of the potential markers was determined in the groundwater samples across the sampling months.

### Preparation of DNA standards for quantification of human fecal pollution

The HF183 primer pair (HF183F and Bac708R) (Bernhard & Field 2000) that flanks the fragments in the 16S rRNA gene of human *Bacteroidales* populations was used to prepare standards for quantification of human-based fecal pollution. First, a conventional PCR assay was performed to generate a 500 bp product. The amplified product was



then cloned into a vector which served as standard in qPCR assay. For conventional PCR, 50.0  $\mu\text{L}$  reaction volumes were prepared. The reaction mix contained 5.0  $\mu\text{L}$  of Thermopol Buffer ( $10\times$ , NEB), 1.0  $\mu\text{L}$  (10.0  $\mu\text{M}$  each) each of the forward (HF183F: 5'-ATCATGAGTTCACATGTCCG-3') and reverse primer (Bac708R: 5'-CAATCGGAGTTCCTTCGTG-3'), 1.0  $\mu\text{L}$  of dNTP mix (NEB, 10 mM), 1.0  $\mu\text{L}$  of 5 U *Taq* Polymerase (NEB), and 2.0  $\mu\text{L}$  of DNA template. All PCR assays were run with the conditions as follows: initial denaturation for 5 min at 95 °C, 35 cycles of denaturation at 95 °C for 30 s, annealing at 52 °C, and extension at 72 °C for 1 min, and a final extension step at 72 °C for 6 min. PCR products were loaded onto 1% agarose gels and amplicons were extracted from the gel using the Nucleospin PCR and Gel Cleanup Kit (Macherey-Nagel, Bethelhem, PA, USA). To generate plasmid DNA standards, the TOPO TA Cloning Reaction (Invitrogen, Burlington, ON, Canada) was used. Briefly, 2.0  $\mu\text{L}$  of the amplicons were ligated into the TOPO vector, according to the manufacturer's instructions. Transformants were isolated using blue-white screening technique and subsequently purified. Purified colonies were then inoculated into 3.0 mL of LB broth supplemented with 50.0  $\mu\text{g}/\text{mL}$  kanamycin for overnight incubation at 37 °C with shaking at 220 rpm. DNA was extracted from 1.5 mL aliquots of the cultures with a Plasmid DNA Miniprep Kit (Norgen Biotek, Thorold, ON, USA). DNA concentrations were quantified using the Qubit Fluorometer 2.0 (Invitrogen, Burlington, ON, USA). DNA extracts were diluted such that a range of  $10^1$ – $10^7$  plasmid copies of the marker was present in the standards.

#### Quantification of human *Bacteroidales* marker in sewage and groundwater samples

Quantification of human specific *Bacteroidales* marker was performed using the same forward primer used for the conventional PCR (HF183F) and a different reverse primer (5'-TACCCCGCTACTATCTAATG-3') (Seurinck et al. 2005). The newly amplified product had a length of 82 bp. qPCR assays were conducted using 2.0  $\mu\text{L}$  of groundwater DNA extracts and plasmid standard in 10.0  $\mu\text{L}$  reaction mix volumes. All DNA templates were diluted 10-fold to minimize qPCR inhibition. Each reaction mix contained 1.0  $\mu\text{L}$  of DNA template, 5.0  $\mu\text{L}$  of the SsoFast EvaGreen SuperMix

(Bio-Rad), and 0.5  $\mu\text{L}$  each of the forward and reverse primer (10  $\mu\text{M}$ ). Thermal cycling conditions included the following steps: Initial denaturation at 95 °C for 30 s, 40 cycles of denaturation at 95 °C for 30 s and 60 °C for 10 s, and then a melt curve analysis from 65 to 95 °C at increments of 0.5 °C for 5 s per increment. The CFX Analyzer (Bio-Rad, Mississauga, ON, Canada) was used to generate a standard curve and a melt curve and quantify qPCR copy numbers.

## RESULTS

### Detection of FIB in groundwater wells

As a preliminary assessment of fecal contamination levels in Wainfleet's well waters, we examined FIB detection frequency in tap water collected from private wells using Health Canada recommended guidelines (see above under 'Methods'). Sixty-one per cent of the groundwater samples collected from 21 wells within the boil water advisory in 2015 contained either *E. coli* or *Enterococcus* (Figure 2).

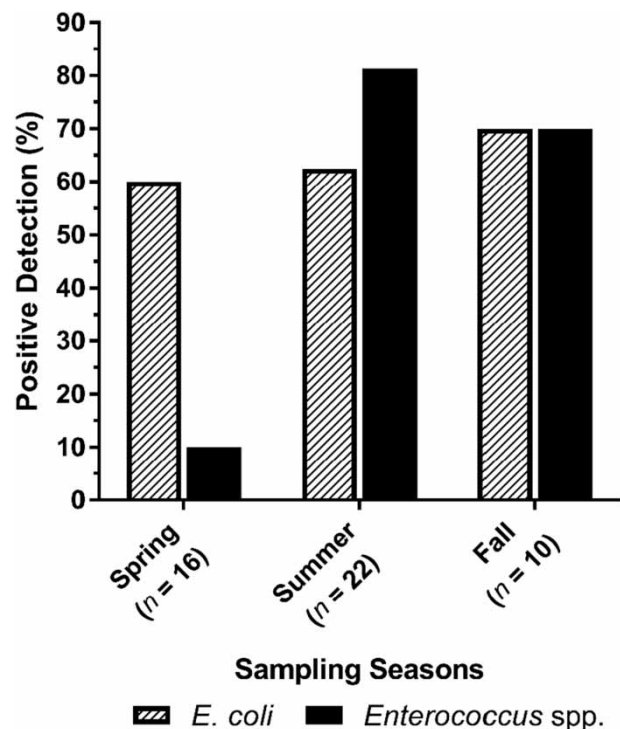
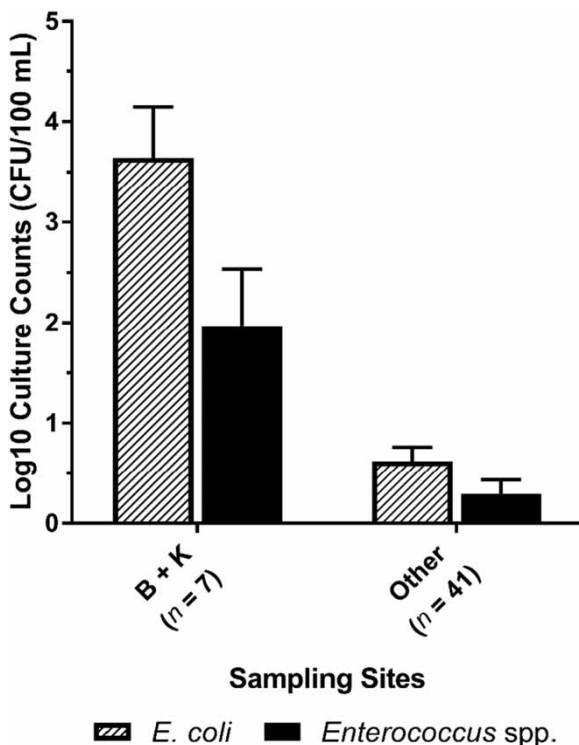


Figure 2 | Detection of *E. coli* and *Enterococcus* in well water samples ( $n = 48$ ) at locations in the boil water advisory zone in 2015.

Positive *E. coli* detection ranged from 60.0% to 70.0%. *Enterococcus* detection rates had a large seasonal range from 10.0% to 81.0%, with the detection rate lowest in the spring and highest in the summer (Figure 2). Groundwater samples collected from two residential wells, B and K, also contained much higher mean *E. coli* and *Enterococcus* than the other groundwater wells (Figure 3).

### Basic sequencing data

As a first step towards identifying potential fecal contamination sources in the FIB-positive well waters, we profiled microbial communities from 21 groundwater samples, six samples from two STEs, and 12 manure samples from four different animal farms using 16S rRNA sequencing. Sequence reads obtained from septic tank samples were classified into 972 ( $\pm 89$ ) OTUs. The number of reads obtained from animal manure and groundwater samples were higher than septic tank samples and reads from animal manure and groundwater wells were classified into 2,987 ( $\pm 315$ ) and 3,101 ( $\pm 287$ ) OTUs respectively.



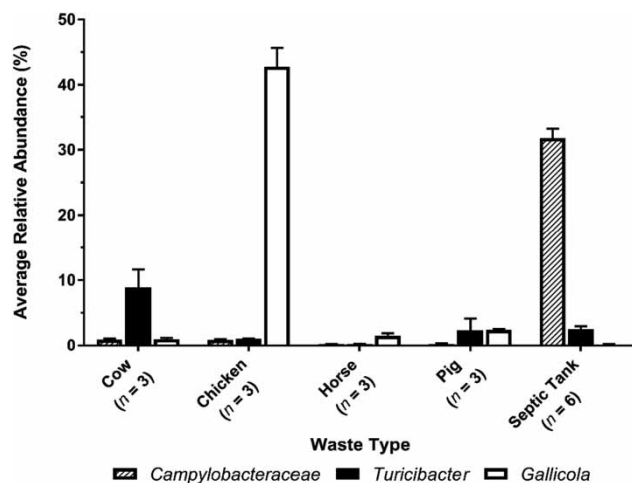
**Figure 3** | *E. coli* and *Enterococcus* levels in wells waters B and K compared with the other well water sites. Error bars represent standard deviation of the mean.

### Identification of potential STE and animal manure contamination markers

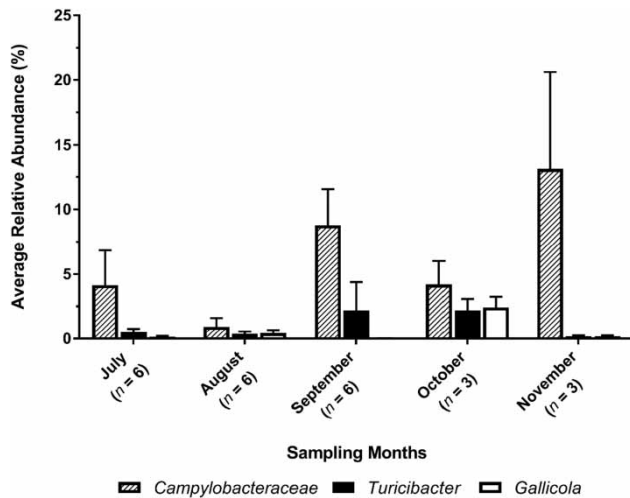
To identify potential fecal contamination markers in groundwater, we parsed for sewage and manure-associated bacteria in the 16S rRNA sequencing data. We defined a host-specific fecal marker as a microbial group detected at  $\geq 10.0\%$  relative abundance in one type of fecal matter but detected at  $\leq 1.0\%$  abundance in the other fecal sources. In the two STE samples, the mean abundance of *Campylobacteraceae* was 32.5% (Figure 4). At the genus level, sequences annotated to *Sulphospirillum* and *Arcobacter* were the most abundant members of *Campylobacteraceae*. In contrast, the average abundance of *Campylobacteraceae* was  $\leq 1\%$  for all four animal manure samples. *Gallicola* and *Turicibacter* were the most abundant genera in chicken and cow manure, comprising 42.2% and 9.4% of 16S rRNA sequences respectively. In contrast, these markers were detected at  $\leq 1.0\%$  abundance among the pig and horse manure and STE samples (Figure 4). No genetic marker was identified in pig and horse manure based on the classification criterion.

### Abundance of STE and manure-based OTUs in groundwater samples

To identify possible fecal contaminants in well water, we determined the relative abundance of the human and



**Figure 4** | Relative abundance of potential manure and sewage markers in waste samples collected in Wainfleet. Error bars represent standard deviation of the mean. Three biological replicates for each of the samples was used to calculate mean and standard deviation.



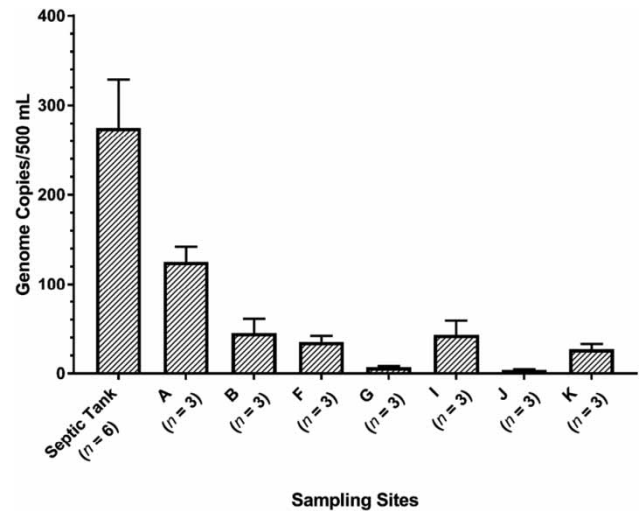
**Figure 5** | Relative abundance of potential STE and animal-specific contamination markers in well water sites within the active boil water advisory zone. Error bars represent standard deviation of the mean.

animal-associated markers we identified in the reference fecal samples. Groundwater wells sampled in July, September, and November had the highest relative abundance of STE-based *Campylobacteraceae* sequences (Figure 5). None of the chicken and cow markers had a relative abundance above 2.0%, with the relative abundance below 1.0% in wells collected in July, August, and November (Figure 5).

### Identification of human fecal contamination

To determine whether the human-specific fecal contamination was present in the groundwater samples, we conducted qPCR assays of the HF183 human *Bacteroidales* marker in selected wells with or without FIB detection. We prepared a qPCR standard curve that quantifies the human *Bacteroidales* marker in each run (Figure S1A). Each standard curve was robust for HF183 quantification (Figure S1, available with the online version of this paper). Melt curve analysis of the standard curve and all groundwater samples that were examined after the qPCR assay yielded a single peak at 84 °C (Figure S1B).

The HF183 marker genome was more abundant in STE samples than in groundwater wells (Figure 6). qPCR of the human *Bacteroidales* marker in groundwater wells B, F and K resulted in positive detection, containing 30–50 genome copies/100 mL (Figure 6). A subset of groundwater



**Figure 6** | Presence of human *Bacteroidales* marker in selected septic tank and groundwater wells within the boil water advisory zone in Wainfleet. Error bars represent standard deviation of the mean.

wells without *E. coli* detection also tested positive for the HF183 marker. Groundwater well A, treated with UV light before sampling, contained more HF183 copies than other groundwater wells (Figure 6). Groundwater well I had a slightly lower HF183 marker level than B, F, and K.

### DISCUSSION

The presence of FIBs in drinking waters is a major health concern. Using national guidelines for drinking water and the results of a previous study as a reference, we sought to determine whether high levels of FIBs are still present in the Wainfleet well waters. Our analyses suggest that FIBs are still present in over half of the tested well water sites and, therefore, the quality of drinking water may still be a concern in the majority of the households within the township. Many of the sampled groundwater wells are located near the shores of Lake Erie, a low-lying region where most of the *E. coli* and total coliform exceedances were previously observed (Niagara Region Report 2007). Some well water samples also contained far higher FIB counts than others. Mean *E. coli* and *Enterococcus* levels in some groundwater wells were three and two orders of magnitude higher than the average FIB counts for the groundwater wells collected across all the other locations. These wells with higher than average FIB counts are located near

septic tanks that may leak raw sewage into the aquifer (Niagara Region Report 2007). Poor well maintenance also facilitates sewage leaching into the groundwater (Howard *et al.* 2003). Altogether, chronic FIB contamination in individual groundwater wells like B and K require further investigation to confirm potential fecal contamination sources.

### Characterization of fecal microbiota and identification of potential fecal markers

The presence of host-associated fecal contamination may help to trace fecal contamination sources in complex environmental samples. As a proof-of-principle experiment, we examined whether we could complement FIB detection with 16S rRNA sequencing methods by identifying sewage and manure-associated markers. We selected a maximum detection rate of 1.0% limit to ensure the specificity of the marker to its host. In cow manure, we identified *Turicibacter* spp., a member of the Firmicutes phylum present in high abundance at the genus level, agreeing with previous cow microbiome profiles (Kim *et al.* 2011). In chicken manure, *Gallicola* – also a member of the Firmicutes phylum – was the most abundant genus. Both markers were also detected at  $\leq 0.1\%$  abundance in other manure types and sewage samples. These genera may act as DNA markers of host-associated fecal contamination in the town's groundwater wells.

In the two STE sites, we identified the *Campylobacteraceae* family (member of the Proteobacteria phylum) as fecal pollution marker, comprising 32.5% of 16S rRNA gene sequences at the family level. The predominance of *Campylobacteraceae* 16S rRNA gene sequences in STEs is identified in other septic tank systems (Tomaras *et al.* 2009). Although we only collected STEs from two septic tanks, both septic tank microbiomes contained a similar abundance of *Campylobacteraceae* sequences. *Campylobacteraceae* was also absent in the animal manure samples, suggesting that the *Campylobacteraceae* microbes are likely exclusive to STEs in the township.

While we identified host-associated markers for chicken manure, cow manure, and STEs, we could not establish the presence of host-associated markers for horse and pig manure. Although *Clostridium* and *Turicibacter* were more abundant in horse and pig manure samples, these sequences

were also present in other manure types. However, analyses of additional samples that include both biological and technical replicates are required to confirm the findings of this proof-of-principle study. A higher sequencing depth, in conjunction with shotgun metagenomic sequencing, may facilitate the identification of novel or rare (but nonetheless important) taxa that can be used to identify contamination at the species level. DNA markers unique to horse or pig manure are yet to be established. The robust resolution of horse and pig-associated markers with next-generation sequencing represents opportunities for future investigation.

### Fecal source identification in residential groundwater sites using 16S rRNA sequencing

The co-detection of animal and/or human associated markers in drinking water provides evidence of possible fecal contamination by that source. After identifying human and animal-associated markers in our reference fecal samples, we screened for possible fecal contamination in private well waters by profiling the well water microbial communities. The mean *Campylobacteraceae* relative abundance was highest in November and September, but lowest in August. The large standard deviations in *Campylobacteraceae* abundances within the sampling months suggest that some groundwater sites contain more fecal contamination than others. The relative abundance of *Campylobacteraceae* sequences still far exceeded the relative abundance of chicken and cow fecal markers, *Gallicola* and *Turicibacter*. While we cannot exclude the possibility of cow or chicken farm contamination in September and October, the higher mean relative abundance of *Campylobacteraceae* across the groundwater wells in every month except August raises the likelihood of STE contamination in the residential well water.

Interestingly, well waters collected in August contained the lowest abundance of the three host-associated markers. Except for well water site K, *E. coli* counts did not exceed 100 CFUs/100 mL (raw data), suggesting diffuse fecal contamination among the well water samples collected in August. In other sampling months, there was a far higher abundance of *Campylobacteraceae* sequences in the autumn. The detection of *E. coli* counts could be due to leaking septic tanks from individual sites.



## Quantification of human-based fecal contamination

To confirm the presence of STE contamination in selected well sites, we used the HF183 marker to quantify human-based contamination in DNA extracted from STE and groundwater samples. We first validated the use of the HF183 marker by preparing standard and melt curves of the HF183 qPCR assay. Although qPCR assays were conducted with 100 bp fragments to minimize spurious fluorescence, we observed high  $R^2$  value and robust E-value, suggesting the quality of the amplification reactions (Figure S1A). Furthermore, the melt curve analysis shows that a single product was detected by the qPCR assay (Figure S1B). Furthermore, we detected the HF183 marker in the two STE sites at 10-fold higher concentration than groundwater samples. These results validate quantifying the entire HF183 amplicon in DNA extracted from the town's well water as a proxy of STE contamination.

We also detected the HF183 DNA marker in B and K, groundwater wells containing the highest *E. coli* CFUs. Curiously, groundwater wells I and F, which did not contain *E. coli*, were also positive for the HF183 marker. A weak correlation between *E. coli* counts and HF183 marker concentrations was established in residential areas (Nshimiyimana et al. 2014) and the Great Lakes beach sands (Staley et al. 2015). The weak correlation (Pearson correlation;  $r = 0.33$ ) between *E. coli* counts and HF183 marker concentrations indicates that *E. coli* levels are not a reliable indicator of human fecal contamination in well waters. This is primarily due to the factors that affect viability of *E. coli* in well water and contamination of well waters through sources other than humans. Within Wainfleet, groundwater wells B and K may receive *E. coli* from Lake Erie where surface water flows into the aquifer. However, these groundwater wells also receive human contamination loads that may come from leaking septic tanks, evidenced by the positive HF183 detection in septic tanks and the wells.

Interestingly, the groundwater well sample collected from well A contained almost five times the HF183 marker concentration as groundwater wells B and K. Well A's homeowner installed a UV treatment system in residence to inactivate fecal microbes. While *E. coli* and *Enterococcus* were absent in the UV treated well water, the HF183 marker was still present in UV treated well water. This could be due to the

presence of DNA from dead (or inactivated) *Bacteroidales* (inactivated through UV treatment) which was amplified during the qPCR assay. Other groundwater wells, like G and J, did not contain a detectable HF183 marker, removing the possibility of STE contamination in those wells. Culturable FIBs were also absent in G and J, corroborating the absence of STE-based contamination.

## CONCLUSIONS

A prior study found extensive FIB contamination throughout a rural community (Niagara Region Report 2007). In this study, we complemented cultured-based methods with 16S sequencing and qPCR to re-assess fecal contamination levels and, further, to identify potential contamination sources in Wainfleet. We found FIB contamination in the well waters we tested, with *E. coli* counts as high as  $10^6$  CFU/mL. In addition, the identification of additional STE DNA markers to *Campylobacteraceae* and the human *Bacteroidales* is consistent with the idea that human waste may impact residential groundwater wells. The low abundance (and absence) of animal manure-associated DNA markers – *Turicibacter* in cow manure and *Gallicola* in chicken manure – reduces the possibility that observed contamination is due to animal sources. These results indicate that profiling microbial communities using 16S rRNA sequencing can augment culture-based methods in contamination analysis studies. The use of next-generation sequencing methods can specifically facilitate the assessment of groundwater quality by detecting host-associated markers and quantifying relative contributions of likely fecal sources to groundwater contamination.

To trace fecal contamination sources in water sources more robustly, amplicon sequencing at higher depths (more reads per sample) may be useful in identification of novel or rare taxa. Shotgun metagenomic sequencing may also be used which, at a higher depth, may facilitate the identification of FIB at the species level. While deeper sequencing depth may facilitate species identification, such as the *Campylobacter* spp. that comprise STEs and possible pig and horse-associated markers within *Clostridium* and other genera, the cost of such in-depth analyses are fairly high. Highly focused DNA sampling programs that target

well water sites having high FIB levels may allow detailed identification contamination inputs that include potential pathogens that, unlike *E. coli*, are difficult to culture.

## ACKNOWLEDGEMENTS

We thank members of the Schellhorn Lab for their help and comments on the manuscript. We thank Trevor Imhoff (Wainfleet Township) for advice and support of this project. We gratefully acknowledge the financial support through the Niagara Region WaterSmart Program, the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Program (CRDP), and NSERC operating grants to HES. The authors declare no conflict of interest.

## REFERENCES

- APHA 2012 *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington, DC.
- Bernhard, A. E. & Field, K. G. 2000 *A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA*. *Appl. Environ. Microbiol.* **66** (10), 4571–4574.
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A. & Caporaso, J. G. 2013 *Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing*. *Nat. Meth.* **10** (1), 57–59.
- Cao, Y., Van De Werfhorst, L. C., Dubinsky, E. A., Badgley, B. D., Sadowsky, M. J., Andersen, G. L., Griffith, J. F. & Holden, P. A. 2013 *Evaluation of molecular community analysis methods for discerning fecal sources and human waste*. *Water Res.* **47** (18), 6862–6872.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Tumbaugh, P. J., Walters, W. A., Widmann, J., Yatsunencko, T., Zaneveld, J. & Knight, R. 2010 *QIIME allows analysis of high-throughput community sequencing data*. *Nat. Meth.* **7** (5), 335–336.
- Edgar, R. C. 2010 *Search and clustering orders of magnitude faster than BLAST*. *Bioinformatics* **26** (19), 2460–2461.
- Field, K. G. & Samadpour, M. 2007 *Fecal source tracking, the indicator paradigm, and managing water quality*. *Water Res.* **41** (16), 3517–3538.
- Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. 2014 *Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes*. *FEMS Microbiol. Rev.* **38** (1), 1–40.
- Health-Canada 2013 *Guidance on the Use of the Microbiological Drinking Water Quality Guidelines*. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada.
- Howard, G., Pedley, S., Barrett, M., Nalubega, M. & Johal, K. 2003 *Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda*. *Water Res.* **37** (14), 3421–3429.
- Kim, M., Morrison, M. & Yu, Z. 2011 *Status of the phylogenetic diversity census of ruminal microbiomes*. *FEMS Microbiol. Ecol.* **76** (1), 49–63.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glockner, F. O. 2013 *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies*. *Nucleic Acids Res.* **41** (1), e1.
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R. & Hugenholtz, P. 2012 *An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea*. *ISME J.* **6** (3), 610–618.
- McLellan, S. L., Newton, R. J., Vandewalle, J. L., Shanks, O. C., Huse, S. M., Eren, A. M. & Sogin, M. L. 2013 *Sewage reflects the distribution of human faecal Lachnospiraceae*. *Environ. Microbiol.* **15** (8), 2213–2227.
- Mohiuddin, M. M., Salama, Y., Schellhorn, H. E. & Golding, G. B. 2017 *Shotgun metagenomic sequencing reveals freshwater beach sands as reservoir of bacterial pathogens*. *Water Res.* **115**, 360–369.
- Mohiuddin, M. M., Botts, S. R., Paschos, A. & Schellhorn, H. E. 2018 *Temporal and spatial changes in bacterial diversity in mixed use watersheds of the Great Lakes region*. *J. Great Lakes Res.* **45** (1), 109–118.
- Niagara Region Report 2007 *Wainfleet Township – Lakeshore Area: Water and Wastewater Servicing; Final Report on Alternatives*. Wainfleet, Ontario, Canada.
- Nshimiyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L. H. & Thompson, J. R. 2014 *Distribution and abundance of human-specific Bacteroides and relation to traditional indicators in an urban tropical catchment*. *J. Appl. Microbiol.* **116** (5), 1369–1383.
- Seurinck, S., Defoirdt, T., Verstraete, W. & Siciliano, S. D. 2005 *Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater*. *Environ. Microbiol.* **7** (2), 249–259.
- Staley, Z. R., Chase, E., Mitraki, C., Crisman, T. L. & Harwood, V. J. 2013 *Microbial water quality in freshwater lakes with different land use*. *J. Appl. Microbiol.* **115** (5), 1240–1250.

- Staley, Z. R., Vogel, L., Robinson, C. & Edge, T. A. 2015 Differential occurrence of *Escherichia coli* and human *Bacteroidales* at two great lakes beaches. *J. Great Lakes Res.* **41** (2), 530–535.
- Tomaras, J., Sahl, J. W., Siegrist, R. L. & Spear, J. R. 2009 Microbial diversity of septic tank effluent and a soil biomat. *Appl. Environ. Microbiol.* **75** (10), 3348–3351.
- Unno, T., Jang, J., Han, D., Kim, J. H., Sadowsky, M. J., Kim, O.-S., Chun, J. & Hur, H.-G. 2010 Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environ. Sci. Technol.* **44** (20), 7777–7782.
- USEPA 1993 *Wellhead Protection: A Guide for Small Communities*. USEPA, Office of Research and Development Office for Water, Washington, DC.

First received 3 December 2018; accepted in revised form 14 January 2019. Available online 1 March 2019