Microbial Source Tracking in the Love Creek Watershed, Delaware (USA)

Christopher R. Main, PhD;[1] Robin Tyler, PhD;[1] Sergio Huerta, MD[2]

1. Environmental Laboratory Section, Division of Water, Delaware Department of Natural Resources and Environmental Control

2. Laboratory Director, Public Health and Environmental Laboratories, State of Delaware

Abstract

Fecal contamination of waterways in Delaware pose an ongoing problem for environmental and public health. For monitoring efforts, *Enterococcus* has been widely adopted by the state to indicate the presence of fecal matter from warm-blooded animals and to establish Primary and Secondary Contact Recreation criteria. In this study, we examined sites within the Love Creek watershed, a tributary of the Rehoboth bay, using next-generation sequencing and SourceTracker to determine sources of potential fecal contamination and compared to bacterial communities to chemical and nutrient concentrations. Microbial community from fecal samples of ten different types of animals and one human sample were used to generate a fecal library for community-based microbial source tracking. Orthophosphate and total dissolved solids were among the major factors associated with community composition. SourceTracker analysis of the monthly samples from the Love Creek watershed indicated the majority of the microbial community were attributed to "unknown" sources, i.e. wildlife. Those that attribute to known sources were primarily domestic animals, i.e. cat and dog. These results suggest that at the state level these methods are capable of giving the start for source tracking as a means to understanding bacterial contamination.

Introduction

The Department of Natural Resources and Environmental Control (DNREC) Environmental Laboratory Section (ELS) has been monitoring the waters of Delaware for several decades as a requirement of the Clean Water Act (CWA) (USEPA 1987 - Sections 106, 303, 304 and 305). Section 303(d) identifies "impaired" waters as those that do not meet the Water Quality Standards laid out in Section 304(a). It includes processes for determining the degree of pollution reduction from human related sources likely to result in attainment of the Total Maximum Daily Load (TMDL) and recommending and informing land and water management practices necessary to achieve the targeted pollution reduction. Routine long-term monitoring at established stations tracks water quality status and documents how State waters are responding to environmental stewardship efforts.

One measurable type of potential human-related water pollution is fecal bacteria. These bacteria may derive from humans, various domesticated animals (i.e. dogs, cats, cows, etc.) or from wild animals. Bacteria within the genus *Enterococcus* has been widely adopted to indicate the presence of fecal matter from warm-blooded animals in 305(b) monitoring efforts. Delaware has risk-based numeric criteria for Primary Contact Recreation (PCR) and Secondary Contact Recreation (SCR) in freshwater and saltwater "determined by the Department (DNREC) to be of non-wildlife origin based on best scientific judgment using available information."[1] Important among the limitations of the *Enterococcus* test is that it does not differentiate as to what type of

animal the detected bacteria are from. It could be any mix of warm-blooded animal types that might be in the watershed – i.e. mammals and birds, domestic and wildlife.

While these *Enterococcus* criteria are generally recognized to be protective of human health and the continuation of routine monitoring of waters is necessary to track status and trends, this test does not help in locating sources of bacterial contamination, which may be abatable. A supplemental, complimentary addition to the existing monitoring format is needed to facilitate where to apply pollution control practices and maximize the frugal utilization of the increasingly scarce resources available to bring about the environmental improvements intended under the CWA, such that waters meet standards criteria and attain designated uses,[1] for example, PCR, SCR, fish, aquatic life and wildlife and harvestable shellfish waters.

Delaware's inland coastal bays (DIB) consist of three interconnected water bodies, Rehoboth, Indian River and Little Assawoman bays that drain approximately 300 square miles of mixed land use. Eutrophication of the DIB has increased over the last several decades with inputs from agricultural and urban sources.[2,3] The Love Creek watershed (Figure 1) is a tributary of the Rehoboth bay and part of the National Estuary Program for over 20 years. The watershed has undergone extensive human development of various types in its tidal and non-tidal segments with the inevitability of substantially more, making for an ideal study site. Its environmental condition, aquatic and terrestrial, could benefit substantially from improved precision in identifying human-related pollution sources. Additionally, considerable periodic seabird activity in the tidal segment and relatively wide forested stream corridors in the non-tidal segment are evidence of a robust complement of indigenous wildlife (e.g. deer, raccoons).

Figure 1. Septic System density in the Love Creek watershed.

Microbial source tracking (MST) methods use specific bacterial profiles associated with hosts, e.g. human, farm animal, bird, to determine sources of fecal contamination in the environment. The most common MST method uses quantitative real-time polymerase chain reaction (qPCR) and primers that target the 16S rRNA genes of host associated bacteria,[4,5] however, qPCR may suffer from specificity and sensitivity issues.[6–8] In the last decade, advancements in high-throughput DNA sequencing has led to large-scale microbial community studies. Community based MST analysis is generated from unique microbial community profiles of environmental and fecal sources leveraged from next generation sequencing (NGS) techniques,[9] Two methods may be used to define the microbial community: 1) assignment of operational taxonomic units (OTUs) via a clustering algorithm, usually to a 97% similarity between sequences, or 2) amplicon sequence variant (ASV) which can resolve difference in gene regions to a single nucleotide.[10]

SourceTracker is a Bayesian classification program, which provides an estimated percentage of the sequenced microbial community from an environmental samples can be attributed to a specific fecal source.[6,11] SourceTracker uses a Gibbs sampling algorithm to examine the likely prior distribution of OTUs/ASVs within user-defined sources. This distribution within the sources can be used to determine the affiliation within sample (referred to as sinks) communities and the contribution of each source is then determined based on this distribution. The SourceTracker algorithm has been shown to be more accurate than random forest analysis, or naive Bayesian classification.[11] SourceTracker has been used to determine contamination sources in the Russian River,[12] ATM keypads,[13] recreational beaches in Australia,[4,14–16] lakes in St. Paul, Minnesota,[6] and the upper Mississippi River.[17,18]

Our goals were to use next generation sequencing to create community based microbial profiles of fecal sources and monthly samples of sites within the Love Creek watershed. Using SourceTracker we examined the bacterial inputs from various fecal sources into the monthly samples. In addition, we examined the impact of a significant rain event, within 24 hours, at the tidally driven sites. The results of this pilot study are broadly relevant to assessing fecal pollution within the waterways of Delaware.

Material and Methods

Study Site and Sample Collection

The Love Creek watershed is drains approximately 24 square miles of land into the Rehoboth Bay (Figure 1) and is tidally driven to the dam at Goslee Pond (STORET: 308291).[19] In the last two decades urban development has increased approximately 80% with a loss of forested uplands and agriculture.[19] Water samples were collected monthly from March to October 2017 from seven sites in the Love Creek watershed (Table 1), with triplicate samples collected at the Rt. 24 Marina (STORET: 308021). Sites were considered marine waters when salinity was greater than 5.[1] To examine the influence of rainfall on microbial transport, rainfall was monitored using the Delaware Environmental Observing System at the Millsboro Long Neck station (DLNK). Samples at the tidally driven sites (STORET: 308015, 308018, 308021, and 308024) were collected 24 hours after a rainfall event of approximately 62 mm of rain in two hours. Temperature, salinity, dissolved oxygen (mg l^{-1}), pH, and specific conductivity were measured using a YSI 650 (YSI Inc., Yellow Springs, OH). Dissolved nutrients (NO$_3$ plus NO$_2$ [NO$_X$], OP, organic carbon, and dissolved solids), and total nutrients (chlorophyll a, nitrogen, phosphorus, organic carbon, turbidity and suspended solids) concentrations were determined by APHA Standard Methods.[20] *Enterococcus* concentrations were determined for each sample using Enterolert© (IDEXX, Westbrook, ME).

Table 1. Location Descriptions, STORET Numbers and Latitude and Longitude of Sampling Sites

Location	STORET Number	Latitude	Longitude
Love Creek mouth, Rehoboth Bay Community Dock	308015	38.68637	-75.12705
Love Creek lower, dock at end of West Lane	308018	38.69265	-75.14222
Love Creek middle, at Rt. 24 bridge	308021	38.70323	-75.16156
Love Creek upper, dock off Misty Lane	308024	38.71282	-75.17635
Love Creek at Robinsville Rd. (Goslee Pond outlet)	308291	38.71775	-75.18684
Bundicks Branch at Rt. 23 (Bill Ullman's Station)	308371	38.72151	-75.20627
Goslee Creek at Jimtown Rd.	308411	38.73050	-75.18430

Sample Filtering and Processing

Water samples were filtered within six hours of sampling under gentle vacuum (~380 mm Hg) on 0.45-µm polycarbonate filters (Millipore Isopore, Billerica, MA). Filters were immediately placed into CTAB buffer (100 mM Tris-HCl (pH 8), 1.4 M NaCl, 2% (wt/vol) cetyltrimethylammonium bromide (CTAB), 0.4% (vol/vol) 2-mercaptoethanol, 1% (wt/vol) polyvinylpyrrolidone, and 20 mM EDTA)[21] and stored at -80°C until extraction. Fecal samples were collected in July 2016 during the Delaware State Fair. Bovine, chicken, goat, horse, pig,

sheep, domesticated duck and goose samples were collected with as much metadata information as possible, i.e. location, feed, sex, etc. by Christopher Main and Karen Lopez, Deputy State Veterinarian for Delaware. Cat, dog and human samples were collected by or from Christopher Main and all samples were stored at -80°C until extraction.

Before extraction, filters were heated at 65°C for 10 minutes. Following incubation, 700 µl of a 24:1 isoamyl:chloroform solution was added to each sample and briefly vortexed. Samples were gently rocked for 20 minutes and centrifuged at $14000 \times g$ for 15 minutes. The top aqueous phase was used for further extraction using GeneJet Plant Extraction kit according to manufacturer's instructions (Thermo Scientific, Waltham, MA). Fecal samples were extracted using a PureLink™ Microbiome DNA Purification Kit (Thermo Scientific) using a small portion of each fecal sample. All samples, water and fecal, were eluted to 100 µl of elution buffer and stored at -20°C until analysis. The V4 region of the bacterial rRNA gene was selected for community analysis.[22] Extracted samples were sent to Molecular Research LP (Shallowater, TX) for library prep and amplicon sequencing.

Microbial community analysis

Raw samples were processed using FASTQ Processor to generate forward, reverse and barcode fastq files for QIIME2 analysis. Pair-end reads were demultiplexed using the q2-demux emp-paired method and denoised using q2-dada2's denoise-paired method.[23] Sequencing of samples occurred in multiple runs requiring QIIME2 analysis occurring in tandem until after DADA2 analysis, which were then combined for further downstream analysis. Taxonomic composition was generated using a pre-trained Naive Bayes classifier trained on the Greengenes 13_8 99% OTUs trimmed to include the V4 (515F/806R primer pairs). Prior to further downstream analysis, nontarget ASVs, i.e. chloroplast and mitochondria, were removed from sample table and representative sequences.

Constrained analysis of principal coordinates (CAP) was carried out using the vegan R package[24] to identify factors contributing to differences between sites with respect to environmental factors. Collinearity of environmental factors was examined using a Pearson's correlation with factors considered collinear with an r^2 of 0.8 and $P < 0.05$. Weighted Unifrac and Bray-Curtis dissimilarity matrices were used for CAP analysis for environmental samples including total N and P and without total N and P, models were selected by pseudo-AIC using a stepwise algorithm. Alpha and beta diversity, including weighted and unweighted Unifrac principal component analysis, analyses of ASVs was carried out using the phyloseq R package.[25] Dissimlarity between groups was tested using non-parametric PERMANOVA tests with weighted and unweighted Unifrac beta diversity distances.[13] All statistical analyses were evaluated at an $\alpha = 0.05$.

An ASV table was exported for use with SourceTracker[11] using a parallel version SourceTracker to decrease computational time. Default parameters established in the sourcetracker_for_qiime.py pipeline were used with five runs being conducted. For each sample, the mean proportion (%) for each run were averaged for determination of fecal source contribution.

Results

Study Site and Field Samples

A total of 2,261 septic systems, with a density of 94.2 systems per square mile, are located within the Love Creek Watershed, with gravity systems being the dominant system type (Figure 1). Previous analysis by Homsey et al.[19] showed a density of active septic permits of 55.5 permits per square mile, with 1,340 septic systems within the watershed, an increase of approximately 70% of septic systems within the watershed.

Altogether, 74 samples were analyzed from seven sites within the Love Creek watershed between March and October 2017. Water temperatures during the collection period ranged from 8.3 to 29.7°C, salinity from 0.1 to 41.2, chlorophyll a from 0.3 to 382 mg l^{-1}, dissolved oxygen from 5.01 to 17.3 mg l^{-1}, turbidity from 1 to 44 mg l^{-1}, total suspended solids from 1 to 75.9 mg l^{-1} and total dissolved solids from 70 to 32,300 mg l^{-1}. Organic carbon concentrations ranged from 1.8 to 10.9 mg C l^{-1}, dissolved NO_X from 0.004 to 7.13 mg N l^{-1}, NH_3 from 0.01 to 0.281 mg N l^{-1}, PO_4 from 0.004 to 0.047 mg P l^{-1}, total nitrogen from 0.242 to 7.35 mg N l^{-1}, and total phosphorus 0.014 to 0.272 mg P l^{-1}. Total *Enterococcus* levels ranged from non-detects to 6,130 mpn 100 ml^{-1} with 27 of 74 samples being above primary contact recreation levels. At Jimtown Road, five samples were above PCR, Bundicks Branch four samples above PCR, Misty Lane three samples above PCR, Route 24 eleven of twenty-four samples above PCR, and West Lane four of nine samples were above PCR. No samples for total *Enterococcus* at Goslee Pond and the mouth of Love Creek were above PCR levels (Figure 2).

Figure 2. *Enterococcus* levels as determined by Enterolert for each sampling site over the sampling period. PCR for freshwater (blue line) and marine waters (>5 ppt, red line) are also shown.
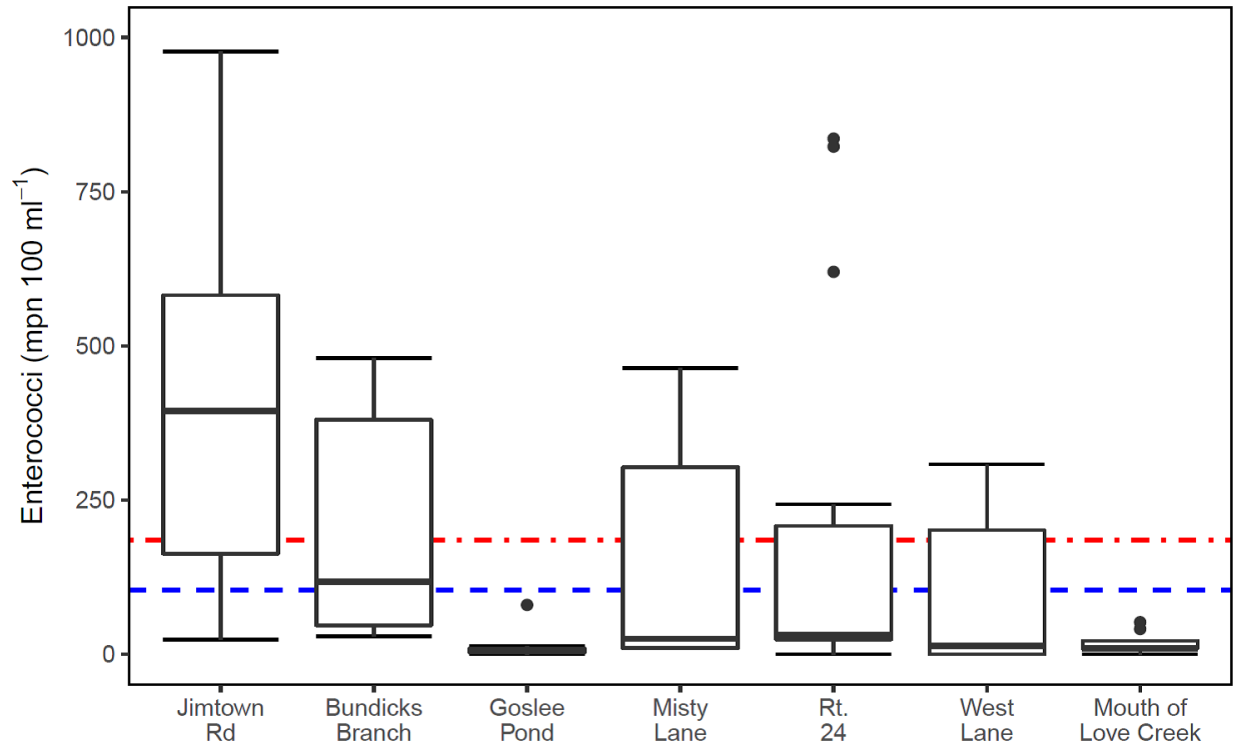
Microbial Community Analysis

Taxonomic analysis for 16S data showed that major taxa were largely consistent across sampling sites (Figure 3). The most abundant phyla across most sample locations were *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria* (Figure 3a). *Firmicutes*, one of the most abundant gut microbes,[26] was one of the top phyla for all sampling sites. At Jimtown Road and Bundicks Branch, *Acidobacteria* and *OP3*, were also in high abundance across all sampling time points. Both the *Acidobacteria*[27] and *OP3*[26] have been identified from anoxic sediment samples suggesting a potential resuspension of sediments at these sites. At the class level, *Gammaproteobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Flavobacteriia* showed the highest relative abundance across most of the sampling sites (Figure 3b).

Figure 3. Relative Abundance of bacterial groups at the phylum level (A) and class level (B), showing the top 15 most abundant taxa from monthly samples.
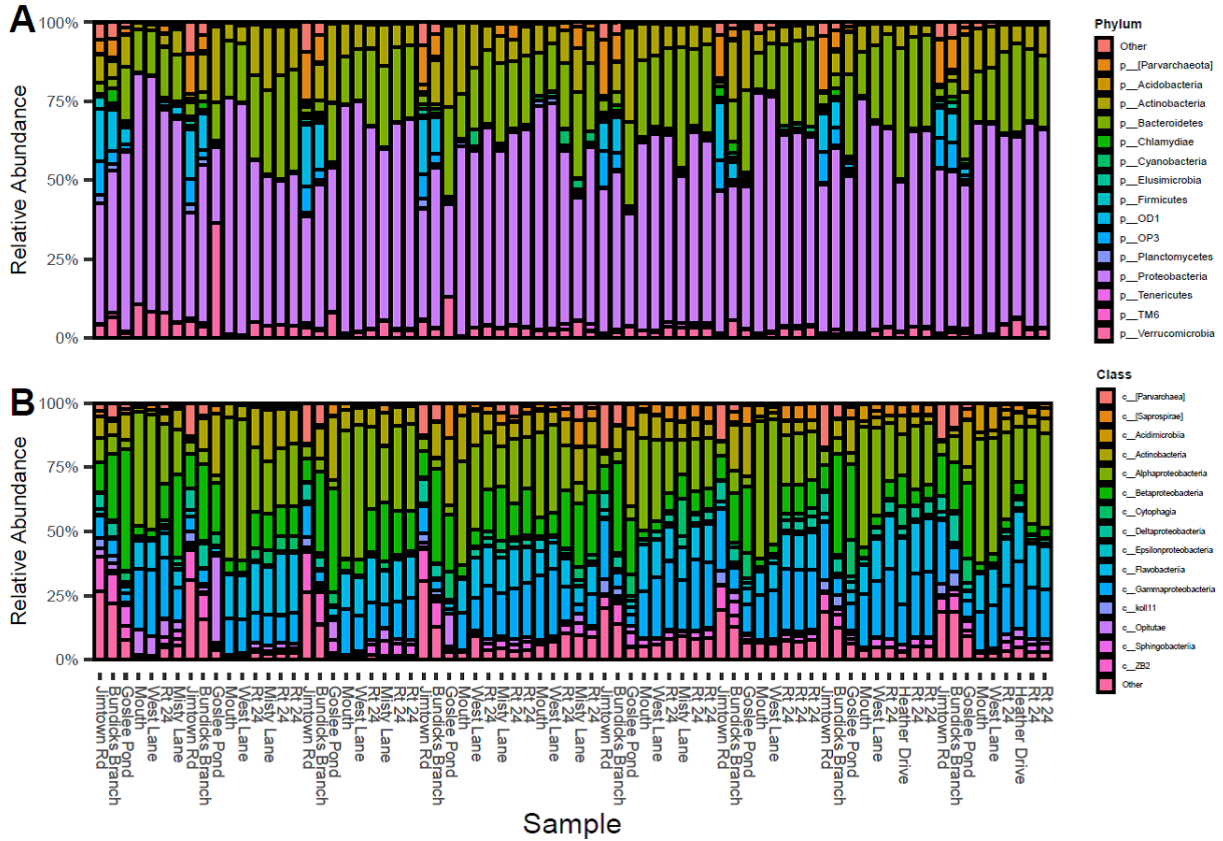
For fecal samples at the phylum level, the most abundant phyla were *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Euryarchaeota* (Figure 4a). *Clostridia*, *Bacilli*, *Bacteroidia*, and *Methanobacteria* were the dominant classes for fecal samples (Figure 4b).

Figure 4. Relative Abundance of bacterial groups at the phylum level (A) and class level (B), showing the top 15 most abundant taxa from fecal samples.
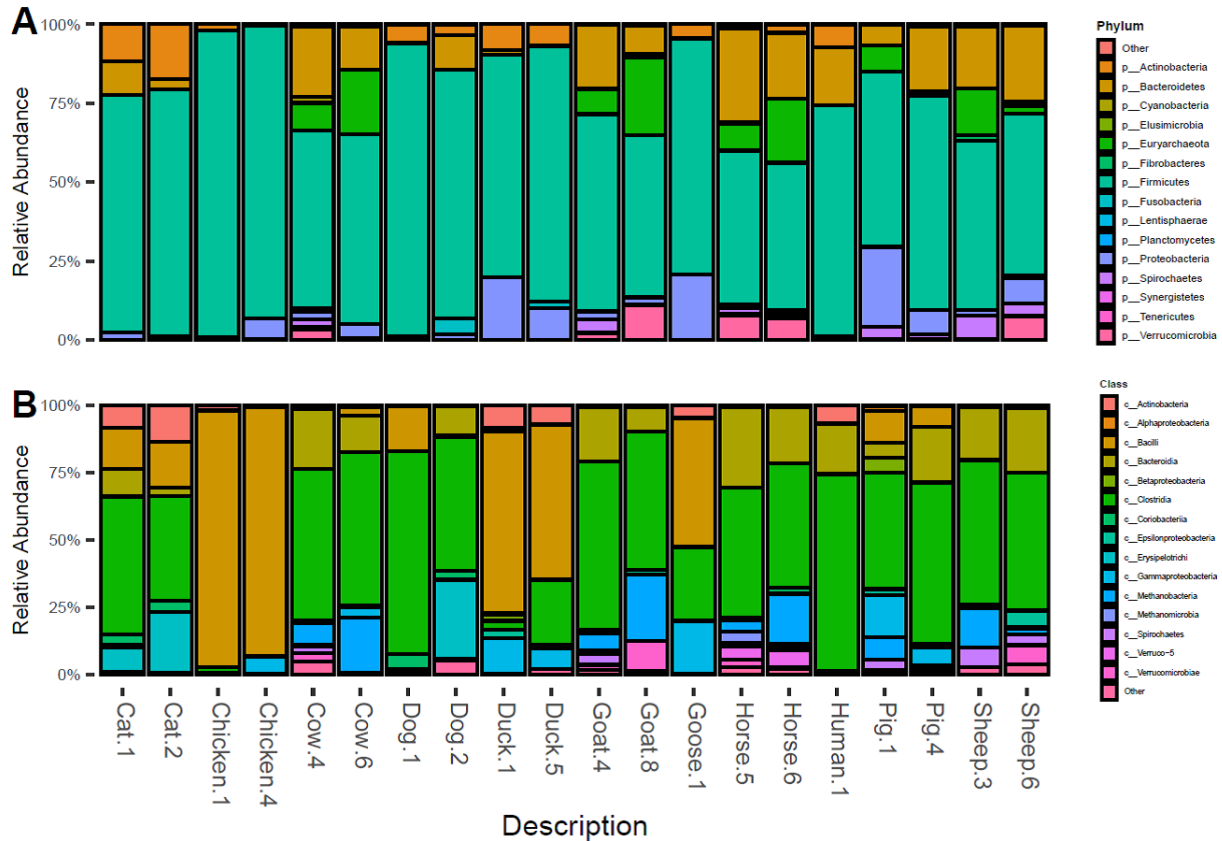
Chlorophyll *a*, dissolved oxygen saturation, NO$_X$, specific conductivity, total dissolved solids (TDS) and turbidity were determined to be collinear (Total P, DO, Total N, Salinity, Salinity and total suspended solids, respectively). Previous research has shown a significant correlation to microbial community structure and total dissolved solids,[17] therefore all CAP evaluations included total dissolved solids. Model selection by stepwise algorithm for weighted Unifrac on all selected factors were TDS, temperature, OP, pH, Salinity, NH$_3$, total phosphorus and dissolved oxygen and without total N and P: TDS, temperature, OP, pH, Salinity, NH$_3$, chlorophyll *a* and dissolved oxygen (Figure 5a) with 51.2% of the variation explained on CAP1 and 8.3% explained on CAP2. For Bray-Curtis factors were: salinity, temperature, OP, *Enterococcus*, total N, NH$_3$, total organic C, total P, pH, chlorophyll *a*, and dissolved oxygen and without total N and P: salinity, temperature, OP, *Enterococcus*, NO$_X$, total organic C, NH$_3$, pH, chlorophyll *a*, and dissolved oxygen (Figure 5b) with 21% of the variation accounted for on CAP1 and 9.7% accounted for on CAP2.

Figure 5. Constrained analysis of principal coordinates showing relationships among species compositions for sites and environmental factors (arrows) for weighted Unifrac (A) and Bray-Curtis (B) for all non-collinearity factors within the watershed (color) and sampling month (shape).
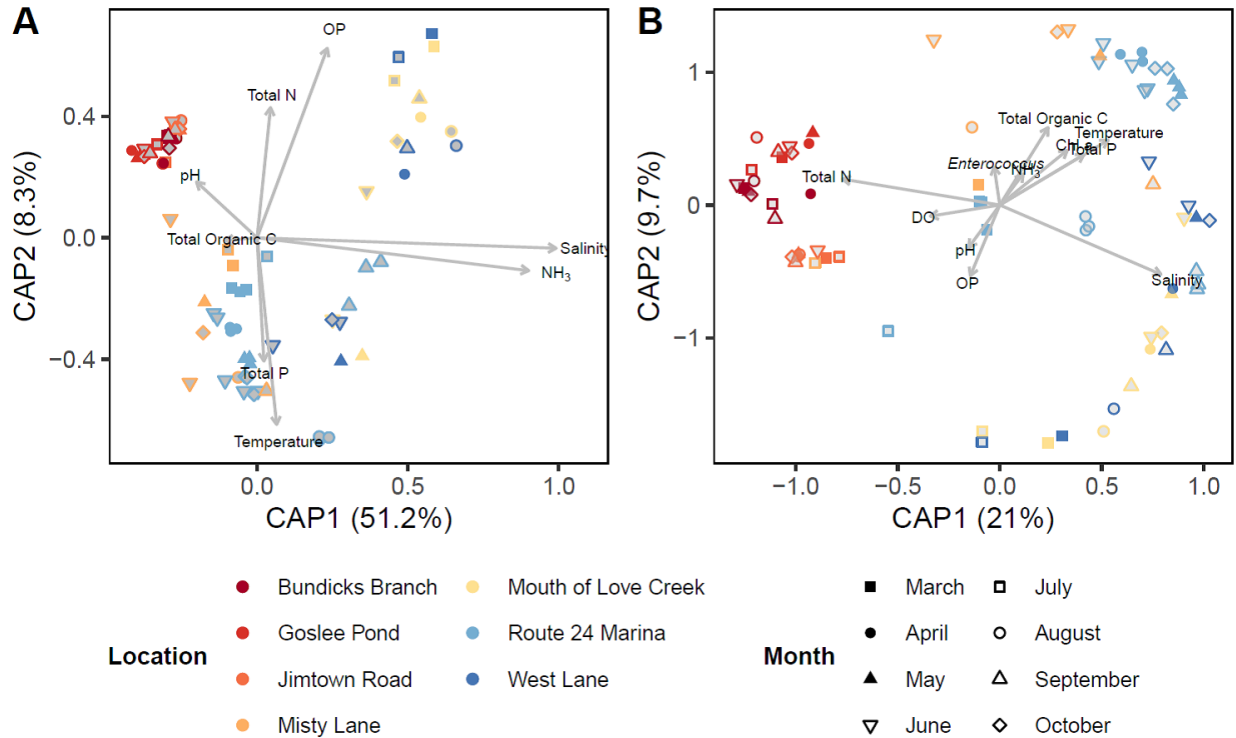
Beta diversity analysis indicated a strong clustering of one distinct group with non-tidal sites (Jimtown Road, Bundicks Branch and Goslee Pond) separated from tidally driven sites for both weighted Unifrac with 56.9% of the variation accounted for on PCoA1 and 15.7% accounted for on PCoA2 and Bray-Curtis (Figure 6a and 6b respectively) with 24.1% of the variation on PCoA1 and 11.6% on PCoA2.

Figure 6. Beta diversity analyses of microbial community for weighted Unifrac PCoA (A) and Bray-Curtis (B) showing sampling location within the watershed (color) and sampling month (shape).
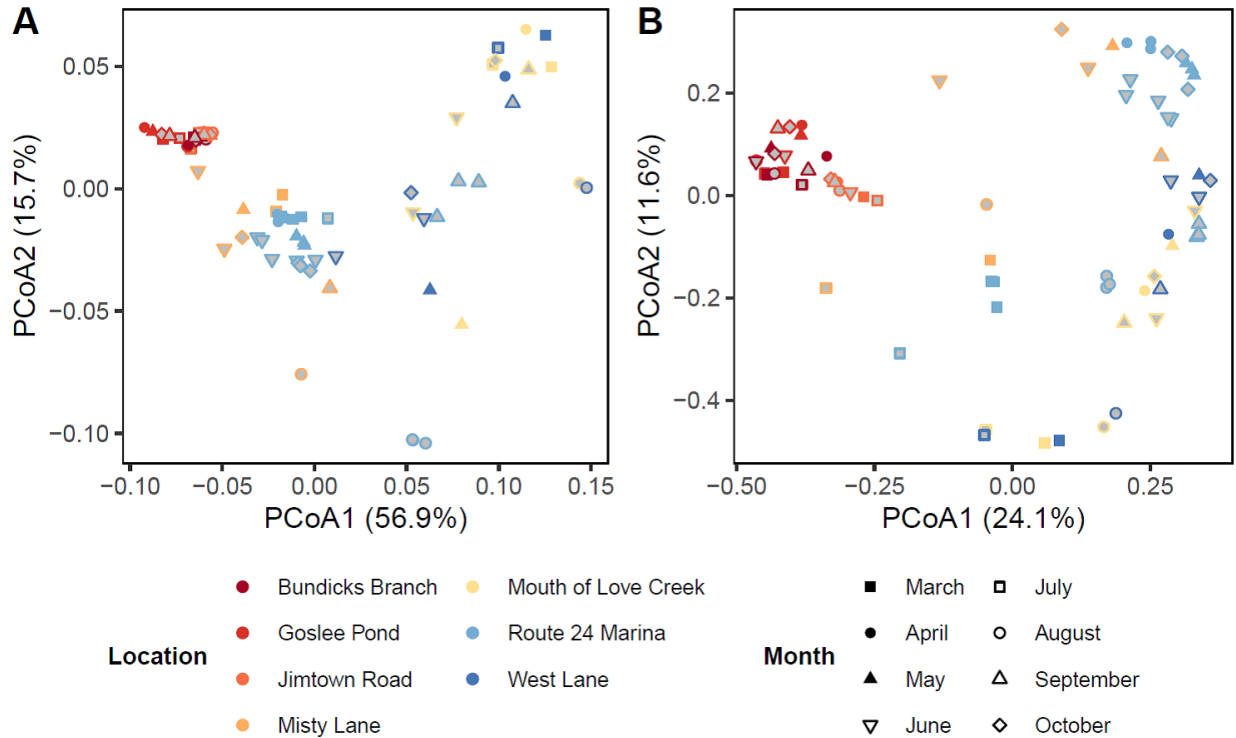
Samples collected at the Route 24 marina and Misty Lane clustered together depending on sampling time salinity and temperature (Figures 5 and 6).

PERMANOVA testing showed significant groupings for most environmental factors for both weighted Unifrac and Bray-Curtis (Table 2). Previous research[13,28] has shown that a putative technical artifact resulting in splitting of 16S sequencing runs. For weighted Unifrac PERMANOVA analysis this artifact was not statistically significant for sample grouping (P = 0.0694), whereas Bray-Curtis was statistically significant (P = 0.0001). For fecal samples, the ruminant species clustered together for Bray-Curtis but were more dispersed in weighted Unifrac analysis. Additionally, avian samples were clustered in the Bray-Curtis analysis but not weighted Unifrac.

Table 2. PERMANOVA test for statistical significance (Bold values represent P < 0.05)

Environmental Factor	Weighted UniFrac P Value	Bray-Curtis P Value
Monthly Sampling	**0.0001**	**0.0001**
Site Location	**0.0001**	**0.0001**
Temperature	**0.0317**	**0.0001**
Salinity	**0.0001**	**0.0001**
Total dissolved solids	**0.0001**	**0.0001**
NH$_3$	0.1908	**0.0084**
Chlorophyll a	0.1305	**0.0001**
Dissolved oxygen	0.0657	**0.0057**
Total P	0.1379	**0.0001**
Orthophosphate	**0.0042**	**0.0002**
pH	0.1607	**0.0159**
Total N		**0.0001**
Enterococcus		**0.0137**

|		
Total organic C		**0.0002**
NO$_x$		**0.0001**
Sequence Run	0.0694	**0.0001**

Microbial Sources

SourceTracker analysis on the 16S data indicated that the majority of the bacterial community from all samples were derived from unknown sources (Figure 7, Figure 8). For the majority of sites, <40% of the microbial community was assigned to an identified source. At Jimtown Road, the majority of samples were from unknown sources, i.e. wildlife, with samples from April, July, August, September and October above PCR and SCR levels for *Enterococcus* (Figure 7a). Of those known samples, there was an increase in the amount of cat associated fecal bacteria as the summer progressed. Bundicks Branch had a greater proportion of the microbial community associated with known samples (Figure 7b) with July through September being above PCR and SCR levels. A decrease of known sources occurred from July to August and from September to October to approximately 5% of the microbial community associated to known sources. Samples collected from Goslee Mill Pond also showed an increase in known sources associated with cat fecal matter (Figure 7c), however none of the samples were near PCR during the sampling period (Figure 7).

Figure 7. SourceTracker analysis of bacterial assemblages from freshwater sites at Jimtown Road (A), Bundicks Branch (B) and Goslee Mill Pond (C). Sample times that were above PCR are indicated with a blue rectangle and times above both PCR and SCR are indicated with a red rectangle.
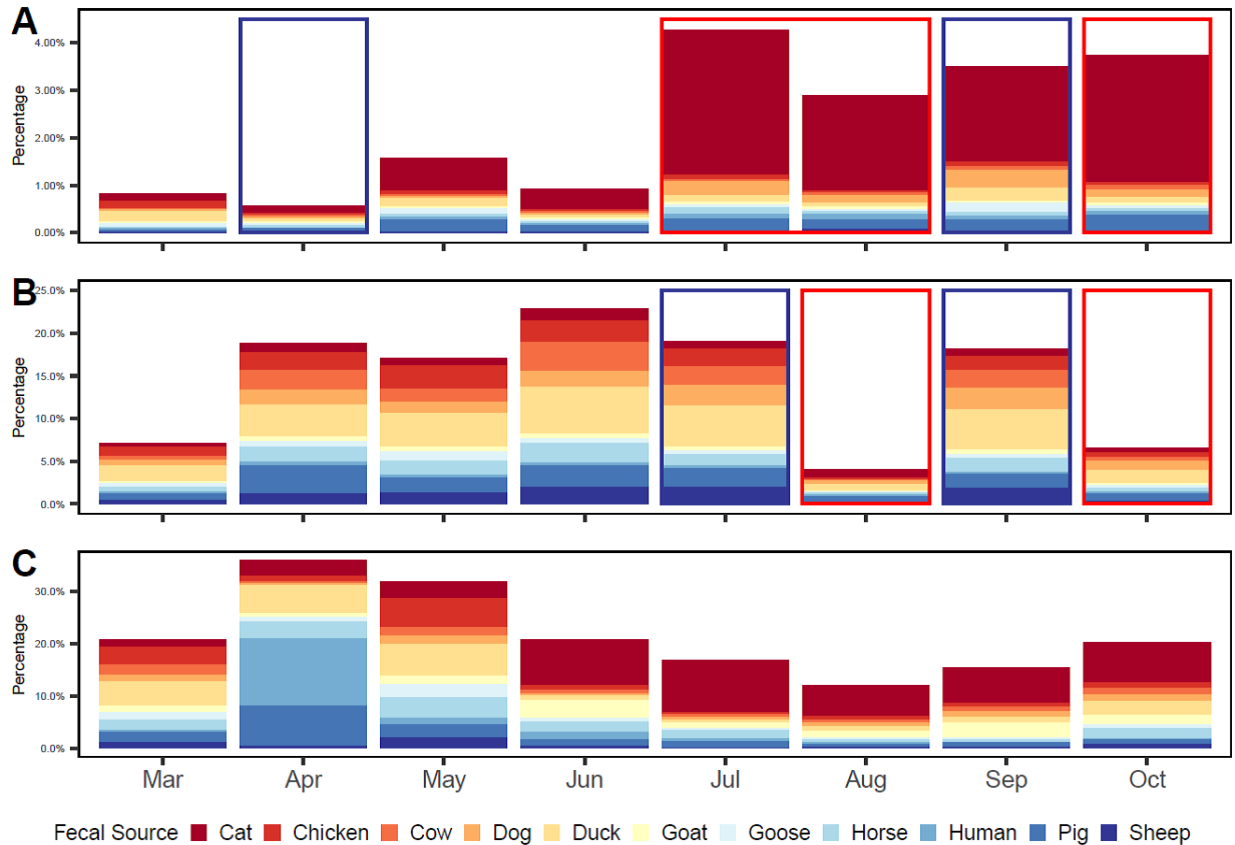
Three sampling periods, June rain event, September and October, at Misty Lane were above PCR levels for *Enterococcus* (Figure 8a). As with Jimtown Road, proportions of the microbial community associated with cat fecal matter increased as the summer progressed. In addition to cat, dog proportions increased over the summer as well, with both of these sources a large portion of the known sources in the microbial community. During the beginning of the sampling period, the housing unit was not occupied, beginning in July onwards the house was occupied with dogs evident on the premises. The Route 24 marina had four separate samples above PCR and SCR during the June rain event, and August through October. Both cat and dog sources were a large portion of the known sources at the marina (Figure 8b). West Lane showed four sampling periods above PCR levels and similar trends in known sources with large proportions of the microbial community associated with dog and cat fecal matter (Figure 8c). The mouth of Love Creek did not have any samples above contact levels, but did show increases in both cat and dog fecal matter associated microbes as the summer progressed (Figure 8d).

Figure 8. SourceTracker analysis of bacterial assemblages from marine sites at Misty Lane (A), Route 24 marina (B), West Lane (C) and the mouth of Love Creek (D). Sample times that were above PCR are indicated with a blue rectangle and times above both PCR and SCR are indicated with a red rectangle.
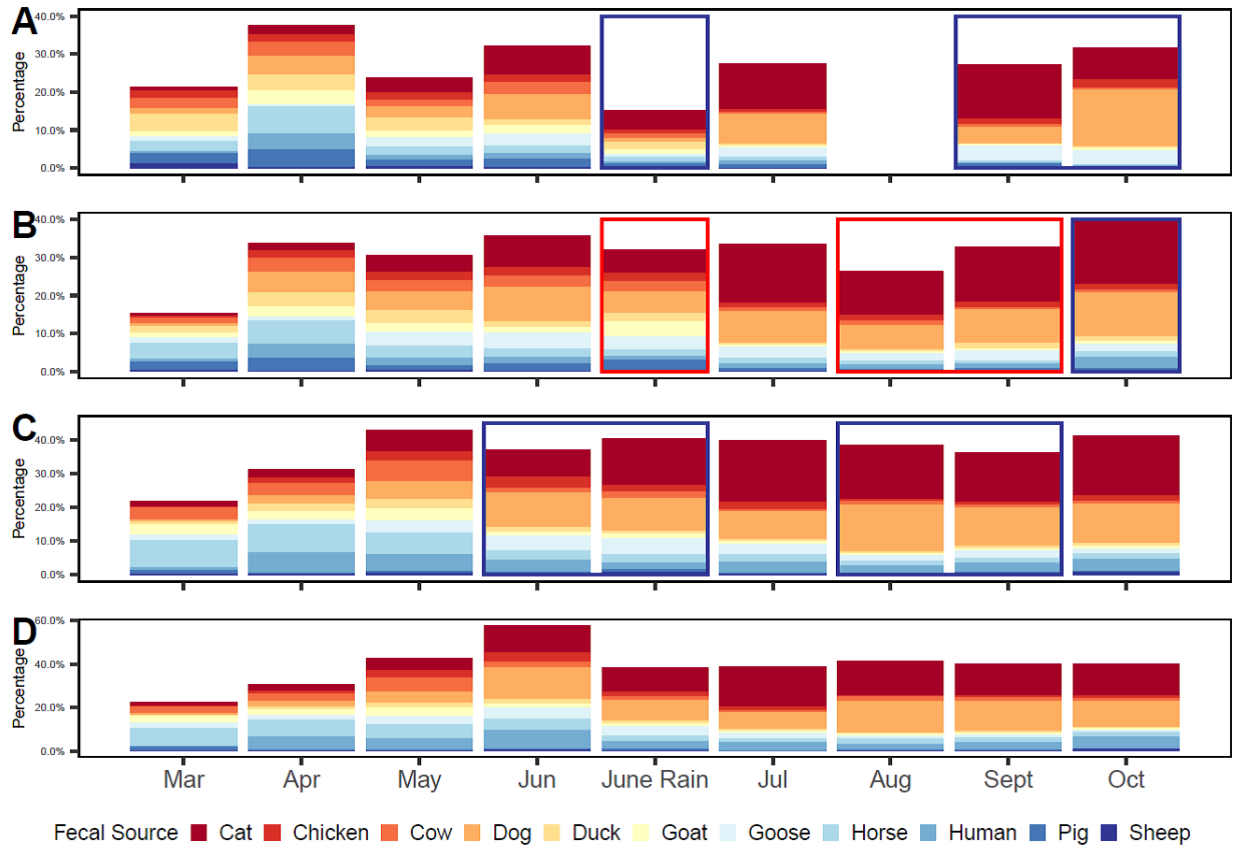
Discussion

Here we evaluated the microbial community and the relationships between environmental and chemical parameters, and the potential usage of next-generation sequencing (NGS) techniques for microbial source tracking (MST) in the Love Creek watershed. The watershed has undergone extensive development over the last several decades resulting in increased the number of potential human-related pollution sources, e.g. septic systems or domestic animals. In the non-tidal sites of Jimtown Road and Bundicks Branch high bacterial levels occurred during the warmer late summer months. Both of these sites are highly wooded with low population densities. At Jimtown Road, there is a high density of aging septic systems, however all samples were below 4% of mapping to known sources suggesting that most of the high bacteria levels are potentially from wildlife sources. Two samples at Bundicks Branch were above PCR followed by above SCR the next month. Interestingly, both points above SCR approximately 95% of the community were identified as unknown sources. Unlike Jimtown Road, Bundicks Branch the highest proportion comes from duck. In contrast, Goslee Mill Pond did not have any high levels of *Enterococcus* during any of the sampling periods suggesting that the pond may act similar to a settling point. Therefore potentially preventing upstream fecal contamination from moving further into the more populated waterways.

Tidally driven sites, with the exception of the mouth of Love Creek, also experienced high *Enterococcus* levels during the warmer months. Land use at Misty Lane is similar to Jimtown Road, with primarily woods with a few smaller vacation homes. A large proportion of known sources are from both cat and dog for July and September. The Route 24 Marina and West Lane demonstrated similar trends for cat and dog and had three periods of *Enterococcus* levels above

SCR and one above PCR. The collection site for Route 24 was directly at a public boat ramp which may increase exposure and health risks for recreational boaters. Additionally, the marina is located within a small development with older septic systems which are primarily used during the summer months with few year round residents. West Lane contains more established homes but at a higher density of septic systems. No high levels of *Enterococcus* were found at the mouth of Love Creek. Sampling occurred at a pier near the edge of the mouth, and may not be indicative of the whole mouth of Love Creek. As with many of the sites within the watershed, cat and dog signals increased during the warmer months, with cat being the largest proportion for both Route 24 and West Lane. The State of Delaware has seen the population of outdoor cats increase over the last several years. As evident by large portions of the known sources from cat signatures. Other sources, e.g. goat and horse, are also in high proportions in many of the sites. However, observations during collection do not indicate likely locations for these sources, requiring a "boots on the ground" approach for determining likely locations of sources.

Both *Bacteroidetes* and *Firmicutes*, two of the major phyla within human fecal matter,[26,29] were present in all our samples but not as prevalent as *Proteobacteria*. Other studies have shown that the prevalence of *Proteobacteria* in samples may be an indicator of run-off or storm water[30,31] or from sewage infrastructure, i.e. a transient population.[30] SourceTracker analysis indicated that almost all samples from the Love Creek watershed had human signatures from <0.1% to 13%. However, the human signature represents a single individual and does not constitute a representative library. Presently, it is not known how many samples would constitute a representative library[14] and large variations may occur among human and other fecal samples.[32] In contrast, Staley, et al.[33] suggests that a minimum of ten individuals are sufficient for accurate fecal contamination detection. However, when few source are used and/or available, results may be better interpreted using broad category classifications, e.g. livestock rather than specific organism. Additionally, Staley, et al.[33] demonstrated to increase accuracy of the SourceTracker analysis, geographically associated fecal samples are required. All samples used in this study were geographically located in Delaware. A large library and rarefaction depth may improve identification of sources by SourceTracker.[14] However, previous research has suggested, from a statistical standpoint, rarefying data was inadmissible,[34] as part of our workflow samples were not rarefied.

In this study, we showed that using NGS methods and bacterial community structure can be combined to identify potential sources of fecal pollution in the Love Creek watershed. The use of SourceTracker and NGS are still part of emerging techniques for the tracking of fecal sources. Increasing the library size of known sources, sequencing depth, and greater diversity of sources will require further analysis.[14] Nonetheless, this study is proof that at the state level these methods are capable of giving the start of source tracking. We recommend using these methods as a starting pointing for ground-truthing, as sources may not be evident. One limitation of these methods are the time to develop a library and the computational requirements for large datasets can be a significant challenge. With open-source tools such as QIIME2, SourceTracker and others, are efficient and simple to use, and with the advent of cloud computing large computational clusters are no longer required.

Public Health Implications

Current methodologies for determining bacterial contamination in the water bodies of Delaware use testing developed to measure the genus *Enterococcus* via Enterolert©. Primary Contact

Recreation (PCR) and Secondary Contact Recreation (SCR) risk-based numeric criteria have been developed by DNREC "to be of non-wildlife origin based on best scientific judgment using available information."[1] However, one limitation of the use of *Enterococcus* as a fecal indicator species is the inability to differentiate between fecal sources, i.e. human-related sources and wildlife. Common MST methods use specific bacterial primers from hosts to measure quantities via quantitative real-time PCR (qPCR).[4,5] However, the development of primers for fecal sources of interest may be time consuming and previous research has shown potential issues with specificity and sensitivity of qPCR.[6–8] The usage of NGS methods to define the microbial community of aquatic and fecal samples using amplicon sequence variant (ASV), which can resolve the difference of sequences to a single nucleotide, increases the power of determining potential fecal contamination sources.[9,10] SourceTracker uses Bayesian models to derive proportions of sources within sink samples, however it has been reported that sources are low concentrations have high variability in estimates.[11] From a public health standpoint, even at low concentrations of fecal contamination may pose public health risks,[5,15] thus the high variability in estimates may inhibit detection. Henry et al.[15] demonstrated that at low proportions (as low as 0.1%), successful detection by SourceTracker can occur.

Correspondence: Dr. Christopher Main, Christopher.main@delaware.gov

References

1. DNREC. Surface Water Quality Standards (2014). Retrieved from http://regulations.delaware.gov/AdminCode/title7/7000/7400/7401.pdf

2. Price, K. (1998). A framework for a Delaware Inland Bays environmental classification. *Environmental Monitoring and Assessment*, *51*(1/2), 285–298. Retrieved from http://www.springerlink.com/index/X207104L477X9255.pdf https://doi.org/10.1023/A:1005951706152

3. Sallade, Y. E., & Sims, J. T. (1997). Phosphorus transformations in the sediments of Delaware's agricultural drainageways: II. Effect of reducing conditions on phosphorus release. *Journal of Environmental Quality*, *26*(6), 1579. https://doi.org/10.2134/jeq1997.00472425002600060018x

4. Brown, C. M., Staley, C., Wang, P., Dalzell, B., Chun, C. L., & Sadowsky, M. J. (2017, August 1). A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary. *Environmental Science & Technology*, *51*(15), 8263–8271. PubMedhttps://doi.org/10.1021/acs.est.7b01353

5. Harwood, V. J., Staley, C., Badgley, B. D., Borges, K., & Korajkic, A. (2014, January). Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiology Reviews*, *38*(1), 1–40. PubMedhttps://doi.org/10.1111/1574-6976.12031

6. Brown, C. M., Mathai, P. P., Loesekann, T., Staley, C., & Sadowsky, M. J. (2019). Influence of Library Composition on SourceTracker Predictions for Community-Based Microbial Source Tracking. [research-article. https://doi.org/10.1021/acs.est.8b04707]. *Environmental Science & Technology*, *53*(1), 60–68.

7. Green, H. C., Haugland, R. A., Varma, M., Millen, H. T., Borchardt, M. A., Field, K. G., . . . Shanks, O. C. (2014, May). Improved HF183 quantitative real-time PCR assay for

characterization of human fecal pollution in ambient surface water samples. *Applied and Environmental Microbiology*, *80*(10), 3086–3094.
PubMedhttps://doi.org/10.1128/AEM.04137-13

8. Stewart, J. R., Boehm, A. B., Dubinsky, E. A., Fong, T. T., Goodwin, K. D., Griffith, J. F., . . . Weisberg, S. B. (2013, November 15). Recommendations following a multi-laboratory comparison of microbial source tracking methods. *Water Research*, *47*(18), 6829–6838.
PubMedhttps://doi.org/10.1016/j.watres.2013.04.063

9. Unno, T., Staley, C., Brown, C. M., Han, D., Sadowsky, M. J., & Hur, H.-G. (2018, September). Fecal pollution: New trends and challenges in microbial source tracking using next-generation sequencing. *Environmental Microbiology*, *20*(9), 3132–3140.
PubMedhttps://doi.org/10.1111/1462-2920.14281

10. Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017, December). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643. PubMedhttps://doi.org/10.1038/ismej.2017.119

11. Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., . . . Kelley, S. T. (2011, July 17). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, *8*(9), 761–763.
PubMedhttps://doi.org/10.1038/nmeth.1650

12. Dubinsky, E. A., Butkus, S. R., & Andersen, G. L. (2016, November 15). Microbial source tracking in impaired watersheds using PhyloChip and machine-learning classification. *Water Research*, *105*, 56–64. PubMedhttps://doi.org/10.1016/j.watres.2016.08.035

13. Bik, H. M., Maritz, J. M., Luong, A., Shin, H., Dominguez-Bello, M. G., & Carlton, J. M. (2016, November 16). microbial community patterns associated with automated teller machine keypads in New York City. *MSphere*, *1*(6), e00226–e16.
PubMedhttps://doi.org/10.1128/mSphere.00226-16

14. Ahmed, W., Staley, C., Sadowsky, M. J., Gyawali, P., Sidhu, J. P. S., Palmer, A., . . . Toze, S. (2015, October). Toolbox approaches using molecular markers and 16S rRNA gene amplicon data sets for identification of fecal pollution in surface water. *Applied and Environmental Microbiology*, *81*(20), 7067–7077.
PubMedhttps://doi.org/10.1128/AEM.02032-15

15. Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., . . . McCarthy, D. (2016, April 15). Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Research*, *93*(March), 242–253.
PubMedhttps://doi.org/10.1016/j.watres.2016.02.029

16. McCarthy, D. T., Jovanovic, D., Lintern, A., Teakle, I., Barnes, M., Deletic, A., . . . Henry, R. (2017, February 1). Source tracking using microbial community fingerprints: Method comparison with hydrodynamic modelling. *Water Research*, *109*, 253–265.
PubMedhttps://doi.org/10.1016/j.watres.2016.11.043

17. Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2014, October 8). Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. *Frontiers in Microbiology*, *5*(OCT), 524.
PubMedhttps://doi.org/10.3389/fmicb.2014.00524

18. Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2015, February 1). Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River. *The Science of the Total Environment*, *505*, 435–445. PubMedhttps://doi.org/10.1016/j.scitotenv.2014.10.012

19. Homsey, A., Walch, M., Boswell, S., & Bason, C. (2015). *The State of Your Creek: Love Creek on Rehoboth Bay*.

20. American Public Health Association. (2012). *Standard methods for the examination of water and wastewater* (22nd ed.). Washington, DC: American Public Health Association.

21. Dempster, E. L., Pryor, K. V., Francis, D., Young, J. E., & Rogers, H. J. (1999, July). Rapid DNA extraction from ferns for PCR-based analyses. *BioTechniques*, *27*(1), 66–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10407666 PubMedhttps://doi.org/10.2144/99271bm13

22. Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016, May). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. PubMedhttps://doi.org/10.1111/1462-2920.13023

23. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2018). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ*. https://doi.org/https://doi.org/10.7287/peerj.preprints.27295v2

24. Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2019). vegan: Community Ecology Package. Retrieved from https://cran.r-project.org/package=vegan

25. McMurdie, P. J., & Holmes, S. (2013, April 22). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), e61217. PubMedhttps://doi.org/10.1371/journal.pone.0061217

26. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., ... Bork, P., & the MetaHIT Consortium. (2011, May 12). Enterotypes of the human gut microbiome. *Nature, 473*(7346), 174–180. PubMed

27. Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., & Kuramae, E. E. (2016, May 31). The ecology of acidobacteria: Moving beyond genes and genomes. *Frontiers in Microbiology*, *7*(MAY), 744. PubMedhttps://doi.org/10.3389/fmicb.2016.00744

28. Chase, J., Fouquier, J., Zare, M., Sonderegger, D. L., Knight, R., Kelley, S. T., . . . Caporaso, J. G. (2016, April 19). Geography and location are the primary drivers of office microbiome composition. *mSystems*, *1*(2), 1–18. PubMedhttps://doi.org/10.1128/mSystems.00022-16

29. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., . . . Relman, D. A. (2005, June 10). Diversity of the human intestinal microbial flora. *Science*, *308*(5728), 1635–1638. PubMedhttps://doi.org/10.1126/science.1110591

30. Shanks, O. C., Newton, R. J., Kelty, C. A., Huse, S. M., Sogin, M. L., & McLellan, S. L. (2013, May). Comparison of the microbial community structures of untreated wastewaters

from different geographic locales. *Applied and Environmental Microbiology*, *79*(9), 2906–2913. [PubMedhttps://doi.org/10.1128/AEM.03448-12](https://doi.org/10.1128/AEM.03448-12)

31. García-Aljaro, C., Blanch, A. R., Campos, C., Jofre, J., & Lucena, F. (2019, March). Pathogens, faecal indicators and human-specific microbial source-tracking markers in sewage. *Journal of Applied Microbiology*, *126*(3), 701–717. [PubMedhttps://doi.org/10.1111/jam.14112](https://doi.org/10.1111/jam.14112)

32. Huse, S. M., Ye, Y., Zhou, Y., & Fodor, A. A. (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One*, *7*(6), e34242. [PubMedhttps://doi.org/10.1371/journal.pone.0034242](https://doi.org/10.1371/journal.pone.0034242)

33. Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V. J., Brown, C. M., & Sadowsky, M. J. (2018, April 3). Application of SourceTracker for accurate identification of fecal pollution in recreational freshwater: A double-blinded study. *Environmental Science & Technology*, *52*(7), 4207–4217. [PubMedhttps://doi.org/10.1021/acs.est.7b05401](https://doi.org/10.1021/acs.est.7b05401)

34. McMurdie, P. J., & Holmes, S. (2014, April 3). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. [PubMedhttps://doi.org/10.1371/journal.pcbi.1003531](https://doi.org/10.1371/journal.pcbi.1003531)