Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds

Bernadette J. Connors, PhD
Dominican College
Science Department
Bernadette.connors@dc.edu

Abstract
In New York State, 792 waterbodies are on the “Impaired/TMDL” list, which identifies bodies of water that might require remediation. The Sparkill Creek, placed on the list first in 2010, empties into the Hudson River. It faces issues with stormwater runoff, causing elevated levels of pathogens in the creek and decreased oxygen availability. According to published data, the levels of the indicator bacteria are 24 times higher than EPA standards. The Pocantico River, located in Westchester County, faces many of these same issues. Both were tested for microbial and coliphage loads and diversity as related to weather events. Coliform and E. coli levels were measured, with a significant increase found with a rainfall event for both sampling locations. The coliphage numbers and diversity were also significantly different with a rain event. Microbial community analyses were also completed. Of note was increased prevalence of Enterobacter and Escherichia three miles from the mouth of the creek. Future studies include a more thorough analysis of the microbial community data in both time and space, along with further testing of the fungal, bacterial, and algal populations.
Three Summary Points of Interest

- Microbial community profiling yields a more comprehensive view of challenges faced by aquatic ecosystems.
- Coliphage load, in addition to coliform levels, provide valuable insights when source tracking.
- Educational benefits from a project that centers on a local problem allows undergraduate researchers an opportunity to connect with a broader community.

Keywords

Microbial source tracking, coliphage, microbiome, Sparkill Creek, Pocantico River
Introduction
Microbial contamination of water sources is an important issue that faces urbanized areas across the world. The presence of pathogenic microbes poses a risk to individuals due, not only to their innate abilities to cause disease, but to the increased prevalence of multidrug resistant bacteria (Bhardwa et al., 2013). In terms of environmental health, non-native microbes can negatively impact the macroinvertebrates, plants, and animals that reside in the area.

In the current study, the microbial and phage loads were monitored in the Sparkill Creek and Pocantico River in summer 2017. The Sparkill Creek spans parts of Rockland County, NY and Bergen County, NJ. The creek begins in New York, flowing into northern New Jersey (Bergen County) through suburban and industrialized areas, eventually returning back to NY and emptying into the Hudson River at Piermont Marsh. The Pocantico River has its headwaters at Echo Lake, discharging into the Hudson River in Sleepy Hollow, NY. These two watersheds are located on opposite sides of the Hudson River, at similar latitudes and have been monitored for many years by citizen science groups, in conjunction with Riverkeeper (available at www.riverkeeper.org).

The standard indicator organism used in testing water is Enterococcus spp., although more injurious bacteria, fungi, protozoa, and viruses may be present (Water Quality Standards, 2015). Enterococcus is a Gram positive, facultative anaerobe that is found in both human and nonhuman-animal gastrointestinal tracts (Devriese et al., 1987, Manero et al., 1999). According to published data (riverkeeper.org), levels of the fecal bacterial indicator Enterococcus are greater than those found in the reference site (Hudson River) in both the Sparkill Creek and Pocantico River (Table 1). These data show that the Sparkill Creek experiences biological pollution such that Enterococcus levels are 24 times higher than EPA standards for issuing a beach advisory, while the levels are 11 times higher in the Pocantico River. The counts are routinely elevated following a rain event, with the levels sometimes exceeding the capability of the monitoring system itself.

### Table 1. Sites sampled in the current study. Data collected by Riverkeeper in partnership with residents of the Hudson Valley, O’Mullan GD, Juhl AR, and Lipscomb J. (available at www.riverkeeper.org)

<table>
<thead>
<tr>
<th>Sparkill Creek Site</th>
<th>Latitude</th>
<th>Fail rates</th>
<th>Distance from mouth</th>
<th>Tidal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tackamack</td>
<td>41.06273</td>
<td>88%</td>
<td>8mi</td>
<td>No</td>
</tr>
<tr>
<td>Spruce Street</td>
<td>41.06192</td>
<td>76%</td>
<td>7.2mi</td>
<td>No</td>
</tr>
<tr>
<td>Clausland</td>
<td>41.05872</td>
<td>94%</td>
<td>6.8mi</td>
<td>No</td>
</tr>
<tr>
<td>Blauvelt</td>
<td>41.05985</td>
<td>97%</td>
<td>6.8mi</td>
<td>No</td>
</tr>
<tr>
<td>Confluence</td>
<td>41.05712</td>
<td>NA</td>
<td>6.8mi</td>
<td>No</td>
</tr>
<tr>
<td>303/340</td>
<td>41.04367</td>
<td>94%</td>
<td>6.21mi</td>
<td>No</td>
</tr>
<tr>
<td>Rockleigh (trib)</td>
<td>41.00762</td>
<td>94%</td>
<td>3.2mi</td>
<td>No</td>
</tr>
<tr>
<td>Motoris</td>
<td>41.01792</td>
<td>97%</td>
<td>2.8mi</td>
<td>No</td>
</tr>
<tr>
<td>Skating Pond</td>
<td>41.03094</td>
<td>97%</td>
<td>1.8mi</td>
<td>No</td>
</tr>
<tr>
<td>Marsh</td>
<td>41.03832</td>
<td>NA</td>
<td>0</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pocantico River Site</th>
<th>Latitude</th>
<th>Fail rates</th>
<th>Distance from mouth</th>
<th>Tidal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above Pocantico</td>
<td>41.12711</td>
<td>79%</td>
<td>5.2mi</td>
<td>No</td>
</tr>
<tr>
<td>Caney Brook (trib)</td>
<td>41.13357</td>
<td>90%</td>
<td>4.2mi</td>
<td>No</td>
</tr>
<tr>
<td>Rockefeller Brook (trib)</td>
<td>41.09623</td>
<td>82%</td>
<td>2.4mi</td>
<td>No</td>
</tr>
<tr>
<td>Cemetery</td>
<td>41.09042</td>
<td>85%</td>
<td>0.6mi</td>
<td>No</td>
</tr>
<tr>
<td>Kingsland Pt/Hudson</td>
<td>41.09158</td>
<td>50%</td>
<td>0</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The use of Enterococcus to determine the point source of contamination has been challenged because enterococci are found in fecal material from warm-blooded animals, on plants and decaying plant matter, and in the soil itself (Devriese et al., 1987). In addition, standardized water testing protocols have been reported to have a bias toward detecting E. faecalis and E. faecium or to yield false positive results due to the presence of other microbes in the sample (Kinzelman et al., 2003, Ferguson et al., 2013). With the use of 16S rDNA sequencing the entire microbial population can be examined en masse (Garza and Dutilh, 2015). This can be done using uncultured bacteria, captured on a porous
Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds

filter, or using cultured bacteria, grown on undefined or selective media (Pascual et al., 2016). The benefit to using uncultured bacteria is that both fastidious organisms and anaerobes can be assessed for presence and prevalence, while the limitation is the amount of DNA extracted from manageable volumes of water. The benefit to using cultured bacteria is that the DNA quantity is sufficient for sequencing, while the downside is that not all bacteria are captured in this analysis. Prior research in the Sparkill Creek utilized uncultured bacteria to monitor the diversity in the water at selected sites for 2015 and 2016. Those data indicated the presence of bacterial genera associated with fecal matter, namely, Bacteroides (unpublished results).

Coliphage are bacteriophage that utilize Escherichia coli as a host. These viral particles belong to one of two types, based upon their infection strategy, namely, somatic or F⁺ coliphages. Studies have been conducted that use somatic coliphages to source track fecal contamination (Brezina and Baldini 2008, Gantzer et al., 1998, Skraber et al., 2004). There is a strong correlation between coliforms and coliphages (Brezina and Baldini 2008). By deduction, if the Enterococcus counts are high so, too, should be the coliforms and, therefore, the coliphages. A correlation has also been shown between coliphages and pathogenic viruses, such as Enterovirus and Norovirus (Skraber et al., 2004), so we can assume these should correlate with Enterococcus counts, as well.

The aim of the research study was to understand the bacterial and coliphage diversity of two polluted waterways in the Hudson Valley, the Sparkill Creek watershed and the Pocantico River. It is our expectation that these data will inform the community as to the health of these two bodies of water, and help us to better understand if current practices are satisfactory in monitoring these challenged ecosystems (Guang et al., 2004, Stoeckel and Harwood 2007, Water Quality Standards, 2015).

Results & Discussion

Determination of bacterial and coliphage loads in the Pocantico River and Sparkill Creek

According to the NYS DEC website, the Sparkill Creek encounters input from stormwater runoff, which likely contributes to its elevated Enterococcus counts. The Pocantico River has been observed to have similarly increased levels of indicator bacteria. Following several days without rainfall, referred to as a “dry collection”, and a period of rainfall (>0.25 inches in a four day period), referred to as a “wet collection”, viable E. coli, coliform, and somatic coliphage levels were determined. The levels of each increased, with the wet collection experiencing significantly higher levels of E. coli (p(Sparkill)=0.0687; p(Pocantico)=9.3E-5), coliform (p(Sparkill)=0.0365; p(Pocantico)=0.0042), and coliphage (p(Sparkill)=5.2E-10; p(Pocantico)=0.0026) (Figure 1).

Coliphage type has been weakly linked to pollutant sources. In the current study, somatic coliphage using the E. coli host strain CN13 (ATCC 700609) were used rather than F⁺ coliphage because of its utility. The range in plaque sizes was determined and compared for both dry and wet collections (Figure 2). As can be seen, following a rain event, the range in sizes is broadened, but the pattern in observable plaque diameters is not different than seen in the dry collection. The numbers of these coliphage increased in wet conditions, while the increases per site were in line with the fluctuations observed for E. coli and coliform (Figure 3).

Measurement of Nitrate and Phosphate Levels

Nitrate and phosphate levels are correlated with human activity. Both were measured in the current study, and correlated to E. coli, coliform, and coliphage levels. Using these testing methods, no correlation between E. coli, coliphage, nitrates, or phosphates were observed. In the future, a more sensitive and accurate means of testing for nitrates and phosphates would be advised.

Determination of the presence of Bacteroides

Bacteroides spp. is well recognized to be an indicator of pollution with fecal origin, and specific PCR primers have been developed that recognize human-specific Bacteroides (Bernhard and Field 2000). Total DNA isolated from 0.22 µm filters, which retained bacteria from the samples, was used as a template in PCR for both Bacteroides (general) and Bacteroides (human-specific) (Figure 4). As can be seen, Bacteroides were observed in both the Sparkill Creek sites and Pocantico River. Human-specific Bacteroides, however, was only observed in the Rockleigh site of the Sparkill Creek in dry weather.
Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds

This report was prepared for the New York State Water Resources Institute (WRI) and the Hudson River Estuary program of the New York State Department of Environmental Conservation, with support from the NYS Environmental Protection Fund.

Figure 1. Correlation between *E. coli* and coliform and rain events. Samples taken at selected sites were analyzed using Petrifilm® in duplicate. Levels per 100ml are recorded. (A, C: Sparkill and Pocantico, respectively, in wet weather; B, D: Sparkill and Pocantico, respectively, in dry weather)

Figure 2. Plaque size distribution of coliphage in the Sparkill Creek and Pocantico River. Diameters of all plaques present on triplicate plates were measured. (A: Sparkill Creek; B: Pocantico River)

Figure 3. Correlation between coliphage and *E. coli* levels. The fluctuations in the levels of *E. coli* and coliphage are apparent in the Sparkill Creek in both wet (A) and dry (B) weather. These same patterns were observed in the Pocantico River (data not shown).
Figure 4. Presence of Bacteroides spp was detected in selected samples in both the Sparkill Creek and Pocantico River. (Testing positive for Bacteroides were the following samples: Lane 1 Blauvelt (Sparkill, Dry), 2 Rockefeller Brook (Pocantico, Dry), 5 Caney Brook (Pocantico, Dry), 11 Rockleigh (Sparkill, Dry)*, 13 Motoris (Sparkill, Dry), Hudson River at Devries (data not shown), Caney Brook (Pocatico, Wet) (data not shown)). The Rockleigh site on the Sparkill Creek also tested positive for human Bacteroides via PCR in its dry collection (data not shown).

Coliphage Identification using Family-specific Primers
In order to determine the family type of the isolated coliphage, selected plaques were purified. Plaques of different morphologies were chosen for purification, with lytic, lysogenic, small, large, and those showing a “bullseye” morphology, being chosen (Figure 5). Primers were designed according to Lee (2009), and were selected to amplify Siphoviridae (HK, JK, and Lambda phage), Microviridae, Myoviridae (T4 and Mu phage), and Podoviridae (N4, 933, and K1F phage). At the current time, 50% of phages identified as T4 from the dry collection in both Sparkill Creek and Pocantico River belonged to the Myoviridae family of somatic coliphages, which is the predominant type in human sewage (Muniesa et al., 1999). Sites at which these phage were identified include sites in both the Sparkill Creek and Pocantico River.

Isolation and Analysis of DNA from Cultured Bacteria
Due to low yield of DNA from the filters, DNA was extracted from Gram negative, aerobic bacteria that had been cultured on eosin methylene blue (EMB) media. Gram negative bacteria represent species often associated with the guts of warm-blooded animals, such as those that belong to the family Enterobacteriaceae. The V3 region of the 16S rDNA gene was sequenced to determine family and genus identities of these bacteria. Variations in the levels of these bacteria between wet and dry days can help to pinpoint source and type of contamination. Figure 6 shows the variations in bacterial genera present along the Sparkill Creek in both wet and dry conditions. Similar data are available for the Pocantico River, as well as bacteria present in the soil at selected sites along the Sparkill Creek.

Future studies
Often overlooked is the population of algae present in these ecosystems, specifically the microalgae. Past studies have used these primary producers as pollution indicators (Barinova et al., 2010), and there are several pathogenic forms associated with human sewage. This upcoming year our research group will use both computational and taxonomical approaches to examine the microalgae present in the creek. Water will be collected in late summer (August) for analysis, as well as late fall (2018) and sent for sequencing of the eukaryotic population present in the samples. Microscopic analysis of unicellular algae will be carried out at that time, with culturing to follow. A botanical survey will also be done, with plants identified and preserved through plant pressing for curation. We expect these analyses to be completed by June 2019, with significant contributions made by undergraduate student researchers.
Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds

Figure 6. Microbial profiling of cultured bacteria from EMB media. Bacteria identified at the genus level were identified through 16S rDNA sequencing and those bacteria introduced or depleted from selected sites along the Sparkill Creek are shown. Genera and family level identifications are available for both the Sparkill Creek and Pocantico River, and will be published at MG-RAST (hosted by Argonne National Laboratories) for public access.

Methods

Water Collection
Sites in both the Pocantico River and Sparkill Creek were chosen based on data obtained from Riverkeeper. Those sites that showed the greatest number of fail rates in the past, based on the geometric mean of collections from 2011-2016, were selected for this study. Water was collected in early June on what is considered a “dry” day, that is, less than 0.25 inches of rainfall in a four day period. A second collection, if possible, was done on a “wet” day, in July. Rainfall amounts were obtained from the USGS unheated rain gauge in Nanuet, NY (41.05’18”). Water was collected in sterile 1L Nalgene bottles, with a total of 2L collected per site. Samples were transported directly back to the laboratory where they were then processed. Site locations are given in Table 1, along with percent fail rates for each site.

Determination of Coliform and E. coli levels
Prior to filtration, Petrifilm® (Coliform/E. coli) was inoculated with 1ml of collected water (duplicates). The films were incubated at 37°C for 24 hours, at which time they were observed. The presence of coliform was indicated by red colonies that exhibited gas formation, while blue colonies and gas formation was indicative of E. coli.

Isolation and Analysis of DNA from Cultured Bacteria
In addition to culturing on Petrifilm®, a small inoculum was plated on EMB media, which selects for the growth of Gram negative bacteria. Plates were inoculated with 0.2 ml of unfiltered water, and incubated at 37°C for 24 hours. Culturing at this temperature allowed for the growth of potential human pathogens. The bacterial growth on these plates was removed using a sterile swab, and DNA was extracted from this growth using the Qiagen PureGene Yeast/Bacteria Kit (Qiagen). Total genomic DNA was observed on an 0.8% agarose gel, and
~5 μg was sent away for 16S rDNA sequencing (Seqmatic, LaJolla, CA). The V3 region of the 16S rDNA was analyzed. Results from Seqmatic were investigated for genera common to sites, as well as those that were introduced or depleted from nearby locations.

**Coliphage Isolation**

Water was filtered using Nalge Nunc PES Rapid Flow filtration device, with a 0.22 μm pore size. Water was examined for coliphage, with 3-4 plates prepared per site. The host, *Escherichia coli* CN13 (ATCC 700609), was grown at 37°C to an OD660 of 1.0 in Luria broth and nalidixic acid (100mg/L) in the presence of 1 mM CaCl₂, to enhance phage attachment to the host. In order to isolate coliphage, 1ml of *E. coli* CN13 was incubated with filtrate (0.5-1ml) for 10 minutes at room temperature without disturbance. To this suspension, 8ml of molten top agar (Luria broth with 0.4% agar, 1mM CaCl₂ kept at 65°C) was added, and the mixture was plated onto Luria agar containing 1mM CaCl₂. Plates were incubated at 37°C overnight. Plaques were counted and size of plaques was measured, and a note was made about its morphology (lytic or lysogenic, as well as the presence of a “bullseye” appearance). Selected plaques were “picked” by touching a sterile pipette tip to the plaque, which was then moved to 100 µL of sterile phage isolation buffer. A 1:10 serial dilution in phage isolation buffer was carried out, and 10 µl of selected dilutions (10⁻², 10⁻³, and 10⁻⁴) were plated with 250 µl host *E. coli* CN13 as above. Webbed plates, indicated by nearly confluent growth of the coliphage, were flooded with 8 ml of phage isolation buffer, and refrigerated overnight. The lysate, containing phage that diffused from the top agar into the phage isolation buffer, was filtered through a 0.22 μM filter into sterile culture tubes to remove host cells. Lysates remained refrigerated until further use.

**Coliphage Identification using Family-specific Primers**

DNA from the lysate of purified coliphage was extracted using the Wizard DNA kit (Promega). In brief, 500 μl of lysate was subjected to guanidium thiocyanate extraction, followed by visualization on a 0.8% agarose gel. PCR, using primers developed by Lee (2009), was completed with each phage DNA sample, and amplicons were visualized on agarose gels (1-2% agarose). Negative controls were included for each round of amplification, which contained no DNA template. In total, DNA from 44 coliphages were analyzed.

**Determination of the presence of Bacteroides**

The filter from the Nalge Nunc PES Rapid Flow filtration device unit was used to extract DNA from uncultured bacteria. The Purelink Microbiome DNA Extraction Kit (Thermo Fisher) was utilized to extract the genomic material. DNA was then subjected to PCR using primers specific to general *Bacteroides* and human-specific *Bacteroides*, as described by Seurinck et al (2005). Amplicons were visualized on a 1% agarose gel.

**Water Chemistry**

Nitrate and phosphate levels were determined for each water sample using the Lamotte Water Chemistry kit, according to manufacturer’s protocols.

**Outreach Comments**

Site selection was done with the advice of Larry Vail, president of the Sparkill Creek Watershed Alliance.

**Student Training**

Undergraduate and high school students were critical to completing this work. In total, four undergraduate students and ten high school students worked directly on the project over the course of the funding period. The undergraduate students were given the freedom to modify protocols and advise on how the project should move forward, as well as to develop ideas for their own individual projects moving forward. It is expected that the genomic data will supply future students with data for their own research projects.

**Publications/Presentations**

Bacterial and Viral Source Tracking in the Sparkill Creek Watershed (Acevedo, K, Annual Biomedical Conference for Minority Students, Phoenix, AZ (Nov 2017); Acevedo, K and Joseph, C, Eastern Colleges Science Conference, Ithaca College, Ithaca, NY (2018))

Collective Impact Practices in STEM Research and Teaching (Clemson University, Invited lecture)

**Acknowledgments**

This work would not have been possible without the work of Christina Joseph (Polgar Fellow 2017) and Kimberly Acevedo (American Society of Microbiology Undergraduate Research Fellow 2017). Guidance of the students in the use of statistical software was provided by Dr Kasie Farlow (Dominican College).
Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds

References


U.S. Environmental Protection Agency. 2000. Proposed ground water rule EPA 815-F-00-003. Office of Ground

This report was prepared for the New York State Water Resources Institute (WRI) and the Hudson River Estuary program of the New York State Department of Environmental Conservation, with support from the NYS Environmental Protection Fund
This report was prepared for the New York State Water Resources Institute (WRI) and the Hudson River Estuary program of the New York State Department of Environmental Conservation, with support from the NYS Environmental Protection Fund.