BACTERIAL AND VIRAL SOURCE TRACKING IN THE POCANTICO AND SPARKILL CREEK WATERSHEDS

A Final Report of the Tibor T. Polgar Fellowship Program

Christina Joseph

Polgar Fellow

Science Department
Dominican College of Blauvelt
Orangeburg, NY 10962

Project Advisor:

Bernadette Connors, PhD
Science Department
Dominican College of Blauvelt
Orangeburg, NY 10962

Joseph, C. and B. Connors. 2020. Bacterial and Viral Source Tracking in the Pocantico and Sparkill Creek Watersheds. Section VII:1-28 pp. *In* D.J. Yozzo, S.H. Fernald, and H. Andreyko (eds.), Final Reports of the Tibor T. Polgar Fellowship Program, 2017. Hudson River Foundation.

ABSTRACT

Microbial and viral contamination of critical water sources is an important issue facing many communities. In New York State alone, 792 streams, rivers, and lakes are on the "Impaired/TMDL" list, which identifies bodies of water that might require remediation. In the lower Hudson River drainage basin, Sparkill Creek, placed on the list in 2010, faces issues with urban/storm water runoff, causing elevated levels of pathogens in the creek and decreased oxygen availability. According to Riverkeeper data, the levels of the indicator bacteria, *Enterococcus*, are 24 times higher than EPA standards allowable before a beach advisory is issued. The Pocantico River, although not on the DEC Impaired Waters List, also has higher than acceptable levels of these bacteria, exceeding the EPA standard. Both were examined for microbial and coliphage diversity and loads as related to weather events. Coliform and *E. coli* levels were measured, with significant increases found with a rainfall event for both sampling locations at most sites. The coliphage numbers were also significantly higher with a rain event. Family typing of coliphage by PCR showed a prevalence of *Myoviridae,* a family most often associated with wastewater. Genomic DNA from Gram-negative cultured bacteria was isolated and subjected to 16S rDNA sequencing. It was noted that there were higher counts of *Stenotrophomonas* in dry weather, seen less often after a rainfall event, within Sparkill Creek. Extensions of these studies have included algal and fungal source tracking, refinement of community profiling at specific sites along Sparkill Creek, and metabolomics within Piermont Marsh. The work presented in this report has launched numerous additional studies in these watersheds.

TABLE OF CONTENTS

Abstract VII-2

Table of Contents VII-3

Lists of Figures and Tables VII-4

Introduction VII-5

Methods VII-10

Results VII-14

Discussion VII-21

Acknowledgements VII-24

References VII-25

LIST OF FIGURES AND TABLES

Figure 1- Correlation between coliphage and *E. coli* levels VII-16

Figure 2- Predominant bacterial genera present at each site along Sparkill Creek and the Pocantico River VII-19

Table 1- Sites sampled in the current study VII-7

Table 2- Primer sequences utilized in the current study VII-14

Table 3- Characterization of predominant bacterial phyla at selected sites in dry and wet collections... VII-18

INTRODUCTION

Microbial contamination of water sources is an important issue that faces all urbanized areas. Of particular concern is the presence of human viral pathogens and bacteria such as *Escherichia coli*, *Salmonella* spp., *Giardia lamblia*, and *Cryptosporidium parvum*. These microbes are associated with fecal pollution and may cause disease through the fecal-oral route of transmission in humans. Their presence poses a risk to individuals using these water sources for both consumption and recreation. In terms of environmental health, significant disruptions to microbial communities may adversely affect aquatic and terrestrial macroinvertebrates, plants, and animals. With increasing impacts to waterways and shorelines as a result of climate change and pollution through human activity, understanding the bearing of changes in microbial communities is of the utmost importance because of the fundamental significance of microbiota on the normal functioning of any ecosystem.

In the current study, microbial diversity was monitored in Sparkill Creek and the Pocantico River. Sparkill Creek spans parts of Rockland County, NY, and Bergen County, NJ. Its headwaters are located at the northern end of the watershed in a sparsely populated suburban neighborhood. It then flows into northern New Jersey (Bergen County) through residential and lightly-industrialized areas, eventually turning back northwards and emptying into the Hudson River at Piermont Marsh (Table 1). The Pocantico River begins at Echo Lake and flows through very populated suburban areas, including many parks and tourist attractions in Westchester County, emptying finally into the Hudson River in Sleepy Hollow, NY. These two watersheds are located on opposite sides of the Hudson River, at similar latitudes and with similar climates, and both have

been monitored for many years by citizen science groups, in conjunction with Riverkeeper. Sparkill Creek was placed on the "Impaired/TMDL list" in 2010 by the New York State Department of Environmental Conservation (NYS DEC), which identifies bodies of water that might require remediation. It remains on this list due to increased levels of pathogens and reduced oxygen availability due to urban and storm water runoff (NYSDEC 2019). Although the Pocantico River and its tributaries are not on the list, they, too, likely face similar issues as evidenced by the elevated levels of fecal indicator bacteria.

The standard indicator organism used to test for the presence of fecal contamination in surface waters is *Enterococcus*. According to published data (Riverkeeper.org), levels of these bacteria are greater than those found in the reference site (Hudson River) in both Sparkill Creek and the Pocantico River, with counts routinely elevated following a rain event. The use of these Gram-positive bacteria has been challenged by some as a reliable source for tracking 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 that detect enterococci 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).

Site	Latitude	Fail rates	Distance from mouth	Tidal
Tackamack	41.06273	88%	8mi	No
Spruce Street	41.06192	76%	7.2mi	No
Clausland	41.05872	94%	6.8mi	No
Blauvelt	41.05985	97%	6.8mi	No
Confluence	41.05712	NA	6.8mi	No
303/340	41.04367	94%	6.21mi	No
Rockleigh (trib)	41.00762	94%	3.2mi	No
Motoris	41.01792	97%	2.8mi	No
Skating Pond	41.03094	97%	1.8mi	No
Above Pocantico	41.12711	79%	5.2mi	No
Caney Brook (trib)	41.13357	90%	4.2mi	No
Rockefeller Brook (trib)	41.09623	82%	2.4mi	No
Cemetery	41.09042	85%	0.6mi	No
Kingsland Pt/Hudson	41.09158	50%	0	Yes

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)**

With the use of alternative detection and tracking techniques, the scientific community can better understand ecosystems that are differentially stressed, rather than those facing biological pollution alone. Each method has its own value and limitations, which argues for a mixed-methods approach to source tracking (McLellan 2004; Plummer and Long 2009; Stewart et al. 2003). In the current study, community profiling and enumeration of viable microbes was done. Although not exhaustive, approaching the problem using two methods extends the research carried out in these waterways previously.

Bacterial community profiling relies on 16S rRNA sequencing of the entire microbial population (Garza and Dutilh 2015) with subsequent identification and enumeration of the bacteria. The 16S rRNA gene is currently the most commonly used molecular marker for profiling studies (Cardenas and Tiedje 2008; Clarridge 2004). The gene, which is 1,550 base pairs in length, is present in all bacteria and comprised of both conserved and hypervariable regions (V1-V9) (Chakravorty et al. 2007). The taxonomic identification of bacteria relies on differences between the hypervariable regions of this gene, with sequences from two or more of these regions providing sufficient data to distinguish distinctive species. One significant disadvantage to its use is its lack of correlation to phenotype, disallowing its contribution to metabolomic studies (Schloss et al. 2011). There are a number of rRNA databases available to which sequences can be compared, including the Ribosomal Database Project (Cole et al. 2009) and Greengenes 2.0 (DeSantis et al. 2006). Both can be used with partial rRNA sequences, that is, using only a single hypervariable region as the query source.

In order to complete profiling analyses, uncultured bacteria, captured on porous filters, or cultured bacteria, grown on undefined or selective media (Pascual et al. 2016) can be used. 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 value 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. In either case, analyses of identity and prevalence can reveal which taxonomic units are introduced, depleted, or unchanged at a particular location, relative to one another. Their best use in microbial

source tracking is seen, however, when multiple samples taken at regular intervals are evaluated, and correlation of these data to water chemistry, and hydrological and meteorological data, are done.

Coliphage are double-stranded DNA viruses that utilize *E. coli* as a host and are classified by the means of infection, namely as somatic or F^+ coliphages. Studies have shown the successful use of somatic coliphages to source track fecal contamination (Brezina and Baldini 2008; Gantzer et al. 1998; Skraber et al. 2004), and they can be recommended as source indicators in surface water and ground water (US EPA 2006). They originate from the feces of warm-blooded animals and are present in wastewater at higher densities than in the fecal material itself (Dhillon et al. 1976; Gantzer et al. 2002; Grabow 2001; Long et al. 2005). Since coliphage cannot replicate unless inside of their host nor persist in surface waters, undergoing limited replication only under specific circumstances, they can be used to track point sources of pollution. They are also known to occur differentially based on temperature, rainfall and salinity (Reyes and Jiang 2010). Of importance is the strong correlation between fecal coliforms and coliphages (Brezina and Baldini 2008), and between coliphages and pathogenic viruses, such as *Enterovirus* and *Norovirus* (Skraber et al. 2004). By deduction, if the *Enterococcus* counts are high so, too, should be the coliforms and, therefore, the coliphages.

Prior research in Sparkill Creek utilized uncultured bacteria to monitor the diversity of microbes in the water at selected sites from 2014-2016. Coliphage were isolated as a proof of concept during one of these collections. Those data indicated the presence of elevated levels of bacteria associated with fecal pollution, including *Bacteroides* and *E. coli*. Other research showed a high percentage of bacteria that have

antibiotic resistance to multiple drugs (five or more). Multiplex PCR analysis of individual *Enterococcus* isolates indicated a high proportion of *E. faecalis* and *E. faecium*, two commonly found species in the gut of warm-blooded animals. The research described herein aims to understand the microbial communities in the watershed to provide baseline data by which impacts related to pollution can be measured. In order to achieve this goal, the objectives of this study were to: (1) collect, identify, and analyze bacteria and somatic coliphage at selected sites in Sparkill Creek and the Pocantico River; (2) correlate identity and prevalence of bacteria and coliphage to weather events and water chemistry; and (3) track differential presence of bacteria at selected sites along each body of water. The hypothesis tested was that the microbial and coliphage diversities were likely to differ between the two locations but increases in both prevalence and diversity would increase after a rainfall event. Additionally, it was predicted that tracking of bacterial communities would allow for identification of source of contamination.

METHODS

Water Collection

Sites in both the Pocantico River and Sparkill Creek were chosen based on tracking 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 2017 on what was considered a "dry" day, that is, less than 0.25 inches of rainfall in a four-day period prior to collecting. A second collection was done on a "wet" day, in July 2017. 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 2 L 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 as reported by Riverkeeper.

Determination of Coliform and *E. coli* levels

Duplicate Petrifilm® (Coliform/*E. coli*) were inoculated with 1 ml of collected water. 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

A small inoculum was plated on eosin methylene blue (EMB) media, which selects for the growth of Gram-negative bacteria. Duplicate plates were inoculated with 0.2 ml of unfiltered water and incubated at 37°C for 24 hours in an aerobic chamber. Colonies were removed using a sterile swab and DNA was extracted using the Qiagen PureGene Yeast/Bacteria Kit (Qiagen) according to the manufacturer's recommended protocol for Gram-negative bacteria. Bacteria captured on filters were also subjected to the extraction protocol. Total genomic DNA was observed on an 0.8% agarose gel and ~5 µg was subjected to 16S rDNA sequencing (Seqmatic, LaJolla, CA). The V3 region of the 16S rDNA was analyzed.

Coliphage Isolation and Purification

Water was filtered using Nalge Nunc PES Rapid Flow filtration device, with a 0.22 µm pore size. Filters were saved for genomic DNA extraction of captured bacteria. The filtrate was examined for coliphage, with 3-4 plates prepared per site, using the double-layer agar method. The host, *Escherichia coli* CN13 (ATCC 700609), was grown at 37°C to an OD_{660} of 1.0 in Luria broth and nalidixic acid (100 mg/L) in the presence of 1 mM $CaCl_2$, to enhance phage attachment to the host. In order to isolate coliphage, 1 ml of *E. coli* CN13 was incubated with filtrate (0.5-1 ml) for 10 minutes at room temperature without disturbance. To this suspension, 3 ml of molten top agar (Luria broth with 0.4% agar, 1 mM $CaCl_2$ kept at 65°C) was added, and the mixture was plated onto Luria agar containing 1 mM $CaCl_2$. 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}) was 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 thiocyante extraction, followed by visualization on a 0.8% agarose gel. PCR, using primers developed by Lee (2009) (Table 2), 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.

Determination of the presence of *Bacteroides*

Water collected was filtered using Nalge Nunc PES Rapid Flow filtration device. DNA was extracted from the filters using the Purelink Microbiome DNA Extraction Kit (Thermo Fisher). DNA was then subjected to PCR using primers specific to general *Bacteroides* and human-specific *Bacteroides*. Research has shown that the overall specificity of the HF183 markers to differentiate between humans and animals was 99% (Ahmed et al. 2009).

Primer name	Primer sequence	Reference
GenBacF	AAC GCT AGC TAC AGC CTT	Bernhard and Field 2000
GenBacR	CAA TCG GAG TTC TTC GTG	Bernhard and Field 2000
HF183F	ATC ATG AGT TCA CAT GTC CG	Bernhard and Field 2000
Myoviridae T4 set	GATATTTGTGGYGTTCAGCC (FW) GTCAAATACACCAGCTTTAGAACC (RV)	Lee 2009
Myoviridae Mu set	GAAAACGACTCAATCCTTGCC (FW) TCATCAGGTCTTTTGTTGTGG (RV)	Lee 2009
Microviridae	GCTGCCGTCATTGCTTATTATGTTC (FW) GYTAYCGBMMCATYAAYTAHTCACG(RV)	Lee 2009
Siphoviridae HK set	CACAGCGAGAAATTGATCGC (FW) CTAATCGGACTGATGTCTG (RV)	Lee 2009
Siphoviridae JK set	GYGAYCAGATGGTTCC (FW) CAATRTCYTCYTARTTG (RV)	Lee 2009
Siphoviridae Lambda set	TGGGCGTACTTTATGGGGCG (FW) CGGACCTGCTGGGCAAAAAT (RV)	Lee 2009
Podoviridae 933 set	GCAATACATCAAACGCCG (FW) GCGAATGCCAGCGGCG (RV)	Lee 2009
Podoviridae K1F set	TGGAAGCCCGTGAGAC (FW) GCAGCGTCAATCGCTCGG (RV)	Lee 2009
Podoviridae N4 set	GCACATGCAGAATAAGGTTG (FW) CCATTAGTAACACCATCTGC 20 (RV)	Lee 2009

Table 2. Primer sequences utilized in the current study.

RESULTS

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

According to the NYS DEC, Sparkill Creek receives input from storm water runoff, which likely contributes to its elevated *Enterococcus* counts. The Pocantico River has similarly increased levels of indicator bacteria, although it is not currently on the Impaired/TMDL Waters list. 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," levels of viable *E. coli,* coliform, and somatic coliphage were

determined. As can be seen in Figure 1, 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) (data not shown), and coliphage (p(Sparkill)=$5.2e^{-10}$; p(Pocantico)=0.0026). Elevated numbers of *E. coli* and coliphage were observed at Rockleigh and Motoris sites in dry collections, and highest in wet collections at Motoris. Levels of coliform were also elevated at the Motoris site, regardless of rainfall event. Of importance, is the detection of *Bacteroides* at this same site (below).

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*. Total genomic DNA isolated from 0.22 µm filters, which retained bacteria from the collected water samples, was used as a template in PCR for both *Bacteroides* (general) and *Bacteroides* (human-specific). *Bacteroides* was observed in both Sparkill Creek sites and the Pocantico River at selected sites during the dry collection (Blauvelt, Rockleigh, Motoris (Sparkill); Rockefeller Brook, Caney Brook (Pocantico)). Human-specific *Bacteroides*, however, was only observed at the Rockleigh site of Sparkill Creek in dry weather (data not shown).

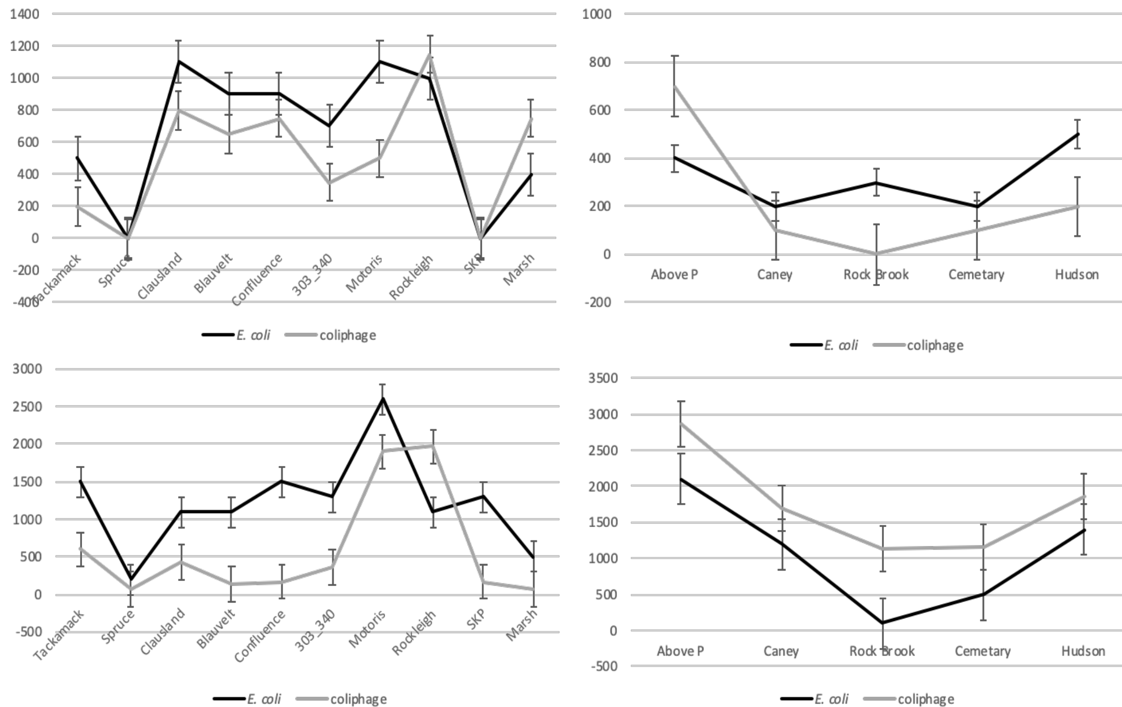

Figure 1. **Correlation between coliphage and *E. coli* levels. The fluctuations in the levels of *E. coli* and coliphage are shown in Sparkill Creek (left) and the Pocantico River (right) in both wet (bottom) and dry (top) weather.**

Coliphage identification using family-specific primers

Somatic coliphages can be detected and their prevalence quantified by simple, inexpensive, and reliable means (Gerba 1987; Havelaar 1987). EPA methods 1601 and 1602 (or approved equivalent methods) for coliphage detection rely on the host bacterium *Escherichia coli* strain CN13, and both allow for the visualization of plaques produced as a result of host infection and subsequent phage release. In order to determine the family type of the isolated coliphage, selected plaques were purified and amplified. Plaques of different morphologies were chosen at random, with lytic, lysogenic, small, and large chosen. Primers were designed according to Lee (2009) and used to amplify and identify coliphages (Table 2). Nearly 50% of phages were identified as *Myoviridiae* T4 from the

dry collection in both Sparkill Creek and the Pocantico River, and 6% were identified as *Siphoviridae* JK phage at one site only. PCR amplifications of the remaining phage were unsuccessful.

Isolation and analysis of DNA from cultured bacteria

Due to low yield of DNA from the filters, insufficient for Illumina sequencing, genomic DNA was extracted from Gram-negative, aerobic bacteria that had been cultured on EMB media. The V3 region of the 16S rDNA gene was sequenced to determine family and genus identities. Sequences were compared against the Greengenes 2.0 database, and identification to the genus level was done. Relative abundances were determined and normalized against the total number of hits retrieved from the sequencing runs. Figure 2 shows bacteria that are introduced, depleted or unchanged between dry and wet collections. Focusing on the sites from which *Bacteroides* were amplified, an analysis at the phylum level shows differences in composition between wet and dry collections (Table 3). Of particular interest is the decrease in the presence of *Bacteroidetes* in wet weather at the Blauvelt and Rockleigh sites, as well as the significant number of *Bacteroidetes* at the Rockleigh site in dry conditions.
Particular to Sparkill Creek, members of the genus *Stenotrophomonas* were present in dry collections at almost every site and found not at all or in significantly diminished numbers during wet collections. This was not true for the Pocantico River samples, with *Stenotrophomonas* found in both wet and dry collections or introduced during wet weather. At the Rockleigh and Motoris sites, *Stenotrophomonas* constituted 33% and 20% of the bacteria identified on dry days. At Piermont Marsh, although

Stenotrophomonas was found in low numbers on wet days, it constituted 49% of the bacteria identified on dry days.

Site	Classification	% Dry reads	% Wet reads
Blauvelt	*Proteobacteria*	91.74 %	99.58%
	Bacteroidetes	8.04%	0.19%
	Firmicutes	0.18%	0.18%
Motoris	*Proteobacteria*	99.11%	99.37%
	Bacteroidetes	0.82%	0.51%
	Firmicutes	0.03%	0.09%
Rockleigh	*Proteobacteria*	83.29%	99.81%
	Bacteroidetes	16.51%	0.07%
	Firmicutes	0.31%	0.09%
Rockefeller Brook	*Proteobacteria*	99.69%	99.33%
	Firmicutes	0.17%	0.10%
	Bacteroidetes	0.10%	0.49%
Caney Brook	*Proteobacteria*	97.97%	98.49%
	Bacteroidetes	1.87%	1.34%
	Firmicutes	0.08%	0.12%

Table 3. Characterization of predominant bacterial phyla at selected sites in dry and wet collections.

Tackamack Dry (6 genera)	**Tackamack Wet (7 genera)**	**Clausland Dry (7 genera)**	**Clausland Wet (7 genera)**
Acinetobacter, Enterobacter, Pseudomonas, Pseudoxanthomonas, Stenotrophomonas,Yersinia	**Unchanged:** *Acinetobacter, Enterobacter, Pseudomonas* **Introduced:** *Plesoimonas, Serratia, Shewanella, Vogesella* **Depleted:** *Pseudoxanthomonas, Stenotrophomonas,Yersinia*	*Comamonas, Enterobacter, Escherichia, Oxakobacter, Pseudomonas, Pseudoxanthomonas, Stenotrophomonas*	**Unchanged:** *Enterobacter, Pseudomonas* **Introduced:** *Acinetobacter, Chromobacterium,Citrobacter, Klebsiella, Vogesella.* **Depleted:** *Comamonas,Escherichia, Oxalobacter, Pseudoxanthomonas, Stenotrophomonas*
Spruce Dry (7 genera)	**Spruce Wet (6 genera)**	**Confluence Dry (6 genera)**	**Confluence Wet (8 genera)**
Acinetobacter, Enterobacter, Erwinia, Klebsiella, Pseudomonas, Serratia, Yersina	**Unchanged:** *Pseudomonas, Serratia, Yersina* **Introduced:** *Chromobacterium, Escherichia, Plesoimonas* **Depleted:** *Acinetobacter, Enterobacter, Erwinia, Klebsiella*	*Acinetobacter, Achromobacter, Delftia, Enterobacter, Pseudomonas,Stenotrophomonas*	**Unchanged:** *Acinetobacter, Delftia, Pseudomonas, Stenotrophomonas* **Introduced:** *Comamonas Escherichia, Serratia, Vogesella* Depleted: *Achromobacter, Enterobacter*
Blauvelt Dry (7 genera)	**Blauvelt Wet (6 genera)**	**Motoris Dry (7 genera)**	**Motoris Wet (7 genera)**
Acinetobacter,Comamonas, Diaphorobacter,Pseudomonas, Pseudoxanthomonas, Sphingobacterium, Stenotrophomonas	**Unchanged:** *Acinetobacter, Comamonas,Pseudomonas* **Introduced:** *Enterobacter, Erwinia, Serratia* **Depleted:** *Diaphorobacter, Pseudoxanthomonas, Sphingobacterium,Stenotrophomonas*	*Acinetobacter,Citrobacter, Comamonas,Enterobacter,Erwinia, Pseudomonas, Stenotrophomonas*	**Unchanged:** *Acinetobacter, Comamonas, Enterobacter,Erwinia, Pseudomonas* **Introduced:** *Shewanella, Vogesella* **Depleted:** *Citrobacter, Stenotrophomonas*
Rockleigh Dry (7 genera)	**Rockleigh Wet (8 genera)**	**Marsh Dry (7 genera)**	**Marsh Wet (6 genera)**
Acinetobacter,Comamonas Delftia,Myroides,Pseudomonas, Pseudoxanthomonas, Stenotrophomonas	**Unchanged:** *Acinetobacter, Pseudomonas* **Introduced:** *Enterobacter,Erwinia, Escherichia,Providencia,Shewanella, Vogesella* **Depleted:** *Comamonas, Delftia, Myroides, Pseudoxanthomonas, Stenotrophomonas*	*Citrobacter, Delftia, Enterobacter, Plesoimonas, Pseudomonas, Shewanella, Stenotrophomonas*	**Unchanged:** *Enterobacter, Pseudomonas, Shewanella, Stenotrophomonas* **Introduced:** *Erwinia, Acinetobacter* **Depleted:** *Citrobacter, Delftia, Plesoimonas*
303/340 Dry (8 genera)	**303/340 Wet (5 genera)**	**Skating Pond Dry (6 genera)**	**Skating Pond Wet (6 genera)**
Acinetobacter, Chromobacterium,, Myroides, Pseudomonas, Pseudoxanthomonas,Plesoimonas, Stenotrophomonas, Yersinia	**Unchanged:** *Pseudomonas* **Introduced:** *Enterobacter, Escherichia, Ewrinia, Shewanella* **Depleted:** *Acinetobacter, Chromobacterium, Myroides, Pseudoxanthomonas,Plesoimonas, Stenotrophomonas, Yersinia*	*Acidrorax, Enterobacter, Pseudomonas,Pseudoxanthomonas, Plesoimonas, Wautersiella, Yersiania*	**Unchanged:** *Enterobacter, Pseudomonas,Pseudoxanthomonas, Plesoimonas* **Introduced:** *Acinobacter,Escherichia, Vogesella* **Depleted:** *Acidrorax,Wautersiella, Yersiania*

(A)

<table>
<tr><th>Hudson Dry (7 genera)</th><th>Hudson Wet (6 genera)</th><th>Caney Brook Dry (6 genera)</th><th>Caney Brook Wet (7 genera)</th></tr>
<tr><td rowspan="3">Acinetobacter,Comamonas, Myroides, Plesoimonas, Pseudomonas, Stenotrophomonas,Yersinia</td><td>Unchanged:Acinetobacter, Pseudomonas</td><td rowspan="3">Caulobacter,Enterobacter, Janthinobacterium,Plesoimonas, Pseudomonas, Stenotrophomonas</td><td>Unchanged:Pseudomonas, Stenotrophomonas</td></tr>
<tr><td>Introduced:Citrobracter,Enterobacter, Shewanella,Vogesella</td><td>Introduced:Acinetobacter Chromobacterium, Dickeya,Providencia, Vogesella</td></tr>
<tr><td>Depleted:Comamonas,Myroides, Plesoimonas,Stenotrophomonas, Yersinia</td><td>Depleted:Caulobacter,Enterobacter, Janthinobacterium,Plesoimonas</td></tr>
<tr><th>Rockefeller Brook Dry (5 genera)</th><th>Rockefeller Brook Wet (8 genera)</th><th>Cemetery Dry (5 genera)</th><th>Cemetery Wet (7 genera)</th></tr>
<tr><td rowspan="3">Acinetobacter,Comamonas,Delftia, Plesoimonas,Yersinia</td><td>Unchanged:Comamonas,Delftia</td><td rowspan="3">Acinetobacter,Plesoimonas, Pseudomonas,Serratia, Wohlfahrtiimonas</td><td>Unchanged:Acinetobacter, Plesoimonas,Pseudomonas</td></tr>
<tr><td>Introduced:Citrobracter,Enterobacter Pseudomonas,Stenotrophomonas Vogesella, Serratia</td><td>Introduced:Citrobracter,Shewanella, Stenotrophomonas,Vogesella</td></tr>
<tr><td>Depleted:Acinetobacter, Plesoimonas,Yersinia</td><td>Depleted:Serratia, Wohlfahrtimonas</td></tr>
<tr><th>Above Pocantico Dry (6 genera)</th><th>Above Pocantico Wet (7 genera)</th><th></th><th></th></tr>
<tr><td rowspan="3">Achromobacter,Enterobacter, Pseudomonas,Serratia, Stenotrophomonas, Yersinia</td><td>Unchanged:Enterobacter, Pseudomonas,Serratia, Stenotrophomonas</td><td></td><td></td></tr>
<tr><td>Introduced: Acinetobacter,Comamonas, Vogesella,</td><td></td><td></td></tr>
<tr><td>Depleted:Achromobacter, Yersinia</td><td></td><td></td></tr>
</table>

(B)

Figure 2. Predominant bacterial genera present at each site along Sparkill Creek (A) and Pocantico River (B). Bacteria that are introduced, depleted, or unchanged are indicated.

DISCUSSION

Bacterial and coliphage loads are elevated after rainfall events

As expected, the microbial and coliphage loads increased following a rainfall event. The abundances of *E. coli*, coliform, and coliphage are seen to increase and decrease alongside the other indicators measured. A site by site comparison revealed that levels of coliform were significantly higher at Rockleigh and Motoris ($p=0.05$) in Sparkill Creek. *Bacteroides,* as well, was detected in both Sparkill Creek and the Pocantico River. Prior community profiling in the Creek pointed to their presence, although it was unclear if they were human-specific *Bacteroides.* Using PCR to amplify general *Bacteroides* has proven their presence at Rockleigh and Motoris, with human-specific forms found at the Rockleigh site. Along with the elevated levels of coliform, coliphage, and *E. coli*, these sites deserve further investigation as a possible source of biological pollution in Sparkill Creek.

Coliphage Identification using Family-specific Primers

Coliphage type has been weakly linked to pollutant sources. In the current study, somatic coliphage using the *E. coli* host strain CN13 were used rather than F^+ coliphage because of its utility. *Siphoviridae* (HK, JK, and Lambda phage), *Microviridae*, *Myoviridae* (T4 and Mu phage), and *Podoviridae* (N4, 933, and K1F phage) (Table 2) were detected. The *Myoviridae* family of somatic coliphages is the predominant type in human sewage (Muniesa et al. 2011) and represented the majority of coliphages found in the dry collection in both Sparkill Creek and the Pocantico River. *Siphoviridae*, which are most common in surface waters, represented only 6% of the phage.

Persistence studies have shown that the phages are highly subject to decay when exposed to ultraviolet light and freezing temperatures (Gilpin et al. 2013; Lebarbenchon et al. 2011). They are generally incapable of infection in water due to the dilution of their hosts. The population of somatic coliphage is subject to changes, as well. In controlled experiments, it has been shown that members of the *Microviridae* become more predominant with time, surpassing *Myoviridiae* (Lee 2009). As such, depending on time of introduction and meteorological factors, the coliphage detected may not accurately represent the input population.

Isolation and Analysis of DNA from Cultured Bacteria

Figure 3 tracks Gram-negative aerobic bacteria cultured from both wet and dry collections. These bacteria represent organisms often associated with the gastrointestinal tract of warm-blooded animals and, as such, some fecal indicator bacteria are expected to be cultured. Differential presence and abundance of these bacteria can help to pinpoint source and type of contamination. The results presented indicate a source of *Bacteroidetes* at the Rockleigh and Blauvelt sites. *Bacteroidetes* is a diverse group of organisms, comprised of Gram-negative anaerobic and aerobic bacteria. There are over 7,000 species in this phylum, having colonized myriad niches in the environment (Ryan et al. 2009). A dissection of the types of this phyla present at the Rockleigh site is warranted. Those identified in the current study are only the aerobic members of *Bacteroidetes.*

The genus *Stenotrophomonas* is composed of Gram-negative motile bacilli, containing over ten species (Palleroni and Bradbury 1993). These bacteria normally

reside in soil and on plants (Ryan et al. 2009); however, there are species of *Stenotrophomonas* that are opportunistic human or plant pathogens, and one species that is considered an emerging antibiotic resistant healthcare-acquired infectious agent (*S. maltophilia*). Studies have shown that some species can be involved in biofilm formation in the soil (Di Bonaventura et al. 2003). Expression studies have determined the relevance of certain genes to this biofilm formation in clinical isolates of *S. maltophilia.* The presence of *Stenotrophomonas* almost exclusively in dry conditions in Sparkill Creek implies the prevalence of biofilms, disturbed with rainfall events. It was not found in abundance at the Spruce and Skating Pond sites that, in dry conditions, exhibit the lowest numbers of coliphage, coliform, and *E. coli.* Biofilm formation using this organism as an indicator will be the focus of future research.

Continuing Studies

Subsequent studies along Sparkill Creek have focused on research initiated in the current study. Algal and fungal source tracking has been done at Motoris and Rockleigh sites in Sparkill Creek, comparing eukaryotic communities in dry and wet weather, and bacterial communities in colder weather. Isolation of *Stenotrophomonas* has also been attempted, albeit unsuccessfully. Markers known to be associated with biofilm formation in *Stenotrophomonas* were amplified from the DNA isolated in this study, subcloned and then sequenced. Finally, all analyses described have been done to examine the relationship between soil and water microbiota.

ACKNOWLEDGEMENTS

Thank you to the Hudson River Foundation for supporting this study through the Tibor T. Polgar Fellowship Program. A special thanks to Dr. Kasie Farlow and Kimberly Acevedo for their significant contributions to this research and the analyses. This work was funded by NYS WRI and the NYSDEC HREP, with support from the NYS Environmental Protection Fund.

REFERENCES

Ahmed, W., A. Goonetilleke, D. Powell, and T. Gardner. 2009. Evaluation of multiple sewage-associated *Bacteroides* PCR markers for sewage pollution tracking. Water Research 43:4872-4877.

Bernhard, A.E., and K.G. Field. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Applied Environmental Microbiology 66: 4571–4574.

Brezina, S.S. and M.D. Baldini. 2008. Detection of somatic coliphages as indicators of faecal contamination in estuarine waters. Revista Argentina de Microbiologica 40: 72-74.

Cardenas, E., and J.M. Tiedje. 2008. New tools for discovering and characterizing microbial diversity. Current Opinion in Biotechnology 19: 544–549.

Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods 69: 330–339.

Clarridge, J. E. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clinical Microbiology Reviews 17: 840–862.

Cole, J.R, Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, and J.M. Tiedje. 2009. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37 doi:10.1093/nar/gkn879.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G.L. Andersen. 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology 72: 5069–5072.

Devriese L.A., A. Van De Kerckhove, I.R. Kilpper-Balz, and K.H. Schleifer. 1987. Characterization and identification of *Enterococcus* species isolated from the intestines of animals. International Journal of Systematic Bacteriology 37: 257-259.

Dhillon, T.S., E.K. Dhillon, H.C. Chau, W.K. Li, and A.H. Tsang. 1976. Studies on bacteriophage distribution: Virulent and temperate bacteriophage content of mammalian feces. Applied and Environmental Microbiology 32: 68-74.

Di Bonaventura, G., I. Spedicato, D. D'Antonio, I. Robuffo, and R. Piccolomini. 2003. Biofilm formation by *Stenotrophomonas maltophilia*: Modulation by quinolones, trimethoprim-sulfamethoxazole, and ceftazidime. Antimicrobial Agents and Chemotherapy 48: 151–160.

Ferguson, D. M., J.F. Griffith, C.D. McGee, S.B. Weisberg, and C. Hagedorn. 2013. Comparison of *Enterococcus* species diversity in marine water and wastewater using Enterolert and EPA Method 1600. Journal of Environmental and Public Health. Article ID 848049.

Gantzer, C., A. Maul, J.M. Audic, and L. Schwartzbrod. 1998. Detection of infectious enteroviruses, enterovirus genomes, somatic coliphages, and *Bacteroides fragilis* phages in treated wastewater. Applied and Environmental Microbiology 64:4307-4312.

Gantzer, C., J. Henny, and L. Schwartzbrod. 2002. *Bacteroides fragilis* and *Escherichia coli* bacteriophages in human faeces. International Journal of Hygiene and Environmental Health 205: 325-328.

Garza, D.R., and B. E. Dutilh. 2015. From cultured to uncultured genome sequences: Metagenomics and modeling microbial ecosystems. Cellular and Molecular Life Sciences 72: 4287–4308.

Gerba, C.P. 1987. Phage as indicators of fecal pollution, in Phage Ecology. Eds. Goyal, S.M., Gerba, C.P., Bitton, G. Wiley Interscience: New York, NY.

Gilpin, B.J., M. Devane, B. Robson, F. Nourozi, P. Scholes, S. Lin, D.R. Wood, and L.W. Sinton. 2013. Sunlight inactivation of human polymerase chain reaction markers and cultured fecal indicators in river and saline waters. Water Environment Research 85: 743–750.

Grabow, W.O. 2001. Bacteriophages: Update on application as models for viruses in water. Water SA 27: 251-268.

Havelaar, A.H. 1987. Bacteriophages as model organisms in water treatment. Microbiological Sciences 4: 362-364.

Kinzelman, J., C. Ng, E. Jackson, S. Gradus, and R. Bagley. 2003. Enterococci as indicators of Lake Michigan recreational Water quality: Comparison of two methodologies and their impacts on public health regulatory events. Applied and Environmental Microbiology 69: 92–96.

Lebarbenchon, C., M. Yang, S.P. Keeler, M.A. Ramakrishnan, J.D. Brown, D.E. Stallknecht, and S. Sreevatsan. 2011. Viral replication, persistence in water and genetic characterization of two Influenza A viruses isolated from surface lake water. PLoS ONE 6(10):e26566.

Lee, H.S. 2009. Somatic coliphage families as potential indicators of enteric viruses in water and methods for their detection. PhD Thesis. University of North Carolina, Chapel Hill, NC.

Long, S.C, S.S. El-Khoury, S. Oudejans, M.D. Sobsey, and J. Vinjé. 2005 Assessment of sources and diversity of male-specific coliphages for source tracking. Environmental Engineering Science 22: 367-377.

McLellan, S. L. 2004. Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. Applied and Environmental Microbiology 70: 4658-4665.

Muniesa, M. 2011. Bacteriophages and genetic mobilization in sewage and faecally polluted environments. Microbial Biotechnology 4:725–734.

New York State Department of Environmental Conservation (NYSDEC). 2019. NYS Section 303(d) List of Impaired/TMDL Waters. NYS Section 303(d) List of Impaired/TMDL Waters: www.dec.ny.gov/chemical/31290.html. Retrieved April 2019.

Palleroni, N. J., and J. F. Bradbury. 1993. *Stenotrophomonas*, a new bacterial genus for *Xanthomonas maltophilia*. International Journal of Systematic Bacteriology 43: 606–609.

Pascual, J., S. Blanco, M. García-López, A. García-Salamanca, S.A. Bursakov, O. Genilloud, G.F. Bills, J.L. Ramos, and P. van Dillewijn. 2016. Assessing bacterial diversity in the rhizosphere of *Thymus zygis* growing in the Sierra Nevada National Park (Spain) through culture-dependent and independent approaches. PloS ONE 11(1): e0146558.

Plummer, J. D. and S.C. Long. 2009. Identifying sources of surface water pollution: A toolbox approach. Journal of the American Water Works Association 101:75-88.

Reyes V. and S. Jiang. 2010. Ecology of coliphages in southern California coastal waters. Journal of Applied Microbiology 109: 431-440.

Ryan R.P., S. Monchy, M. Cardinale, S. Taghavi, L. Crossman, M.B. Avison, G. Berg, D. van der Lelie, and J.M. Dow. 2009. The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. Nature Reviews Microbiology 7: 514–525.

Schloss, P.D., D. Gevers, and S.L. Westcott. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE 6(12):e27310

Skraber S., B. Gassilloud, and C. Gantzer. 2004. Comparison of coliforms and coliphages as tools for assessment of viral contamination in river water. Applied Environmental Microbiology 70: 3644-3649.

Stewart, J. R., R.D. Ellender, J.A. Gooch, S. Jiang, S.P. Myoda, and S.B. Weisberg. 2003. Recommendations for microbial source tracking: Lessons from a methods comparison study. Journal of Water and Health 1:225-231.

United States Environmental Protection Agency (US EPA). 2006. Final Ground Water Rule 71 FR 65574. Website accessed December 18, 2018: www.epa.gov/dwreginfo/ground-water-rule. Office of Ground Water, U.S. Environmental Protection Agency, Washington, DC.