Vol. 77: 23–36, 2016
doi: 10.3354/ame01786

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online June 7

Site-specific environmental factors control bacterial and viral diversity in stormwater retention ponds

Matthew A. Saxton[1,5], Nuha S. Naqvi[2], Faraz Rahman[2], Charleton P. Thompson[2], Randolph M. Chambers[1,2], James M. Kaste[1,3], Kurt E. Williamson[2,4,*]

[1]Environmental Science and Policy Program, College of William & Mary, Williamsburg, VA 23185, USA
[2]Department of Biology, College of William & Mary, Williamsburg, VA 23185, USA
[3]Department of Geology, College of William & Mary, Williamsburg, VA 23185, USA
[4]2127 Integrated Science Center, College of William & Mary, Williamsburg, VA 23185, USA
[5]*Present address:* Department of Marine Sciences, University of Georgia, Athens, GA 30602, USA

ABSTRACT: Stormwater retention ponds are ubiquitous in urban and suburban landscapes of the United States. Most studies of the microbiology of these ponds have focused on the abundance and removal efficiency of fecal indicator bacteria. Here we provide the first comprehensive study of microbial community diversity and activity in these ponds, and assess how different environmental and engineering factors influence these communities. Watershed land-use had no grouping effect on pond geochemistry or biology. Instead, we found that microbial community composition and activity were best explained by site-specific environmental variables. 16S rRNA gene sequence analysis indicated that bacterial community structure varied greatly across sampled ponds, and appeared to be the result of autochthonous bacterial growth and not simply surface runoff. Metagenomic sequencing of pond viral communities suggested that viral taxa were influenced by external inputs, with viromes composed of bacteriophage that infect environmental bacteria, as well as viruses that infect eukaryotes and phages that infect bacteria pathogenic to eukaryotes. Collectively, our results indicate that the environmental conditions and microbial communities of these ponds vary greatly, even among ponds in close spatial proximity, and that pond microbial communities appear to be shaped by site-specific environmental factors.

KEY WORDS: Virus · Bacteria · Microbial ecology · Stormwater · Runoff · Retention ponds · Community composition

INTRODUCTION

Stormwater retention ponds are ubiquitous features in modern urban and suburban landscapes, and are designed to reduce the peak flows of runoff from impermeable surfaces, extending the time to discharge downstream. These ponds are additionally engineered to retain sediment, typically laden with nutrients such as phosphate and ammonium, metals, and biological materials, thereby protecting downstream ecosystems (Yousef et al. 1986, Livingston et al. 1997). Studies investigating the performance of these ponds in controlling water, sediment, and nutrient flow indicate a high level of variability in the ability of ponds to perform these functions (Schueler 1997a,b, Baldwin et al. 2009, Hancock et al. 2010).

Considering their widespread use, studies of the biological impact of stormwater retention ponds are currently insufficient. Investigation of the microbiology of these systems has largely been limited to

*Corresponding author: kewilliamson@wm.edu

studies focusing on the efficacy of these ponds in the removal of potential human pathogens using indicator species, including *Escherichia coli* (Struck et al. 2008, Hathaway et al. 2009a,b), fecal coliforms (Jeng et al. 2005), and West Nile virus (Gingrich et al. 2006, Rajal et al. 2007, Jackson et al. 2009). Results from these studies have been mixed, though some studies suggest that human pathogens are at times present and some may persist in these ponds for up to weeks (Davies et al. 2008, Pettersson & Åström 2010, Kinnaman et al. 2012).

A recent study investigated the water column biogeochemistry of stormwater retention ponds in southern Ontario, Canada (Williams et al. 2013). This work used dissolved organic matter and total suspended solids as proxies for internally cycled and externally loaded C, respectively. This study indicated that microbially driven internal nutrient cycling shapes the geochemical state of the pond more than external inputs, with most ponds having a low proportion of terrestrially derived humic matter in the overall dissolved organic matter pool and active internal microbial communities. A predominance of autochthonous C species was observed in the dissolved organic carbon (DOC) pools of most sampled ponds despite 2 notable precipitation events (>25 and >10 mm, respectively) over the course of the study, suggesting that the cycling of terrestrial C may occur rapidly following a rain event. These results suggest that while exogenous inputs have significant and measurable impacts on pond functions, internal processes and dynamics may be more important in determining overall pond activity.

In the present study, we undertook an initial survey of the water column microbiology of wet stormwater retention ponds. Our primary goal was to determine the bacterial and viral community composition of these ponds. Because of the focus on the efficacy of stormwater retention ponds in controlling the flow of potentially pathogenic taxa, a more thorough description of the microbial communities naturally arising in these engineered environments is important, but currently lacking. An important secondary goal was to identify how environmental factors and land use might explain the microbiology (i.e. community composition, activity) of the ponds. Additionally, we sought to identify common characteristics or correlations to describe the relationship between the geochemistry and microbiology of these ponds. Our specific hypotheses were: (1) ponds with similar land use will have more similar microbial community composition; and (2) community composition similarity between ponds will scale with geographic distance between ponds. To this end, we sampled 20 wet stormwater ponds in James City and York counties and the town of Williamsburg, Virginia, USA, in July 2012. Ten of the sampled ponds drained commercial areas, with land use characterized by the presence of large parking lots and roadways. The other 10 ponds drained residential areas, with land use including privately owned houses, lawns, and woodlands. We combined measures of bacterial and viral abundances, bacterial productivity, and nucleic acid-based community fingerprinting methods with basic geochemical measures to examine relationships between the geochemical state of these ponds and their microbiology. We followed this work with high-throughput DNA sequencing targeting the bacterial community via 16S amplicon sequencing and viruses via metagenomic DNA sequencing to characterize and compare the taxonomic composition of a subset of these ponds.

MATERIALS AND METHODS

Sample collection

Samples were collected from 20 stormwater re tention ponds in the town of Williamsburg as well as adjacent James City and York counties in Virginia, USA, between 16 and 27 July 2012 (Fig. 1). The ponds were all roughly the same size in terms of area and shared a similar geologic foundation on relatively flat unconsolidated Quaternary-aged

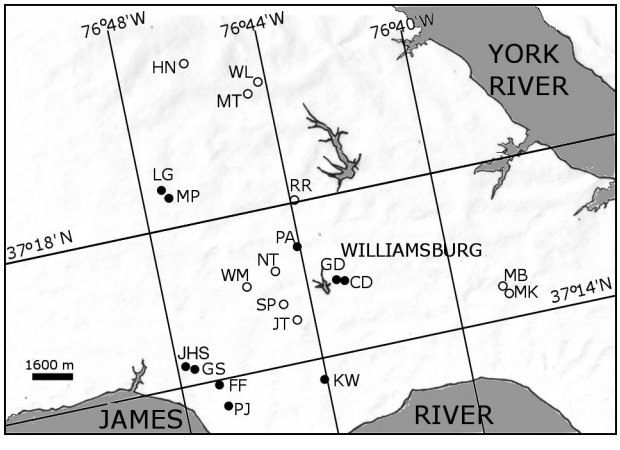

Fig. 1. Locations of sampling sites in James City and York counties as well as the city of Williamsburg, Virginia, USA. ○: stormwater ponds draining commercial areas; ●: ponds draining residential areas

sand, silt, and clay deposits. Ponds were grouped into 10 that drained commercial areas and 10 that drained residential areas, and were classified based on municipal zoning. Ponds in areas zoned as general residential, multifamily residential, and public land were classified as 'residential' and ponds in areas zoned for commercial, business, or mixed use were classified as 'commercial' (www.williamsburgva.gov, www.jamescitycountyva.gov/447/Zoning-Enforcement/zoning/, www.yorkcounty.gov/Portals/0/gis/Zoning_Map.pdf). Surface water was collected from open pond water by wading past surrounding vegetation and allowing disturbed sediment to settle for roughly 5 min. Water samples were collected by hand via surface grab using acid-washed 2 l polycarbonate bottles (Nalge Nunc International). At each site, 4 l were collected and immediately returned to the lab where approximately 2 l of each sample were immediately processed as described below and 2 l were frozen at −80°C for later analyses.

Water chemistry, bacterial production, and chlorophyll *a*

Temperature and conductivity were measured using a YSI 30 hand-held probe and dissolved oxygen (DO) was measured using a YSI 55 probe *in situ* at the time of sample collection. pH was determined once samples were returned to the laboratory using an UltraBasic pH meter (Denver Instruments). Dissolved nutrient concentrations (NO_2 + NO_3, NH_4, and PO_4) were determined using standard colorimetric assays on water filtered through Whatman GF/C glass fiber filters (Parsons et al. 1984). Estimation of bacterial production via ^{3}H-labled leucine incorporation was performed using the microcentrifuge method (Kirchman 2001). Chlorophyll *a* (chl *a*) samples were collected on 47 mm Whatman GF/C filters and frozen at −20°C until analysis. Chl *a* was extracted in 90% acetone at 4°C for 24 h. Following extraction, chl *a* was determined spectrophotometrically (Lorenzen 1967) on a Genesys 20 spectrophotometer. Calculations were corrected for the presence of phaeopigments. DOC samples were filtered through Whatman GF/C glass fiber filters and stored frozen at −20°C in polycarbonate bottles until processing. The filtrate for each sample was divided and stored frozen in glass or polycarbonate bottles, and non-purgable carbon (DOC) concentrations were determined using a Shimadzu TOC-V-L.

Bacterial and viral enumeration

Water samples (~1.5 ml) for bacterial and viral abundance determinations were dispensed into cryovials, flash frozen in liquid nitrogen, and stored at −80°C until processing (Wen et al. 2004). Bacterial and viral abundance were determined in duplicate subsamples from each pond via epifluorescence microscopy using a method adapted from Noble & Fuhrman (1998) as has been described previously (Hardbower et al. 2012).

Viral community profile analysis via random amplified polymorphic DNA-PCR

Viral concentrates for analysis of viral community composition and richness were generated from whole water samples, collected in 50 ml Falcon tubes and frozen at −80°C until further processing. Samples were thawed and syringe-filtered through a 0.2 µm pore-size sterivex filter into polyallomer ultracentrifuge tubes. Samples were then spun at 100 000 × *g* for 2 h at 4°C in a Beckman L8-M ultracentrifuge fitted with a Beckman SW41Ti swinging bucket rotor. Following centrifugation, supernatants were decanted and viral pellets were resuspended in a total of 80 µl of TMG buffer (10 mM Tris-Cl, 10 mM $MgSO_4$, and 1% v/v glycerol) and sodium azide was added (0.1% final conc.) to inhibit growth of bacterial contamination. Viral concentrates were stored in 1.5 ml microcentrifuge tubes at 4°C, as described in Hardbower et al. (2012).

Random amplified polymorphic DNA-PCR (RAPD-PCR) reactions were set up using the primer CRA-22 (5′-CCG CAG CCA A-3′), and thermocycler conditions were programmed as described by Winget & Wommack (2008). A 1 µl aliquot of virus concentrate, diluted to contain 10^5 virus particles, was used to provide template DNA, as previously described (Hardbower et al. 2012). RAPD-PCR products from the complete 25 µl reaction were separated by gel electrophoresis on 13 × 16 cm 1.8% metaphor agarose gels (Lonza) in 0.5× TBE buffer, run at 4 V cm^{-1}. Gels were stained with 1× SYBR Safe (Life Technologies) in 0.5× TBE buffer for 1 h prior to visualization of bands using a Kodak Gel Logic 100 imaging system. Banding patterns were analyzed using ImageQuant TL software (GE Healthcare Life Sciences) to perform densitometry analysis and converted to binary matrix format.

Bacterial community profile analysis via terminal restriction fragment length polymorphism

Total microbial community DNA extracts for downstream bacterial community fingerprinting and 16S sequence analysis were generated using 50–100 ml of whole water filtered through a 47 mm diameter 0.2 µm pore-size nitrocellulose filter (Millipore). DNA extraction was performed using the phenol-chloroform method as described in Hardbower et al. (2012). DNA extracts were stored in 0.2 µm filter-sterilized TE buffer at −20°C until analysis. A fragment of the 16S rRNA gene was amplified by PCR for each sample and bacterial community composition was characterized using terminal restriction fragment length polymorphism (T-RFLP) as described in Hardbower et al. (2012), except only *Hinf*I was used for restriction digestion.

Preparation of 16S rRNA gene amplicons

Eight ponds were selected for further investigation via 16S rRNA gene sequence analysis. Ponds were chosen to allow for analysis over several close-proximity pairs of both commercial and residential ponds distributed throughout the sample area. 16S rRNA gene fragments targeting hypervariable regions V1-2 were PCR amplified using primers 8F (Eden et al. 1991) and 336R (Weidner et al. 1996); these primers included multiplex identifier tags and adaptors for downstream sample multiplexing and sequencing. These primers were selected to cover the same region of the 16S gene investigated in the T-RFLP method described above. PCR amplification was performed on a Bio-Rad MJ Mini thermocycler using GoTaq Flexi DNA polymerase (Promega), using the following protocol: initial denaturation at 94°C for 3 min, 30 cycles of denaturation at 94°C for 1 min, primer annealing at 55°C for 1 min, and primer extension at 72°C for 2 min, with final extension at 72°C for 10 min. No fewer than 3 replicate PCR reactions were pooled and purified using the QIAquick PCR Purification Kit (QIAgen). Sequencing was performed at Selah Genomics using titanium chemistry on a Roche Applied Sciences GS-FLX+ platform (454 Life Sciences). Sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under ascension number SRP072803.

Bacterial sequence analysis

Sample demultiplexing, denoising, filtering of low-quality reads, and initial sequence analyses were performed using the QIIME 1.8.0 (Caporaso et al. 2010) bioinformatics package. Denoising was carried out using the script denoise_wrapper.py. Operational taxonomic unit (OTU) grouping and chimera checking and removal were carried out in QIIME using USEARCH (Edgar 2010), as were taxonomic assignments of OTUs (determined at 97% similarity) using UCLUST against the Greengenes 13.8 database (DeSantis et al. 2006) at a sequence similarity threshold of 0.97. The identity of taxa discussed explicitly in the text was confirmed via BLAST against the nr database (Altschul et al. 1990). Alpha-diversity metrics Chao1 (Chao 1984), inverse Simpson index (Simpson 1949), and Faith's phylogenetic diversity (Faith 1992) were also determined in QIIME. Libraries were normalized to the lowest read abundance (9411 reads, see Table S3 in the Supplement at www.int-res.com/articles/suppl/a077p023_supp.pdf) in all alpha-diversity measurements. Analyses were performed in R (RCoreTeam 2012) using the package phyloseq (McMurdie & Holmes 2013). OTU similarity was investigated using the package NeatMap (Rajaram & Oono 2010), in which OTU arrangement in a heatmap is determined by non-metric multidimensional scaling (NMDS) ordination using Bray-Curtis distance. Figures were generated using the R package ggplot2 (Wickham 2009).

Viral DNA metagenomes

Frozen water samples (2 l) were thawed within 6 mo of collection and processed according to Thurber et al. (2009). Briefly, samples were first filtered through 0.45 µm Stericup filters (Millipore) to remove cells. Virus particles in the filtrate were precipitated by the addition of NaCl (1 M final conc.) and polyethylene glycol (PEG) 8000 (10% w/v final conc.) and samples were incubated overnight at 4°C before centrifuging at 11 000 × *g* for 30 min. PEG pellets containing recovered virus particles were dissolved in TE buffer and virus particles were further purified using CsCl density gradient centrifugation, treated with DNase to remove free nucleic acids, and viral nucleic acids were extracted using formamide as previously described (Thurber et al. 2009). Viral nucleic acids were amplified in duplicate reactions using the illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences), and duplicate reactions were pooled and purified using the Qiagen DNeasy Kit (Qiagen) before sequencing at Selah Genomics on a Roche Applied Sciences GS-FLX+ platform (454 Life Sciences).

Viral sequence analysis

All libraries were dereplicated prior to analysis. Viral metagenomes (viromes) were annotated using MG-RAST v3.3.6 (Meyer et al. 2008) and Metavir v2.0 (Roux et al. 2012) using an Expect-value (E-value) cutoff of 10^{-5}. MG-RAST generates taxonomic assignments based on BLASTx searches against the M5NR database (which includes SEED, KEGG, NCBI nr, Phantome, GO, EBI, JGI, UniProt, VBI, and egg NOG), and functional assignments based on BLASTx searches against the SEED-Subsystem database. Metavir generates taxonomic assignments of virus-related sequences based on BLASTx searches against RefSeq Virus database. MG-RAST accession numbers are 4523577.3–4523580.3; Metavir identifiers are Crim Dell (CD), Jamestown High School (JH), Greensprings (GS), and John Tyler–199 (JT).

Viromes were assembled using the GS De Novo Assembler v2.8 (Roche Diagnostics) using default settings, and contigs were analyzed using contig mapping and visualization tools included in Metavir. Cross-assembly of the viromes was accomplished by concatenating the 4 libraries into a single fasta file, performing de novo assembly of this file using the GS De Novo Assembler (default settings), and then comparing each of the individual library sequences to the cross-assembly using crAss (Dutilh et al. 2012). Based on previous reports (Dutilh et al. 2012, de Cárcer et al. 2015) Wooter's distance metric was used.

Contig spectra were generated by Circonspect, using the default parameters of 35 bp overlap with 98% similarity (Angly et al. 2009). Community structure and alpha-diversity were modeled in PHACCS (Angly et al. 2005, 2006) using the contig spectra from Circonspect and average genome size from Genome relative Abundance and Average Size (GAAS); diversity statistics were also computed based on an assumed average genome size of 50 000 bp, typical of most phages (Steward et al. 2000). All rank-abundance models were tested (power law, exponential, logarithmic, log-normal, broken stick, and niche preemption) and the best model was selected based on lowest error values. PHACCS was also used to estimate gamma diversity based on the mixed contig spectrum of the combined viromes (Angly et al. 2006).

Statistical analyses

A 2-way ANOVA was used to investigate the impact of drainage area land-use on environmental and biological variables. Bivariate correlations using Spearman's rank were used to determine relationships between pond engineering, environmental, and biological factors. Correlation analyses and 2-way ANOVAs were performed in SPSS v20. Significant differences were qualified by a threshold of $p < 0.05$ unless otherwise indicated. Cluster dendrograms and NMDS ordination were used to investigate the similarity between pond bacterial and viral communities, respectively. RAPD and T-RFLP results were converted to similarity matrices (Dice coefficient) and ordination analysis were performed using PAST v.3.04 (Hammer et al. 2009). Mantel test was used to test for relationships between bacterial and viral profiles within ponds using the ecodist package in R (Goslee & Urban 2007). BioEnv was used to test for relationships between microbial community composition (using Bray-Curtis distance for T-RFLP and RAPD data, and weighted UniFrac for 16S sequence data; Bray & Curtis 1957, Lozupone & Knight 2005) and environmental factors using the vegan package in R executed through QIIME 1.9.1 (Clarke & Ainsworth 1993, Caporaso et al. 2010, Oksanen et al. 2015).

RESULTS

Water chemistry

A summary of water chemistry results including dissolved nutrients and physical characteristics is provided in Table S1 in the Supplement at www.int-res.com/articles/suppl/a077p023_supp.pdf. Notably, many of the measured parameters varied greatly among the 20 sampled ponds. For example, DO ranged from 0.4 to 11.5 mg l^{-1}, pH from 5.14 to 8.02, and DOC from 2.72 to 13.71 ppm. No significant differences were observed in any geochemical or physical variables via 2-way ANOVA comparing ponds draining residential areas and ponds draining commercial areas. Drainage area (43.5–1.8 ha), impervious cover area (28.3–0.6 ha), and percent impervious cover (90–10.62%) all varied greatly between ponds. Ponds draining commercial areas had significantly higher ($p = 0.008$) percent impervious cover, but total drainage area ($p = 0.152$) and total impervious area ($p = 0.110$) were not significantly different between commercial and residential ponds.

Bacterial abundance, production, and viral abundance

Bacterial and viral abundance are shown in Fig. 2A and Table S1. Bacterial abundance ranged from

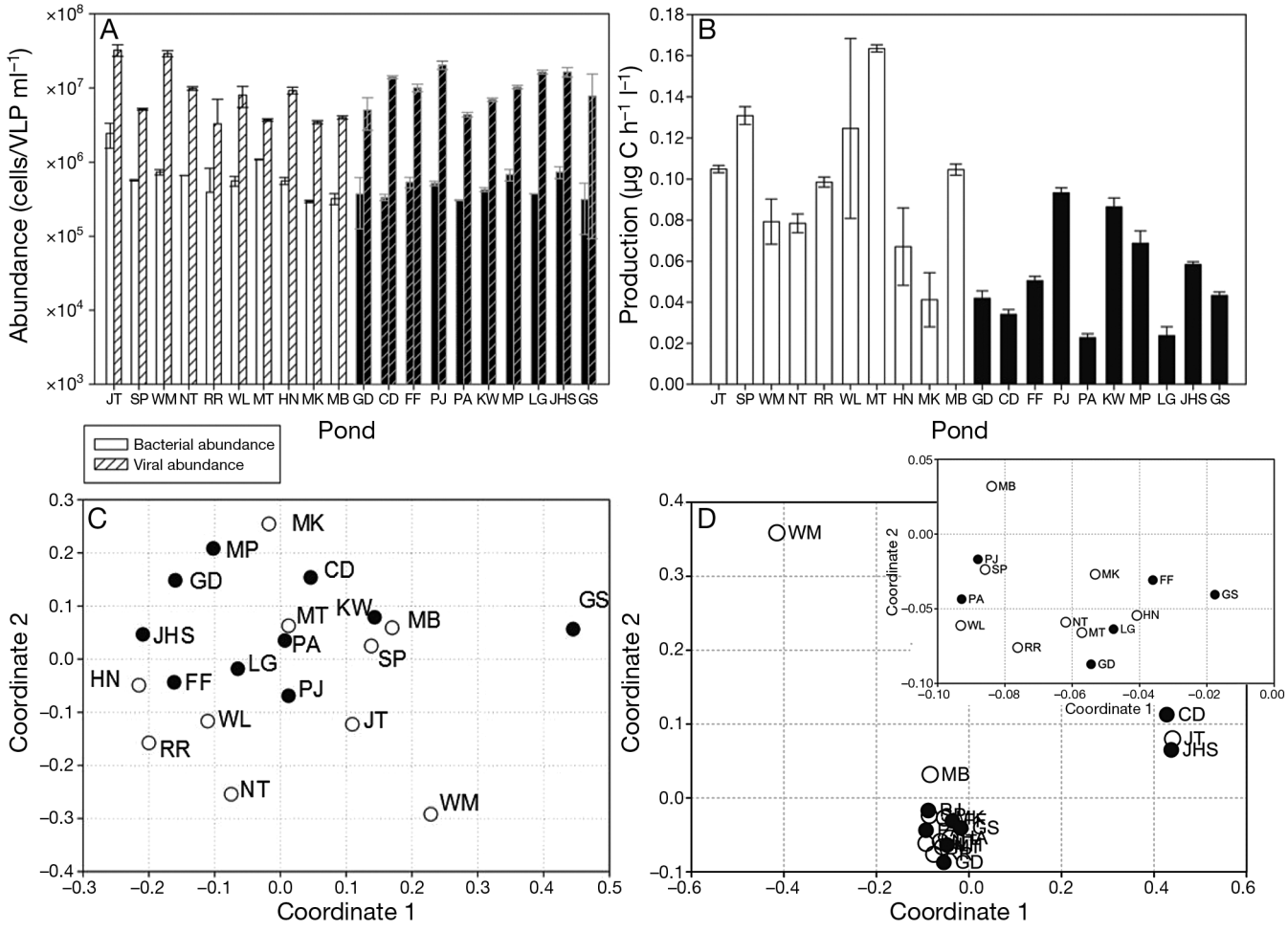

Fig. 2. Biological variables measured from stormwater ponds in July 2011. White bars and circles denote ponds that drain commercial areas and black bars and circles denote ponds that drain residential areas. (A) Bacterial (solid bars) and viral (hatched bars) abundance as measured by epifluorescence microscopy. All abundance values are the mean (±SD) of duplicate counts. (B) Bacterial production via ^{3}H leucine incorporation. All production values are the mean (±SD) of triplicate incubations. (C) Non-metric multidimensional scaling (NMDS) ordination plot of viral community structure determined by random amplified polymorphic DNA-PCR based on Dice similarity. (D) NMDS ordination plot of bacterial community structure determined by terminal restriction fragment length polymorphism based on Dice similarity. Inset details pond cluster (no data for KW)

2.95×10^{5} to 2.43×10^{6} cells ml^{-1}. Abundance of virus-like particles (VLPs) ranged from 3.28×10^{6} to 3.25×10^{7} ml^{-1}. The virus to bacterium ratio ranged from 3.40 to 43.83. No significant differences in bacterial ($p = 0.054$) or viral ($p = 0.87$) abundance were observed via 2-way ANOVA comparing ponds draining residential areas and ponds draining commercial areas. Bacterial production as measured by ^{3}H leucine incorporation varied from 2.28×10^{-2} to 1.64×10^{-1} µg C l^{-1} h^{-1} across the sampled ponds (Fig. 2B, Table S1). Production in ponds draining commercial areas was significantly higher ($p < 0.001$) via 2-way ANOVA than ponds draining residential areas.

Viral and bacterial community fingerprinting

Viral richness as estimated by RAPD-PCR band abundance ranged from 5 to 16 OTUs (Table S1). NMDS of the RAPD-PCR viral community matrix showed no obvious grouping of communities collected from residential or commercial ponds, nor was there clustering of ponds based on geographic distance, though the stress on the figure was relatively high (Fig. 2C, stress = 0.3244). Bacterial richness was determined using T-RFLP. Bacterial richness as estimated by the number of terminal fragments ranged from 9 to 80 OTUs. Similarity between ponds was investigated using a NMDS of

the T-RFLP-derived bacterial community matrix (Fig. 2D). A large grouping of 14 ponds with 40–60% fragment occurrence similarity was observed, but communities did not group by residential or commercial pond type, nor was there clustering of ponds based on geographic distance. RAPD and T-RFLP data matrices were used in a Mantel test based on Spearman's rank correlation to test for relationships between viral and bacterial community composition across ponds. Results of this Mantel test suggested that pond viral community composition was not correlated with bacterial community composition ($r = 0.105$, $p = 0.287$). Mantel tests were also used to test for relationships between geographical distance between ponds and viral and bacterial community composition. However, no significant relationship was observed (T-RFLP, $r = -0.153$, $p = 0.270$; RAPD, $r = -0.077$, $p = 0.504$). The set of environmental factors best describing viral and bacterial community differences as determined by RAPD and T-RFLP, respectively, was determined using BIOENV (Table S2 in the Supplement). Differences in viral communities as determined by RAPD were best explained by a combination of viral abundance, bacterial richness as measured by T-RFLP, and the percentage of pond watershed impervious cover area ($r = 0.2648$). Differences in pond bacterial communities as determined by T-RFLP were best explained by a combination of conductivity, virus abundance, and bacterial abundance ($r = 0.4858$).

Sequence analysis of bacterial 16S amplicon libraries

A total of 91 790 16S reads were generated (post-quality control) across the 8 stormwater pond libraries sequenced (Table S3 in the Supplement). These sequences were grouped into a total of 1076 97% similarity OTUs across all samples. Corrected chao1 richness ranged between 304.3 and 512.0. The alpha diversity measures presented in Table S1 indicate that Ponds GD and JT were the least diverse among those sampled (inverse Simpson = 21.8 and 14.6, respectively) and Pond GS was the most diverse (inverse Simpson = 65.0). Faith's phylogenetic diversity metric, which is the sum of phylogenetic tree branch lengths for a given library (Faith 1992), indicates that Pond CD (24.4) was the most diverse library while Pond MP (15.7) was the least diverse.

Bacterial community structure of selected ponds as determined by 16S rRNA amplicon sequencing is summarized in Fig. S1 in the Supplement at www.int-res.com/articles/suppl/a077p023_supp.pdf. The majority of OTUs collected across all ponds were affiliated with taxa commonly observed in freshwater habitats, including betaproteobacteria of the genera *Limnohabitans* and *Poynucleobacter*, a variety of cyanobacterial taxa, and chloroplasts originating from diatoms and green algae. Taxa associated with soil, including those affiliated with the class Gemmatimonadetes and of the family Microbacteriaceae, were also frequently observed. Frequency of shared OTUs between ponds was low (Fig. 3), and communities were often dominated by characteristic taxa not observed at high abundance in any other pond. The set of environmental factors best describing bacterial community differences as determined by 16S rRNA gene sequencing was also determined using BIOENV (Table S2). These factors were identified as water pH and bacterial richness as determined by T-RFLP ($r = 0.7619$).

Viral taxonomy

A total of 75.6 Mbp were generated from the 4 pond samples, corresponding to 169 962 reads post-

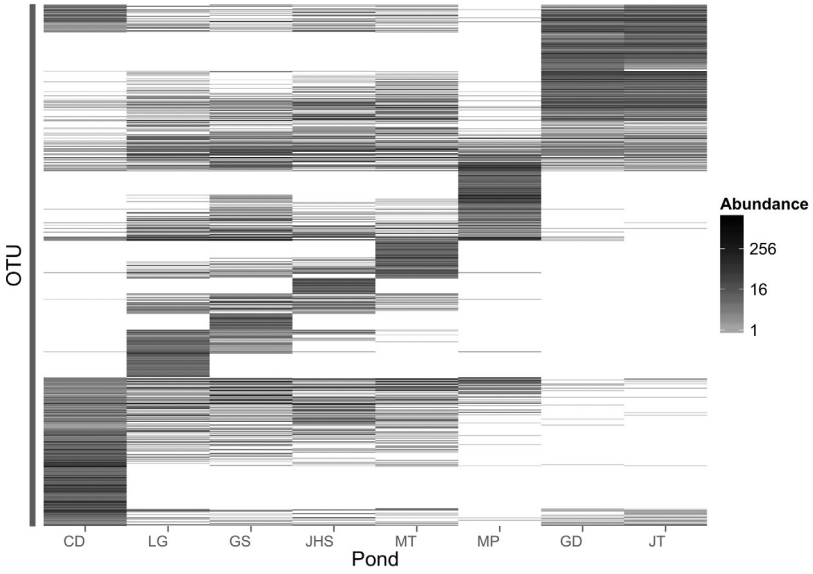

Fig. 3. Bacterial 16S rRNA gene sequence data from selected ponds (see Fig. 1) as 97% similar operational taxonomic units (OTUs) ordered using non-metric multidimensional scaling ordination of Jaccard distance. OTUs with an abundance lower than 2 across the ponds were omitted

quality control (Pond CD, 59 079; Pond JH, 50 920; Pond GS, 62 046; Pond JT, 22 571; a full list of ponds is provided in Table S1 in the Supplement) with an average read length of 412 bp. Roughly 10% (10.4–13.6%) of reads failed quality checks based on Duplicate Read Inferred Sequencing Error Estimation (DRISEE, Keegan et al. 2012), k-mer profiles, and nucleotide bias within reads, and were removed from subsequent analysis (Fig. S2a in the Supplement). Reads containing ribosomal RNA sequences comprised less than 2% of any library (0.8–2.0%) and were also removed prior to further analysis (Fig. S2a). Annotation of reads using MG-RAST (E-value cutoff of 10^{-5}) indicated that 29.4–60.0% of reads had no significant homology to known sequences (Fig. S2a). Reads were passed to Metavir for in-depth analysis of library taxonomic composition using the GAAS tool (Angly et al. 2009). Of the reads with homology to known sequences, 12.3–23.4% were classified as viruses, while the majority of reads in each library (66.8–80.3%) were classified as bacteria (Fig. S2b). Dominance of different virus groups varied by pond. The majority of reads from Ponds JH and GS were from double-stranded DNA (dsDNA) viruses (64.06 and 79.37%, respectively), with 56.52 and 66.54% affiliated with *Caudovirales*, respectively (Table S4 in the Supplement). Within these 2 metagenomes, *Podoviridae* sequences were the most highly represented of the tailed phages, followed by *Siphoviridae* and *Myoviridae*. In contrast, the majority of reads from Ponds CD and JT were from single-stranded DNA (ssDNA) viruses (75.85 and 70.82%, respectively), with 65.18 and 38.10% affiliated with *Microviridae*, respectively (Table S4). In addition to these general trends, almost one-third of the reads from the JH library were affiliated with the ssDNA family *Microviridae* (30.53%, Table S4), and almost one-quarter of reads from the JT library were affiliated with the ssDNA family *Circoviridae* (24.34%, Table S4). Very few reads were found to be associated with ssRNA viruses, and these were only observed in the JT library (0.67%, Table S4).

The most highly represented viral genotypes in 3 of the 4 libraries (CD, JH, JT) belonged to Chlamydia phages and comprised 23–51% of each library (Table S5 in the Supplement). The most represented viral genotype in the GS library belonged to Myxococcus phage Mx8 (17%), with Chlamydia phages as the second most abundant genotype (12%, Table S5). For Ponds JH and GS, the vast majority of known genotypes represented in each library were phages (Table S5), generally consistent with the dominance of tailed, dsDNA phages observed in Table S4. While the CD library was dominated by Chlamydia phages (>50%), this library contained relatively fewer sequences affiliated with tailed dsDNA phage genomes. Instead, the remaining viral genotypes comprising ≥1% of reads in the library belonged to circoviruses that infect eukaryotic hosts, and a plant virus (cassava mosaic virus, Table S5). The most represented genotypes in the JT library were markedly different from the other libraries, with reads affiliated with only one dsDNA phage (Bdellovibriophage phiMH2K, 15% of reads, Table S5) and ssDNA Chlamydia phages (23% of reads, Table S5). The remaining viral genotypes included cycloviruses and circoviruses infecting eukaryotic hosts such as bats and birds (Table S5).

While no bacterial cells were detected in viral concentrates via epifluorescence or electron microscopy, 66.8–80.3% of reads in the viral metagenomes were classified as bacterial (Fig. S2b). Based on GAAS analysis, sequences related to *Proteobacteria* dominated the CD, JH, and GS libraries, while sequences related to *Firmicutes* dominated the JT library (Table S6 in the Supplement).

Viral functional annotation

MG-RAST was used to generate metabolic profiles for each library using best BLASTx hit against the M5NR database with an E-value cutoff of 10^{-5}. Using this approach, 8.2–13.6% of reads could be classified by function. The most highly represented functional categories included phages, prophages, plasmids, and mobile genetic elements; nucleotide metabolism; cofactor, vitamin, pigment, and prosthetic group biosynthesis; membrane transport; and carbohydrate and protein metabolism (Fig. S3 in the Supplement). Pairwise comparisons of metabolic profiles using *t*-tests showed that cell division, RNA metabolism, and phage-related subsystems were more highly represented in Pond CD than in the other ponds ($p < 0.05$), and that photosynthesis-related subsystems were underrepresented in Pond GS relative to the other ponds ($p < 0.05$; Fig. S3).

Assembly and contig analysis

Average contig length varied by library from 1360 to 1683 bp, and the largest contigs produced for each library varied from 16 048 to 36 026 bp (Table S7 in the Supplement). Between 35.37 and 66.33% of all reads could be assembled, depending on the library

(Table S7). Mapping reads against the genomes of the most abundant viruses in each metagenome resulted in variable coverage from 18.17 to 76.7% (Tables S8 & S9 in the Supplement), with higher coverage of small ssDNA phages and lower coverage of larger (>3500 bp) dsDNA phage genomes. Predicted open reading frames (ORFs) in contigs >500 bp were compared with the GenBank nr database using BLASTx via the contig analysis tools in Metavir 2. Contigs contained recognizable phage structural, packaging, and nucleotide metabolism genes, as well as several conserved hypothetical phage proteins of unknown function (Table S10 in the Supplement). Contigs were generally AT-rich (54–59%) and contained high proportions of ORFans (73–78% of predicted ORFs).

Community structure and diversity

Based on PHACCS results, the community structures of the pond viral metagenomes — defined by richness (R), evenness (E), and Shannon-Wiener diversity index (H') — were best modeled by the power law, and graphically represented as rank-abundance curves (Fig. S4 in the Supplement). PHACCS results were quite different depending on whether an assumed average genome size of 50 000 bp was used for each metagenome, or the average genome size specifically calculated by GAAS was used (Table S11 in the Supplement). We believe that the numbers based on GAAS-estimated average genome size are more accurate; thus, only values based on GAAS were used in further analysis. Pond JT contained the highest richness of viral genotypes and overall viral diversity (R = 16 847, H' = 9.39), followed by Pond CD (R = 5312, H' = 8.04), and Pond JH (R = 3117, H' = 7.65), while Pond GS had the lowest richness and viral diversity (R = 756, H' = 6.17) (Table S11). By assembling a mixed sample across the 4 ponds, the gamma diversity represented by the pond viral communities was estimated at 5898 genotypes with H' of 8.23 (Table S11). Cross-assembly was used to assess dissimilarities across the pond viromes. The 4 viromes were cross-assembled using the GS De Novo Assembler and then each individual virome was compared with the cross-assembly using crAss to create a distance matrix that quantifies the dissimilarity across each pair of viromes (Dutilh et al. 2012). Values range from 0 (completely identical reads between libraries) to 1 (no shared reads between libraries). Our results (0.953–0.994) indicated genetically distinct viral assemblages within each pond (Table 1).

Table 1. Wooter's distance matrix comparison of viromes. Viromes were cross-assembled into 1 file using the GS De Novo Assembler and then each virome was compared with the cross-assembled file using crAss. CD: Crim Dell; GS: Greensprings; JH: Jamestown High School; JT: John Tyler. Values range from 0 (all reads identical) to 1 (no shared reads)

Pond	CD	GS	JH	JT
CD	0	0.976	0.956	0.975
GS		0	0.953	0.994
JH			0	0.960
JT				0

DISCUSSION

The primary goal of the present study was to provide a general survey of the bacterial and viral communities that develop in stormwater retention ponds to determine how environmental conditions may influence community structure. Understanding the microbial communities that develop naturally in these engineered ecosystems is important because of the ubiquity of these ponds in many urban and suburban settings (Yousef et al. 1986, Livingston et al. 1997) and the high potential that resident microbial communities will be exported to downstream ecosystems due to poor regulation and flow control (Hancock et al. 2010). A secondary goal was to investigate relationships between various geochemical, biological, and physical factors, including nutrient concentration and drainage area land use, and microbial community structure and activity.

The measurements collected in this study provide single time points, providing snapshot views of these individual ponds. Large-scale temporal changes in pond community composition are highly likely due to storm- and precipitation-driven perturbations. Indeed, at least one study has shown that aquatic viral communities can change rapidly and dramatically in response to major (tropical storms and hurricanes) weather-driven perturbations (Williamson et al. 2014). In the present study, rainfall conditions during the sampling period were average for the region, with 5 cm falling in the 2 wk prior to sampling, and another 7 cm falling within our sampling window. While this rain event likely impacted the microbial communities and geochemistry of the sampled ponds, no microbial community convergence or divergence (as measured by T-RFLP or RAPD, Fig. 2C,D) was observed among ponds sampled after this rainfall. Among those ponds selected for high-throughput DNA sequencing, no ponds with sequenced viral

metagenomes were sampled following this rainfall, and 2 of the 8 ponds on which 16S rRNA gene sequencing was conducted were sampled following this rainfall. Additionally, 16S rRNA gene libraries from Ponds LG and MP (the ponds sampled following the rainfall) were no more similar or different than the other sampled ponds (Figs. 3 & S1).

Geochemical and biological variability between ponds

High variability was observed in most measurements, both within land-use subgroups and the dataset as a whole. Particularly striking was the variability in bacterial and viral abundance (2.95×10^5 to 2.43×10^6 cells ml^{-1} and 3.28×10^6 to 3.25×10^7 VLPs ml^{-1}, respectively), both of which had a range of approximately one order of magnitude (Fig. 2A, Table S1). These values span much of the range observed in natural freshwaters, roughly 10^5–10^8 (Bird & Kalff 1984, Maranger & Bird 1995, DeBruyn et al. 2004). Chl *a* concentration also varied appreciably (28.43 to 2060.72 µg ml^{-1}), with values ranging from those expected in eutrophic lakes to those observed during a mild algal bloom (DeBruyn et al. 2004). Bacterial production ranged from 0.03 to 0.325 µg C l^{-1} h^{-1}, comparable to values observed in mesotrophic lakes and eutrophic Lake Erie (DeBruyn et al. 2004), but less than those measured concurrently in a nearby eutrophic lake (10 µg C l^{-1} h^{-1}, M. A. Saxton & K. E. Williamson unpubl.). This high variability across geochemical variables is similar to that observed in southern Ontario (Canada) stormwater ponds by Williams et al. (2013).

NMDS analysis showed no clustering of sites based on land-use type in either the viral or the bacterial community compositions (Fig. 2C,D). Based on viral community composition, the samples were scattered more or less evenly across the NMDS plot, with no apparent grouping by land-use type or geographic proximity (Fig. 2C). Based on bacterial community composition, samples broke out into 3 discrete clusters, one of which included 14 of the 20 sampled ponds with a similarity of between 40 and 60% shared restriction fragments (Fig. 2D). No correlation between bacterial and viral richness across the 20 ponds (Table S12 in the Supplement), or between bacterial and viral community composition via Mantel test ($r = 0.105$, $p = 0.287$) was observed. This result contrasts with a previous report that described viral and bacterial dynamics in a nearby lake (Hardbower et al. 2012) in which bacterial and viral richness and community composition were significantly correlated. This contrast may be explained by 2 important differences: (1) the lake in the cited study was ~100-fold larger in volume than the ponds in the present study, providing for a larger autochthonous community to buffer against storm inputs; and (2) the lake study involved repeated measures of community composition of a single location over time whereas the present study involved single time points of multiple locations. Overall, the magnitude of variability across ponds is striking, especially given the small geographic distance between ponds (maximum 16 km) and the similar geology and weather conditions experienced by these ponds.

Differences in observed patterns of bacterial and viral diversity

While a core of shared bacterial OTUs (~20%) was observed across ponds (most significantly those affiliating with the family *Comamonadaceae*), pond bacterial communities were largely defined by high abundances (>10% of total pond OTUs) of OTUs that are only present in that individual site (Fig. 3). Microbial diversity appears to be strongly influenced, both between and within environments, by differences in defining key variables such as oxygen concentration, pH, and depth (Martiny et al. 2006, Hanson et al. 2012). This appears to be the case among the bacterial communities in the stormwater ponds sampled in this study, with pH identified as the key environmental factor that best explains differences in bacterial communities among ponds based on 16S rRNA gene sequencing ($r = 0.7619$, Table S2). Pond MP exemplifies this 'key variable' effect, as this pond is notable for its low DO concentration (0.4 mg l^{-1}) and the high relative abundance of sequences affiliating with many bacterial taxa identified as obligate anaerobes, including those from the families *Rhodocyclaceae* (13.8%) and *Chlorobiaceae* (2.93%) and the genus *Geothrix* (6.5%) (Fig. S1). This finding suggests that bacterial communities develop independently in each pond in response to the geochemical and physical conditions of the pond and are not necessarily a direct reflection of the community imported during a storm event. Further, no significant relationship was observed between the geographical distance between ponds and bacterial or viral community richness. Thus, we found no support for the idea that pond microbial communities are impacted by geographic proximity. However, time series data, which are currently lacking for these environments, would

be extremely helpful in further supporting or refuting this interpretation.

Viral communities, as determined by metagenomic analysis, were highly divergent across the sampled ponds, with each virome (viral metagenome) sharing very little genetic overlap with any other virome (Table 1). Based on viral genotypes that could be identified through database matches, pond viromes appeared to be determined not only by autochthonous viral production via bacterial infection and lysis, but also via watershed animal and plant populations. The influence of local eukaryotic populations on pond viromes can be observed by the presence of viruses that infect eukaryotes as well as viruses that infect bacterial pathogens of eukaryotes. Sequences affiliating with viruses of the family *Microviridae* were present in all ponds and were made up largely of viruses infecting eukaryotic hosts or bacteria that infect eukaryotes, including Chlamydia phages (Table S5). While the high relative abundance of *Microviridae* (>30% of reads affiliating with known viral sequences in 3 of the 4 metagenomes; Table S5) may be the result of the bias of the phi29 DNA polymerase toward the amplification of ssDNA genomes (Angly et al. 2006, López-Bueno et al. 2009, Roux et al. 2012), these phages are also likely present and abundant at these sites. The influence of local fauna on pond viral communities can be clearly illustrated by the metagenome of Pond JT. A notable feature of this pond is the presence of a flock (20–30 individuals) of Canada geese throughout much of the year. Interestingly, the virome of this pond was enriched for sequences grouping with viruses that infect birds, including beak and feather disease virus (Table S5). No combination of sampled environmental and biological variables strongly explains the observed variation in viral communities (Table S2). This is likely because the sampled environmental variables focused on those impacting bacterial communities (pH, DO, nutrient concentrations, etc.) and these variables are unlikely to directly impact the observed eukaryote-associated viruses. This highlights the importance of including additional measured variables in future metavirome studies of wet retention ponds, such as surveys of potential host populations (e.g. plants, insects, and animals).

All pond viromes included notable populations of tailed bacteriophage (*Caudovirales*), many of which likely infect the endemic bacterial population of the pond. Viruses that infect freshwater bacteria were poorly represented in all viromes, largely because there are few sequenced viruses of this type in extant databases. This result highlights the need for further study, isolation, and genomic sequencing of bacteriophages endemic to freshwaters.

Few significant differences were observed among functional gene categories between the ponds. The fact that the Pond GS virome had significantly fewer hits to photosynthetic genes (Fig. S3) was not surprising in the context of the larger data set. This pond had the lowest chl *a* concentration (118.88 µg l^{-1}, Table S1) and was the only pond to have no OTUs of photosynthetic taxa (cyanobacteria or chloroplast) higher than 1% of the total 16S reads among the ponds with sequenced viromes (Fig. S1).

Similar to the trends observed for bacterial communities, no clear relationship was observed between spatial proximity of ponds and genetic similarity between viral communities. For example, despite the close geographic proximity shared by Ponds JH and GS (Fig. 1), cross-contig comparisons suggested that the viral communities are only about 5% similar (Table 1). Analyses based on 16S amplicon sequences suggested that the bacterial communities of these ponds also have low degrees of overlap (23%, Figs. 3 & S1). Additional studies and larger sample sizes will be required to draw definitive conclusions on the factors that drive microbial dynamics, and the relationships between bacterial and viral community structure in these systems. However, in the present study, we found no support for the hypotheses that ponds with similar land use contained similar microbial communities, or that community composition similarity between ponds scaled with geographic distance between ponds.

Comparison with other freshwater viral metagenomes

The 4 pond viromes from our study were compared with those from other freshwater studies, including aquaculture ponds (Rodriguez-Brito et al. 2010), a wastewater treatment plant (Tamaki et al. 2012), 2 French lakes (Roux et al. 2012), 4 Saharan gueltas (Fancello et al. 2013), as well as 3 marine viromes (Angly et al. 2006) as an outgroup, using the Compare Viromes function in Metavir. Briefly, sampled reads from each metagenome were compared with those in every other selected metagenome using tBLASTx. Samples of 50 000 sequences of 100 bp read lengths were used to obtain comparable results across metagenomes. A similarity score between every pair of metagenomes was computed as the sum of best BLAST hit scores between each pair. The resulting score matrix for all metagenome pairs was used to

generate an NMDS plot that graphically represented global sequence similarity across multiple viromes. The metagenomes appeared to cluster according to broad habitat types, with marine samples in the upper left quadrant, wastewater samples to the far right along the first axis, and freshwater samples in the lower left quadrant (Fig. 4). The aquaculture samples were somewhat scattered within this quadrant, but the Saharan gueltas, French lakes, and stormwater retention ponds appeared to form tighter clusters. The observed clusters formed based on habitat type, independent of the large geographic distance between the sampling sites. These results suggest specific viral genotypes or that viral sequences are associated with particular freshwater biomes.

Considered together, these results suggest different models for how bacterial and viral communities develop in stormwater retention ponds. It is highly likely that microbial communities of wet retention ponds change drastically in response to stormwater influx. However, these temporal effects are beyond the scope of the present study. Even if initial communities are deposited by storm flows, our results suggest that pond bacterial communities develop autochthonously in response to the specific environmental conditions that arise within a given pond. Because this specific geochemistry varies so notably across different ponds, the resulting bacterial communities differed greatly from pond to pond. In contrast, viral communities appeared to be a combination of bacteriophage produced internally (and subject to the same pond-specific environmental factors that drive bacterial communities), viruses infecting eukaryotic hosts associated with the pond, and potentially, viruses persisting in the water column following runoff from the environment. While pond viral communities show a different set of factors influencing structure than those observed for bacteria, great variability is observed between ponds in both cases. In both bacterial and viral communities, geography and drainage basin usage do not seem to be a factor pushing pond microbial community structure together. These observations further support the idea that microbial communities in each pond develop independently, each with its own set of driving factors. This reality makes it difficult to draw overall conclusions about specific factors that determine pond microbial communities generally, but illustrate the importance of further study of stormwater pond microbiology, and the potential of these systems as models of microbial biogeography.

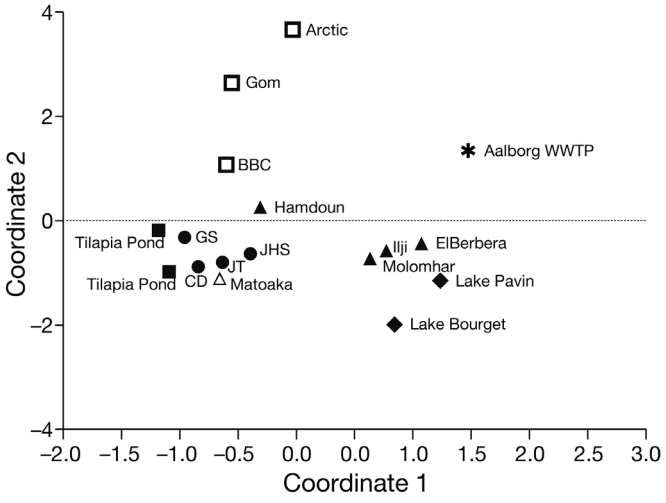

Fig. 4. Principal coordinate analysis of pond viromes based on BLASTx comparisons in MG-RAST. Symbols indicate different sample types. ●: stormwater retention ponds (present study); △: Matoaka, unpublished freshwater lake virome (Green et al. 2015); ■: aquaculture ponds; ◆: French lakes; □: marine samples; ✱: wastewater treatment plant; ▲: Saharan gueltas

Acknowledgements. The authors declare no conflict of interest. We thank Timothy Russell at the Keck Environmental Field Laboratory, College of William & Mary, for water chemical analysis. Stormwater pond drainage and impervious area information were provided by James City and York counties. Funding was provided by the Andrew W. Mellon Foundation and the Jeffress Memorial Trust (J-988).

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Angly F, Rodriguez-Brito B, Bangor D, McNairnie P and others (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics 6: 1–9

Angly FE, Felts B, Breitbart M, Salamon P and others (2006) The marine viromes of four oceanic regions. PLoS Biol 4: e368

Angly FE, Willner D, Prieto-Davó A, Edwards RA and others (2009) The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLOS Comput Biol 5:e1000593

Baldwin AH, Simpson TW, Weammert SE (2009) Urban wet ponds and wetlands best management practice. Final Report December 2009, Department of Environmental Science and Technology, University of Maryland, College Park, MD, p 540–577

Bird D, Kalff J (1984) Empirical relationships between bacterial abundance and chlorophyll concentration in fresh and marine waters. Can J Fish Aquat Sci 41:1015–1023

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27: 325–349

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

Chao A (1984) Nonparametric estimation of the number of classes in a population. Scand J Stat 11(4):265–270

Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. Mar Ecol Prog Ser 92:205–219

Davies CM, Petterson SM, Kaucner C, Ashbolt NJ, Mitchell VG, Taylor GD, Lewis J (2008) Microbial challenge-testing of treatment processes for quantifying stormwater recycling risks and management. Water Sci Technol 57: 843–847

de Cárcer DA, López-Bueno A, Pearce DA, Alcamí A (2015) Biodiversity and distribution of polar freshwater DNA viruses. Sci Adv 1:e1400127

DeBruyn JM, Leigh-Bell JA, McKay RML, Bourbonniere RA, Wilhelm SW (2004) Microbial distributions and the impact of phosphorus on bacterial activity in Lake Erie. J Gt Lakes Res 30:166–183

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M and others (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072

Dutilh BE, Schmieder R, Nulton J, Felts B, Salamon P, Edwards RA, Mokili JL (2012) Reference-independent comparative metagenomics using cross-assembly: crAss. Bioinformatics 28:3225–3231

Eden PA, Schmidt TM, Blakemore RP, Pace NR (1991) Phylogenetic analysis of *Aquaspirillum magnetotacticum* using polymerase chain reaction-amplified 16S rRNA-specific DNA. Int J Syst Bacteriol 41:324–325

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10

Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D, Desnues C (2013) Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. ISME J 7:359–369

Gingrich JB, Anderson RD, Williams GM, O'Connor L, Harkins K (2006) Stormwater ponds, constructed wetlands, and other best management practices as potential breeding sites for West Nile virus vectors in Delaware during 2004. J Am Mosq Control Assoc 22:282–291

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw 22: 1–19

Green JC, Rahman F, Saxton MA, Williamson KE (2015) Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA. Aquat Microb Ecol 75:117–128

Hammer Ø, Harper D, Ryan P (2009) PAST-PAlaeontological STatistics, ver. 1.89. University of Oslo

Hancock GS, Holley JW, Chambers RM (2010) A field-based evaluation of wet retention ponds: How effective are ponds at water quantity control? J Am Water Resour Assoc 46:1145–1158

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10:497–506

Hardbower DM, Dolman JL, Glasner DR, Kendra JA, Williamson KE (2012) Optimization of viral profiling approaches reveals strong links between viral and bacterial communities in a eutrophic freshwater lake. Aquat Microb Ecol 67:59–76

Hathaway A, Hunt W, Wright J, Jadlocki S (2009a) Field evaluation of indicator bacteria removal by stormwater BMPs in North Carolina. In: Starrett S (ed) Proc World Environmental and Water Resources Congress 2009: Great Rivers. May 17–21, 2009. ASCE, New York, NY, p 1123–1132

Hathaway J, Hunt W, Jadlocki S (2009b) Indicator bacteria removal in storm-water best management practices in Charlotte, North Carolina. J Environ Eng 135:1275–1285

Jackson MJ, Gow JL, Evelyn MJ, Meikleham NE and others (2009) Culex mosquitoes, West Nile virus, and the application of innovative management in the design and management of stormwater retention ponds in Canada. Water Qual Res J Canada 44:103

Jeng HC, England AJ, Bradford HB (2005) Indicator organisms associated with stormwater suspended particles and estuarine sediment. J Environ Sci Health Part A Tox Hazard Subst Environ Eng 40:779–791

Keegan KP, Trimble WL, Wilkening J, Wilke A, Harrison T, D'Souza M, Meyer F (2012) A platform-independent method for detecting errors in metagenomic sequencing data: DRISEE. PLOS Comput Biol 8:e1002541

Kinnaman AR, Surbeck CQ, Usner DC (2012) Coliform bacteria: the effect of sediments on decay rates and on required detention times in stormwater BMPs. J Environ Prot 3:787–797

Kirchman D (2001) Measuring bacterial biomass production and growth rates from leucine incorporation in natural aquatic environments. Methods Microbiol 30:227–237

Livingston E, Shaver E, Skupien J, Horner R (1997) Operation, maintenance, and management of stormwater management systems. Watershed Management Institute, Crawfordville, FL

López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A (2009) High diversity of the viral community from an Antarctic lake. Science 326:858–861

Lorenzen CJ (1967) Determination of chlorophyll and pheo-pigments: spectrophotometric equations. Limnol Oceanogr 12:343–346

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235

Maranger R, Bird DF (1995) Viral abundance in aquatic systems: a comparison between marine and fresh waters. Mar Ecol Prog Ser 121:217–226

Martiny JB, Bohannan BJ, Brown JH, Colwell RK and others (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112

McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217

Meyer F, Paarmann D, D'Souza M, Olson R and others (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386

Noble RT, Fuhrman JA (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. Aquat Microb Ecol 14:113–118

Oksanen J, Blanchet FG, Kindt R, Legendre P and others (2015) Package 'vegan'. Community ecology package, version 2.2-1. http://CRAN.R-project.org/package=vegan

Parsons TR, Maita Y, Lalli CM (1984) A manual of chemical and biological methods for seawater analysis. Pergamon Press, New York, NY

Pettersson TJ, Åström J (2010) Microbial contaminants removal in a stormwater pond. In: Rauch S, Morrison GM, Monzón A (eds) Highway and urban environment. Springer, Dordrecht, p 177–185

Rajal VB, McSwain BS, Thompson DE, Leutenegger CM, Wuertz S (2007) Molecular quantitative analysis of human viruses in California stormwater. Water Res 41: 4287–4298

Rajaram S, Oono Y (2010) NeatMap — non-clustering heat map alternatives in R. BMC Bioinformatics 11:45

RCoreTeam (2012) R: a language and environment for statistical computing. www.r-project.org

Rodriguez-Brito B, Li L, Wegley L, Furlan M and others (2010) Viral and microbial community dynamics in four aquatic environments. ISME J 4:739–751

Roux S, Enault F, Robin A, Ravet V and others (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7:e33641

Schueler T (1997a) Comparative pollutant removal capability of urban BMPs: a reanalysis. Watershed Protection Techniques 2:515–520

Schueler T (1997b) Influence of groundwater on performance of stormwater ponds in Florida. Watershed Protection Techniques 2:525–528

Simpson EH (1949) Measurement of diversity. Nature 163: 688

Steward GF, Montiel JL, Azam F (2000) Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. Limnol Oceanogr 45:1697–1706

Struck SD, Selvakumar A, Borst M (2008) Prediction of effluent quality from retention ponds and constructed wetlands for managing bacterial stressors in storm-water runoff. J Irrig Drain Eng 134:567–578

Tamaki H, Zhang R, Angly FE, Nakamura S and others (2012) Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ Microbiol 14:441–452

Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. Nat Protoc 4:470–483

Weidner S, Arnold W, Puhler A (1996) Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. Appl Environ Microbiol 62:766–771

Wen K, Ortmann AC, Suttle CA (2004) Accurate estimation of viral abundance by epifluorescence microscopy. Appl Environ Microbiol 70:3862–3867

Wickham H (2009) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, NY

Williams CJ, Frost PC, Xenopoulos MA (2013) Beyond best management practices: pelagic biogeochemical dynamics in urban stormwater ponds. Ecol Appl 23:1384–1395

Williamson KE, Harris JV, Green JC, Rahman F, Chambers RM (2014) Stormwater runoff drives viral community composition changes in inland freshwaters. Front Microbiol 5:105

Winget DM, Wommack KE (2008) Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. Appl Environ Microbiol 74:2612–2618

Yousef Y, Wanielista M, Harper H (1986) Design and effectiveness of urban retention basins. In: Urbonas BR, Roesner LA (eds) Proc Urban Runoff Quality: Impact and Quality Enhancement Technology. ASCE, New York, NY, p 338–350

Editorial responsibility: Curtis Suttle, Vancouver, British Columbia, Canada

Submitted: October 21, 2015; Accepted: April 18, 2016
Proofs received from author(s): May 24, 2016