CENTER for PRODUCE SAFETY

CPS 2016 RFP
FINAL PROJECT REPORT

Project Title
Developing cross-assembly phage as a viral indicator for irrigation waters

Project Period
January 1, 2017 – December 31, 2018 (extended to January 31, 2019)

Principal Investigator
Kyle Bibby
University of Notre Dame
Dept. of Civil & Environmental Engineering & Earth Sciences (CEEES)
171 Fitzpatrick Hall
Notre Dame, IN 46556
T: 574-631-1130
E: Kyle.J.Bibby.1@nd.edu

Objectives
1. *Collect irrigation water samples.*
   Irrigation waters will be collected and processed from western Pennsylvania (high sample coverage), as well as Arizona and California (once annual sampling). We will specifically target sampling impacted surface waters utilized for irrigation. Additional regions and sampling sites will be explored for sampling.
2. *Measure and determine correlation of crAssphage, viral pathogen, and FIB levels in irrigation water samples.*
   Previously developed methods for crAssphage detection, viral pathogens, and FIB will be used on collected irrigation water samples. A statistical analysis will then be employed to determine the correlation between crAssphage, pathogens, and existing indicators.
3. *Determine crAssphage limit of detection.*
   A critical question in the evaluation of crAssphage as a viral indicator for irrigation waters is the amount of sample volume necessary. To address this question, the limit of detection will be determined.

Funding for this project provided by the Center for Produce Safety through:
Florida SCBGP grant# 024072

FINAL REPORT

Abstract
Surface irrigation water represents a potential transmission source of viruses to produce, so ensuring irrigation water quality is critical to limiting produce-associated disease transmission. The greatest infectious risk from exposure to contaminated water is due to viral pathogens. All existing indicators of viral pollution are challenged by low abundance in the environment. Irrigation waters are typically monitored for biological quality utilizing fecal indicator bacteria (FIB). FIB are inadequate representatives of viruses due to differential occurrence, persistence, source, and abundance in the environment. Recently, a highly abundant bacteriophage, named cross-assembly phage (abbreviated as crAssphage), was found in the human gut. Initial investigations in the PI's group have found this viral indicator to be highly abundant in sewage and specific to sewage. In this study, 106 irrigation water samples (2 L/sample) were collected from four states (Pennsylvania, California, Arizona, and New York). Samples were analyzed for culturable fecal pollution indicators (*E. coli*, enterococci, and somatic coliphage); and for molecular fecal pollution indicators, including two assays targeting crAssphage (CPQ56 and CPQ64), a general bacterial fecal pollution marker assay (HF183/BacR287), and molecular assays for the viral pathogens adenovirus and polyomavirus. For the molecular indicators, CPQ56 had the highest detection rate (37.74%), followed by HF183/BacR287 (32.08%). CPQ56 also had the lowest method limit of quantification for the assays tested. Adenovirus and polyomavirus were not detected in any tested sample, demonstrating the utility and need for an improved viral indicator. CrAssphage had low correlation with the culturable indicators *E. coli,* enterococci, and somatic coliphage (R^2 = 0.04, 0.17, and 0.04, respectively), suggesting differential source characteristics. The best correlation for crAssphage with other indicators was between CPQ56 and HF183 (R^2 = 0.23). Ultimately, this work demonstrated that:

1. crAssphage can be detected in low-volume irrigation water samples;
2. crAssphage CPQ56 assay has improved sensitivity compared with other leading molecular fecal source tracking assays (HF183 and CPQ64); and
3. crAssphage detection has not yet been connected with pathogen co-occurrence or health outcomes in an agricultural setting.

Background
Controlling viral contamination from irrigation waters is essential to limit produce-borne outbreaks. The majority of infectious risk from exposure to contaminated water in certain exposure scenarios is believed to be due to viral infections, and indicators of viral pathogens have strongly correlated with increased risk of infection. Crops with the highest likelihood of contamination are those that are typically consumed raw or with limited processing. Low levels of virus are predicted to account for unacceptable levels of infectious risk from produce consumption. For example, a concentration of 10^{-1} norovirus per liter would represent an unacceptably high annual per-person likelihood of infection of >0.1% [1]. Similarly, a hepatitis A virus concentration of 10^{-2} per liter would be necessary to meet the acceptable annual risk likelihood of infection of 0.01% [2].

Fecal indicator bacteria (FIB) are currently utilized to determine irrigation water quality, based largely on their use in other applications (e.g., recreational water) and the lack of a widely-accepted alternative. FIB have different physiology, persistence, and transport than viral pathogens [3, 4]. Alternative pollution markers specific to viruses, such as non-pathogenic virus and bacteriophage-based approaches, have been proposed. All previously proposed viral markers are challenged by a low abundance in the environment. Preliminary results suggest crAssphage to be significantly more abundant and specific to human sewage contamination,

facilitating measurement of a viral indicator for risk-based decision-making and irrigation water management.

CrAssphage is a bacteriophage (i.e., a virus that infects bacteria) found in the human gut, and that was recently identified via metagenomic data mining [5]. CrAssphage was six times more abundant in tested fecal samples than all other known bacteriophages combined, comprising ~1.7% of all fecal metagenome reads [5]. Additionally, crAssphage was found to be near ubiquitous in human gut metagenomes [5]. The genome sequence of crAssphage was confirmed by PCR-based analyses [5]. In response to these findings, the PI's research group developed a qPCR assay to detect crAssphage [6]. This assay is highly sensitive (identified in 100% of wastewater samples) and largely human specific [6]. Follow-on studies have demonstrated the successful detection of crAssphage in sewage-polluted environments, and the correlation between crAssphage and viral pathogen occurrence [7]. Limited work has been done on the potential to apply crAssphage in less-impacted environments or agricultural settings.

The goal of the current study was to sample >100 irrigation water samples nationally to evaluate the suitability of crAssphage molecular markers for application as a fecal pollution indicator in an agricultural setting. In addition to crAssphage, both culturable and molecular indicators were quantified to determine their co-occurrence and correlation with the crAssphage indicator.

Research Methods

Study sites and sample collection
Irrigation water samples (2 L/sample) were collected from sites distributed across the United States over a two-year period starting in September 2016. Sites were located in Arizona (AZ; n=31), California (CA; n=24), New York (NY; n=29), and Pennsylvania (PA; n=22). Air and water temperature and pH were obtained for all water samples at the sampling sites prior to overnight shipment to our laboratory. Upon receipt, 1 L of water was used for total carbon and total inorganic carbon testing, and for culture enumerations (i.e., FIB). The remaining 1 L was used for DNA concentration, extraction, and qPCR analysis.

FIB testing
Water samples were analyzed for enterococci and *E. coli* presence according to EPA methods 1600 [8] and 1603 [9], respectively. Somatic coliphage was identified using EPA method 1602 [10]. Negative controls were included in all culture experiments to monitor potential contamination.

Sample concentration
Water samples were filtered in duplicate (500 mL per filter) according to previously published methods [11]. Briefly, the pH was adjusted to ~3.5 with 1 M HCl, and the sample was shaken vigorously prior to loading into a filtration device equipped with a 0.45-µm (47 mm dia.) nitrocellulose membrane filter (Thermo Fisher Scientific, Waltham, MA). Most filters (~ 98%) received the designated 500 mL water volume; however, due to clogged filters, about 2% of the filters received lower volumes of water. Calculations were adjusted to account for reduced filtration volumes. Four blanks were included as filtration controls. Following filtration, all filters were aseptically transferred into pre-loaded bead tubes (GeneRite, North Brunswick, NJ), and stored at −20°C (for <6 months) for downstream applications.

DNA extraction and molecular analyses
The QIAGEN DNeasy PowerWater Kit was used to extract DNA from the nitrocellulose membrane filters, following manufacturer instructions with some modifications: (i) Bead tubes were vigorously shaken in a FastPrep-24 Homogenizer (MP Biomedicals) at 6.0 m/s for 30 s,

and, (ii) DNA was eluted in 50 µL EB (provided in the kit). A total of eight method extraction blanks were included for DNA extraction quality control purposes.

DNA fragments containing PCR products for crAssphage CPQ56 and CPQ64 [6], human polyomavirus (HPyV) [11], human adenovirus (HAdV) [12], and the human-associated *Bacteroides* cluster HF183/BacR287 [13] were synthesized by IDT. The fragments were diluted in 10 mM Tris with 0.1 mM EDTA generating 10, 10^2, 10^3, 10^4, 10^5 copies/2-µL calibration standards.

Five human-specific TaqMan qPCR assays were utilized in this study: HF183/BacR287, HPyV, HAdV, and crAssphage CPQ56 and CPQ64. First, DNA was tested for HF183/BacR287 and crAssphage CPQ56. Given than only <50% of the samples yielded positive results for crAssphage CPQ56, we tested the crAssphage CPQ56-positive samples for the remaining crAssphage CPQ64, HPyV, and HAdV detection assays to confirm (i) the correlation between HF183/BacR287, crAssphage CPQ56 and CPQ64, and (ii) co-occurrence of the crAssphage virus with the HPyV and HAdV. Each qPCR reaction contained the TaqMan Environmental Master Mix, 0.2 mg/mL bovine serum albumin, 1 µM of each F and R primers, and 80 nM 6-carboxyfluorescein (FAM) probe with corresponding 2 µL template DNA (calibration standards or unknown DNA samples), plus DNase-free water to generate a 25-µL reaction [6, 13].

Assay data is summarized in Table 1. Three independent calibration standard runs (containing triplicates of 10, 10^2, 10^3, 10^4 and 10^5 copies/2-µL standards in each run) were performed to generate an "average" calibration standard curve for each of the five targets. Five lower limit of quantification (LLOQ; 95% CI associated with 10 copies/2 µL) values were calculated using Rstudio (Version 1.1.442), assuming a normal distribution of 10 copies/2-µL quantification cycle (Cq). The standard curves for all five assays were subject to linear calibration model and amplification efficiency ($E = 10^{(-1/slope)} - 1$) acceptance criteria ($R^2 \geq 0.98$ and $0.87 \leq E \leq 1.1$). All qPCR reactions were performed in triplicate under the thermal cycling profile: 10 min at 95°C followed by 40 cycles of 15 s at 95°C, and 1 min at 60°C in a CFX96 Touch™ Real-Time PCR detection system. Data was analyzed using the CFX Maestro data analyzing software. The relative fluorescent unit (RFU) was manually fixed to a value of 100 [14]. Cq values smaller than the LLOQ were classified as in the range of quantification (ROQ); those greater than the LLOQ were categorized into the most probable number (MPN) group. In every qPCR instrumental run, six no-template controls were incorporated by adding DNase-free water instead of DNA extract. The raw Cq values were exported to Microsoft Excel, and all unknown sample concentrations were calculated from raw Cq values with five corresponding standard curves.

Research Results

Sample Collection

In total, 106 irrigation water samples were collected and analyzed. We note that collaborators were instructed to target sites with expected contamination to maximize the likelihood of crAssphage detection. Irrigation water samples were collected at air temperatures of 24.37±7.18 °C (n=95, mean±sd); water temperature was 20.29 ± 5.42 °C (n=95, mean±sd), pH was 7.78±0.67 (n=98, mean±sd), total carbon was 2.41±1.85 mg/100 mL (n=51, mean±sd), and total inorganic carbon was 2.12±1.88 mg/100 mL (n=51, mean±sd).

Culturable Indicators

Concentrations of *E. coli* (range = 0.1–3.8 log CFU/100 mL; median = 1.8 log CFU/100 mL), enterococci (range = 0.1–3.9 log CFU/100 mL; median = 1.8 log CFU/100 mL), and somatic coliphage (range = -0.5–3.7 log CFU/100 mL; median = 1.3 log CFU/100 mL) were determined using standard culture methods as described above. Concentration data for these indicators is shown in Figure 1. Enterococci were detected in 93 samples, *E. coli* was detected in 85 samples, and somatic coliphage was detected in 90 samples.

Molecular Indicators
Concentrations of the molecular indicators CPQ56, CPQ64, and HF183/BacR287 are shown in Figure 2. CPQ56 was detected in 40 samples, HF183/BacR287 in 34 samples, and CPQ64 in 14 samples. Human adenovirus and human polyomavirus were not detected in any sample.

Sensitivity of CPQ56 Assay
Our results indicate that crAssphage CPQ56 had the lowest LLOQ value, making this the most "sensitive" water quality assay in our study. The crAssphage CPQ56 detection assay showed the highest detection rate (37.74%), followed by HF183/BacR287 (32.08%). The combined result of the crAssphage and bacterial molecular markers in irrigation waters clearly illustrates that CPQ56 is a more abundant and robust indicator. Two crAssphage indicator candidates were tested in our study: CPQ56 and CPQ64. From these, the CPQ56 assay (with lower LLOQ) can more reliably detect lower target concentrations from environmental water samples with higher statistical confidence than CPQ64. This finding is consistent with observations by Ahmed et al. [15], that the CPQ56 assay performs better than CPQ64 in non-irrigation water samples.

Correlation of CrAssphage with Water Quality Parameters
No correlation was observed between crAssphage occurrence (+/-) with total carbon ($p = 0.83$) or total inorganic carbon ($p = 0.81$). Weak but statistically significant correlations were observed between crAssphage occurrence (+/-) with pH ($p = 0.03$) and water temperature ($p = 0.04$). Lower pH was associated with positive crAssphage occurrence, and lower water temperature was associated with positive crAssphage occurrence. We hypothesize that this is reflective of the transport and/or persistence properties of crAssphage, as both pH and water temperature are known to influence viral fate and transport. However, additional controlled experiments will be necessary to confirm these hypotheses.

Correlation of CrAssphage with Culturable Indicators
Regression between CPQ56 and *E. coli*, enterococci, and somatic coliphage concentrations demonstrated limited correlation, as shown in Figure 3. We note that this is a highly censored data set, as <50% of samples were CPQ56 positive. The observed R^2 values for enterococci, *E. coli,* and somatic coliphage were 0.17, 0.04, and 0.04, respectively. Pairwise comparison of the concentrations of culturable indicators in CPQ56 positive and CPQ56 negative samples similarly demonstrated limited correlation between CPQ56 detection and culturable indicator concentration. P-values for pairwise comparison were $p = 0.86$, 0.89, and 0.66 for enterococci, *E. coli*, and somatic coliphage, respectively.

Correlation of Molecular Indicators
Correlations between molecular indicators measured in this study were weak (Figure 4). Regression between the CPQ56 and HF183/BacR287 markers had $R^2 = 0.23$. Regression between the CPQ56 and CPQ64 markers had $R^2 = 0.07$. These datasets are severely censored (i.e., the majority of samples were non-detects), necessitating that these correlations are built upon a small subset of total samples. Correlations between crAssphage and human adenovirus and human polyomavirus were not possible as these markers were not detected in any sample.

Limit of Detection
The assay limit of detection was estimated using standard protocols for study concentration (500 mL sample), DNA extraction (50 µL elution), and qPCR (2 µL DNA per reaction, 10 copies within the quantifiable range), which resulted in a standard detection limit of 50 CPQ56 copies/100 mL sample. This was then confirmed via dilution of CPQ56-positive samples, and no inhibitory effects or other factors that may increase the detection limit were observed.

Study Limitations

Our study had several important limitations that should be considered when interpreting the results. First, the data for all molecular indicators was highly censored (i.e., majority of samples were non-detects) and many molecular measurements were near the method detection limit. In addition, samples were necessarily transported prior to analysis, generally via overnight shipping. This sample transport time may have altered the microbiological properties of the samples. Samples were collected by a variety of collaborators, and while steps were taken to standardize the collection procedures, minor site-specific differences may have impacted study results. Finally, the project PI moved institutions in the middle of the project period (University of Pittsburgh to University of Notre Dame). This resulted in the shipment of samples, change in project personnel, and change in specific analytical instruments, all of which have the potential to influence observed project outcomes.

Outcomes and Accomplishments

- Successful collection and analysis of >100 samples.
- The crAssphage CPQ56 assay was judged to have the greatest sensitivity, as measured by the highest lower limit of quantification (LLOQ) and highest detection rate of molecular indicators (37.7% of samples).
- CrAssphage occurrence was weakly correlated with water temperature and pH.
- CrAssphage occurrence was not significantly correlated with *E. coli*, enterococci, or somatic coliphage.
- Pathogenic viruses (human adenovirus and human polyomavirus) were not detected.
- Limited correlation was observed between evaluated indicators, and correlation between indicators was challenged by a highly censored dataset.

Summary of Findings and Recommendations

CrAssphage (as the CPQ56 assay) was successfully detected in 37.7% of samples; however, limited correlation between this marker and established sources of fecal contamination (*E. coli*, enterococci, somatic coliphage, HF183 molecular assay) was observed. The poor correlation between crAssphage occurrence and abundance with more established fecal pollution sources suggests differing sources and/or fate and transport characteristics than these sources. In addition, the poor correlation between crAssphage and more established fecal pollution sources limits the connection between crAssphage occurrence at low concentrations and health effects. Nearly all crAssphage studies to date have focused on environments with high levels of fecal contamination (e.g., sewage, sewage-contaminated waters). This is one of the first studies to investigate crAssphage characteristics in waters that are not known to be sewage impacted. Future studies should look to connect crAssphage occurrence with health effects, e.g. occurrence on produce with known contamination. In addition, future studies should seek to clarify crAssphage transport and persistence characteristics to inform its application in sites without a known direct sewage impact source.

Acknowledgements

The study authors gratefully acknowledge the help of several collaborators in sample collection, including: Martin Wiedmann and Dan Weller from Cornell University; the staff of the Laboratory of Environmental Microbiology at the Water and Energy Sustainable Technology Center (University of Arizona); Michael Cooley and Lisa Gorski from the USDA in California; and Tom Ford at Penn State Extension.

The CPQ56 and CPQ64 primers used in this study are the subject of a patent application entitled “Cross-Assembly Phage DNA Sequences, Primers and Probes for PCR-based Identification of Human Fecal Pollution Sources” (Application Number: 62/386,532). Universities and non-profit researchers interested in using this technology must obtain a research license from the US EPA. To apply for a research license, please request additional information from ftta@epa.gov.

APPENDICES

Publications and Presentations

- X. Li, N. Morales-Soto, K. Crank, Z. Wu, J. Greaves, N. Aquino-Carvalho, C. Ference, and K. Bibby et al. Evaluating Cross-Assembly Phage as a Viral Indicator for Irrigation Waters. *In preparation*.

  Note: External collaborators are not listed on the manuscript in preparation as they have not yet had the opportunity to review or comment on study results, but they will be included as study authors.

- The PI presented interim results for the project at the 2017 and 2018 CPS Research Symposium, and will present final results at the 2019 CPS Research Symposium.

Budget Summary

Total funds awarded were $149,385.00, and all funds will be used by the end of the project period.

Tables and Figures

(see below)

Table 1 and Figures 1–4

Table 1. qPCR parameters for assays used in this study. The LLOQ (lower limit of quantification) is displayed as the highest Cq, or qPCR cycle number, that allowed quantification. A higher value here allows more cycles to be included within the quantifiable range, and thus lower quantities of initial target could be quantified (i.e., more sensitive assay).

Assay	Slope	Y-Intercept	*E*	R^2	LLOQ (Cq)
HF183/BacR287	-3.468	40.108	0.943	0.985	37.639
crAssphage CPQ56	-3.604	42.669	0.894	0.989	39.539
crAssphage CPQ64	-3.487	45.403	0.936	0.992	38.897
Human Adenovirus	-3.695	43.178	0.865	0.995	36.186
Human Polyomavirus	-3.475	42.165	0.940	0.996	39.172

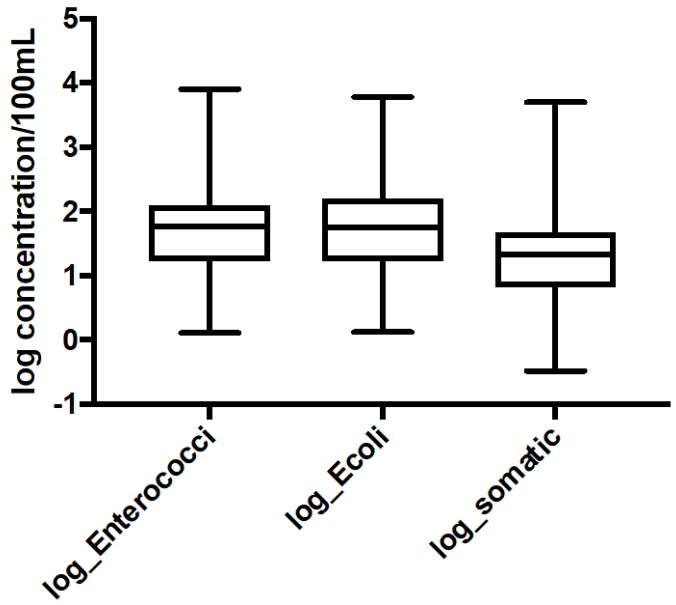

Figure 1. Comparison of concentrations for culturable indicators. Enterococci were detected in 93 samples, *E. coli* was detected in 85 samples, and somatic coliphage was detected in 90 samples.

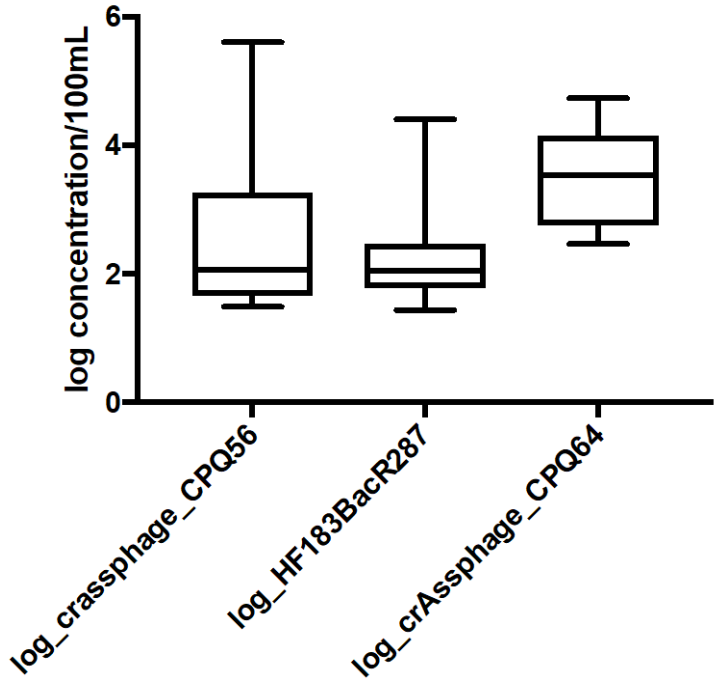

Figure 2. Comparison of concentrations for molecular indicators. CPQ56 was detected in 40 samples, HF183/BacR287 was detected in 34 samples, and CPQ64 was detected in 14 samples.

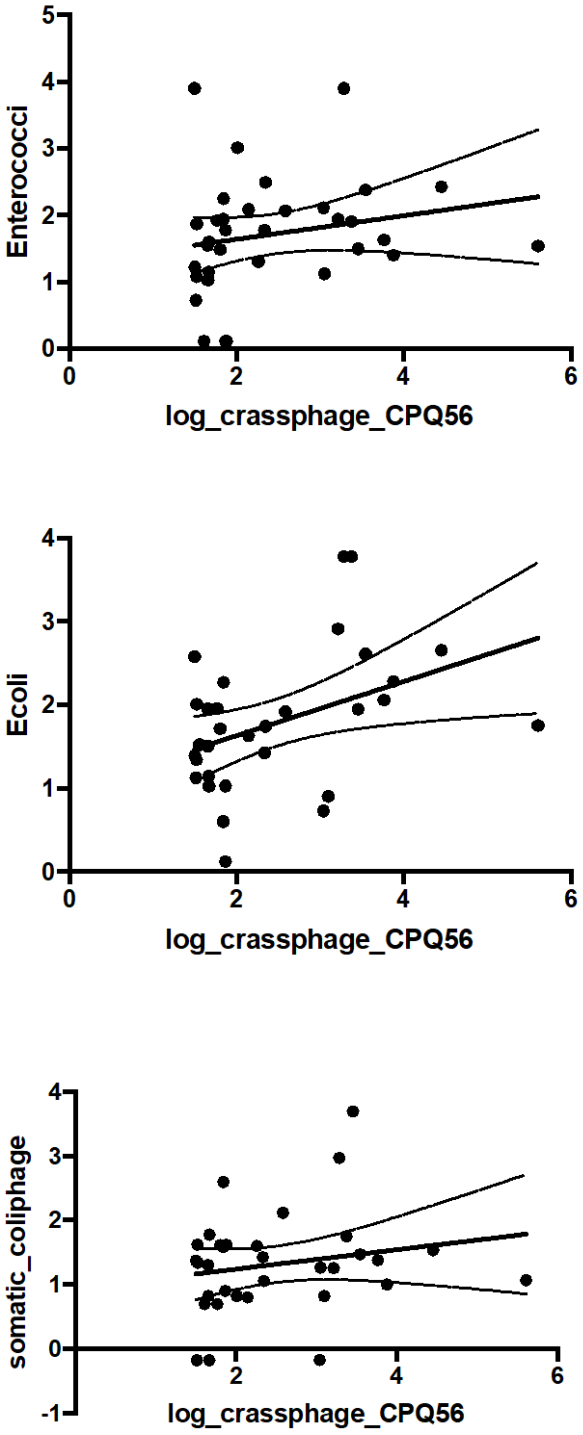

Figure 3. Correlation of the crAssphage CPQ56 assay with culturable indicators enterococci, *E. coli*, and somatic coliphage. The observed R^2 values for enterococci, *E. coli,* and somatic coliphage were 0.17, 0.04, and 0.04, respectively.

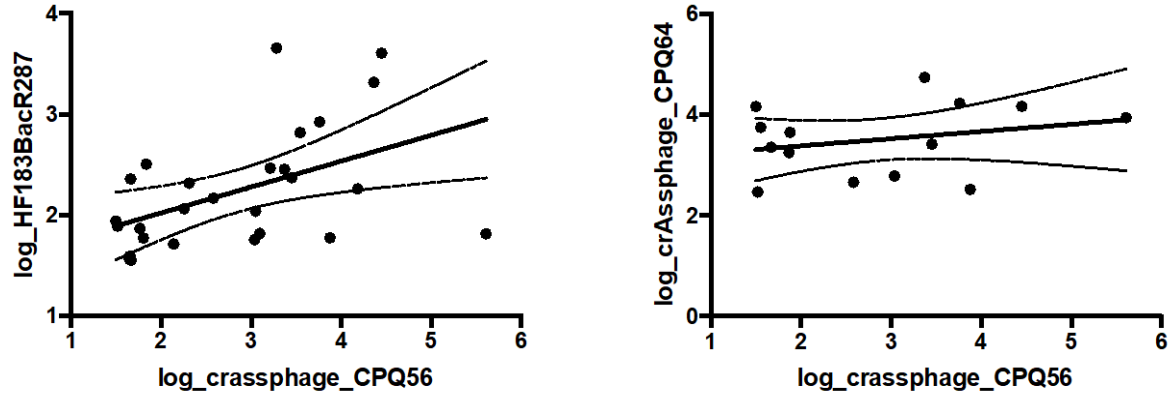

Figure 4. Observed correlations between log HF183/BacR287 and log CPQ56 (R^2 = 0.23; Y = 0.257*X + 1.508) and CPQ64 and CPQ56 (R^2 = 0.07; Y = 0.1426*X + 3.094).

References cited

1. Mara, D.; Sleigh, A., Estimation of norovirus infection risks to consumers of wastewater-irrigated food crops eaten raw. *Journal of Water and Health* **2010,** *8*, (1), 39-43.
2. Stine, S. W.; Song, I.; Choi, C. Y.; Gerba, C. P., Application of microbial risk assessment to the development of standards for enteric pathogens in water used to irrigate fresh produce. *Journal of Food Protection* **2005,** *68*, (5), 913-918.
3. Harwood, V. J.; Levine, A. D.; Scott, T. M.; Chivukula, V.; Lukasik, J.; Farrah, S. R.; Rose, J. B., Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl Environ Microbiol* **2005**, *71*, (6), 3163-3170.
4. Payment, P.; Locas, A., Pathogens in water: value and limits of correlation with microbial indicators. *Groundwater* **2011,** *49*, (1), 4-11.
5. Dutilh, B. E.; Cassman, N.; McNair, K.; Sanchez, S. E.; Silva, G. G. Z.; Boling, L.; Barr, J. J.; Speth, D. R.; Seguritan, V.; Aziz, R. K.; Felts, B.; Dinsdale, E. A.; Mokili, J. L.; Edwards, R. A., A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* **2014,** *5*, 4498.
6. Stachler, E.; Kelty, C.; Sivaganesan, M.; Li, X.; Bibby, K.; Shanks, O. C., Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environmental Science & Technology* **2017,** *51*, (16), 9146-9154.
7. Stachler, E.; Akyon, B.; Aquino de Carvalho, N.; Ference, C.; Bibby, K., Correlation of crAssphage-based qPCR markers with culturable and molecular indicators of human fecal pollution in an impacted urban watershed. *Environ Sci Technol* **2018**, *52*, (13), 7505-7512.
8. US EPA, Method 1600: Enterococci in Water by Membrane Filtration Using membrane-*Enterococcus* Indozyl-B-D-Glucoside Agar (mEI). In USEPA: Washington, DC, 2006.
9. US EPA, Method 1603: *Escherichia coli* (*E. coli*) in Water by Membrane Filtration Using Modified membrane-Thermotolerant *Escherichia coli* Agar (Modified mTEC). In USEPA: Washington, DC, 2014.
10. US EPA, Method 1602: Male-specific (F+) and Somatic Coliphage in Water by Single Agar Layer (SAL) Procedure. In USEPA: Washington, DC, 2001.
11. Staley, C.; Gordon, K. V.; Schoen, M. E.; Harwood, V. J., Performance of two quantitative PCR methods for microbial source tracking of human sewage and implications for microbial risk assessment in recreational waters. *Appl Environ Microbiol* **2012,** *78*, (20), 7317-7326.
12. Jothikumar, N.; Cromeans, T. L.; Hill, V. R.; Lu, X.; Sobsey, M. D.; Erdman, D. D., Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Appl Environ Microbiol* **2005,** *71*, (6), 3131-3136.
13. Green, H. C.; Haugland, R. A.; Varma, M.; Millen, H. T.; Borchardt, M. A.; Field, K. G.; Walters, W. A.; Knight, R.; Sivaganesan, M.; Kelty, C. A.; Shanks, O. C., Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl Environ Microbiol* **2014,** *80*, (10), 3086-3094.
14. Cao, Y.; Sivaganesan, M.; Kelty, C. A.; Wang, D.; Boehm, A. B.; Griffith, J. F.; Weisberg, S. B.; Shanks, O. C., A human fecal contamination score for ranking recreational sites using the HF183/BacR287 quantitative real-time PCR method. *Water Res* **2018,** *128*, 148-156.
15. Ahmed, W.; Lobos, A.; Senkbeil, J.; Peraud, J.; Gallard, J.; Harwood, V. J., Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida. *Water Res* **2018,** *131*, 142-150.